

The Evolution of Pholcidae (Araneae) Spiders

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Guanliang Meng
aus
GUANGXI, CHINA

Bonn, 04.2025

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen
Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer: Prof. Dr. Bernhard Misof
Gutachter: PD. Dr. Lars Podsiadlowski
Tag der Promotion: 25.06.2025
Erscheinungsjahr: 2025

Summary

The daddy long-legs spiders, Pholcidae C.L. Koch, 1850, also known as cellar spiders, are among the most species-rich families of spiders (Araneae), encompassing more than 2,000 formally described species and 97 genera (WSC 2025). They are distributed worldwide but are mainly found in tropical and subtropical areas, with some species being synanthropic (Huber, **Meng**, Král, Ávila Herrera et al. 2023). These spiders exhibit a variety of morphological changes associated with the different microhabitats they inhabit, which range from leaf litter, under or between rocks, the downside of leaves, to areas among vegetation. Currently, the Pholcidae family includes five recognized subfamilies: Ninetinae Simon, 1890, Arteminae Simon, 1893, Modisiminae Simon, 1893, Smeringopinae Simon, 1893, and Pholcinae C.L. Koch, 1850 (Huber 2011). Within Pholcidae, Pholcinae and Modisiminae are the two largest subfamilies, with approximately 900 and 500 described species, respectively, followed by Smeringopinae with about 130 species, Arteminae with around 100 species, and finally, Ninetinae with approximately 80 species (Huber, Eberle et al. 2018; Huber, **Meng**, Acurio et al. 2022; Huber, **Meng**, Král, Ávila Herrera et al. 2023; Huber, **Meng**, Král, Herrera et al. 2024). The mechanisms responsible for the notable differences in species diversity among subfamilies are not well understood. According to a prior study, the frequent changes in microhabitat within Pholcinae and Modisiminae may have played a crucial role in shaping this pattern of diversity (Eberle et al. 2018). Over the past two decades, several studies have consistently supported the sister-group relationships of both Pholcinae to Smeringopinae and Arteminae to Modisiminae, based on a limited number of genetic markers (Astrin et al. 2007; Bruvo-Madaric et al. 2005; Dimitrov et al. 2013; Eberle et al. 2018; Huber, Eberle et al. 2018). However, there is no consensus among these studies regarding the phylogenetic placements of the subfamily Ninetinae and the genera *Artema* (Walckenaer, 1837) and *Priscula* (Simon, 1893) (Astrin et al. 2007; Bruvo-Madaric et al. 2005; Dimitrov et al. 2013; Eberle et al. 2018; Huber, Eberle et al. 2018), likely due to insufficient sampling of taxa and genes. Furthermore, while the influence of microhabitats on pholcid diversification has been evaluated (Eberle et al. 2018), the significance of two particularly variable traits—body size and leg length—remains untested. These traits are in general closely related to an organism's fitness, ecology, behavior, physiology, and life history (Griethuijsen et al. 2014; Holmes et al. 2006; LaBarbera 1989; Labonte et al. 2013; Silva et al. 2021; Smith et al. 2018). The aforementioned knowledge gap in Pholcidae phylogeny has impeded our ability to examine how body size and leg length might affect pholcid diversification.

Improvements in sequencing technology have greatly increased access to molecular data, which has important applications in phylogenetic systematics—a discipline focused on resolving the relationships among different organisms (Hennig 1999; Wägele 2005). However, the complexities of sequencing technology and evolutionary histories of different organisms, combined with a multitude of models and algorithms developed to address these complexities, has posed significant challenges for understanding and performing both molecular experiment and data analysis. Therefore, before applying these new technolo-

gies and analyses to Pholcidae spiders, in Chapter 2, I first provide guidance on sample collection, storage, and transportation, followed by a brief overview of the processes involved in DNA extraction, library construction, and sequencing in the context of High throughput Sequencing (HTS) and reduced representation sequencing (RRS) methods (Good 2011). I introduce methods for data filtering, contig assembly, contamination detection & filtering, and ortholog inference, as well as methods for multiple sequence alignments and quality control. I then present techniques for evaluating systematic bias in the data and summarize key aspects of a variety of molecular evolution models, highlighting the strengths and weaknesses of different phylogenetic inference algorithms. Additionally, I summarize the methods for evaluating the reliability of the resulting phylogenetic trees. Furthermore, I explain the causes behind gene tree incongruence and the discordance between gene tree and species tree, which typically result from incomplete lineage sorting, gene flow, intra-locus recombination, taxon/gene coverage, low phylogenetic information and technical errors (e.g., hidden paralogs), before introducing the tip-dating and node-dating methods for estimating species divergence times.

Subsequently, in Chapter 3, using the spider-specific probe set for ultra-conserved elements (UCE) developed by Kulkarni et al. (2020), we applied the target capture sequencing (TCS) method to 96 Pholcidae samples, which included representatives from the five recognized subfamilies, as well as the two genera *Artema* and *Priscula*. Our analyses demonstrated that substitution saturation, model violation, low phylogenetic signal and compositional heterogeneity have little impact on our data and analysis. Additionally, few sequences exhibit intra-locus recombination, suggesting that our datasets are suitable for phylogenetic inference. Consistent subfamily phylogenetic relationships were obtained from analyses based on 10 different data matrices (with the exception of two short matrices) and various tree construction methods (concatenation vs. coalescence; maximum likelihood vs Bayesian inference). Our results further substantiate the sister relationship between Smeringopinae and Pholcinae, as well as between “other Arteminae” (Arteminae excluding the genus *Artema*) and Modisiminae, as reported in previous studies (Bruvo-Madaric et al. 2005; Dimitrov et al. 2013; Eberle et al. 2018; Huber 2011; Huber, Eberle et al. 2018). Therefore, our results refute the monophyly of the traditional concept of the subfamily Arteminae. We support a sister relationship between *Artema* and *Priscula*, which forms a sister group (gene support frequency: 43%, site support frequency: 61%). Together, they form a sister to the subfamily Ninetinae (gene support frequency: 65%). Previous studies suggested that Ninetinae was the “basal” group, i.e. sister to all other subfamilies; however, our results indicate that the clade (Ninetinae, (*Artema*, *Priscula*)) forms a sister group with (“other Arteminae”, Modisiminae). They all together then sister to the clade consisting of Pholcinae and Smeringopinae. Our phylogenetic analysis suggests that synspermia, the sperm transfer form that characterizes the closest relatives of Pholcidae, may have emerged independently in Ninetinae. Alternatively, Pholcidae experienced at least four times of convergent evolution from synspermia toward cleistospermia, a derived sperm transfer form. Our findings also suggest that the unique morphology of the male procursus (intromittent organ) present in both *Artema* and “other Arteminae” that is related to asymmetric insertion of the male genitalia, may have evolved independently in each of the two groups or been lost repeatedly in the other groups. Finally, we found that the most recent common ancestor (MRCA) of Pholcidae dates back to around 117 million years ago, which is consistent with a recent study on related spider families (Magalhaes et al. 2020) but significantly later than suggested for Pholcidae by a previous study (Dimitrov et al. 2013).

Building upon the Pholcidae phylogeny outlined above, we investigated the influence of the evolution of body size and leg length on pholcid diversification by constructing the largest Pholcidae phylogeny to date, encompassing 991 species (including 35 outgroup species) using both newly sequenced and publicly available genetic sequences (Chapter 4). Through a series of comparative analyses, we assessed the dynamics

of speciation and extinction, and the evolutionary flexibility of traits, measured by the standard variance or disparity of these traits, as well as their interactions. Our findings revealed that increased evolutionary flexibility is linked to accelerated rates in speciation and trait evolution, as well as greater species richness. Additionally, we observed a near-isometric relationship between body size and leg length in species-rich lineages, suggesting stabilizing selection, while positive allometry in species-poor lineages indicates directional selection. We also discovered a positive correlation between trait evolution rates and microhabitat shifts, suggesting that environmental heterogeneity and frequent microhabitat transitions could play a significant role in fostering and maintaining evolutionary flexibility, which enhances the organisms' abilities to exploit new resources and habitats. Conversely, our study implies that a lack of evolutionary flexibility, such as that seen in taxa with consistent dwarfism or gigantism, respectively, might lead to an evolutionary dead end. Furthermore, our results challenge the view that taxa consistently adhere to Cope's rule, which posits that body size increases with lineage age (Stanley 1973). Overall, this research enhances our understanding of the mechanisms underlying diversification, demonstrating that the evolutionary flexibility of morphological traits can vary among closely related taxa and likely contributes to the uneven distribution of biodiversity across the tree of life.

Extensive taxon and gene sampling can often discover unanticipated species diversity. In Chapter 5, we present two new species derived from collections in the poorly sampled northeastern part of Brazil: GENUS-1 sp1 and GENUS-1 sp2 (**The taxonomic names will be made available in a journal publication, and I will not apply the formal taxonomic names in this thesis in order to avoid any possible confusion about validity and availability; for the same reason, new subfamily names below are written in a way that clearly shows that they are not made available by this thesis**). Morphologically, the genus GENUS-1 Huber gen. n. displays similarities to the Chilean genus *Aucana* Huber (2000) of the group "other Arteminae," leading us to expect that they may be closely related. To determine their phylogenetic positions, we applied low-coverage whole genome sequencing (lcWGS) to the two new species (three specimens) along with four representatives of *Aucana*. We then combined these lcWGS data with the TCS data of 66 Pholcidae species and four outgroup taxa, as discussed in Chapter 3, for our phylogenetic analysis. The results consistently revealed the same subfamily-level phylogenetic relationships as those presented in Chapter 3. Moreover, strong support was found for the sister-group relationship between the two new species; thus, we classified them together in the single genus GENUS-1 Huber gen. n., despite their pronounced morphological differences. This genus is positioned as sister to a large clade that includes several pholcid subfamilies, necessitating the creation of a new subfamily: Cai***inae. Additionally, we confirm the separation of the genus *Artema* from "the other Arteminae," a distinction previously suggested in Chapter 3 and now strongly reinforced by our new genomic data. The subfamily Arteminae has now been redefined to include only *Artema* and *Priscula*, while we propose the name Phy***inae for "other Arteminae." As a result, the Pholcidae family is now divided into seven subfamilies, with the following relationships indicated by genomic data: (Pholcinae + Smeringopinae) + (Cai***inae + ((Arteminae + Ninetinae) + (Phy***inae + Modisiminae))). Finally, the hypothesis that the Chilean genus *Aucana* is the closest relative of the new Brazilian species (Cai***inae) is strongly refuted, as *Aucana* was found to be the only known South American representative of Phy***inae.

Introgression is one of the important factors contributing to gene tree heterogeneity and to the incongruence between gene trees and species trees. Recent studies have also shown that introgression may have important roles in the speciation and adaptation of organisms. However, we have little knowledge about the extent of introgression and its evolutionary significance in spiders. In Chapter 6, we combined 90 additional pholcid samples sequenced with the lcWGS method with the data from Chapter 3 and Chapter 5, and investigated the prevalence of introgression and its influence on phylogenetic inference in the spider

family Pholcidae. We identified multiple inter-subfamily and inter-genus introgressions within Pholcidae.

Mitochondrial genomes (mitogenomes) have important evolutionary, taxonomic, and ecological implications (Biot-Pelletier et al. 2023; Lattimore et al. 2011; Tang et al. 2015). Compared to other clades, such as insects and vertebrates, the scarcity of spider mitogenomes has limited our understanding of their evolutionary dynamics and interactions with nuclear genomes. In Chapter 6, we also present the mitogenomes of 193 pholcids. Importantly, 98 out of these mitogenomes are complete or near-complete, representing an 18% increase in spider mitogenomes longer than 10 Kbp and publicly available on NCBI. We found that some protein-coding genes (PCGs), including ATP8, ND4, ND4L, and ND5, as well as tRNA genes in pholcids, are difficult to annotate with existing pipelines. Our results suggest that the gene order of mitochondrial PCGs is highly conserved within Pholcidae. Additionally, mitochondrial PCGs exhibit strong codon usage bias (CUB) and significant purifying selection. Selection pressure, instead of mutation bias, seems to be responsible for this CUB. Large genetic distances of mitochondrial genes between subfamilies, genera, and species were observed, suggesting fast evolution. Therefore, we argue that introgression and the rapid evolution of mitogenomes contribute to the discordance observed between published pholcid phylogenies constructed from different data. Finally, our results highlight the importance of considering biological processes, including introgression, when resolving the phylogenetic relationships among organisms. It also emphasizes the need for enhancements in the annotation of spider mitochondrial genomes.

References

- Astrin, J. J., B. Misof and B. A. Huber (2007), *The pitfalls of exaggeration: molecular and morphological evidence suggests Kaliana is a synonym of Mesabolivar (Araneae: Pholcidae)*, *Zootaxa* **1646** 17.
- Biot-Pelletier, D., S. Bettinazzi, I. Gagnon-Arsenault et al. (2023), *Evolutionary trajectories are contingent on mitonuclear interactions*, *Mol Biol Evol* **40** msad061.
- Bruvo-Madaric, B., B. A. Huber, A. Steinacher and G. Pass (2005), *Phylogeny of pholcid spiders (Araneae: Pholcidae): combined analysis using morphology and molecules*, *Mol Phylogenet Evol* **37** 661.
- Dimitrov, D., J. J. Astrin and B. A. Huber (2013), *Pholcid spider molecular systematics revisited, with new insights into the biogeography and the evolution of the group*, *Cladistics* **29** 132.
- Eberle, J., D. Dimitrov, A. Valdez-Mondragon and B. A. Huber (2018), *Microhabitat change drives diversification in pholcid spiders*, *BMC Evol Biol* **18** 141.
- Good, J. M. (2011), *Reduced representation methods for subgenomic enrichment and next-generation sequencing*, *Molecular methods for evolutionary genetics* **772** 85.
- Griethuijsen, L. I. van and B. A. Trimmer (2014), *Locomotion in caterpillars*, *Biol Rev Camb Philos Soc* **89** 656.
- Hennig, W. (1999), *Phylogenetic systematics*, University of Illinois Press.
- Holmes, P., R. J. Full, D. Koditschek and J. Guckenheimer (2006), *The Dynamics of Legged Locomotion: Models, Analyses, and Challenges*, *SIAM Review* **48** 207.
- Huber, B. A. (2011), *Phylogeny and classification of Pholcidae (Araneae): an update*, *J Arachnol* **39** 211.
- Huber, B. A., J. Eberle and D. Dimitrov (2018), *The phylogeny of pholcid spiders: a critical evaluation of relationships suggested by molecular data (Araneae, Pholcidae)*, *Zookeys* 51.

- Huber, B. A., **G. Meng**, A. E. Acurio et al. (2022), *Metagonia spiders of Galápagos: blind cave-dwellers and their epigean relatives (Araneae, Pholcidae)*, *Invertebr Syst* **36** 647, DOI: [10.1071/IS21082](https://doi.org/10.1071/IS21082).
- Huber, B. A., **G. Meng**, J. Král, I. M. Ávila Herrera et al. (2023), *High and dry: integrative taxonomy of the Andean spider genus Nerudia (Araneae: Pholcidae)*, *Zool J Linn Soc* **198** 534, DOI: [10.1093/zoolinnean/zlac100](https://doi.org/10.1093/zoolinnean/zlac100).
- Huber, B. A., **G. Meng**, J. Král, I. M. Á. Herrera and L. S. Carvalho (2024), *Diamonds in the rough: Ibotyporanga (Araneae, Pholcidae) spiders in semi-arid Neotropical environments*, *Eur J Taxon* **963** 1, DOI: [10.5852/ejt.2024.963.2687](https://doi.org/10.5852/ejt.2024.963.2687).
- Kulkarni, S., H. Wood, M. Lloyd and G. Hormiga (2020), *Spider-specific probe set for ultraconserved elements offers new perspectives on the evolutionary history of spiders (Arachnida, Araneae)*, *Mol Ecol Resour* **20** 185.
- LaBarbera, M. (1989), *Analyzing body size as a factor in ecology and evolution*, *Annu Rev Ecol Syst* **20** 97.
- Labonte, D. and W. Federle (2013), *Functionally different pads on the same foot allow control of attachment: stick insects have load-sensitive "heel" pads for friction and shear-sensitive "toe" pads for adhesion*, *PloS One* **8** e81943.
- Lattimore, V. L., C. J. Vink, A. M. Paterson and R. H. Cruickshank (2011), *Unidirectional introgression within the genus Dolomedes (Araneae: Pisauridae) in southern New Zealand*, *Invertebr Syst* **25** 70.
- Magalhaes, I. L. F., G. H. Azevedo, P. Michalik and M. J. Ramírez (2020), *The fossil record of spiders revisited: implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic*, *Biological Reviews* **95** 184.
- Silva, V., C. Biancardi, C. Perafan et al. (2021), *Giant steps: adhesion and locomotion in theraphosid tarantulas*, *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **207** 179.
- Smith, A. N. and M. C. Belk (2018), *Does body size affect fitness the same way in males and females? A test of multiple fitness components*, *Biol J Linn Soc Lond* **124** 47.
- Stanley, S. M. (1973), *An explanation for Cope's rule*, *Evolution* **27** 1.
- Tang, M., C. J. Hardman, Y. Ji et al. (2015), *High-throughput monitoring of wild bee diversity and abundance via mitogenomics*, *Methods Ecol Evol* **6** 1034.
- Wägele, J.-W. (2005), *Phylogenetic systematics*, Verlag Dr. Friedrich Pfeil, München, Germany **3** 9.
- WSC (2025), *World Spider Catalog*, DOI: [10.24436/2](https://doi.org/10.24436/2), (visited on 20/01/2025).

List of Publications and Contributions

Publications as part of this thesis

1. **G. Meng**, C. Zhou, S. Liu and X. Zhou (2022), “Phylogenomics based on transcriptomics and short-read sequences”, *Zoological Systematics and Evolution Protocol eBook*, ed. by G. Qiao, M. Zhu, W. Wang, Y. Qishen, D. Ren, S. He, F. Lei, W. Cai and W. Piao, 1st ed., Bio-protocol LLC 793, DOI: [10.21769/BioProtoc.1010649](https://doi.org/10.21769/BioProtoc.1010649) (Chapter 2)
2. **G. Meng**, L. Podsiadlowski, D. Dimitrov and B. A. Huber (2025), *The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders*, submitted (currently under review by *Systematic Biology*) (Chapter 3 and 4)
3. **G. Meng**, L. Podsiadlowski and B. A. Huber (2025b), *Low coverage genome sequencing reveals a new subfamily of daddy long-legs spiders from Brazilian Caatinga (Araneae, Pholcidae)*, in preparation (to be submitted) (Chapter 5)
4. **G. Meng**, L. Podsiadlowski and B. A. Huber (2025a), *Exploring Phylogenetic Discordance: The Influence of Introgression and Insights into the Evolution of the Pholcidae (Araneae) Mitochondrial Genomes*, in preparation (to be submitted) (Chapter 6)

Additional publications and contributions

Pholcidae taxonomy and biodiversity related studies

Along with Chapter 5, during my PhD study period, I also contributed to the following publications related to Pholcidae taxonomy, systematics and biodiversity. More specifically, I was responsible for the molecular analysis and the corresponding writing in these publications. The molecular analyses mainly involved with sequence assembly, contamination checking, tree construction, genetic distance calculation and molecular species delimitation. These publications totally described **73 new species of 16 genera** and **revised 5 genera**.

1. B. A. Huber, **G. Meng**, A. E. Acurio, J. J. Astrin, D. J. Inclán, M. Izquierdo and A. Valdez-Mondragón (2022), *Metagonia spiders of Galápagos: blind cave-dwellers and their epigean relatives (Araneae, Pholcidae)*, *Invertebr Syst* **36.7** 647, DOI: [10.1071/IS21082](https://doi.org/10.1071/IS21082)
2. B. A. Huber, **G. Meng**, H. L. Clark and G. Cazanove (2023), *First blind daddy long-legs spiders from Australia and Réunion (Araneae, Pholcidae)*, *Subterr Biol* **46** 1, DOI: [10.3897/subtbiol.46.105798](https://doi.org/10.3897/subtbiol.46.105798)

3. B. A. Huber, **G. Meng**, J. Král, I. M. Ávila Herrera, M. A. Izquierdo and L. S. Carvalho (2023), *High and dry: integrative taxonomy of the Andean spider genus Nerudia (Araneae: Pholcidae)*, Zool J Linn Soc **198.2** 534, doi: [10.1093/zoolinnean/zlac100](https://doi.org/10.1093/zoolinnean/zlac100)
4. B. A. Huber, **G. Meng**, N. Dupérré, M. Herrera, D. J. Inclán and B. Wipfler (2023), *Hump-back spiders from Ecuador: relationships, prosoma ‘inflation’ and genital asymmetry (Araneae: Pholcidae: Mecolaesthus)*, Invertebr Syst **37.2** 117, doi: [10.1071/IS22052](https://doi.org/10.1071/IS22052)
5. B. A. Huber and **G. Meng** (2023), *On the mysterious Seychellois endemic spider genus Cenemus (Araneae, Pholcidae)*, Arthropod Syst Phylogeny **81** 179, doi: [10.3897/asp.81.e86793](https://doi.org/10.3897/asp.81.e86793)
6. B. A. Huber, **G. Meng**, J. Král, I. M. Ávila Herrera and M. A. Izquierdo (2023), *Revision of the South American Ninetinae genus Guaranita (Araneae, Pholcidae)*, Eur J Taxon **900.1** 32, doi: [10.5852/ejt.2023.900.2301](https://doi.org/10.5852/ejt.2023.900.2301)
7. B. A. Huber, **G. Meng**, A. Valdez-Mondragón, J. Král, I. M. Á. Herrera and L. S. Carvalho (2023), *Short-legged daddy-long-leg spiders in North America: the genera Pholcophora and Tolteca (Araneae, Pholcidae)*, Eur J Taxon **880** 1, doi: [10.5852/ejt.2023.880.2173](https://doi.org/10.5852/ejt.2023.880.2173)
8. B. A. Huber, **G. Meng**, N. Dupérré, J. Astrin and M. Herrera (2023), *Andean giants: Priscula spiders from Ecuador, with notes on species groups and egg-sac troglomorphism (Araneae: Pholcidae)*, Eur J Taxon **909** 1, doi: [10.5852/ejt.2023.909.2351](https://doi.org/10.5852/ejt.2023.909.2351)
9. B. A. Huber, **G. Meng**, T. M. Dederichs, P. Michalik, M. Forman and J. Král (2024), *Castaways: the Leeward Antilles endemic spider genus Papiamenta (Araneae: Pholcidae)*, Invertebr Syst **38.2** IS23052, doi: [10.1071/IS23052](https://doi.org/10.1071/IS23052)
10. B. A. Huber, **G. Meng**, J. Král, I. M. Á. Herrera and L. S. Carvalho (2024), *Diamonds in the rough: Ibotyporanga (Araneae, Pholcidae) spiders in semi-arid Neotropical environments*, Eur J Taxon **963** 1, doi: [10.5852/ejt.2024.963.2687](https://doi.org/10.5852/ejt.2024.963.2687)
11. B. A. Huber, **G. Meng**, J. C. garcía and L. S. Carvalho (2024), *Thriving in dry conditions: on the Neotropical spider genus Galapa (Araneae: Pholcidae)*, Zootaxa **5419.3** 301, doi: [10.11646/zootaxa.5419.3.1](https://doi.org/10.11646/zootaxa.5419.3.1)
12. B. A. Huber, **G. Meng** and A. Valdez-Mondragn (2024), *Notes on Chisosa (Araneae, Pholcidae), with the description of a new species from Mexico.*, Zootaxa **5419.2** 217, doi: [10.11646/zootaxa.5419.2.3](https://doi.org/10.11646/zootaxa.5419.2.3)
13. B. A. Huber and **G. Meng** (2024), *Old World Micropholcus spiders, with first records of ac-rocerid parasitoids in Pholcidae (Araneae)*, ZooKeys **1213** 95, doi: [10.3897/zookeys.1213.133178](https://doi.org/10.3897/zookeys.1213.133178)
14. B. A. Huber and **G. Meng** (2025), *Like grains of sand: Ninetis spiders on the Arabian Peninsula (Araneae, Pholcidae)*, Zootaxa **5563.1** 290, doi: [10.11646/zootaxa.5563.1.19](https://doi.org/10.11646/zootaxa.5563.1.19)

Non-Pholcidae-related studies

The following publications are related to phylogenomics or biodiversity of non-Pholcidae species, in which I made contributions outside of this thesis:

1. S. Liu, C. Zhou, **G. Meng**, T. Wan, M. Tang, C. Yang, R. W. Murphy, Z. Fan, Y. Liu and T. Zeng (2022), *Evolution and diversification of Mountain voles (Rodentia: Cricetidae)*, Commun Biol **5.1** 1417, DOI: [10.1038/s42003-022-04371-z](https://doi.org/10.1038/s42003-022-04371-z)
2. S. Liu, D. Lang, **G. Meng**, J. Hu, M. Tang and X. Zhou (2022), *Tracing the origin of honey products based on metagenomics and machine learning*, Food chemistry **371** 131066, DOI: [10.1016/j.foodchem.2021.131066](https://doi.org/10.1016/j.foodchem.2021.131066)
3. A. Klimova†, R. Rodríguez-Estrella, **G. Meng**†, J. N. Gutiérrez-Rivera, M. L. Jimenez-Jimenez and S. Liu (2023), *Metabarcoding reveals seasonal and spatial patterns of arthropod community assemblages in two contrasting habitats: Desert and oasis of the Baja California Peninsula, Mexico*, Diversity and Distributions **29.3** 438, DOI: [10.1111/ddi.13672](https://doi.org/10.1111/ddi.13672) († **contributed equally to this work**)
4. L. Zhou†, **G. Meng**†, L. Zhu, L. Ma and K. Chen (2024), *Insect Antimicrobial Peptides as Guardians of Immunity and Beyond: A Review*, Int J Mol Sci **25.7** 3835, DOI: [10.3390/ijms25073835](https://doi.org/10.3390/ijms25073835) († **contributed equally to this work**)

Conference contribution

I presented a talk on our findings of Chapter 3 and Chapter 4:

- **G. Meng**, L. Podsiadlowski, B. A. Huber and D. Dimitrov (2024), “The Complex Interplay Between Evolutionary Flexibility and Diversification in a Family of Spiders”, *The 116th Annual Meeting of the German Zoological Society*, Stuttgart, Germany

Contents

Summary	iii
List of Publications and Contributions	ix
1 Introduction	1
1.1 A general introduction to spiders	1
1.2 The daddy long-legs spiders (Pholcidae)	4
1.2.1 Morphology	4
1.2.2 Taxonomy	6
1.2.3 Biology and biogeography	7
1.3 Research objectives	9
1.3.1 Phylogenetic systematics	9
1.3.2 Factors affecting species diversification	10
1.3.3 Exploring hidden biodiversity in spiders	12
1.3.4 Investigating phylogenetic discordance and the evolution of mitochondrial genomes	14
1.4 Research questions	15
1.4.1 The knowledge gap in Pholcidae phylogeny	15
1.4.2 The role of body size and leg length evolution in Pholcidae diversification	16
1.4.3 Revealing hidden pholcid diversity by extensive taxon and gene sampling	16
1.4.4 Introgression and mitochondrial evolution of Pholcidae spiders	18
1.5 Research methods	18
1.6 Thesis structure	19
1.7 Expected impact	19
2 Methods for Phylogenomic Analysis	31
2.1 Summary	31
2.2 Introduction	33
2.3 Sample preparation and data acquisition	33
2.4 High-throughput sequencing	36
2.5 Data quality control and sequence assembly	36
2.6 Ortholog inference	40
2.7 Multiple sequence alignment (MSA)	42
2.8 Systematic bias assessment of MSA	43
2.9 Model selection	47
2.10 Phylogenetic tree/network reconstruction	49

2.11	Phylogenetic post-analysis	55
2.12	Estimation of species divergence times	58
2.13	Perspectives	59
3	Unraveling the Phylogeny of the Pholcidae Family Using Targeted Capture Sequencing	77
3.1	Summary	77
3.2	Introduction	79
3.3	Materials and methods	80
3.3.1	Taxonomic Sampling and Target Capture Sequencing	80
3.3.2	Data Matrix Construction	81
3.3.3	Systematic Bias Assessment	82
3.3.4	Phylogenetic Inference	83
3.4	Results	84
3.5	Discussion	88
3.6	Conclusion	89
3.7	Supplementary material	89
3.8	Funding	90
3.9	Acknowledgments	90
4	Evolutionary Flexibility and Diversification of Pholcidae	95
4.1	Summary	95
4.2	Introduction	97
4.3	Materials and methods	99
4.3.1	Molecular methods	99
4.3.2	Comparative analysis	101
4.4	Results	103
4.4.1	Speciation and Extinction Rates	103
4.4.2	Trait Evolution	103
4.4.3	Relationships between Trait Variation and Diversification	109
4.5	Discussion	109
4.5.1	Subfamily Species Diversity and Morphological Trait Evolution in Pholcids	109
4.5.2	Evolutionary Flexibility is Positively Correlated with Trait Evolution Rate and Speciation Rate	112
4.5.3	The Interplay between Allometry and Evolutionary Flexibility	113
4.6	Conclusion	115
4.7	Supplementary material	115
5	A New Subfamily of Pholcidae from Brazilian Caatinga	127
5.1	Summary	127
5.2	Introduction	129
5.3	Materials and methods	129
5.3.1	Taxonomy and Morphology	129
5.3.2	Abbreviations	130
5.3.3	Molecular phylogeny	130

5.4	Results	132
5.4.1	Molecular data and phylogeny	132
5.4.2	Taxonomy	132
5.5	Discussion	153
5.5.1	How reliable is the molecular phylogeny?	153
5.5.2	Morphological aspects	154
5.5.3	Notes on the biogeography of Cai****inae	155
5.6	Conclusion	157
5.7	Acknowledgements	158
6	Exploring Phylogenetic Discordance: The Influence of Introgression and Insights into the Evolution of the Pholcidae (Araneae) Mitochondrial Genomes	165
6.1	Summary	165
6.2	Introduction	166
6.3	Materials and methods	168
6.3.1	Taxonomic sampling	168
6.3.2	Molecular methods	168
6.4	Results	171
6.5	Discussion	179
6.5.1	Reconciling phylogeny with introgression: challenges and considerations	179
6.5.2	The need to improve mitochondrial gene annotation in spiders	185
6.5.3	Rapid evolution of mitogenomes in Pholcidae	186
6.5.4	Selection pressure drives codon usage bias in Pholcidae	187
6.6	Conclusions	187
6.7	Supplementary data	187
7	Discussion and Perspectives	197
7.1	Review of research goals and questions	197
7.2	Summary of main findings	197
7.3	Evaluation of methodology	198
7.4	Theoretical and practical significance	199
7.5	Study Limitations	200
7.6	Future research directions	201
	Acknowledgements	205
A	Supplementary information for Chapter 3	211
B	Supplementary information for Chapter 4	237
C	Supplementary information for Chapter 5	279
D	Supplementary information for Chapter 6	295
	List of Figures	335
	List of Tables	339

Introduction

1.1 A general introduction to spiders

Spiders (Arachnida: Araneae), whose history dates back to the Carboniferous (Dunlop 2010), are important terrestrial predators belonging to chelicerates (Arthropoda: Chelicerata) (Foelix 2010; Nyffeler et al. 2017; Wheeler et al. 2017). Unlike insects, chelicerates' body consists of two parts, the anterior cephalothorax (or prosoma) and the posterior abdomen (or opisthosoma), which are connected by a narrow stalk called pedicel (Platnick 2020). While the cephalothorax is primarily responsible for locomotion, food intake, and housing the central nervous system, the opisthosoma is mainly responsible for vegetative functions such as digestion, circulation, breathing, waste excretion, reproduction, and silk generation (Foelix 2010). The cephalothorax is shielded by a dorsal (upper) carapace, a ventral (lower) labium at the front, and a sternum at the back (Kaston 1978). While insects have three pairs of walking legs, chelicerates have four, in addition to a pair of pedipalps and a pair of chelicerae at the mouthpart (Platnick 2020). Spiders differ from other chelicerates by having silk glands in the abdomen and modified pedipalps in adult male spiders (Dimitrov and Hormiga 2021; Platnick 2020).

Eyes. In general, spiders have eight simple eyes located at the front of the cephalothorax, and these eyes are mostly arranged in two rows (Platnick 2020). However, a few species, such as wolf spiders (Lycosidae), have their eyes arranged in three rows. The pattern of eye arrangement and their relative size are important features for the identification of different spider families (Kaston 1978).

Chelicerae and venom. The chelicerae are the first appendages of the cephalothorax, followed by pedipalps and then legs (Magalhaes et al. 2020). A chelicera is composed of two parts: the basal segment and the fang. Near the tip of the fang, there is a tiny opening through which venom is delivered to the prey, immobilizing it during capture (Magalhaes et al. 2020).

Legs and movement. Locomotion and dispersal enable spiders to capture prey, search for mates more effectively, inhabit various habitats, and spread. For spiders, this can be achieved with either silk or legs. Technically, three types of locomotion can be observed when silk is involved, but a specific species may not utilize all of them (B. A. Huber 2023): (1) Ballooning. Spiders climb up to some places and then release one or several silk threads which caught by wind. The air currents and electric fields work together to lift the spider up to 4 km high and may carry it 1,000 km away (Bell et al. 2005; Glick 1939; Morley and Gorham 2020; Morley and Robert 2018; Narimanov et al. 2021). (2) Bridging. Airborne silk thread is utilized to span the gap between two objects which assists the movement of spiders through complex environments (Wolff, Schneider et al. 2014). It has been suggested that this behavior is prevalent

in araneomorph spiders and is probably a typical way of locomotion for arboreal web-weaving spiders, which are physically adapted for movement along silk threads (Moya-Larano et al. 2008). (3) Dropping on a line, which is also a common behavior observed in spiders (Bristowe 1939, 1958). On the other side, spiders may walk and run, jump and even cartwheel (Cloudsley-Thompson 1995; Nentwig et al. 2022). The spider walking legs (4 pairs) are composed of seven segments in the following order, from the body to the tip: coxa, trochanter, femur, patella, tibia, metatarsus, and tarsus (Magalhaes et al. 2020). A pedipalp has the same order of segments as walk legs, but it lacks a metatarsus. The tarsus may have two or three claws at its tip, which are often equipped with teeth (Magalhaes et al. 2020).

Respiratory systems. Different spiders may rely on distinct respiratory systems. The species from the Mesothelae, Mygalomorphae, and certain early-diverged Araneomorphae groups (Figure 1.1), such as Gradungulidae and Hypochilidae, possess two pairs of book lungs located on the second and third abdominal segments, respectively (Ramírez et al. 2021; Ramírez 2000; Schmitz 2012). In most spiders (Araneomorphae), however, the second pair of book lungs has evolved into tubular tracheae (Ramírez 2000; Schmitz 2012). The tracheae tubes may be short or very long branching into cephalothorax or even legs (Bromhall 1987). In certain species, such as those in the Oonopidae and Symphytognathidae families, the book lungs have been heavily reduced, and tracheae alone are used for respiration (Kulkarni, Wood et al. 2023; Schmitz 2012). In contrast, in the lost tracheae clade (belonging to Synspermiata), which consists of Diguettidae, Pacullidae, Pholcidae, Plectreuridae, and Tetrablemmidae, the derived tracheae have been lost, and only a single pair of book lungs is retained (Figure 1.1) (Kulkarni, Wood et al. 2023; Schmitz 2012).

Taxonomy. There are more than 130 nominated families within Araneae (WSC 2025). Based on the presence of segmented abdomen and the location of spinnerets, the order Araneae is divided into two extant suborders: Mesothelae and Opisthothelae (Figure 1.1; Platnick and Gertsch 1976). Mesothelae spiders have multiple plates on their abdomen and possess spinnerets located near the front of the abdomen, which are considered as primitive (plesiomorphic) characters. Within Mesothelae, for a long time, only one family Liphistiidae Thorell, 1869 was recognized, whose monophyly is well-supported by both morphological and molecular data (Coddington and Levi 1991; Platnick and Gertsch 1976; Wheeler et al. 2017; Xu, F. Liu, R.-C. Cheng et al. 2015). Liphistiidae has two subfamilies, Liphistiinae Thorell, 1869 and Heptathelinae Kishida, 1923, which have one genus (59 spp.) and seven genera (107 spp.) (S. Li 2022), respectively. Recently, these two subfamilies have been elevated to family-level (thus, Liphistiidae and Heptathelidae) (S. Li 2022) based on the results of several studies (Xu, F. Liu, Chen et al. 2015; Xu, Su et al. 2021), but received debates (Breitling 2022).

On the other hand, Opisthothelae lack abdominal plates and have spinnerets located at the rear of the abdomen (Platnick 2020; Xu, Su et al. 2021). Opisthothelae are further divided into Mygalomorphae and Araneomorphae (Figure 1.1; Platnick and Gertsch 1976). In Mygalomorphae, which includes more than 20 families (Opatova et al. 2020), the chelicerae are in parallel alignment and move in an up-and-down manner, while the remaining spiders (around 90% of all spiders) belong to Araneomorphae, the “true spiders,” whose chelicerae are facing toward each other and move from side to side (Foelix 2010). In the past, it was suggested Araneomorphae should be classified as Cribellatae and Ecribellatae spiders, depending on the presence or absence of a cribellum (Kaston 1964). A cribellum is a small, sieve-like plate in front of the spinnerets, on the underside of the abdomen (Kaston 1964). But recent studies have shown that the cribellum is a plesiomorphic character which is homologous to the anterior median spinnerets in Mesothelae. Thus, during evolution, some spiders reduced or completely lost their cribella independently (Coddington 2005; Foelix 2010).

The Ecribellatae spiders were once divided into Haplogynae and Entelegynae according to the complex-

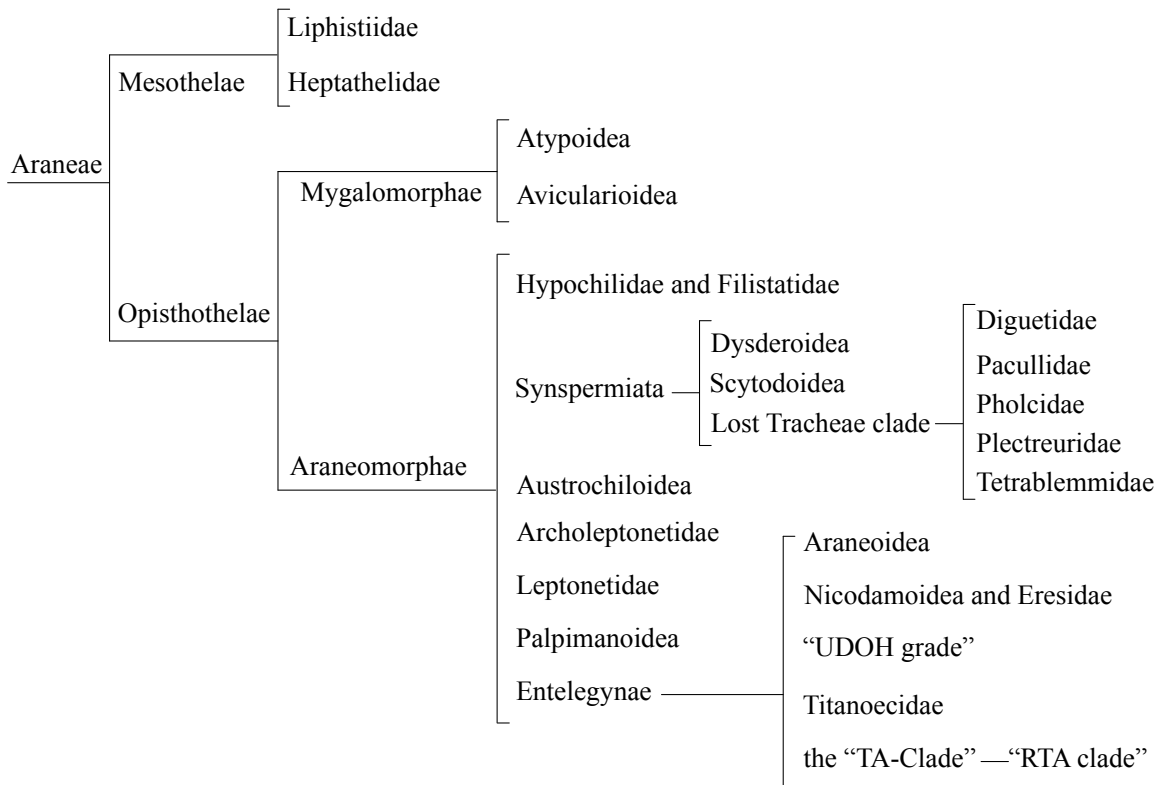


Figure 1.1. Spider tree of life. Based on studies Fernández et al. (2018), Kulkarni, Kallal et al. (2021), Kulkarni, Wood et al. (2023), S. Li (2022), Ramírez et al. (2021) and Wheeler et al. (2017). TA, tibial apophysis. RTA, Retrolateral tibial apophysis.

ity of the female’s reproductive tracts (Coddington 2005; Simon 1893). Haplogynae females have simple genitalia, with copulation and fertilization occurring through a single genital opening, allowing sperm to move both into and out of the spermathecae. In contrast, Entelegyne spiders have a more advanced reproductive structure, featuring an additional pair of copulatory ducts that directly connect the external genital opening (epigynum) to the spermathecae; therefore, the fertilization duct and copulatory ducts are independent. Haplogynae spiders include Filistatidae, Synspermiata and Leptonetidae (Coddington 2005; Magalhaes et al. 2020; Michalik et al. 2014). In the past 50 years, however, some researchers have doubted the monophyly of Haplogynae (Brignoli 1975; Lehtinen 1975; Platnick 1975; Platnick, Coddington et al. 1991). More recently, studies based on molecular data further confirmed the non-monophyly of Haplogynae (Bond, Garrison et al. 2014; Fernández et al. 2018; Garrison et al. 2016; Ramírez et al. 2021). Moreover, Entelegynae was once divided into Dionycha and Trionycha based on the number of tarsal claws on their legs, with Dionycha possessing two and Trionycha having three. In Trionycha, the middle small tarsal claw is important for web spiders to grasp the silk thread (Foelix 1970). However, it has been found that the three tarsal claws is the plesiomorphic state (Ramírez 2014).

In the past few years, studies based on molecular data and large-scale genomic sampling have further remodelled the major subgroups within Araneomorphae as follows (Figure 1.1): (1) Hypochilidae and Filistatidae, (2) Austrochiloidea, which typically possess two pairs of book lungs similar to the tarantulas

(Mygalomorphae), (3) Palpimanoidea, characterized by peg-like teeth on their chelicerae, (4) Synspermiata, distinguished by fused sperm cells, (5) Leptonetidae, which are recovered as monophyletic in some studies but not in others, and (6) Entelegynae. Furthermore, Entelegynae consists of (1) Araneoidea, (2) Nicodamoidea and Eresidae, (3) the paraphyletic group “UDOH grade,” including the families Uloboridae, Deinopidae, Oecobiidae, and Hersiliidae, (4) Titanoecidae, and (5) the tibial apophysis (TA) Clade, which includes Phyxelididae and the “RTA clade,” the latter defined by the presence of a retrolateral tibial apophysis (RTA) on the tibia of the male palp (Fernández et al. 2018; Kulkarni, Kallal et al. 2021; Kulkarni, Wood et al. 2023; Ramírez et al. 2021; Wheeler et al. 2017).

1.2 The daddy long-legs spiders (Pholcidae)

Pholcidae C.L. Koch, 1850, commonly known as daddy long-legs spiders or cellar spiders, is one of the most species-rich families, with over 2,000 described species divided into 97 genera (WSC 2025). It is also the most diverse family within the lost tracheae clade (Synspermiata) (WSC 2025). Pholcids belong to the ecribellate “haplogynes” and possess three claws at the tip of each tarsus (Ubick et al. 2005). Most of the members are web-weavers with irregular space webs or sheets, being able to capture relatively prey (but harmless to human) (B. A. Huber 2012; Platnick 2020). These spiders are globally distributed and some of them are synanthropic (B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023).

1.2.1 Morphology

Pholcids generally have either six or eight eyes (Kaston 1978), but a few species may have only two eyes and some cave species are blind (B. A. Huber 2000, 2018; B. A. Huber, Meng, Acurio et al. 2022). In eight-eyed species, the anterior median eyes (AME) are small, while the remaining eyes are arranged into two groups laterally, each containing three eyes (a triad) (Kaston 1978; Saaristo 2001). Species with six eyes have been found in genera such as *Belisana*, *Modisimus*, *Spermophorides*, *Spermophora* (B. A. Huber 2005a; B. A. Huber and Schaller 2005; Tong et al. 2008), whose AMEs between the two triads are absent. It has been suggested that the eight eyes are the plesiomorphic character of pholcids. In certain genera, the eyes are situated on prominent elevations, either on a median turret or lateral eye stalks (B. A. Huber and Nuñez 2015).

Compared to other spiders, pholcids are characterized by the following features: (1) The male palp possesses a prominent retrolateral paracymbium (“procursus”), which is rarely reduced to a small, indistinct lobe (B. A. Huber 2011a). (2) Except in a few species, the male chelicerae are modified and functionally associated with the female external genitalia (B. A. Huber 1993, 1995, 1996, 1998, 2002a). The external genitalia in females may have or lack a complex sclerotized plate (referred to as the “epigynum”), which can have pockets, grooves, apophyses, and other structures (B. A. Huber 2011a). (3) The clypeus, the region between the anterior row of eyes and the chelicerae (Kaston 1978), is approximately the same height as the length of the chelicerae, making it a potential synapomorphy (B. A. Huber 2000). (4) The tarsi are usually pseudosegmented; in some genera, the pseudosegments are difficult to discern under light microscopy and may appear secondarily fragmented into smaller sclerites. (5) The tibiae of the walking legs generally contain three trichobothria, though in rare cases, the first leg may have only two (B. A. Huber 2000, 2011a; Uhl et al. 1995).

Their body size are very small to medium, ranging from 1 millimeter (mm) (e.g., many *Ninetinae*, *Spermophorides lascars*) to 11 mm (e.g. some *Smeringopina*, *Artema atlanta*) (B. A. Huber 2013; B. A. Huber and Meng 2025; Saaristo 2001). The carapace is roughly as wide as it is long, and can lack a

median indentation, possess a median groove, or display a circular median depression (pit) (B. A. Huber 2000). In certain species such as the genus *Mecolaesthus* Simon, 1893, the male prosoma is distinctly inflated at the posterior end, which is presumed to function in male–male fights. (B. A. Huber, Meng, Dupérré, M. Herrera et al. 2023).

Their leg length is usually very long and thin, varying from 1 mm to 180 mm (the genus *Smeringopina* Kraus, 1957) (B. A. Huber 2013), and the male leg 1 is usually 2 to 25 times of its body length (B. A. Huber 2000). Female legs are usually shorter and finer compared to those of males. In spiders, the ratio of male tibia 1 length to diameter measures the robustness of the leg. In most pholcids, this ratio ranges from 40 to 80, but it can be as low as 8 and as high as 120 (B. A. Huber 2000). The tarsi of the legs are highly flexible, containing false segments, with their number ranging from 5 to 40 or even more (Kaston 1978; Saaristo 2001). While the leg tibiae usually have three trichobothria (tibia 1 sometimes two), the metatarsi consistently have one trichobothrium. The dorsal and prolateral trichobothria on the tibiae are consistently found at the proximal end, while the placement of the retrolateral trichobothrium varies widely. When expressed as a percentage of the tibia's overall length, its positioning can fall between 1% and 70%, displaying very little variation within the species (B. A. Huber 2000). Additionally, most pholcids follow the leg formula 1243 (leg length), whereas short-legged groups typically have 4123 or 1423.

Pholcids are typically colored in yellow, pale, gray, or light brown, often with brown or black markings (S. Rose 2022). Their chelicerae are fused at the base and feature a lamina that opposes the cheliceral fang, rather than teeth (B. A. Huber 2000). The size and number of eggs per egg sac are variable, with most egg sac shapes being round or oval (B. A. Huber and Eberle 2021). Females always carry the egg sacs using their chelicerae, and the sacs are wrapped in a layer of silk (B. A. Huber and Eberle 2021). These eggs are sometimes parasitized by wasp larvae (B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023; Johnson et al. 2018). The opisthosoma shows considerable variation in shape, ranging from higher-than-long to long and cylindrical or even worm-shaped. It can be spherical, triangular (in lateral view), rectangular, or bifid (in dorsal view) (B. A. Huber 2000). The clypeus, can vary from almost vertical to nearly horizontal and may exhibit sexual modifications in males, which can be ultrastructurally intricate (B. A. Huber 2000). The sternum is about as wide as it is long, or sometimes wider than long, and may have anterior bulges in males (Zhao, Jiang et al. 2023). The labium is fused with the sternum (Ubick et al. 2005).

They rely on a single pair of book lungs for respiration, with the tracheal system entirely lost, despite the apparent presence of a tracheal spiracle (Ubick et al. 2005). The male gonopore (i.e., the external opening of the testes on the abdomen) typically has either four epiandrous spigots or none, although it may sometimes have two or up to eight (B. A. Huber 2013, 2020; B. A. Huber, Deeleman-Reinhold et al. 1999). The female external genitalia may or may not possess a sclerotized plate, referred to as the “epigynum”. This plate is frequently complex, with pockets, grooves, apophyses, and other structures (B. A. Huber 2000). Internally, the uterus externus usually contains a pair of dorsal pore plates, which may be absent, fused into a single pore plate, or merged into a pore field. It also has a valve of varying complexity that marks the entrance to the uterus internus (Myrmidonella et al. 1998). Membranous sacs, typically originating near the valve of the uterus externus, are commonly found, usually as a single median sac, though a lateral pair may rarely occur. The function of these sacs remains unknown (B. A. Huber 2000).

There are three pairs of spinnerets, and no obvious sexual dimorphism in external morphology. The anterior lateral spinnerets (ALS) consistently have an enlarged, widened spigot, along with a smaller, pointed spigot adjacent to it (B. A. Huber 2000). Beyond this standard configuration, certain taxa are marked by the presence of five to seven cylindrical spigots, which are identified as piriform gland spigots,

as outlined by Platnick, Coddington et al. (1991). In contrast, other taxa only have the basic pair of spigots (B. A. Huber 2000). In all species examined, the posterior median spinnerets possess one pair of spigots, whereas the posterior lateral spinnerets are devoid of spigots (B. A. Huber 2000).

1.2.2 Taxonomy

Eugène Simon (Simon 1893) firstly proposed a division of Pholcidae into two subfamilies, Ninetinae and Pholcinae. The latter was further divided into seven subgroups, including Artemeae, Pholceae, Leptopholceae, Smeringopodeae, Prisculeae, Blechrosceleae and Modisimeae. More than a decade ago, based on the phylogenetic analyses of morphological and a few genetic markers, B. A. Huber (2011a) elevated four subgroups of Simon's Pholcinae to the subfamily rank, resulting in five subfamilies in total for the family, including Ninetinae Simon 1890, Arteminae Simon 1893, Modisiminae Simon 1893, Smeringopinae Simon 1893, and Pholcinae C.L. Koch 1850.

Pholcinae. Type genus: *Pholcus* Walckenaer 1805. Synapomorphies include the existence of male lateral proximal cheliceral apophyses (B. A. Huber 1995, 2000) and a single row of tarsus IV comb hairs (B. A. Huber and Fleckenstein 2008). Molecular data further substantiated the monophyly of Pholcinae (J. J. Astrin et al. 2007; Bruvo-Madaric et al. 2005; Dimitrov, J. J. Astrin et al. 2013; Eberle et al. 2018). With more than 900 known species divided into 38 genera, it represents the largest subfamily within Pholcidae (B. A. Huber, Eberle and Dimitrov 2018; B. A. Huber, Meng, Acurio et al. 2022). It also includes the two genera *Metagonia* Simon, 1893 and *Panjange* Deeleman-Reinhold & Deeleman, 1983, which include species whose male genitalia are asymmetric—a rare phenomenon in spiders but common in insects (B. A. Huber 2004; B. A. Huber and Nuñez 2015; Myrmidonella et al. 1998).

Smeringopinae. Type genus: *Smeringopus* Simon 1890. The monophyly of this subfamily lacks strong morphological support, relying on a distinct carapace feature (a large thoracic pit rather than a narrow furrow or an unbroken domed surface) and sperm ultrastructure (Dederichs et al. 2022; B. A. Huber 2011a). However, the monophyly is consistently supported by molecular data from Eberle et al. (2018) and B. A. Huber, Eberle and Dimitrov (2018), and is partially supported by Bruvo-Madaric et al. (2005). Within about 130 known species, Smeringopinae is considered as a small subfamily within Pholcidae.

Arteminae. Type genus: *Artema* Walckenaer 1837. Arteminae is characterized by the presence of two unique structures on the male procurus: a dorsal apophysis and a ventral pocket (B. A. Huber 2000). It has been shown that during mating, one procurus's apophysis fits into the pocket of the other, resulting in an asymmetrical insertion of the procuri (Aharon et al. 2017; B. A. Huber and Eberhard 1997). Initial molecular data supported the grouping of the three genera *Trichocyclus* (Australian), *Physocyclus* (pan-tropical), and *Artema* (Africa, Middle east) (Bruvo-Madaric et al. 2005, J. J. Astrin et al. 2007), and B. A. Huber (2011a) further added *Holocneminus* to the subfamily. The study by Dimitrov, J. J. Astrin et al. (2013) also supported this grouping, adding *Nita* (previously in Ninetinae) and the Australian genus *Wugigarra* (previously placed in Modisiminae) to Arteminae. Eberle et al. (2018) and B. A. Huber, Eberle and Dimitrov (2018) further added a few unnamed genera and transferred *Chisosa* (originally placed in Ninetinae) to Arteminae, however, they also challenged its monophyly. With around 100 known species, Arteminae is among the smaller of Pholcidae (B. A. Huber, Eberle and Dimitrov 2018).

Modisiminae. Type genus: *Modisimus* Simon, 1893. These spiders are grouped together based on the reduction of epiandrous spigots, the big distance between ALE and PME, and the presence of a retrolateral apophysis on the male palpal coxa, which serves to secure the palp in its rotated orientation when copulation begins (B. A. Huber 1998, 2000). The monophyly of this subfamily is also well-supported by molecular data (J. J. Astrin et al. 2007; Bruvo-Madaric et al. 2005; Dimitrov, J. J. Astrin et al. 2013;

Eberle et al. 2018). Comprising around almost 500 species in 24 genera, Modisiminae represents one of the two largest subfamilies of Pholcidae. *Anopsicus* Chamberlin & Ivie, 1938, *Psilochorus* Simon, 1893, *Modisimus* Simon, 1893, *Mesabolivar* and *Carapoia* González-Sponga, 1998 are among the species-diverse genera of the subfamily (B. A. Huber, Eberle and Dimitrov 2018). It remained unclear whether the Andean genus *Priscula* Simon, 1893 should be classified within this subfamily, as molecular analyses have failed to produce consistent results (Dimitrov, J. J. Astrin et al. 2013; Eberle et al. 2018; B. A. Huber, Eberle and Dimitrov 2018). On the other hand, the absence of a retrolateral apophysis on the male palpal coxa and the presence of ALS piriform gland spigots distinguish *Priscula* from other Modisiminae (Huber 2000), indicating that it is either misplaced or preserved plesiomorphic characters (B. A. Huber 2000).

Ninetinae. Type genus: *Ninetis* Simon 1890. B. A. Huber (2000) suggested two possible morphological synapomorphies for this subfamily. Firstly, the capsulate tarsal organ has a narrow opening. Secondly, the retrolateral trichobothrium is situated relatively distally on the leg tibia (beyond 45% of its length). Additionally, a third potential synapomorphy might be the compact ocular region where the AMEs are positioned very high, sometimes even above the anterior lateral eyes (ALE). A fourth potential synapomorphy, though possibly a plesiomorphy, is the presence of short legs, with the male tibia 1 measuring no more than 2.5 times the width of the carapace (B. A. Huber 2011a). The monophyly of Ninetinae was also supported by molecular studies based on a few genetic markers and a limited taxon sampling (J. J. Astrin et al. 2007; Dimitrov, J. J. Astrin et al. 2013; Eberle et al. 2018). With approximately 80 nominated extant species and 14 genera, Ninetinae is among the smallest subfamilies of Pholcidae, comprising only 4% of the extant diversity within this family. However, three out of fifteen (20%) of the known pholcid fossil records belong to Ninetinae (B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023; B. A. Huber, Meng, Král, I. M. Á. Herrera et al. 2024).

1.2.3 Biology and biogeography

Pholcids are distributed worldwide between approximately 56° N and 42° S (or 45° N and 35° S if synanthropic species are excluded), with their highest diversity occurring in humid tropical and subtropical regions, inhabiting various microhabitats, including leaf litter, hidden space among rocks and logs, caves, and the undersides of living leaves (B. A. Huber 2003, 2011b; B. A. Huber, Meng, Acurio et al. 2022; B. A. Huber, Meng, Dederichs et al. 2024). However, some species and even higher taxa are mostly or exclusively found in deserts or semi-arid environments (B. A. Huber and El Hennawy 2007) and high-altitude environments, such as *Smeringopus bujongolo* Huber, 2012 (Smeringopinae; B. A. Huber 2012) and *Nerudia* Huber, 2000 (Ninetinae; B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023), which have been recorded at elevations of up to 4,450 m above sea level. Meanwhile, approximately 25% of pholcids have been discovered in caves, but only 87 species are classified as troglomorphic. Among these, 22 species lack eyes entirely, while 21 species have reduced eyes. Notably, around 45% of these troglomorphic species are found in Mexico, while the remaining species are distributed across several islands, including Jamaica, the Canary Islands, the Galápagos, Cuba, Madagascar, and Réunion (B. A. Huber 2018; B. A. Huber, Meng, Clark et al. 2023).

Pholcinae. As the most speciose subgroup within Pholcidae, Pholcinae spiders exhibit significant variability in body forms and microhabitats, much like the other species-rich subfamily, Modisiminae. They are predominantly (35 of 38 genera) found in the humid tropics and subtropics of the Old World, with a few exceptional taxa occurring in the New World. These include the genus *Metagonia* Simon, 1893, and a few species of *Pholcus* and *Micropholcus*, but only *Metagonia* is limited to the New World within Pholcinae (B. A. Huber 2011b; B. A. Huber, Carvalho and Benjamin 2014; B. A. Huber, Meng, Acurio et al. 2022).

Pholcus is also the most species-rich genus within Pholcidae (Zhao, L. Yang et al. 2023). Furthermore, certain notable groups such as *Pholcus phalangioides* (Fuesslin, 1775) (Schäfer et al. 2001), *Spermophora senoculata* (Dugès, 1836) (B. A. Huber 2005a) and *Micropholcus fauroti* (Simon, 1887) (B. A. Huber, Neumann et al. 2017) are distributed worldwide or have dispersed to different continents (B. A. Huber and Meng 2024). Along with their high diversity, Pholcinae and Modisiminae contain 95% of troglomorphic pholcid species (B. A. Huber 2018). Within the Pholcinae subfamily, the genus *Metagonia* includes 25 troglomorphic species, while another 23 troglomorphic species are found across the genera *Pholcus*, *Spermophorides*, *Zatavua*, *Belisana*, and others (B. A. Huber 2018; B. A. Huber, Meng, Clark et al. 2023). A recent study reported the discovery of the first instance of acrocerid parasitoids within Pholcidae, where larvae from Acroceridae (Diptera) were identified in the book lungs of two species of *Micropholcus*. This occurrence may arise from the fact that these spiders create webs that are closely attached to surfaces beneath them, implying that Pholcidae could potentially serve as relatively common hosts for Acroceridae on a global scale. (B. A. Huber and Meng 2024)

Smeringopinae. All members of this subfamily are eight-eyed, and most of them are medium to large in size, characterized by long legs, elongated or cylindrical abdomens. Smeringopinae share similarities with Ninetinae and Arteminae in two key aspects: (1) species of this subfamily exhibit a relatively uniform body structure, coloration, web types, and microhabitat preferences, which distinguish them from Pholcinae and Modisiminae; and (2) most species of Smeringopinae are found in relatively dry regions (B. A. Huber, Eberle and Dimitrov 2018). But there is a notable exception, the predominantly humid tropical genus *Smeringopina* Kraus, 1957. The body length of this genus vary from 2.5 to 10 mm, and the leg span of some species can be up to 180 mm (B. A. Huber 2013). Additionally, species within the genus are very varialbe, similar to Pholcinae and Modisiminae (B. A. Huber 2013). With three species having greatly expanded their distribution due to human-assisted dispersal, most species of the subfamily are found across Africa, the Mediterranean, and the Middle East (B. A. Huber 2011a). This subfamily is divided into two distinct monophyletic clades: one located in the northern regions, covering Northern Africa, the Mediterranean, the Middle East, and Central Asia, and another in the southern regions, which includes Sub-Saharan Africa (B. A. Huber and Meng 2023). Only three slightly troglomorphic species have been found within Smeringopinae, occurring in the genera *Stygopholcus* and *Hoplopholcus* (B. A. Huber 2018).

Arteminae. This subfamily includes both large spiders with long, robust legs and high rounded abdomens, as well as small or even tiny species. For example, the probably largest (body mass) pholcids belong to the type genus, *Artema* (Aharon et al. 2017). On the other hand, genera such as the North American *Chisosa* and *Nita* include tiny pholcids (B. A. Huber 2011a). Similar to Ninetinae, these spiders are often found in very dry regions and even deserts (e.g., the Australian *Trichocyclus*) (B. A. Huber 2011a). Most genera of the subfamily are distributed in Asia and Australia. Only three genera occur in the New World: *Physocyclus*, *Chisosa*, and *Aucana* (B. A. Huber 2011a; B. A. Huber, Eberle and Dimitrov 2018). Two species are synanthropic and pantropical (*Physocyclus globosus*, *Artema atlanta* Walckenaer 1837).

Modisiminae. These spiders are restricted to the New World, in contrast to the subfamily Pholcinae which has the greatest diversity in the Old World (B. A. Huber, Meng, garcíaA et al. 2024). In the humid Neotropics, Modisiminae are widely distributed across various microhabitats, from the leaf litter to the upper vegetation layers. This adaptability is reflected in their diverse body forms, ranging from small, ground-dwelling species (B. A. Huber and Rheims 2011) to some of the largest pholcids, whose legs can span over 15 cm (B. A. Huber 2015; B. A. Huber and J. J. Astrin 2009). Their ecological and morphological diversity makes them prominent pholcids in these regions (B. A. Huber, Eberle and Dimitrov 2018). Within Modisiminae, 30 troglomorphic species have been found in the genera *Anopsicus* and *Psilochorus*,

respectively (B. A. Huber, Eberle and Dimitrov 2018).

Ninetinae. While most pholcids have body lengths ranging from 2 to 10 mm, Ninetinae spiders are exceptionally small, measuring only 1 to 2 mm (B. A. Huber, Meng, Král, Ávila Herrera and Izquierdo 2023; B. A. Huber, Meng, Král, I. M. Á. Herrera et al. 2024). Ninetinae also have shorter legs, which may explain their fast running as opposed to the slower movements of web-hanging long-legged pholcids. They are most often found in secluded spaces beneath rocks and stones in dry regions (B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023; B. A. Huber, Meng, Král, Ávila Herrera and Izquierdo 2023; B. A. Huber, Meng, Král, I. M. Á. Herrera et al. 2024). Together with the smallest species diversity within Pholcidae, it has been suggested that Ninetinae has been poorly sampled (B. A. Huber, Meng, Král, I. M. Á. Herrera et al. 2024). Only two genera have been documented in the Old World: *Ninetis*, which occurs in Africa and the Arabian Peninsula (B. A. Huber 2000, 2002b; B. A. Huber and Harten 2001;), and *Magana*, known from Oman only (B. A. Huber and Carvalho 2019). In contrast, the majority of this group's diversity is found in the New World (B. A. Huber 2000; B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023; B. A. Huber, Eberle and Dimitrov 2018; B. A. Huber, Meng, garcía et al. 2024). Four genera have been recorded in Brazil, while three genera have been identified in Argentina and Venezuela, but their distribution extend from the southernmost regions of Argentina to the northernmost areas of Canada (B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023). Compared to other pholcids, Ninetinae are more tolerant of extreme environments, with the genus *Nerudia* most often found beneath rocks at an average elevation of 1,750 m, including a species that inhabits the highest known habitat among pholcid species at an altitude of 4,450 m (B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023).

1.3 Research objectives

1.3.1 Phylogenetic systematics

Phylogenetic systematics, also known as phylogenetics, aims to acknowledge and validate relationships between organisms, which allows for placements of organisms in a phylogenetic tree that can be tested intersubjectively (Hennig 1999; Wägele 2005). Species change over time through a process known as “descent with modification” (Darwin 1859). Throughout evolution, species may diverge into distinct descendant species, a phenomenon referred to as speciation (Darwin 1859). According to Hennig (1999), a taxon, or group of organisms, should be monophyletic; this means it includes the ancestral taxon along with all of its descendants. Synapomorphic characters or character states derived and shared by the descendants provide evidence for monophyly, while plesiomorphies refer to primitive characters or character states inherited from the ancestor (Hennig 1999; Wägele 2005). From this perspective, both morphological and molecular data provide evidence for delineating species, establishing higher taxa, and inferring their phylogenetic relationships. With the advancement of sequencing technology, an increasing amount of molecular data has been applied in this field. For instance, over the past two decades, more effective methods for species delineation have been developed, which may be based on morphometric data (Mutanen et al. 2007), molecular data (Luo et al. 2018; Z. Yang and Rannala 2010), or a combination of both molecular and morphological data (Caro et al. 2019; B. Yang et al. 2022). The decreasing cost of sequencing has facilitated the adoption of larger molecular datasets, particularly at the genomic scale, in phylogenetic systematics. This shift has led to the emergence of phylogenomic approaches that leverage these extensive genomic datasets and shed lights into the evolution of many organisms (e.g., Misof et al. 2014; One Thousand Plant Transcriptomes Initiative 2019). However, phylogenomic methods have only

recently been applied to the study of spiders.

A long-standing view regarding the evolution of orb-weaving spiders is that all orb-weavers belong to the clade Orbiculariae, which comprises Deinopoidea, Araneoidea, Linyphioidea, the Spineless Femur clade, and Symphytognathoidea (Coddington 2005). By constructing a backbone phylogeny for major spider lineages using hundreds of orthologs, Bond, Garrison et al. (2014) challenged the monophyly of Orbiculariae. This finding suggests that either the orb web was a primitive characteristic present in the common ancestors of most entelegyne spiders but subsequently lost in some lineages, or that it (along with silk and silk genes) evolved independently multiple times across different lineages. The former hypothesis is supported by Garrison et al. (2016) which was based on transcriptomic data but a limited taxon sampling. Interestingly, a subsequent study by Fernández et al. (2018) based also on transcriptomic data but with a greatly increased taxon sampling rejects the former hypothesis and affirms the latter. Moreover, their results contradict the majority of the araneoid interfamilial relationships suggested by previous studies based on a few genetic markers (Dimitrov, Benavides et al. 2017; Wheeler et al. 2017), highlighting the importance of dense genomic sampling and taxon sampling to resolve phylogenetic relationships between species.

Based on the phylogeny constructed from ultraconserved elements (UCE) data, Ramírez et al. (2021) studied the evolution of the second pair of book lungs in “true spiders,” Araneomorphae. They demonstrated that, following the evolution of new silk gland systems and the emergence of aerial webs, the second book lung independently evolved into morphologically analogous tracheal systems six times. This similarity may arise from the common energy requirements for spinning silk.

By examining the phylogenetic results obtained from different data types (including exons, non-coding regions, or a combination of the two), Kulkarni, Kallal et al. (2021) confirmed the monophyly of the “symphytognathoids,” a group that comprises the families Anapidae, Mysmenidae, Theridiosomatidae, and Symphytognathidae (Griswold et al. 1998) and emphasizes the significance of extensive gene sampling.

The relationships among genera within the family Liphistiidae (Mesothelae), which encompasses both Liphistiinae and Heptathelinae, have not been adequately resolved in the past due to insufficient gene and taxonomic sampling, incomplete lineage sorting, gene flow among related species, and the rapid diversification of Heptathelinae. By analyzing a number of UCE genes, Xu, Su et al. (2021) successfully elucidated the phylogeny and divergence times of Liphistiidae. Their findings suggest that the crown group of Liphistiidae emerged approximately 102 Ma during the mid-Cretaceous period, subsequently diversifying in mainland Southeast and East Asia since Neogene, likely influenced by significant tectonic events (Xu, Su et al. 2021).

In summary, comprehensive genomic and taxon sampling will not only verify the monophyly of existing taxa and help to disentangle the phylogenetic relationships among organisms, but it will also have significant implications for understanding the origins of traits and the diversification of species.

1.3.2 Factors affecting species diversification

Different taxa often exhibit uneven species diversity throughout the tree of life, a phenomenon attributed to variations in speciation and extinction rates, as well as differences in species age (Henao Diaz et al. 2019). Speciation can be driven by ecological factors and/or non-ecological factors (Czekanski-Moir et al. 2019; Nosil, Harmon et al. 2009; Rundell et al. 2009). In ecological speciation, divergent selection (including natural selection and sexual selection) due to environmental heterogeneity acts on different sub-populations of a species, leading to the formation of reproductive isolation (Nosil 2012). In this process, both traits responsible for interacting with the environment (type 1; e.g., obtaining resources, de-

fense) and those responsible for reproductive compatibility (type 2) become divergent and increase their disparity among the sub-populations (Czekanski-Moir et al. 2019). On the other hand, in the process of non-ecological speciation, lineages proliferate via the fixation of different mutations in different sub-populations (either allopatric or sympatric) (Nosil and Flaxman 2011). As a result, only the type 2 traits diverge among the sub-populations (new species). However, these new species will be involved in competition exclusion because they inherit the similar ecological niches (i.e., similar type 1 traits) from their common ancestors (Gause 1932; Rowsey et al. 2020). Geographically, both ecological and non-ecological speciation can happen allopatrically and sympatrically (Czekanski-Moir et al. 2019; Nosil 2012). Even in the case of allopatric speciation, once the geographical barriers disappear, the two previously isolated lineages may contact with each other, leading to various outcomes (Czekanski-Moir et al. 2019). Therefore, studying the origin, maintenance and evolvability of biotic factors – such as traits related to morphology, behavior, physiology and genetics – alongside abiotic factors like climate and their interactions (e.g., via selection), can shed light on the uneven distribution of species diversity in the tree of life.

How may selection drive the evolution of traits? From a selection perspective, different selective forces act at distinct stages of an individual's lifespan, shaping the development of various traits and influencing overall fitness (Czekanski-Moir et al. 2019). During early development, from zygote to adulthood, viability selection plays a key role. Individuals that grow faster, become larger, or mature earlier may gain advantages through enhanced defense, resource competition, or earlier mating opportunities (Hadfield 2008; Jebb et al. 2021). During mate searching, sexual selection favors individuals with certain traits, such as larger body size or stronger dispersal abilities, increasing their likelihood of reproductive success (Dong et al. 2023; Glaudas et al. 2020; Janicke et al. 2021; Oliveira et al. 2024). After copulation, fecundity selection acts on individuals with higher reproductive output, often associated with larger body size (Head 1995; Rocha et al. 2024; Wadgymar et al. 2017). Following this, gametic selection occurs, as different gametes compete for fertilization (Otto et al. 2015; Scott et al. 2024). Finally, compatibility selection determines whether zygotes can successfully develop, completing the cycle of selection across an individual's lifespan (Tregenza et al. 2000). However, the evolution of traits is influenced not only by selection but also by genetic foundations, including the presence of related genes, the evolution of gene expression and regulatory networks, and genetic correlations arising from linkage disequilibrium or pleiotropy (Hoekstra et al. 2007).

Is there any link between species diversification and trait evolution? Key innovations have been found to allow organisms to explore new ecological opportunities and contribute to explosive radiations of many taxa. For instance, anatomical innovations in cetaceans (Gillet et al. 2019) and dinosaur-bird transition (Lee et al. 2014); Trophic innovations in reef fishes (Siqueira et al. 2020); silk structure and web architecture, or the lost of foraging web (biomaterial and behavioral) innovations in spiders (Bond and Opell 1998; Bond, Garrison et al. 2014; Dimitrov and Hormiga 2021; Sensenig et al. 2010). A decoupled relationship between speciation rate and trait evolution has also been observed in diversity-dependent diversification (Crouch et al. 2019; Rabosky 2013). In contrast, many studies highlight the correlation between speciation diversification and trait evolution (e.g., Cardillo et al. 2005; Cerezer et al. 2023; Cooney et al. 2021).

Spider diversification is influenced by various factors, with web architecture being less significant than previously thought (Dimitrov and Hormiga 2021). The transition from web-based foraging to active hunting has occurred across spider lineages but is not the main driver of diversification. Instead, environmental factors and ecological constraints appear to play a more significant role in shaping species diversity (Fernández et al. 2018). Therefore, it highlights the need to identify which traits are involved in interactions with the environment and how they influence those interactions. Spiders exhibit a wide variety of

body sizes and leg lengths, making them an excellent group for examining the evolution of these traits and their impacts on diversification (Shao et al. 2023; Wolff, Wierucka et al. 2022).

Body size is an important trait that affects the interactions between species and their environment, influencing fitness, ecology, physiology, and life history (LaBarbera 1989; Smith et al. 2018). Higher fitness associated with larger body size may be achieved by more effectively escaping predators and capturing prey, outcompeting others in male-male competition to obtain more mating opportunities, and dispersing to new environments with relaxed predation pressure and fewer competitors. The size-grain hypothesis (SGH) posits that smaller organisms perceive the Earth's surface as more uneven than larger organisms do. This increased perceived complexity might allow for greater specialization among these smaller organisms (Kaspari et al. 1999). Interestingly, higher extinction rate has been found in large body size mammals (Cardillo et al. 2005), which may be due to their reduced birth rates and greater demands for resources and energy (Bush 1993) and other factors (Purvis et al. 2000; Van Valen 1973). Increased body size evolution has been found to be strongly correlated with higher speciation rates in South American freshwater fishes (Cerezer et al. 2023). However, some studies also show complex relationships between speciation rate and body size (Cooney et al. 2021), and both speciation rate and extinction rate may decrease with body size (Etienne et al. 2012).

Legs play an important role in adhesion and locomotion in many animals (Griethuijsen et al. 2014; Holmes et al. 2006; Labonte et al. 2013; Silva et al. 2021), influencing stride length, stride frequency and walking speed (Kaspari et al. 1999; Taylor et al. 1980). Short legs, due to a smaller body size, enable organisms to reach small gaps that may offer food and shelter, affecting habitat selection (Kaspari et al. 1999; C. Li et al. 2011). Both body size and leg length affect locomotion costs and dispersal capabilities, with the latter shown to correlate with speciation rates (Cadotte 2006; Claramunt et al. 2012; Czekanski-Moir et al. 2019; Grossi et al. 2016; Pontzer 2007; Venail et al. 2008).

Spider body size and leg length significantly impact their biology. To achieve larger body size, spiders have to undergo more molts and experience longer intervals between molts before reaching sexual maturity (Head 1995). Larger spiders are more likely to win fights with other males (Wells 1988). Average body size increases from cooler, moist environments to warmer, drier ones (Entling et al. 2010). Body size also plays a crucial role in the evolution of web structures and reproductive strategies (Dong et al. 2023; Sensenig et al. 2011). On the other hand, relative leg length varies considerably among spider species; for instance, jumping spiders have shorter legs, while funnel-web and pholcids have longer legs (Pechmann et al. 2010). Spiders that inhabit silk-lined burrows typically have larger bodies and shorter legs, while those that create aerial webs are generally smaller with longer legs. Hunting spiders exhibit intermediate body sizes and structures between these two extremes (Shao et al. 2023).

1.3.3 Exploring hidden biodiversity in spiders

Over the past 260 years, more than 52,200 spider species have been formally described (WSC 2025; Figure 1.2). Surprisingly, 36.8% (19,261 species) of all currently known spider species were described in the last 25 years, with an average of around 740 newly nominated species per year. In 2023 alone, the number of newly described species reached 1,236, highlighting the significant effort dedicated to discovering and describing new spiders (WSC 2025). Over the past several years, many of these newly nominated species have originated from China, Australia, Ecuador, Brazil, Vietnam, and several other countries. For instance, the proportions of newly described species and established genera by Chinese researchers are 41.2% for 2023, 37.0% for 2022, 33.8% for 2021, and an average of 28.1% between 2016 and 2018. In addition to the increase in species, the number of families and genera has also risen, with 3 new families and 700 new

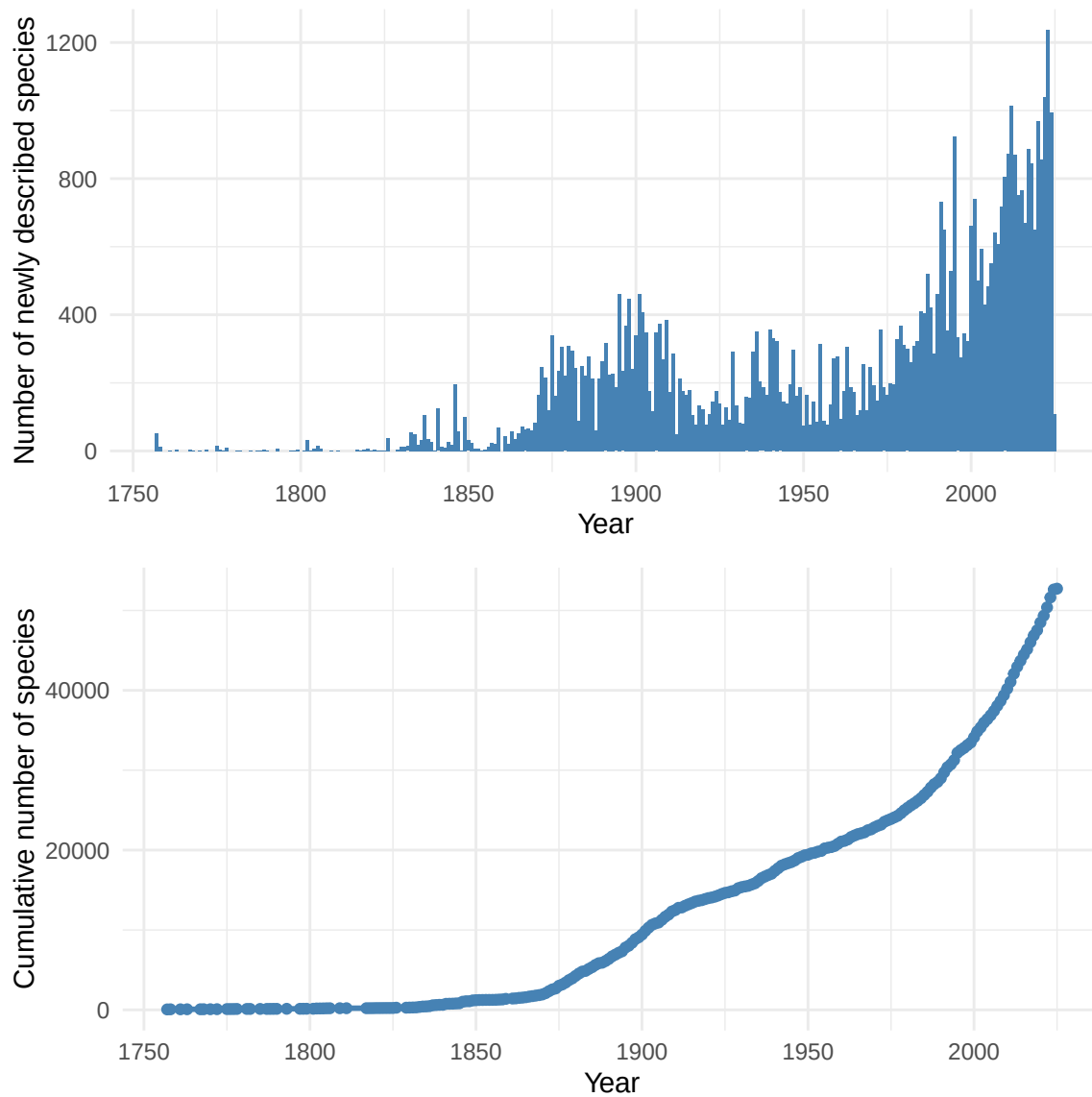


Figure 1.2. Spiders (Araneae) descriptions over the years. Based on data from WSC (2025).

genera added in the past 25 years, respectively (X. Li et al. 2024).

As technology advances, an increasing variety of techniques are being used to describe new taxa. For example, over 80% to 95% of publications now feature digital images for illustration, while around 20% use scanning electron microscope (SEM) images. In contrast, the use of hand-drawings has dropped to about 40%. Additionally, about 18% of studies have started incorporating DNA barcodes for data analysis (R. Yang et al. 2023). Morphological and molecular data provide different lines of evidence for delineating species (e.g., Puillandre, Brouillet et al. (2021), Puillandre, Lambert et al. (2012), Z. Yang (2015) and J. Zhang et al. (2013)), and even for establishing higher taxonomic ranks (e.g., Short et al. (2013)).

Climate change, global warming, rising sea levels, habitat loss and fragmentation, and pollution are serious threats to biodiversity (Bernatchez et al. 2024; Wagner et al. 2021). This makes it increasingly urgent for us to discover and protect unknown species and their ecosystems before they are extinct (Cowie et al. 2022). By extensively sampling taxa and integrating both morphological and molecular information, we can uncover the hidden biodiversity of spiders more efficiently.

1.3.4 Investigating phylogenetic discordance and the evolution of mitochondrial genomes

While a wealth of genomic data has deepened our understanding of the evolution of various organisms in the past two decades, more and more studies have shown that phylogenetic trees constructed from different data often contradict each other. It is now clear that both technical artifacts and biological processes may contribute to such phylogenetic incongruences. Phylogenetic inference methods often make various assumptions regarding the evolution of molecular data; for instance, homologous loci are properly identified and aligned, and different loci share the same evolutionary history. Recombination, incomplete lineage sorting (ILS), and gene flow are biological processes that may lead to inconsistent phylogenies among different loci. Molecular evolution models failing to consider these complexities may lead to misleading species trees.

While a wealth of genomic data has deepened our understanding of the evolution of various organisms in the past two decades, more and more studies have shown that phylogenetic trees constructed from different data often contradict each other. With much effort being dedicated, it is now clear that both technical artifacts and biological processes may contribute to such phylogenetic incongruences. Phylogenetic inference methods often make various assumptions regarding the molecular data; for instance, homologous loci are properly identified and aligned, and different loci share the same evolutionary history. Recombination, incomplete lineage sorting (ILS), and gene flow are the biological processes leading to inconsistent phylogenies among different loci. Molecular evolution models failing to consider these complexities may lead to misleading species trees.

Introgression, the exchange of genetic material between lineages through hybridization and backcrossing, has been found to be more prevalent than expected. It has been documented in various organisms, including plants (H. Wang et al. 2024), fungi (Kinneberg et al. 2023), and vertebrates (Ge et al. 2023; Rancilhac et al. 2024). The introgressed alleles may contribute to speciation and adaptation (Ge et al. 2023; Racimo et al. 2015). Previous studies on introgression in spiders were mostly based on small datasets and limited to a few species of several families at a small scale (e.g., population) (e.g., Garb et al. 2009; Johannesen et al. 2001; Lattimore et al. 2011). For instance, it has been suggested that introgression and Andean uplift both contributed to the diversification of *Gasteracantha cancriformis* (Araneidae) (Salgado-Roa et al. 2018). Other examples include Ivanov et al. (2018) (Lycosidae), Responde et al. (2025) (Theridiidae), and Connolly et al. (2025) (Dolomedidae).

The mitochondrial genome (mitogenome), present in all eukaryotes, is a cellular organelle that has

evolved through intracellular symbiosis (Roger et al. 2017). It has been used in various fields such as taxonomy, phylogenetics, adaptation, and ecology (Biot-Pelletier et al. 2023; Lattimore et al. 2011; Tang et al. 2015). In recent years, there has been growing interest in understanding the architecture and evolutionary history of mitogenomes in various organisms and their interactions with nuclear genomes due to the increased accessibility of genomic data (Biot-Pelletier et al. 2023; Jakovlić et al. 2023). In contrast to other groups like insects or vertebrates that possess numerous publicly available mitogenome sequences (almost 30,000 insect mitogenomes with lengths longer than 10 Kbp on NCBI, access date: March 16, 2025) (Cameron 2024), only about 500 spider mitogenomes longer than 10 Kbp are publicly available on NCBI. This has limited our understanding of mitogenomes in spiders and their evolutionary significance.

1.4 Research questions

1.4.1 The knowledge gap in Pholcidae phylogeny

Pholcidae and Diguettidae form a sister group within Synspermiata (Fernández et al. 2018; Garrison et al. 2016). With much effort invested, Pholcidae is among the best (if not the best) studied of the major spider families. However, insufficient gene and taxon sampling have prevented us from fully understanding the higher ranks (subfamily, genus) of phylogenetic relationships within Pholcidae. In a recent study by Kulkarni, Wood et al. (2023), the authors sampled 24 pholcid species representing the five subfamilies. However, with only four species having UCE data (1 Ninetinae, 1 Modisiminae, 2 Pholcinae) and the remaining species having only Sanger data, their phylogeny had low ultrafast bootstrap values (21-93), and Pholcinae was paraphyletic.

The unification of Arteminae and Modisiminae has been suggested by a joint analysis of four genetic markers (*12S rRNA*, *16S rRNA*, *cytochrome c oxidase subunit I (COI)*, *28S rRNA*) and forty-five morphological traits (Bruvo-Madaric et al. 2005), or phylogenetic analysis based on seven or six genetic markers alone (*wingless*, *Histone H3* and *18S rRNA*, and the four markers just mentioned) (Dimitrov, J. J. Astrin et al. 2013; Eberle et al. 2018). Similarly, molecular data also supports the sister relationship between Pholcinae and Smeringopinae (Dimitrov, J. J. Astrin et al. 2013; Eberle et al. 2018). While the monophyly of Ninetinae is well-supported, several studies have proposed different relationships between Ninetinae and other subfamilies. Bruvo-Madaric et al. (2005) suggested that Ninetinae should be grouped with Pholcinae, while Dimitrov, J. J. Astrin et al. (2013) proposed a sister relationship to all other pholcids. The latter is supported by the presence of synspermia (collections of multiple merged sperm cells creating a syncytium) in most Ninetinae (but see B. A. Huber, Meng, Dederichs et al. (2024) and B. A. Huber, Meng, Valdez-Mondragón et al. (2023)), while other pholcids have cleistospermia (separate individual sperm cells, each encased in its own secretory layer) (B. A. Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023). Later, Eberle et al. 2018 placed Ninetinae together with the genus *Artema* and then as the sister group to remaining pholcids. The phylogenetic positions of both genus *Artema* Walckenaer, 1837 and *Priscula* Simon, 1893 are unstable in various analysis (B. A. Huber, Eberle and Dimitrov 2018). For instance, *Artema* was nested within Smeringopinae (Bruvo-Madaric et al. 2005), or grouped together with other Arteminae (Dimitrov, J. J. Astrin et al. 2013), or sistered to Ninetinae (Eberle et al. 2018). On the other hand, *Priscula* was sister to Modisiminae in some analyses but not in the others (J. J. Astrin et al. 2007; Eberle et al. 2018; B. A. Huber, Eberle and Dimitrov 2018).

Uncertainties in the placement of various Pholcid taxa hinder our ability to address important questions regarding the evolution of related traits and the geographic origins of different pholcids. For example, if the tiny-bodied and short-legged Ninetinae is sister to all other pholcids, then small body size and short

legs may be primitive character states for the entire family. What factors have driven most pholcids to evolve larger body sizes and their characteristically long legs? Therefore, the primary goal of this thesis is to fill this knowledge gap in Pholcidae phylogeny.

1.4.2 The role of body size and leg length evolution in Pholcidae diversification

While sexual selection and microhabitat have been shown to influence the diversification of Pholcidae (Dederichs et al. 2022; Eberle et al. 2018; B. A. Huber 1999, 2005b), the impacts of body size and leg length have yet to be tested. As noted earlier, both body size and leg length exhibit considerable variability among pholcids, with body size ranging from 1 mm to 11 mm and leg length from 1 mm to 180 mm. It has been shown that some spiders may disperse through ballooning using silk (Bell et al. 2005; Morley and Robert 2018), but a recent study challenges this idea for pholcids (B. A. Huber 2023), suggesting that other modes of locomotion such as walking and bridging are responsible for pholcid dispersal. Unlike previous studies that focus on the impact of absolute sizes of body and leg length on diversification, in this thesis, we hypothesize that the evolutionary flexibilities of morphological traits—measured by their variance and disparity—may be more crucial in driving species diversification.

The expression and evolution of many morphological traits are correlated both developmentally and genetically (Armbruster et al. 2014; Conner et al. 2014; Evans et al. 2021); thus, we expect this correlation to extend to their evolutionary flexibilities as well. Therefore, we will further investigate the evolution of allometric scaling (Bolstad et al. 2015) relationships between body size and leg length in pholcids. Specifically, we aim to answer the following questions: Are there any differences in the allometric parameters (i.e., slope and intercept) between species-rich and species-poor sister clades? Do the allometric parameters in different parameter spaces exhibit the same direction and strength of natural selection, and how does this correlate with evolutionary flexibility?

1.4.3 Revealing hidden pholcid diversity by extensive taxon and gene sampling

In the past 25 years, extensive sampling has significantly enhanced our understanding of Pholcidae biodiversity (Figure 1.3). The number of genera has increased from 65 to 97, and the number of species has grown from 586 to 2,028 since the year 2000, with an average of 58 new species being described per year (WSC 2025).

Incomplete and biased sampling can hinder our ability to discover new taxa (B. A. Huber, Meng, Dupérré, J. Astrin et al. 2023). For example, the species-poor subfamily Ninetinae is believed to be underrepresented, partly due to biased sampling practices. This bias may stem from the small to tiny body size of these spiders and the focus on specific geographical regions, often neglecting spiders that inhabit dry tropical regions and habitats (B. A. Huber and Meng 2025; B. A. Huber, Meng, Dederichs et al. 2024; B. A. Huber, Meng, garcía et al. 2024). During my PhD studies, we conducted extensive sampling across various regions, including North America, Central America, and South America, as well as from the Arabian Peninsula. This comprehensive sampling aimed to address the gaps in our understanding of pholcid diversity, particularly in underrepresented habitats.

Therefore, an additional key objective during my PhD period was to assist in species delimitation using molecular data and to incorporate these new taxa into the phylogeny of Pholcidae. In this thesis, we will exemplarily focus on two new species from the semi-arid Brazilian Caatinga, as well as on the phylogenetic affinities of the Chilean genus *Aucana*.

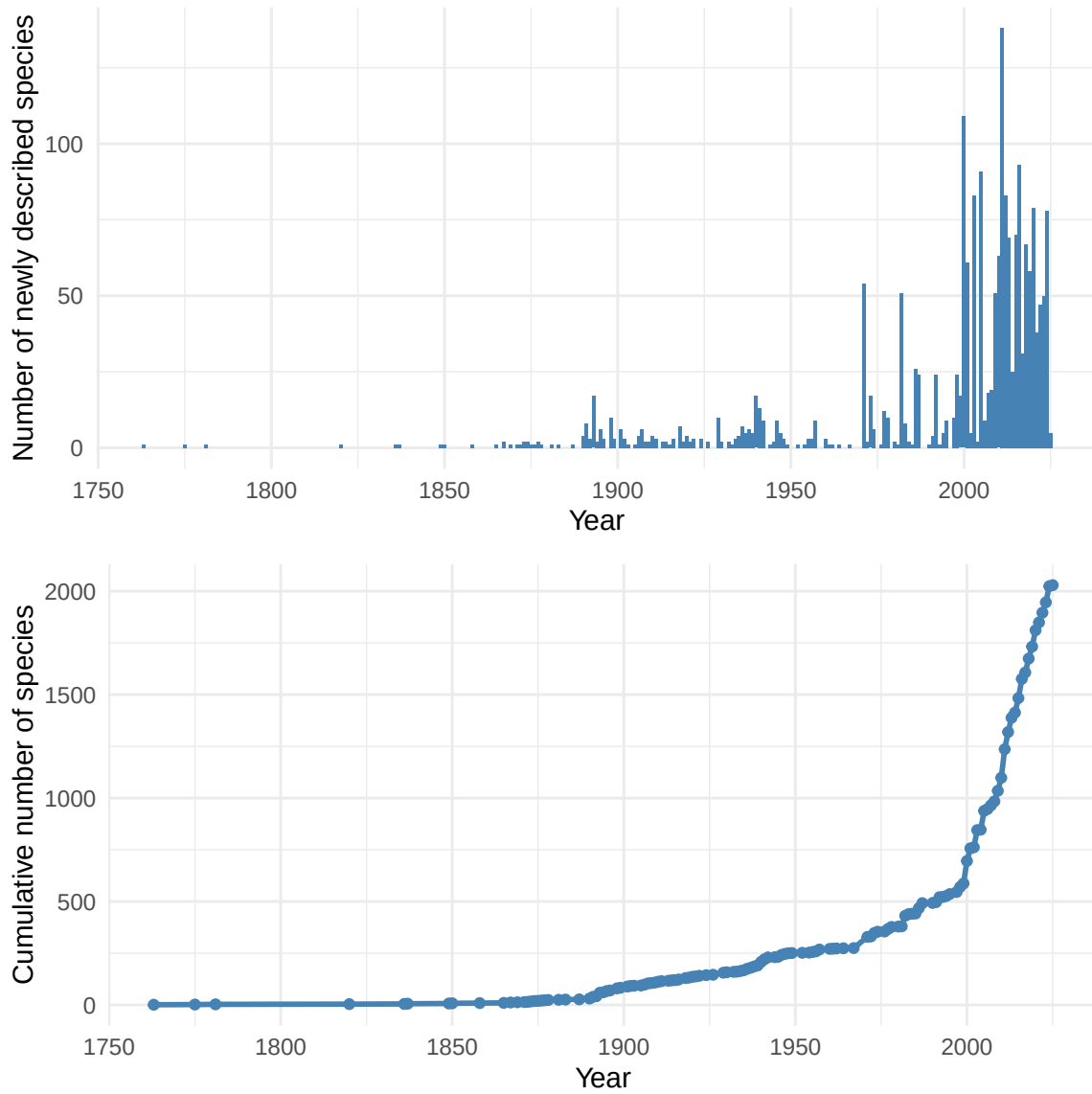


Figure 1.3. Pholcidae spider descriptions over the years. Based on data from WSC (2025).

1.4.4 Introgression and mitochondrial evolution of Pholcidae spiders

Previous studies on Pholcidae phylogenies often contradict each other in the placements of individual problematic groups such as Ninetinae, *Artema* and *Priscula*. To our best knowledge, there is no study addressing the prevalence of introgression using genomic data in Pholcidae yet. Since introgression is a process that may contribute to phylogenetic incongruence, in this thesis, I also aim to evaluate if introgression is present or absent, as well as the extent of it (i.e., involving which groups and the proportion of genomic loci) in the Pholcidae family.

While there are more than 2,000 formally described pholcid species (WSC 2025), there are only 4 mitogenomes with lengths longer than 10 Kbp on NCBI (accessed on March 16, 2025). Little is known about the architecture and evolution of pholcid mitogenomes. Therefore, I will also assemble mitogenomes from our generated genomic datasets, studying the gene rearrangement, selection pressure, and codon usage bias, as well as the phylogenetic relationships among pholcid mitogenomes.

1.5 Research methods

Since the emergence of high-throughput sequencing (HTS) two decades ago, molecular methods and bioinformatics have become essential tools in various fields of biology, particularly evolutionary biology and taxonomy. While many genome sequencing projects have recently been proposed (e.g., the Vertebrate Genomes Project (Rhie et al. 2021), the Earth BioGenome Project (Lewin et al. 2018)), reduced representation sequencing (RRS) offers a more cost-effective solution that allows for large-scale genomic and taxonomic sampling (Davey et al. 2011; Good 2011), playing an increasingly important role in taxonomy and elucidating the evolutionary history of species and their phylogenetic relationships (e.g., Misof et al. 2014; One Thousand Plant Transcriptomes Initiative 2019). However, these techniques can be challenging for many young researchers and taxonomists to learn due to their complexity. To address this gap in the field, I will present a comprehensive overview of the main steps and potential pitfalls associated with RRS methods and phylogenetic analysis as part of my thesis research to encourage more scientists to explore the exciting opportunities that exist within phylogenomics. As a consequence of this, we can better understand how these cutting-edge tools can help address our scientific questions and contribute to advances in evolutionary biology and taxonomy.

To overcome the limitation of having only a small number of loci available for pholcid phylogenetic analysis, we conducted the largest-scale genomic and taxonomic sampling of Pholcidae using two RRS methods: target capture sequencing (TCS) and low-coverage whole genome sequencing (lcWGS).

To investigate the relationship between pholcid diversification and the evolution of body size and leg length, I carried out a series of comparative analyses. These included various methods for modeling trait evolution, assessing phylogenetic signal, reconstructing ancestral states of traits, measuring the evolutionary flexibility and modeling the dynamics of speciation and extinction rates.

To assist with species identification and delimitation, I performed CO1 barcode amplification and sequencing, along with calculations of genetic distances, tree construction, and molecular species delimitation.

1.6 Thesis structure

While the current chapter (Chapter 1) provides an overview of the research background of my thesis, Chapter 2 presents the overview of molecular and bioinformatic methods used in phylogenomics. In Chapter 3, we resolve the subfamily-level phylogenetic relationships among pholcids, while Chapter 4 examines how the inter-specific flexibility of traits, allometry, and microhabitat shifts may be linked to pholcid diversification. Chapter 5 presents two new pholcid species that represent a new subfamily, based on molecular data. Finally, Chapter 6 investigate the introgression and the evolution of mitogenomes among pholcids.

1.7 Expected impact

Our work will enhance our understanding of Pholcidae biodiversity and the interplay between evolutionary flexibility, speciation, trait evolution, and allometry. Furthermore, our newly proposed phylogeny, based on extensive taxon and gene sampling, will not only serve as a framework for organizing the numerous species within the Pholcidae family but also provide a foundation for understanding the evolution of various traits in both Pholcidae and Araneae. Finally, the valuable nuclear and mitochondrial genomic data generated by this thesis will not only contribute to our understanding of the mitogenome evolution of pholcids but also support future studies exploring the evolution of other spider species, thereby expanding our knowledge of spider evolution.

References

- Aharon, S., B. A. Huber and E. Gavish-Regev (2017), *Daddy-long-leg giants: revision of the spider genus *Artema* Walckenaer, 1837 (Araneae, Pholcidae)*, Eur J Taxon **0**.
- Armbruster, W. S., C. Pelabon, G. H. Bolstad and T. F. Hansen (2014), *Integrated phenotypes: understanding trait covariation in plants and animals*, Philos Trans R Soc Lond B Biol Sci **369** 20130245.
- Astrin, J. J., B. Misof and B. A. Huber (2007), *The pitfalls of exaggeration: molecular and morphological evidence suggests *Kaliana* is a synonym of *Mesabolivar* (Araneae: Pholcidae)*, Zootaxa **1646** 17.
- Bell, J. R., D. A. Bohan, E. M. Shaw and G. S. Weyman (2005), *Ballooning dispersal using silk: world fauna, phylogenies, genetics and models*, Bull Entomol Res **95** 69.
- Bernatchez, L., A.-L. Ferchaud, C. S. Berger, C. J. Venney and A. Xuereb (2024), *Genomics for monitoring and understanding species responses to global climate change*, Nat Rev Genet **25** 165.
- Biot-Pelletier, D., S. Bettinazzi, I. Gagnon-Arsenault et al. (2023), *Evolutionary trajectories are contingent on mitonuclear interactions*, Mol Biol Evol **40** msad061.
- Bolstad, G. H., J. A. Cassara, E. Marquez et al. (2015), *Complex constraints on allometry revealed by artificial selection on the wing of *Drosophila melanogaster**, Proc Natl Acad Sci U S A **112** 13284.
- Bond, J. E. and B. D. Opell (1998), *Testing adaptive radiation and key innovation hypotheses in spiders*, Evolution **52** 403.

-
- Bond, J. E., N. L. Garrison, C. A. Hamilton et al. (2014), *Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution*, *Curr Biol* **24** 1765.
- Breitling, R. (2022), *On the taxonomic rank of the major subdivisions of the extant segmented spiders (Arachnida: Araneae: Mesothelae: Liphistiidae s. lat.)*, *Miscellanea Araneologica* **2022** 1.
- Brignoli, P. M. (1975), “Über die Gruppe der Haplogynae (Araneae)”, *Proceedings of the 6th International Arachnology Congress*, vol. 1974, Free University Amsterdam 33.
- Bristowe, W. S. (1939), *The comity of spiders*, vol. 1, London: Ray Society.
- Bristowe, W. S. (1958), *The World of Spiders*, London, Collins.
- Bromhall, C. (1987), *Spider heart-rates and locomotion*, *Journal of Comparative Physiology B* **157** 451.
- Bruvo-Madaric, B., B. A. Huber, A. Steinacher and G. Pass (2005), *Phylogeny of pholcid spiders (Araneae: Pholcidae): combined analysis using morphology and molecules*, *Mol Phylogenet Evol* **37** 661.
- Bush, G. L. (1993), *A reaffirmation of Santa Rosalia, or why are there so many kinds of small animals?*, *Evolutionary patterns and processes* **93** 229.
- Cadotte, M. W. (2006), *Dispersal and species diversity: a meta-analysis*, *Am Nat* **167** 913.
- Cameron, S. L. (2024), *Insect mitochondrial genomics: A decade of progress*, *Annu. Rev. Entomol.* **70** 83.
- Cardillo, M., G. M. Mace, K. E. Jones et al. (2005), *Multiple causes of high extinction risk in large mammal species*, *Science* **309** 1239.
- Caro, A., B. J. Gómez-Moliner and M. J. Madeira (2019), *Integrating multilocus DNA data and 3D geometric morphometrics to elucidate species boundaries in the case of Pyrenaearia (Pulmonata: Hygromiidae)*, *Mol Phylogenet Evol* **132** 194.
- Cerezer, F. O., C. S. Dambros, M. T. P. Coelho et al. (2023), *Accelerated body size evolution in upland environments is correlated with recent speciation in South American freshwater fishes*, *Nat Commun* **14** 6070.
- Claramunt, S., E. P. Derryberry, J. Remsen J. V. and R. T. Brumfield (2012), *High dispersal ability inhibits speciation in a continental radiation of passerine birds*, *Proc Biol Sci* **279** 1567.
- Cloudsley-Thompson, J. L. (1995), *A review of the anti-predator devices of spiders*, *Bull Br Arachnol Soc* **10** 81.
- Coddington, J. A. and H. W. Levi (1991), *Systematics and evolution of spiders (Araneae)*, *Annu Rev Ecol Syst* **22** 565.
- Coddington, J. A. (2005), *Phylogeny and classification of spiders*, American Arachnology Society.
- Conner, J. K., I. A. Cooper, R. J. La Rosa, S. G. Pérez and A. M. Royer (2014), *Patterns of phenotypic correlations among morphological traits across plants and animals*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **369** 20130246.
- Connolly, S. J., K. M. Curtis, C. J. Vink and C. J. Painting (2025), *Mating behaviour influences the direction and geographic extent of introgression in New Zealand fishing spiders (Dolomedes)*, *J Evol Biol* **38** 291.
- Cooney, C. R. and G. H. Thomas (2021), *Heterogeneous relationships between rates of speciation and body size evolution across vertebrate clades*, *Nat Ecol Evol* **5** 101.

- Cowie, R. H., P. Bouchet and B. Fontaine (2022),
The Sixth Mass Extinction: fact, fiction or speculation?, Biological Reviews **97** 640.
- Crouch, N. M. A. and R. E. Ricklefs (2019),
Speciation Rate Is Independent of the Rate of Evolution of Morphological Size, Shape, and Absolute Morphological Specialization in a Large Clade of Birds, Am Nat **193** E78.
- Czekanski-Moir, J. E. and R. J. Rundell (2019),
The Ecology of Nonecological Speciation and Nonadaptive Radiations, Trends Ecol Evol **34** 400.
- Darwin, C. (1859), *On the origin of species*, London: Jonh Murray.
- Davey, J. W., P. A. Hohenlohe, P. D. Etter et al. (2011),
Genome-wide genetic marker discovery and genotyping using next-generation sequencing, Nat Rev Genet **12** 499.
- Dederichs, T. M., B. A. Huber and P. Michalik (2022),
Evolutionary morphology of sperm in pholcid spiders (Pholcidae, Synspermiata), BMC Zoology **7** 52.
- Dimitrov, D., J. J. Astrin and B. A. Huber (2013), *Pholcid spider molecular systematics revisited, with new insights into the biogeography and the evolution of the group*, Cladistics **29** 132.
- Dimitrov, D., L. R. Benavides, M. A. Arnedo et al. (2017), *Rounding up the usual suspects: a standard target-gene approach for resolving the interfamilial phylogenetic relationships of ecribellate orb-weaving spiders with a new family-rank classification (Araneae, Araneoidea)*, Cladistics **33** 221.
- Dimitrov, D. and G. Hormiga (2021), *Spider diversification through space and time*, Annu. Rev. Entomol. **66** 225.
- Dong, Y., J. A. Harvey, R. Steegh, R. Gols and M. Rowe (2023),
The role of male body size in mating success and male–male competition in a false widow spider, Anim Behav **206** 53.
- Dunlop, J. A. (2010), *Geological history and phylogeny of Chelicerata*, Arthropod Struct Dev **39** 124.
- Eberle, J., D. Dimitrov, A. Valdez-Mondragon and B. A. Huber (2018),
Microhabitat change drives diversification in pholcid spiders, BMC Evol Biol **18** 141.
- Entling, W., M. H. Schmidt-Entling, S. Bacher, R. Brandl and W. Nentwig (2010),
Body size–climate relationships of European spiders, J Biogeogr **37** 477.
- Etienne, R. S., S. N. de Visser, T. Janzen et al. (2012),
Can clade age alone explain the relationship between body size and diversity?, Interface Focus **2** 170.
- Evans, K. M., O. Larouche, S.-J. Watson et al. (2021),
Integration drives rapid phenotypic evolution in flatfishes, Proc Natl Acad Sci U S A **118**.
- Fernández, R., R. J. Kallal, D. Dimitrov et al. (2018), *Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life*, Curr Biol **28** 1489.
- Foelix, R. F. (1970), *Structure and function of tarsal sensilla in the spider Araneus diadematus*, J Exp Zool **175** 99.
- Foelix, R. F. (2010), *Biology of spiders*, 3rd, Oxford university press.
- Garb, J. E. and R. G. Gillespie (2009), *Diversity despite dispersal: colonization history and phylogeography of Hawaiian crab spiders inferred from multilocus genetic data*, Mol Ecol **18** 1746.
- Garrison, N. L., J. Rodriguez, I. Agnarsson et al. (2016),
Spider phylogenomics: untangling the Spider Tree of Life, PeerJ **4** e1719.

-
- Gause, G. F. (1932),
Experimental studies on the struggle for existence: I. Mixed population of two species of yeast,
J Exp Biol **9** 389.
- Ge, D., Z. Wen, A. Feijó et al. (2023), *Genomic consequences of and demographic response to pervasive hybridization over time in climate-sensitive pikas*, Mol Biol Evol **40** msac274.
- Gillet, A., B. Frederich and E. Parmentier (2019), *Divergent evolutionary morphology of the axial skeleton as a potential key innovation in modern cetaceans*, Proc Biol Sci **286** 20191771.
- Glaudas, X., S. E. Rice, R. W. Clark and G. J. Alexander (2020), *The intensity of sexual selection, body size and reproductive success in a mating system with male–male combat: is bigger better?*, Oikos **129** 998.
- Glick, P. A. (1939), *The distribution of insects, spiders, and mites in the air*.
- Good, J. M. (2011),
Reduced representation methods for subgenomic enrichment and next-generation sequencing,
Molecular methods for evolutionary genetics **772** 85.
- Griethuijsen, L. I. van and B. A. Trimmer (2014), *Locomotion in caterpillars*,
Biol Rev Camb Philos Soc **89** 656.
- Griswold, C. E., J. A. Coddington, G. Hormiga and N. Scharff (1998),
Phylogeny of the orb-web building spiders (Araneae, Orbiculariae: Deinopoidea, Araneoidea),
Zool J Linn Soc **123** 1.
- Grossi, B., C. Veloso, A. Taucare-Ríos and M. Canals (2016),
Allometry of locomotor organs and sexual size dimorphism in the mygalomorph spider Grammostola rosea (Walckenaer, 1837)(Araneae, Theraphosidae), J Arachnol **44** 99.
- Hadfield, J. D. (2008), *Estimating evolutionary parameters when viability selection is operating*,
Proceedings of the Royal Society B: Biological Sciences **275** 723.
- Head, G. (1995), *Selection on Fecundity and Variation in the Degree of Sexual Size Dimorphism Among Spider Species (Class Araneae)*, Evolution **49** 776.
- Henao Diaz, L. F., L. J. Harmon, M. T. C. Sugawara, E. T. Miller and M. W. Pennell (2019),
Macroevolutionary diversification rates show time dependency,
Proc Natl Acad Sci U S A **116** 7403.
- Hennig, W. (1999), *Phylogenetic systematics*, University of Illinois Press.
- Hoekstra, H. E. and J. A. Coyne (2007),
The locus of evolution: evo devo and the genetics of adaptation, Evolution **61** 995.
- Holmes, P., R. J. Full, D. Koditschek and J. Guckenheimer (2006),
The Dynamics of Legged Locomotion: Models, Analyses, and Challenges, SIAM Review **48** 207.
- Huber, B. A. (1993), *Genital morphology, copulatory mechanism and reproductive biology in Psilochorus simoni (Berland, 1911)(Pholcidae; Araneae)*, Netherlands Journal of Zoology **44** 85.
- Huber, B. A. (1995), *Copulatory mechanism in Holocnemus pluchei and Pholcus opilionoides, with notes on male chelicerapophyses and stridulatory organs in Pholcidae (Araneae)*,
Acta Zoologica **76** 291.
- Huber, B. A. (1996), *On American ‘Micromerys’ and Metagonia (Araneae, Pholcidae), with notes on natural history and genital mechanics*, Zool Scr **25** 341.
- Huber, B. A. (1998), *Genital mechanics in some neotropical pholcid spiders (Araneae: Pholcidae), with implications for systematics*, J Zool (1987) **244** 587.

- Huber, B. A. (1999),
Sexual selection in pholcid spiders (Araneae, Pholcidae): artful chelicerae and forceful genitalia,
 J Arachnol **27** 135.
- Huber, B. A. (2000), *New World pholcid spiders (Araneae: Pholcidae): a revision at generic level*,
 Bull Am Mus Nat Hist **2000** 1.
- Huber, B. A. (2002a), *Functional morphology of the genitalia in the spider *Spermophora senoculata* (Pholcidae, Araneae)*, Zool Anz **241** 105.
- Huber, B. A. (2002b), *Ninetis russellsmithi n. sp., an unusual new pholcid spider species from Malawi (Araneae: Pholcidae)*, Journal of Insect Science **2** 4.
- Huber, B. A. (2003), *High species diversity in one of the dominant groups of spiders in East African montane forests (Araneae: Pholcidae: Buitinga n. gen., Spermophora Hentz)*,
 Zool J Linn Soc **137** 555.
- Huber, B. A. (2004), *Evidence for functional segregation in the directionally asymmetric male genitalia of the spider *Metagonia mariguaitarensis* (González-Sponga) (Pholcidae: Araneae)*,
 Journal of Zoology **262** 317.
- Huber, B. A. (2005a), *Revision of the genus *Spermophora* Hentz in Southeast Asia and on the Pacific Islands, with descriptions of three new genera (Araneae: Pholcidae)*, Zool Meded **79** 61.
- Huber, B. A. (2005b), *Sexual selection research on spiders: progress and biases*,
 Biological Reviews **80** 363.
- Huber, B. A. (2011a), *Phylogeny and classification of Pholcidae (Araneae): an update*,
 J Arachnol **39** 211.
- Huber, B. A. (2011b),
*Revision and cladistic analysis of *Pholcus* and closely related taxa (Araneae, Pholcidae)*,
 Bonn Zool Monogr **58** 1.
- Huber, B. A. (2012), *Revision and cladistic analysis of the Afrotropical endemic genus *Smeringopus* Simon, 1890 (Araneae: Pholcidae)*, Zootaxa **3461** 1.
- Huber, B. A. (2013), *Revision and cladistic analysis of the Guineo-Congolian spider genus *Smeringopina* Kraus (Araneae, Pholcidae)*, Zootaxa **3713** 1.
- Huber, B. A. (2015), *Small scale endemism in Brazil's Atlantic Forest: 14 new species of *Mesabolivar* (Araneae, Pholcidae), each known from a single locality.*, Zootaxa **3942** 1.
- Huber, B. A. (2018), *Cave-dwelling pholcid spiders (Araneae, Pholcidae): a review*,
 Subterr Biol **26** 1.
- Huber, B. A. (2020), *Revision of the spider genus *Hoplopholcus* Kulczyński (Araneae, Pholcidae).*,
 Zootaxa **4726** 1.
- Huber, B. A. (2023), *Do pholcid spiders balloon?*, Arachnology **19** 885.
- Huber, B. A. and J. J. Astrin (2009), *Increased sampling blurs morphological and molecular species limits: revision of the Hispaniolan endemic spider genus *Tainonia* (Araneae: Pholcidae)*,
 Invertebr Syst **23** 281.
- Huber, B. A. and L. S. Carvalho (2019), *Filling the gaps: descriptions of unnamed species included in the latest molecular phylogeny of Pholcidae (Araneae)*, Zootaxa **4546** 1.
- Huber, B. A., L. S. Carvalho and S. P. Benjamin (2014),
*On the New World spiders previously misplaced in *Leptopholcus*: molecular and morphological analyses and descriptions of four new species (Araneae: Pholcidae)*, Invertebr Syst **28** 432.
- Huber, B. A., C. L. Deeleman-Reinhold and A. Pérez González (1999), *The spider genus *Crossopriza* (Araneae, Pholcidae) in the New World. American Museum novitates; no. 3262.*

-
- Huber, B. A. and W. G. Eberhard (1997),
Courtship, copulation, and genital mechanics in Physocyclus globosus (Araneae, Pholcidae),
Can J Zool **75** 905.
- Huber, B. A. and J. Eberle (2021),
Mining a photo library: Eggs and egg sacs in a major spider family, Invertebr Biol **140** e12349.
- Huber, B. A., J. Eberle and D. Dimitrov (2018), *The phylogeny of pholcid spiders: a critical evaluation of relationships suggested by molecular data (Araneae, Pholcidae)*, Zookeys **51**.
- Huber, B. A. and H. El Hennawy (2007), *On Old World ninetine spiders (Araneae: Pholcidae), with a new genus and species and the first record for Madagascar*, Zootaxa **1635** 45.
- Huber, B. A. and N. Fleckenstein (2008),
Comb-hairs on the fourth tarsi in pholcid spiders (Araneae, Pholcidae), J Arachnol **36** 232.
- Huber, B. A. and A. van Harten (2001),
Ninetis subtilissima Simon, 1890 (Araneae: Pholcidae): Redescription and SEM Ultrastructure,
Am Mus Novit **2001** 1.
- Huber, B. A. and **G. Meng** (2023),
On the mysterious Seychellois endemic spider genus Cenemus (Araneae, Pholcidae),
Arthropod Syst Phylogeny **81** 179, doi: [10.3897/asp.81.e86793](https://doi.org/10.3897/asp.81.e86793).
- Huber, B. A. and **G. Meng** (2024), *Old World Micropholcus spiders, with first records of acrocerid parasitoids in Pholcidae (Araneae)*, ZooKeys **1213** 95,
doi: [10.3897/zookeys.1213.133178](https://doi.org/10.3897/zookeys.1213.133178).
- Huber, B. A. and **G. Meng** (2025),
Like grains of sand: Ninetis spiders on the Arabian Peninsula (Araneae, Pholcidae),
Zootaxa **5563** 290, doi: [10.11646/zootaxa.5563.1.19](https://doi.org/10.11646/zootaxa.5563.1.19).
- Huber, B. A., **G. Meng**, A. E. Acurio et al. (2022), *Metagonia spiders of Galápagos: blind cave-dwellers and their epigean relatives (Araneae, Pholcidae)*, Invertebr Syst **36** 647,
doi: [10.1071/IS21082](https://doi.org/10.1071/IS21082).
- Huber, B. A., **G. Meng**, H. L. Clark and G. Cazanove (2023),
First blind daddy long-legs spiders from Australia and Réunion (Araneae, Pholcidae),
Subterr Biol **46** 1, doi: [10.3897/subtbiol.46.105798](https://doi.org/10.3897/subtbiol.46.105798).
- Huber, B. A., **G. Meng**, T. M. Dederichs et al. (2024),
Castaways: the Leeward Antilles endemic spider genus Papiamenta (Araneae: Pholcidae),
Invertebr Syst **38** IS23052, doi: [10.1071/IS23052](https://doi.org/10.1071/IS23052).
- Huber, B. A., **G. Meng**, N. Dupérré, J. Astrin and M. Herrera (2023), *Andean giants: Priscula spiders from Ecuador, with notes on species groups and egg-sac troglomorphism (Araneae: Pholcidae)*,
Eur J Taxon **909** 1, doi: [10.5852/ejt.2023.909.2351](https://doi.org/10.5852/ejt.2023.909.2351).
- Huber, B. A., **G. Meng**, N. Dupérré, M. Herrera et al. (2023), *Humpback spiders from Ecuador: relationships, prosoma 'inflation' and genital asymmetry (Araneae: Pholcidae: Mecolaesthus)*,
Invertebr Syst **37** 117, doi: [10.1071/IS22052](https://doi.org/10.1071/IS22052).
- Huber, B. A., **G. Meng**, J. C. garcía and L. S. Carvalho (2024),
Thriving in dry conditions: on the Neotropical spider genus Galapa (Araneae: Pholcidae),
Zootaxa **5419** 301, doi: [10.11646/zootaxa.5419.3.1](https://doi.org/10.11646/zootaxa.5419.3.1).
- Huber, B. A., **G. Meng**, J. Král, I. M. Ávila Herrera, M. A. Izquierdo and L. S. Carvalho (2023),
High and dry: integrative taxonomy of the Andean spider genus Nerudia (Araneae: Pholcidae),
Zool J Linn Soc **198** 534, doi: [10.1093/zoolinnean/zlac100](https://doi.org/10.1093/zoolinnean/zlac100).

- Huber, B. A., **G. Meng**, J. Král, I. M. Ávila Herrera and M. A. Izquierdo (2023), *Revision of the South American Ninetinae genus Guaranita (Araneae, Pholcidae)*, Eur J Taxon **900** 32, DOI: [10.5852/ejt.2023.900.2301](https://doi.org/10.5852/ejt.2023.900.2301).
- Huber, B. A., **G. Meng**, J. Král, I. M. Á. Herrera and L. S. Carvalho (2024), *Diamonds in the rough: Ibotyporanga (Araneae, Pholcidae) spiders in semi-arid Neotropical environments*, Eur J Taxon **963** 1, DOI: [10.5852/ejt.2024.963.2687](https://doi.org/10.5852/ejt.2024.963.2687).
- Huber, B. A., **G. Meng**, A. Valdez-Mondragón et al. (2023), *Short-legged daddy-long-leg spiders in North America: the genera Pholcophora and Tolteca (Araneae, Pholcidae)*, Eur J Taxon **880** 1, DOI: [10.5852/ejt.2023.880.2173](https://doi.org/10.5852/ejt.2023.880.2173).
- Huber, B. A., J. Neumann, A. Grabolle and V. Hula (2017), *Aliens in Europe: updates on the distributions of Modisimus culicinus and Micropholcus fauroti (Araneae, Pholcidae)*, Arachnol Mitt **0** 12.
- Huber, B. A. and O. M. Nuñeza (2015), *Evolution of genital asymmetry, exaggerated eye stalks, and extreme palpal elongation in Panjange spiders (Araneae: Pholcidae)*, Eur J Taxon.
- Huber, B. A. and C. A. Rheims (2011), *Diversity and endemism of pholcid spiders in Brazil's Atlantic Forest, with descriptions of four new species of the Atlantic Forest endemic genus Tupigea (Araneae: Pholcidae)*, J Nat Hist **45** 275.
- Huber, B. A. and F. Schaller (2005), *High species diversity, male-female coevolution, and metaphyly in Southeast Asian pholcid spiders: the case of Belisana Thorell 1898 (Araneae, Pholcidae)*, vol. 155, Schweizerbart 1.
- Ivanov, V., K. M. Lee and M. Mutanen (2018), *Mitochondrial discordance in wolf spiders: Genomic evidence for species integrity and introgression*, Mol Ecol **27** 1681.
- Jakovlić, I., H. Zou, T. Ye et al. (2023), *Mitogenomic evolutionary rates in bilateria are influenced by parasitic lifestyle and locomotory capacity*, Nat Commun **14** 6307.
- Janicke, T. and S. Fromenteil (2021), *Sexual selection and sexual size dimorphism in animals*, Biol Lett **17** 20210251.
- Jebb, A. H. M., D. T. Blumstein, P. Bize and J. G. A. Martin (2021), *Bigger is not always better: viability selection on body mass varies across life stages in a hibernating mammal*, Ecol Evol **11** 3435.
- Johannessen, J. and M. Veith (2001), *Population history of Eresus cinnabarinus (Araneae: Eresidae) colour variants at a putative species transition*, Heredity **87** 114.
- Johnson, N. F., H. Chen and B. A. Huber (2018), *New species of Idris Förster (Hymenoptera, Platygastroidea) from southeast Asia, parasitoids of the eggs of pholcid spiders (Araneae, Pholcidae)*, ZooKeys **811** 65.
- Kaspari, M. and M. D. Weiser (1999), *The size–grain hypothesis and interspecific scaling in ants*, Functional ecology **13** 530.
- Kaston, B. J. (1964), *The evolution of spider webs*, Am Zool **4** 191.
- Kaston, B. J. (1978), *How to know the spiders*, Ed. 3.
- Kinneberg, V. B., D. S. Lü, D. Peris, M. Ravinet and I. Skrede (2023), *Introgression between highly divergent fungal sister species*, J Evol Biol **36** 1133.
- Kulkarni, S., R. J. Kallal, H. Wood et al. (2021), *Interrogating Genomic-Scale Data to Resolve Recalcitrant Nodes in the Spider Tree of Life*, Mol Biol Evol **38** 891.
- Kulkarni, S., H. M. Wood and G. Hormiga (2023), *Advances in the reconstruction of the Spider Tree of Life: a roadmap for spider systematics and comparative studies*, Cladistics **39** 479.

-
- LaBarbera, M. (1989), *Analyzing body size as a factor in ecology and evolution*, Annu Rev Ecol Syst **20** 97.
- Labonte, D. and W. Federle (2013), *Functionally different pads on the same foot allow control of attachment: stick insects have load-sensitive "heel" pads for friction and shear-sensitive "toe" pads for adhesion*, PloS One **8** e81943.
- Lattimore, V. L., C. J. Vink, A. M. Paterson and R. H. Cruickshank (2011), *Unidirectional introgression within the genus Dolomedes (Araneae: Pisauridae) in southern New Zealand*, Invertebr Syst **25** 70.
- Lee, M. S. Y., A. Cau, D. Naish and G. J. Dyke (2014), *Dinosaur evolution. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds*, Science **345** 562.
- Lehtinen, P. T. (1975), "Notes on the phylogenetic classification of Araneae", *Proceedings of the 6th International Arachnological Congress*, vol. 1974, Free University Amsterdam 26.
- Lewin, H. A., G. E. Robinson, W. J. Kress et al. (2018), *Earth BioGenome Project: Sequencing life for the future of life*, Proc Natl Acad Sci U S A **115** 4325.
- Li, C., X. Lian, J. Bi et al. (2011), *Effects of sand grain size and morphological traits on running speed of toad-headed lizard Phrynocephalus frontalis*, J Arid Environ **75** 1038.
- Li, S. (2022), *On the taxonomy of spiders of the suborder Mesothelae*, Zhuxing Xuebao **31** 71.
- Li, X., Y. Zhang, M. Yan et al. (2024), *New taxa of extant spiders (Araneae) from the world in 2023*, Biodiversity Science **32** 24181.
- Luo, A., C. Ling, S. Y. Ho and C.-D. Zhu (2018), *Comparison of methods for molecular species delimitation across a range of speciation scenarios*, Syst Biol **67** 830.
- Magalhaes, I. L. F., G. H. Azevedo, P. Michalik and M. J. Ramírez (2020), *The fossil record of spiders revisited: implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic*, Biological Reviews **95** 184.
- Michalik, P. and M. J. Ramírez (2014), *Evolutionary morphology of the male reproductive system, spermatozoa and seminal fluid of spiders (Araneae, Arachnida)—Current knowledge and future directions*, Arthropod Struct Dev **43** 291.
- Misof, B., S. Liu, K. Meusemann et al. (2014), *Phylogenomics resolves the timing and pattern of insect evolution*, Science **346** 763.
- Morley, E. L. and P. W. Gorham (2020), *Evidence for nanocoulomb charges on spider ballooning silk*, Physical Review E **102** 012403.
- Morley, E. L. and D. Robert (2018), *Electric fields elicit ballooning in spiders*, Curr Biol **28** 2324.
- Moya-Larano, J., D. Vinković, E. De Mas, G. Corcobado and E. Moreno (2008), *Morphological evolution of spiders predicted by pendulum mechanics*, PloS One **3** e1841.
- Mutanen, M. and E. Pretorius (2007), *Subjective visual evaluation vs. traditional and geometric morphometrics in species delimitation: a comparison of moth genitalia*, Syst Entomol **32** 371.
- Myrmidonella, N. and S. Pholcus (1998), *On the "valve" in the genitalia of female pholcids (Pholcidae, Araneae)*, Bull. Br. Arachnol. Soc **11** 41.

- Narimanov, N., D. Bonte, P. Mason, L. Mestre and M. H. Entling (2021), *Disentangling the roles of electric fields and wind in spider dispersal experiments*, J Arachnol **49** 380.
- Nentwig, W., J. Ansorg, A. Bolzern et al. (2022), *All you need to know about spiders*, Springer Nature.
- Nosil, P. (2012), *Ecological speciation*, Oxford University Press.
- Nosil, P. and S. M. Flaxman (2011), *Conditions for mutation-order speciation*, Proceedings of the Royal Society B: Biological Sciences **278** 399.
- Nosil, P., L. J. Harmon and O. Seehausen (2009), *Ecological explanations for (incomplete) speciation*, Trends Ecol Evol **24** 145.
- Nyffeler, M. and K. Birkhofer (2017), *An estimated 400–800 million tons of prey are annually killed by the global spider community*, The Science of nature **104** 1.
- Oliveira, H. F. M., G. Freire-Jr and F. M. C. Bicalho Domingos (2024), *Digest: Sexually selected traits can affect individual fitness and trait evolution in a butterfly species*.
- One Thousand Plant Transcriptomes Initiative (2019), *One thousand plant transcriptomes and the phylogenomics of green plants*, Nature **574** 679.
- Opatova, V., C. A. Hamilton, M. Hedin et al. (2020), *Phylogenetic systematics and evolution of the spider infraorder Mygalomorphae using genomic scale data*, Syst Biol **69** 671.
- Otto, S. P., M. F. Scott and S. Immler (2015), *Evolution of haploid selection in predominantly diploid organisms*, Proc Natl Acad Sci U S A **112** 15952.
- Pechmann, M., S. Khadjeh, F. Sprenger and N.-M. Prpic (2010), *Patterning mechanisms and morphological diversity of spider appendages and their importance for spider evolution*, Arthropod Struct Dev **39** 453.
- Platnick, N. I. (1975), “On the validity of Haplogynae as a taxonomic grouping in spiders”, *Proceedings of the 6th International Arachnology Congress, Vrije Universiteit, Amsterdam, The Netherlands* 30.
- Platnick, N. I. (2020), *Spiders of the world: a natural history*, Princeton University Press.
- Platnick, N. I., J. A. Coddington, R. R. Forster and C. E. Griswold (1991), *Spinneret morphology and the phylogeny of haplogyne spiders (Araneae, Araneomorphae)*, Am Mus Novit **3016** 1.
- Platnick, N. I. and W. J. Gertsch (1976), *The suborders of spiders: a cladistic analysis (Arachnida, Araneae)*, **2607**.
- Pontzer, H. (2007), *Effective limb length and the scaling of locomotor cost in terrestrial animals*, J Exp Biol **210** 1752.
- Puillandre, N., S. Brouillet and G. Achaz (2021), *ASAP: assemble species by automatic partitioning*, Mol Ecol Resour **21** 609.
- Puillandre, N., A. Lambert, S. Brouillet and G. Achaz (2012), *ABGD, Automatic Barcode Gap Discovery for primary species delimitation*, Mol Ecol **21** 1864.
- Purvis, A., J. L. Gittleman, G. Cowlishaw and G. M. Mace (2000), *Predicting extinction risk in declining species*, Proceedings of the royal society of London. Series B: Biological Sciences **267** 1947.
- Rabosky, D. L. (2013), *Diversity-Dependence, Ecological Speciation, and the Role of Competition in Macroevolution*, Annu Rev Ecol Evol Syst **44** 481.

-
- Racimo, F., S. Sankararaman, R. Nielsen and E. Huerta-Sánchez (2015),
Evidence for archaic adaptive introgression in humans, Nat Rev Genet **16** 359.
- Ramírez, M. J., I. L. Magalhaes, S. Derkarabetian et al. (2021), *Sequence capture phylogenomics of true spiders reveals convergent evolution of respiratory systems*, Syst Biol **70** 14.
- Ramírez, M. J. (2000), *Respiratory system morphology and the phylogeny of haplogyne spiders (Araneae, Araneomorphae)*, J Arachnol **28** 149.
- Ramírez, M. J. (2014),
The morphology and phylogeny of dionychan spiders (Araneae: Araneomorphae), Bull Am Mus Nat Hist **2014** 1.
- Rancilhac, L., E. D. Enbody, R. Harris et al. (2024), *Introgression underlies phylogenetic uncertainty but not parallel plumage evolution in a recent songbird radiation*, Syst Biol **73** 12.
- Responte, M. A., C.-Y. Wu, N. U. Elias et al. (2025), *Recent Range Expansion and Genomic Admixture in a Kleptoparasitic Spider, Argyrodes lanyuensis: A Case of Adaptive Introgression on Small, Isolated Islands of the Taiwan–Philippine Transition Zone?*, Mol Ecol **34** e17630.
- Rhie, A., S. A. McCarthy, O. Fedrigo et al. (2021),
Towards complete and error-free genome assemblies of all vertebrate species, Nature **592** 737.
- Rocha, P. N. and F. M. Gawryszewski (2024),
Foraging strategy as a route for sexual size dimorphism evolution, Ecol Evol **14** e70100.
- Roger, A. J., S. A. Muñoz-Gómez and R. Kamikawa (2017),
The Origin and Diversification of Mitochondria, Curr Biol **27** R1177.
- Rose, S. (2022), *Spiders of North America*, vol. 154, Princeton University Press.
- Rowsey, D. M., R. M. Keenan and S. A. Jansa (2020), *Dietary morphology of two island-endemic murid rodent clades is consistent with persistent, incumbent-imposed competitive interactions*, Proceedings of the Royal Society B **287** 20192746.
- Rundell, R. J. and T. D. Price (2009),
Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation, Trends Ecol Evol **24** 394.
- Saaristo, M. I. (2001), *Pholcid spiders of the granitic Seychelles (Araneae, Pholcidae)*, Phelsuma **9** 9.
- Salgado-Roa, F. C., C. Pardo-Díaz, E. Lasso et al. (2018), *Gene flow and Andean uplift shape the diversification of Gasteracantha cancriformis (Araneae: Araneidae) in Northern South America*, Ecol Evol **8** 7131.
- Schäfer, M. A., A. Hille and G. B. Uhl (2001), *Geographical patterns of genetic subdivision in the cellar spider Pholcus phalangioides (Araneae)*, Heredity **86** 94.
- Schmitz, A. (2012), “Tracheae in spiders: respiratory organs for special functions”,
Spider ecophysiology, Springer 29.
- Scott, M. F., C. Mackintosh and S. Immler (2024), *Gametic selection favours polyandry and selfing*, PLoS Genet **20** e1010660.
- Sensenig, A., I. Agnarsson and T. A. Blackledge (2010),
Behavioural and biomaterial coevolution in spider orb webs, J Evol Biol **23** 1839.
- Sensenig, A., I. Agnarsson and T. A. Blackledge (2011), *Adult spiders use tougher silk: Ontogenetic changes in web architecture and silk biomechanics in the orb-weaver spider*, J Zool (1987) **285** 28.
- Shao, L., Z. Zhao and S. Li (2023),
Is Phenotypic Evolution Affected by Spiders’ Construction Behaviors?, Syst Biol **72** 319.

- Short, A. E. Z. and M. Fikáček (2013),
Molecular phylogeny, evolution and classification of the Hydrophilidae (Coleoptera),
 Syst Entomol **38** 723.
- Silva, V., C. Biancardi, C. Perafan et al. (2021),
Giant steps: adhesion and locomotion in theraphosid tarantulas,
 J Comp Physiol A Neuroethol Sens Neural Behav Physiol **207** 179.
- Simon, E. (1893), *Histoire Naturelle des Araignées*, 2nd, vol. 1, 2, Paris: Roret 256.
- Siqueira, A. C., R. A. Morais, D. R. Bellwood and P. F. Cowman (2020),
Trophic innovations fuel reef fish diversification, Nat Commun **11** 2669.
- Smith, A. N. and M. C. Belk (2018), *Does body size affect fitness the same way in males and females?*
A test of multiple fitness components, Biol J Linn Soc Lond **124** 47.
- Tang, M., C. J. Hardman, Y. Ji et al. (2015),
High-throughput monitoring of wild bee diversity and abundance via mitogenomics,
 Methods Ecol Evol **6** 1034.
- Taylor, C. R., N. C. Heglund, T. A. McMahon and T. R. Looney (1980), *Energetic Cost of Generating*
Muscular Force During Running: A Comparison of Large and Small Animals, J Exp Biol **86** 9.
- Tong, Y. and S. Li (2008),
Four new species of six-eyed pholcid spiders (Araneae: Pholcidae) from Hainan Island, China,
 The Raffles Bulletin of Zoology **56** 45.
- Tregenza, T. and N. Wedell (2000),
Genetic compatibility, mate choice and patterns of parentage: invited review, Mol Ecol **9** 1013.
- Ubick, D., P. Cushing, V. Roth et al., eds. (2005), *Spiders of North America: an identification manual*,
 American Arachnological Society 194.
- Uhl, G., B. Huber, W. Rose et al. (1995), *Male pedipalp morphology and copulatory mechanism in*
Pholcus phalangioides (Fuesslin, 1775)(Araneae, Pholcidae), Bull Br Arachnol Soc **10** 1.
- Van Valen, L. (1973), *Body size and numbers of plants and animals*, Evolution **27** 27.
- Venail, P. A., R. C. MacLean, T. Bouvier et al. (2008),
Diversity and productivity peak at intermediate dispersal rate in evolving metacommunities,
 Nature **452** 210.
- Wadgymar, S. M., S. C. Daws and J. T. Anderson (2017),
Integrating viability and fecundity selection to illuminate the adaptive nature of genetic clines,
 Evol Lett **1** 26.
- Wägele, J.-W. (2005), *Phylogenetic systematics*, Verlag Dr. Friedrich Pfeil, München, Germany **3** 9.
- Wagner, D. L., E. M. Grames, M. L. Forister, M. R. Berenbaum and D. Stopak (2021),
Insect decline in the Anthropocene: Death by a thousand cuts,
 Proceedings of the National Academy of Sciences **118** e2023989118.
- Wang, H., W. Zhang, Y. Yu et al. (2024), *Biased gene introgression and adaptation in the face of*
chloroplast capture in Aquilegia amurensis, Syst Biol **73** 886.
- Wells, M. S. (1988),
Effects of body size and resource value on fighting behaviour in a jumping spider,
 Anim Behav **36** 321.
- Wheeler, W. C., J. A. Coddington, L. M. Crowley et al. (2017), *The spider tree of life: phylogeny of*
Araneae based on target-gene analyses from an extensive taxon sampling, Cladistics **33** 574.

-
- Wolff, J. O., J. M. Schneider and S. N. Gorb (2014),
 “How to Pass the Gap – Functional Morphology and Biomechanics of Spider Bridging Threads”,
Biotechnology of Silk, ed. by T. Asakura and T. Miller, Dordrecht: Springer Netherlands 165.
- Wolff, J. O., K. Wierucka, G. B. Paterno et al. (2022),
Stabilized Morphological Evolution of Spiders Despite Mosaic Changes in Foraging Ecology,
Syst Biol **71** 1487.
- WSC (2025), *World Spider Catalog*, DOI: [10.24436/2](https://doi.org/10.24436/2), (visited on 20/01/2025).
- Xu, X., F. Liu, J. Chen et al. (2015),
A genus-level taxonomic review of primitively segmented spiders (Mesothelae, Liphistiidae),
ZooKeys 121.
- Xu, X., F. Liu, R.-C. Cheng et al. (2015),
Extant primitively segmented spiders have recently diversified from an ancient lineage,
Proceedings of the Royal Society B: Biological Sciences **282** 20142486.
- Xu, X., Y.-C. Su, S. Y. Ho et al. (2021), *Phylogenomic analysis of ultraconserved elements resolves the evolutionary and biogeographic history of segmented trapdoor spiders*, *Syst Biol* **70** 1110.
- Yang, B., Z. Zhang, C.-Q. Yang et al. (2022), *Identification of species by combining molecular and morphological data using convolutional neural networks*, *Syst Biol* **71** 690.
- Yang, R., m. Yan, L. Zhang et al. (2023), *New taxa of spiders (Araneae) from the world in 2022*,
Biodiversity Science **31** 23175.
- Yang, Z. (2015), *The BPP program for species tree estimation and species delimitation*,
Curr Zool **61** 854.
- Yang, Z. and B. Rannala (2010), *Bayesian species delimitation using multilocus sequence data*,
Proc Natl Acad Sci U S A **107** 9264.
- Zhang, J., P. Kapli, P. Pavlidis and A. Stamatakis (2013),
A general species delimitation method with applications to phylogenetic placements,
Bioinformatics **29** 2869.
- Zhao, F., T. Jiang, L. Yang et al. (2023),
Pholcid spiders of the Pholcusphungiformes species-group (Araneae, Pholcidae) from Liaoning Province, China: an overview, with description of a new species, *ZooKeys* **1156** 1.
- Zhao, F., L. Yang, Q. Zou et al. (2023), *Diversity of Pholcus spiders (Araneae: Pholcidae) in China's Lüliang Mountains: an integrated morphological and molecular approach*, *Insects* **14** 364.

Methods for Phylogenomic Analysis

Guanliang Meng¹, Chengran Zhou², Shanlin Liu³, and Xin Zhou^{3,*}

¹ Zoological Research Museum Alexander Koenig, LIB, Bonn 53113, Germany

² BGI-Shenzhen, Shenzhen 518083, China

³ Department of Entomology, China Agricultural University, Beijing 100193, China

* Corresponding author: xinzhou@cau.edu.cn

2.1 Summary

This work was previously published as an book chapter and has since been revised and expanded as a chapter in this thesis:

G. Meng, C. Zhou, S. Liu and X. Zhou (2022), “Phylogenomics based on transcriptomics and short-read sequences”, *Zoological Systematics and Evolution Protocol eBook*, ed. by G. Qiao, M. Zhu, W. Wang, Y. Qishen, D. Ren, S. He, F. Lei, W. Cai and W. Piao, 1st ed., Bio-protocol LLC 793, doi: [10.21769/BioProtoc.1010649](https://doi.org/10.21769/BioProtoc.1010649)

Summary

A phylogeny represents the genealogical relationships between different species. It serves as an essential foundation and prerequisite for taxonomy and the inference of numerous evolutionary events such as ancestral state reconstructions of related traits and biogeographical distribution analysis. Once the phylogenetic relationships among taxa have been established, by analyzing and comparing the evolutionary history of traits in different groups, we can investigate what and when selection pressures have posed to species and address whether different traits are co-evolving due to trait correlation or independent adaptations responding to similar selection pressure. Therefore, elucidating the phylogenetic relationships among species is one of the core tasks in evolutionary biology. Traditionally, the derived morphological characters and character states within a group of species are the main evidence that these species share a more recent common ancestor with each other than with species outside this group. However, the number of morphological characters are very limited, constraining their application to disentangle the elusive phylogenetic relationships of many taxa, such as insects (Misof et al. 2014), plants (One Thousand Plant Transcriptomes Initiative 2019), mammals (Eldridge et al. 2020), spiders (Arakawa et al. 2022) and birds (Stiller, Feng et al. 2024). On the other hand, the number of characters in the molecular data is overwhelming.

This is because a homologous site in the genome between species can be considered as a single character, and a genome typically consists of several hundred million base pairs (M bp) or even several billion base pairs (G bp) in size. The homologous loci inherited by different species in the process of speciation are the main data for molecular phylogenetic inference.

The advent of high-throughput short read sequencing technology has rendered extensive taxonomic sampling and large-scale genomic surveys routine for phylogenetic studies (e.g., Misof et al. 2014; One Thousand Plant Transcriptomes Initiative 2019; Stiller, Feng et al. 2024). However, the prohibitive cost and complexity associated with whole genome sequencing limit its applicability. In this regard, reduced representation sequencing (RRS) methods offer a viable alternative due to their lower costs and high feasibility. Specifically, transcriptome sequencing (i.e., RNA-seq), target capture sequencing (TCS), and low-coverage whole genome sequencing (lcWGS) are cost-effective and exhibit high practicality in phylogenetic investigations. Despite its utility, phylogenetic analysis is a complex process prone to pitfalls such as sample contamination, hidden paralogs, missing data, model misspecification, recombination, incomplete lineage sorting, and gene flow. This chapter provides an overview of the primary steps and potential challenges involved in conducting phylogenetic research employing RNA-Seq, TCS, and lcWGS methods to offer theoretical and practical guidance for researchers working in this field.

First, I briefly summarized the steps and guidelines for sample collection, storage, and transportation. Then, I introduced the major procedures for DNA/RNA isolation, library construction (for RNA-Seq, TCS and lcWGS, respectively), sequencing, and reads quality control. Next, I explained the algorithms for contig assembly (for RNA-Seq, TCS and lcWGS, respectively), contamination detection, ortholog assignment, multiple sequence alignment (MSA), and MSA quality control. Subsequently, I discussed the significance and methods for assessing systematic bias in MSA and model selection. I also introduced the main methods for phylogenetic tree and network inference, gene flow detection, and rogue taxa identification. Finally, I explained how to assess the robustness of the inferred phylogeny and the two strategies (tip-dating and node-dating) for estimating divergence times.

Personal contribution

This study was originally published as a chapter in the book *Zoological Systematics and Evolution Protocol eBook* (in Chinese), which was coordinated by the Key Laboratory of Zoological Systematics and Evolution, Chinese Academy of Sciences. The leader sent invitations to Chinese researchers both domestically and internationally to collaborate on this book. Prof. Dr. Xin Zhou and Shanlin Liu took the invitation and conceptualized the initial idea. I wrote the entire manuscript, with the exception of sections **Transcriptome assembly** (part of section 2.5) and **Comparison of species trees** (part of section 2.11), which were contributed by Dr. Chengran Zhou. The published version of this work only focused on transcriptome sequencing, omitting topics on target capture sequencing and low-coverage whole genome sequencing. It also did not cover **Taxonomic confirmation** (part of 2.5), **Systematic bias assessment of MSA** (part of section 2.8), **Inference of phylogenetic network** (part of section 2.10) and **Inference of gene flow** (part of section 2.10). When adapting this study into my thesis, I extensively revised, enhanced, and expanded the content, thereby providing more comprehensive guidance for researchers in the field. For instance, I recreated the workflow (Figure 2.1), explained in more details on how to model compositional heterogeneity over sites and across lineages, introduced the site-based methods and co-estimation methods for phylogenetic tree inference based on the multiple species coalescent (MSC) theory, and added extensions of the MSC theory to model gene duplication and loss (GDL), and gene flow.

2.2 Introduction

Resolving phylogenetic relationships among species is a crucial step in constructing an evolutionary framework for biology, serving as the foundation to understand the evolution of genes, development mechanisms of traits and geographic distribution patterns. By employing a known species tree, one can undertake several analyses such as studying gene origins, estimating divergence times between species, detecting adaptive evolution, reconstructing ancestral states of traits (including geographical distribution), and advancing taxonomy (Mirarab et al. 2021). The long history of life evolution, multiple massive explosion/extinction events (Hoyal Cuthill et al. 2020), extensive gene flow among lineages (e.g., Arnold 2016; Jensen et al. 2023; X. Wang et al. 2022; Yakimowski et al. 2014) as well as incomplete lineage sorting (ILS) (C. Zhang, Rabiee et al. 2018), have posed significant challenges to accurately representing phylogenetic relationships that align with the actual evolutionary history. In recent years, the sampling of genome-scale gene sets have become a valuable tool in evolutionary biology for reconstructing phylogenetic relationships (e.g., Amemiya et al. 2013; Chaw et al. 2019; Feng, Stiller et al. 2020; Y. Li et al. 2021). However, the application of whole genome data in molecular phylogenetics is still limited due to technological or resource-related difficulties associated with genome sequencing, assembly, and annotation. Furthermore, significant genomic heterogeneity exists among different species, including differences in genome size, heterozygosity, and duplication events, which can impact the identification of orthologs from genomes. Orthologs are genes present in multiple species that share a common ancestry through speciation, making them ideal for inferring species trees. Paralogs, on the other hand, result from gene duplication events. Consequently, reduced representation sequencing (RRS) is more suitable for small labs with limited funding and large-scale taxonomic studies due to its efficient acquisition of many loci for analysis (Davey et al. 2011; J. M. Good 2011).

In this chapter, I will provide an overview of three RRS methods that offer significant cost advantages and increased feasibility: transcriptome sequencing or RNA-Seq, target capture sequencing (TCS), and low-coverage whole genome sequencing (lcWGS). These methods have been successfully applied to a wide range of taxa in recent years, including insects (Misof et al. 2014), plants (One Thousand Plant Transcriptomes Initiative 2019), mammals (Eldridge et al. 2020), and spiders (Arakawa et al. 2022) and have made important contributions to the construction of phylogenetic relationships at higher taxonomic levels. I will summarize the main steps involved in these methods, including sample preparation, sequencing and assembly, ortholog identification, model selection, phylogenetic inference, and divergence time estimation (as shown in Figure 2.1).

2.3 Sample preparation and data acquisition

Transcriptome sequencing (RNA-Seq)

Sample preparation. Ribonucleic acid (RNA) is single-stranded molecules with the biological property of being easily degradable. To obtain high-quality RNA, it is essential to collect fresh tissue samples and preserve them in an appropriate manner. When working with field or laboratory collections, there are two common methods for preserving RNA: (1) liquid nitrogen preservation at about -196 °C under normal pressure, which requires quick placement of the sample into a Dewar flask; and (2) storage solution preservation using specialized solutions such as RNA-later reagent, which effectively inhibits the degradation function of RNA enzymes. Storage solutions can be stored at room temperature and are more portable than liquid nitrogen, making them suitable for field collection work. When employing storage

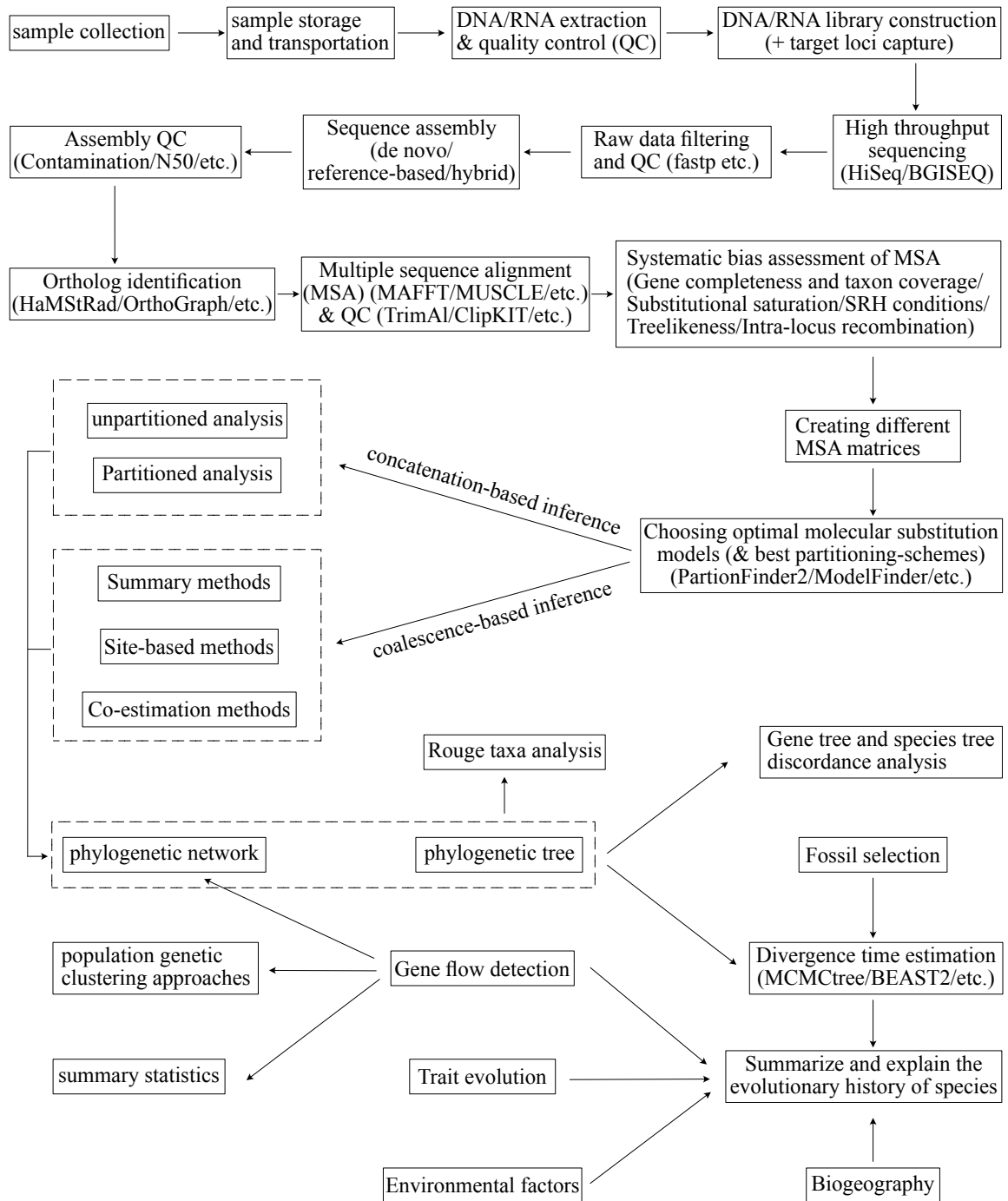


Figure 2.1. Procedures for phylogenomic analysis.

solution preservation, it is important to note that tissue samples should be cut into small pieces quickly (or crushed with a pestle and mortar) to allow the preservation solution to fully penetrate the sample; the ratio of reagent to tissue volume should be maintained at about 5:1 or adjusted according to specific storage solution requirements; and long-term storage should be conducted in an environment of $\leq 4^{\circ}\text{C}$ to extend the lifespan of integrated RNA. It has been demonstrated that samples stored with RNA-later preservation solution at room temperature can still provide high-quality RNA for library construction using High-throughput sequencin (HTS) (or so-called next-generation sequencing, NGS) after approximately one month. To minimize the presence of intestinal microorganisms or prey in sequencing data for small animals like insects or spiders, it is recommended to remove their abdomens/opisthosomas before collection. To reduce the impact of genetic heterogeneity between individuals on subsequent assembly, using a single individual as much as possible for RNA extraction, library construction and transcriptome data generation is advisable, provided that the minimum amount of RNA required for library construction is met. It is essential to emphasize the importance of preventing cross-contamination during sample collection and transportation. For example, it is necessary to avoid using contaminated tools to operate samples, and to use sealing film to seal the instruments to prevent leakage of preservation solution.

RNA extraction. Ribonucleic acid (RNA) extraction methods for eukaryotic organisms are well-established, and various commercial kit products such as TRIzol (Invitrogen) are available. RNA extraction can be performed according to the standard operating procedures provided by the reagent manufacturer. For instance, using the TRIzol kit (Invitrogen), total RNA extraction involves homogenization of tissues or cells, stratification based on density gradient centrifugation, precipitation with isopropanol and ethanol, washing, and dissolving in nuclease-free water. It is crucial to assess the quality of extracted RNA by evaluating its purity (26S/18S ratio), concentration, integrity (RIN value) and mass using instruments such as Qubit, Agilent 2100 Bioanalyzer or NanoDrop.

Library construction. HTS platforms such as Illumina HiSeq and MGI DNBSEQ are commonly in use nowadays. It is essential to follow the product manual for accurate operations during library construction and sequencing processes. The general process of constructing animal transcriptome RNA libraries includes: (1) RNA purification using Oligo (dT) magnetic beads enriched in mRNA; (2) Fragmentation by digesting the purified RNA with fragmentation buffer into target-sized fragments, such as 250 bp; (3) Double-stranded cDNA synthesis from the fragmented RNA template; (4) End repair, adenosine addition (dA-Tailing) and ligation of adaptor molecules; and (5) Amplification to generate sufficient quantities of cDNA for sequencing (Figure 2.1).

Target capture sequencing (TCS)

TCS differs from RNA-Seq in that it focuses on specific genomic regions, such as single copy orthologous genes across various taxa or even historical museum specimens (McCormack et al. 2016). To identify these targeted regions, researchers utilize predefined ortholog groups found in databases like OrthoDB (Kuznetsov et al. 2023; <https://www.orthodb.org/>) and BUSCO (Manni et al. 2021; <https://busco.ezlab.org/>). With the help of genome or transcriptome data closely related to the species of interest, various published pipelines, such as BaitFisher (Mayer et al. 2016), CATCH (Metsky et al. 2019) can be employed to design capturing baits.

For historical museum specimens that have experienced degradation and damage to their DNA due to the use of formalin during fixation and long-term storage (such as depurination or depyrimidation), it is often necessary to perform a FFPE (Formalin-fixed paraffin-embedded) DNA repair procedure in order to address potential DNA damages. This can significantly improve both library quality and the sequencing

data yield, which will contribute to more accurate scientific outcomes (<https://www.biochain.com/blog/ffpe-dna-damage-and-extraction/>, access date: Aug 2, 2024).

The process of constructing a library for target capture sequencing is analogous to that of RNA-Seq, which includes several steps such as DNA fragmentation, end repair, dA-Tailing, and adapter ligation. Following this, the indexed DNA libraries are combined into pools and hybridized with specifically designed target-specific probes. Subsequently, the hybridized DNA is captured using streptavidin-coated beads. The enriched DNA libraries undergo further amplification through a PCR procedure before being purified for sequencing.

Low-coverage whole genome sequencing (lcWGS)

The declining cost of sequencing in recent years has made it increasingly accessible for research groups to sequence multiple specimens using the lcWGS method, regardless of whether they are fresh or old samples. This innovative technology enables simultaneous analysis of thousands of genes and can be applied to a wide variety of materials. The continuous advancements in DNA sequencing technologies have also enabled the starting amount of DNA for library construction to decrease significantly - as low as 1 ng (<https://emea.illumina.com/>, access date: Aug 2, 2024).

2.4 High-throughput sequencing

Qualified libraries from previous experiments can be sequenced on High-throughput sequencing (HTS) platforms such as the Illumina HiSeq or MGI DNBSEQ systems. The amount of data generated for each sample varies depending on factors like genome size, protein-coding gene count, library type, and budget. For RNA-Seq projects with smaller genomes (e.g., 500 Mb), it is suggested to sequence at least 3 Gb per specimen; however, for larger genomes such as fish, a minimum of 8 Gb data should be acquired per specimen. In the case of lcWGA, sequencing 5-15X genome size per sample is recommended. Generating an increased volume of sequencing data improves the acquisition and assembly of low-abundance transcripts and loci; paired-end sequencing with a reading length of 150 bp (150 PE) are typically applied.

2.5 Data quality control and sequence assembly

Tips before the analysis

It is not trivial to handle big data and various analyses, and things are getting more complicated when there are hundreds or thousands of specimens in a project. At the same time, it is critical to ensure all analyses are reproducible. Some aspects should therefore be considered before the analysis.

Sample naming. Assigning a unique, concise name—typically 4–5 letters of the English alphabet—is recommended for each specimen, rather than using lengthy names or names containing special characters (e.g., underscores). Many bioinformatics programs for phylogenetic analysis have limitations regarding name length and the use of special characters. Short names can later be mapped to their full names in the analysis results. Additionally, it may be wise to avoid including species names in the short names, as specimens could be mislabeled or misidentified.

Documenting analytical process. To ensure reproducibility in analysis, it is crucial to record the steps we take. This includes noting the locations of related files and programs, as well as specifying program versions and the parameters used.

Quality control of HTS reads

The reads (i.e., raw data) from HTS platforms is generally saved in fastq format files (Cock et al. 2010). Raw data often contains some low-quality sequences, for instance, adapter sequences, ambiguous bases (“Ns”), or low-quality bases. Adapter sequences and sequencing errors will increase the consumption of computing resources, lead to assembly errors and reads cannot be correctly aligned to the reference genome. Therefore, low quality reads are to be trimmed or filtered out, for instance: (1) Trim off the adapters; (2) Discard reads whose have more than 20% of low quality bases ($\leq Q20$); (3) Discard reads with a high percentage of N bases (e.g., $> 5\%$). Common software used for such analysis includes: SOAPnuke (Chen et al. 2018; <https://github.com/BGI-flexlab/SOAPnuke>), FastQC (Andrews 2010; <https://github.com/s-andrews/FastQC>), Cutadapt (M. Martin 2011; <https://cutadapt.readthedocs.io/en/stable/>), NGS QC Toolkit (Patel et al. 2012), Trimmomatic (Bolger et al. 2014; <http://www.nipgr.ac.in/ngsqctoolkit.html>) and Skewer (Jiang et al. 2014), MultiQC (Ewels et al. 2016; <https://multiqc.info/>). The resulting files from this step is called clean data or clean reads.

Transcriptome assembly

The read length obtained from HTS platforms is generally within the range of 150-300 base pairs. Therefore, it is crucial to perform sequence assembly in order to obtain full-length transcripts or genes. There are three mainstream strategies for this purpose, including reference-based assembly, *de novo* assembly and hybrid assembly (Huang et al. 2016; B. Li et al. 2014; J. A. Martin et al. 2011; C. Zhou et al. 2020).

The reference-based assembly. This strategy relies on a reference genome of the target species or a closely related species. The main steps in this process include: (1) Using software such as TOPHAT2 (Kim, G. Pertea et al. 2013; <http://ccb.jhu.edu/software/tophat/manual.shtml>), STAR (Dobin et al. 2013; <https://github.com/alexdobin/STAR>), HISAT2 (Kim, Paggi et al. 2019; <http://daehwankimlab.github.io/hisat2/>) to align the clean reads to the reference genome. This step can effectively reduce the impact of sample contamination and sequencing errors on the assembly results. Compared with other alignment tools, the main feature of the above software is that they can consider different alternative splicing forms during the alignment process. This is because in eukaryotic organisms, the mRNA precursor (pre-mRNA) produced by the same gene is often adapted into different mature mRNAs (i.e., isoforms) by combining different exons and removing intron. (2) Next, using software such as Cufflinks (Trapnell et al. 2013; <https://github.com/santosjorge/cufflinks>) and StringTie (M. Pertea et al. 2015; <https://github.com/gpertea/stringtie>), construct a “graph” of the related reads mapped to the same locus. Such a graph is composed of all possible isoforms. (3) Then “deconstruct” the graph, that is, traverse different paths, to obtain all possible independent isoforms. Reference-based assembly has a stronger sensitivity to low-expression transcripts, helps to discover new transcripts and also requires less computing resource, lowering the costs. However, its outcomes strongly depends on the quality of the reference genome as well as the similarity between the sample and reference genome. Therefore, it is not suitable for cases where the reference genome quality is poor (e.g., the genome N50 is only a few K bp long, the BUSCO gene completeness is less than 70%) or the reference species is too distantly related (e.g., different orders). It is foreseeable that with the advancement of many large-scale genome projects (such as the 5000 arthropod genomes initiative, i5K; <http://i5k.github.io/>) (Consortium 2013), the Bird 10,000 Genomes Project (B10K; <https://b10k.genomics.cn/>) (G. Zhang 2015), the Earth BioGenome

Project (EBP) (Lewin et al. 2018; <https://www.earthbiogenome.org/>), and the Vertebrate Genomes Project (VGP; <https://vertebrategenomesproject.org/>) (Rhie et al. 2021; Scientists 2009), this reference-based assembly strategy will be applicable to more and more studies.

De novo assembly. When a high quality reference genome of the target species or its closely related species is absent (e.g., many genes are missing or incomplete), a *de novo* assembly strategy based on the *de Bruijn* Graph (DBG) theory (De Bruijn 1946; I. J. Good 1946) can be employed. This approach involves fragmenting reads into k-mers of a specific length (K bp) to construct a DBG graph, where consecutive k-mers overlap with each other with $K - 1$ bp. K-mers represent nodes, while overlapping regions between k-mers serve as edges in the graph. Transcripts are obtained through traversing this graph. Despite the widespread use of DBG graph theory in genome assembly, *de novo* assembly presents unique challenges for transcript assemblers compared to genome assemblers, such as accounting for expression differences among different genes and isoforms of the same gene. Popular transcript assemblers include Trinity (Grabherr et al. 2011), Oases (Schulz et al. 2012), IDBA-Tran (Peng, H. C. M. Leung et al. 2013), SOAPdenovo-trans (Xie et al. 2014), etc. The k-mer size dictates the length of overlap between different nodes, significantly influencing the quality of assembly results. Generally, using a larger k-mer is advised to achieve high-quality transcript assembly. However, an excessively large k-mer may hinder the assembly of low-expression transcripts due to its low coverage of k-mers. Conversely, smaller k-mers can more effectively assemble low-expression transcripts, but they may result in fragmented outcomes where a transcript is divided into multiple contigs. Therefore, some studies (such as De Vries et al., 2017; Robertson et al., 2017) also try to use multiple k-mer sizes for transcript assembly, and finally integrated different results. Related tools include Trans-ABYSS (Robertson et al. 2010), Multiple-k (Surget-Groba et al. 2010) and Rnnotator (J. Martin et al. 2010).

Hybrid assembly. In scenarios where a medium-quality reference genome from the target species or its closely related species (e.g., N50 about 300 Kbps, BUSCO gene integrity of around 80%) is available but requires balancing assembly quality and computing resource consumption, hybrid assembly strategies can be employed to achieve optimal results. Hybrid assembly refers to the simultaneous use of reference assembly and *de novo* assembly strategies for transcript assembly. Specifically, there are two strategies: alignment-then-assembly, and assembly-then-alignment (J. A. Martin et al. 2011). The former method involves mapping all clean reads to the reference genome first, followed by *de novo* assembly of unaligned reads. This approach is advantageous in terms of computing resource requirement. However, if the reference genome has poor quality (e.g., errors, gene missing or incompleteness), the latter method - assembly-then-alignment - can be considered. This is because it allows for detecting novel transcripts and isoforms using *de novo* assembly and utilizes high sensitivity of reference-based assembly to produce more comprehensive results. (J. A. Martin et al. 2011).

Assembly of the TCS and lcWGS data

There are two methods for performing *de novo* assembly from reads generated by either TCS or lcWGS: overlap-layout-consensus (OLC) and DBG theory-based strategies. OLC comprises three stages: identifying Overlaps among clean reads, constructing a Layout of these reads and their overlaps using a graph structure, and deducing the Consensus sequence from the graph. This method was prevalent during the Sanger sequencing era when read lengths were typically 800-1,000 bp but is computationally expensive to implement due to searching for overlap between reads. In contrast, dbG theory uses k-mers derived from reads to construct a DBG graph, with bubbles and tips later removed to generate contigs. Scaffold sequences are then assembled by connecting different contigs based on read pairing information or

BAC/Fosmid clones. Finally, mapping clean data against the scaffolds helps fill gaps in the assembly and fix errors. Compared to OLC method, DBG method efficiently handles large amounts of short reads (150-300 bp) generated by HTS platforms. However, it depends on k-mer size for continuity; larger k-mers increase continuity but reduce abundance while smaller k-mers lead to fragmented assembly due to genomic repeat presence and genome heterozygosity effects. The DBG method is also sensitive to sequencing errors that result in a higher number of unique k-mers. Common software used for this process include ABySS (Simpson et al. 2009), IDBA (Peng, H. C. Leung et al. 2012), SOAPdenovo2 (R. Luo et al. 2012) and SPAdes (Prjibelski et al. 2020). Tools such as GapCloser (R. Li, H. Zhu et al. 2010) can be further used to fill the gaps in the assembly.

Furthermore, reference-based assembly using clean reads can be employed as previously described (S. Liu et al. 2022). This approach entails mapping the short read aligners such as BWA (H. Li and Durbin 2009) or SOAPaligner (R. Li, Y. Li et al. 2008), followed by creating consensus sequences using BCFtools (Danecek et al. 2021). This method is more computationally efficient than the *de novo* assembly approach.

Quality control of the assembly

Performing quality control on assembly results is critical to determine if the sample has been contaminated by exogenous species or if there are cross-contamination issues between samples. Possible sources of exogenous species contamination include intestinal contents (such as microorganisms and food residues), symbiotic/parasitic organisms, or contamination caused by other projects being carried out simultaneously in the lab. Cross-contamination can also occur between different samples of the same project due to improper cleaning of tools during collection and storage, leakage during transportation, or simultaneous DNA/RNA extraction, library construction, and sequencing procedures. Moreover, multiplex sequencing (when pooled lanes are used for sequencing) may result in index cross-talk (Wright et al. 2016), which can lead to incorrect demultiplexing of reads. It is important to note that highly similar sequences between samples do not always indicate authentic cross-contamination; they could also be conserved genomic regions, such as ribosomal DNA (rDNA) genes.

Methods for detecting contamination include (1) identifying marker genes from sources such as mitochondria (*COI*) or chloroplasts (*rbcl*, *matk*), and comparing these marker genes to databases like BOLD (<https://www.boldsystems.org/index.php>) or GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), to ensure that all of them belong to the target species. (2) comparing whole assembly files with databases such as the NT database (<https://www.ncbi.nlm.nih.gov/nucleotide/>) to analyze the species composition in the assembly. (3) analyzing samples, particularly those sequenced in the same lane, for highly similar sequences using BLAST+ (Camacho et al. 2009). (4) calculating the GC-content of all contigs and visualizing them in a plot. Different species generally have different levels of GC percentage, which can result in multiple clusters of data points when contamination is present.

Additionally, VecScreen software (<http://www.ncbi.nlm.nih.gov/tools/vecscreen/>) and UniVec database (<https://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/>), can also be used to identify contamination from vector, linker or adapter sequences. It is important to note that cloning vectors are usually not involved in the HTS experiment process; therefore, detected vector contamination may result from exogenous genes integration into the target organism's genome (e.g., viral or bacterial genes). This may be of great evolutionary significance but might also interfere with subsequent ortholog searches. When evaluating assembly quality, if the 5'- or 3'-end of a contig/scaffold sequence matches vector sequences or linker DNA moderately or highly, it is recommended to delete that sequence.

If the middle part of a sequence matches the vector sequence or linker DNA, it should be deleted from the assembly result and broken into two sub-fragments in the middle. It's also worth noting that when uploading assembly results to NCBI, they will further check for contamination; therefore, we recommend deleting all corresponding detected sequences.

After evaluating for contamination using the methods mentioned earlier, statistical indicators of the assembly can be calculated to gain further insight into the quality and characteristics of the samples. Some examples include (1) total/minimum/maximum sequence length of the assembly, the number of sequences, GC content, N50 length.

Taxonomic confirmation

In addition to contamination, precise taxonomic identification of all specimens is crucial for accurate phylogenetic analysis, as mislabeled or misidentified samples can result in interpretations that are difficult to decipher due to high levels of incomplete lineage sorting (ILS), gene flow (e.g., introgression, horizontal gene transfer or hybridization) and insufficient data. It is highly recommended to verify all samples' taxonomy using marker genes.

Population genomics tools can also be employed to further understand the relationships within and between samples. Principal component analysis (PCA) provides dimension reduction capabilities that visualize population structure by capturing and summarizing genetic variation using principal components. On the other hand, the clustering-based methods, ADMIXTURE (Alexander et al. 2009; <https://dalexander.github.io/admixture/>) and STRUCTURE (Pritchard et al. 2000; <https://web.stanford.edu/group/pritchardlab/structure.html>) are able to define populations and assign individuals to them, and to parsimoniously explain genetic variation among individuals.

Calculating and comparing the genetic distances between different specimens may also help with taxonomic confirmation. Related distance types include the observed distance (p-distance), or model-based distance in which the multiple mutations occurred at the same sites have been corrected.

Finally, molecular species delimitation techniques involving either multiple species coalescent (MSC) theory or exploratory methods (Puillandre, Brouillet et al. 2021) can also be considered. Examples of MSC-based tools include BPP (Z. Yang and Rannala 2014) and BFD (Leaché, Fujita et al. 2014), which compare likelihoods for different partition schemes (i.e., different groupings of samples). Due to their computational demands, these methods are primarily used in studies with limited sequences or species. On the other hand, exploratory tools suggest *de novo* species partitions often based on a single locus and is very fast. It should be kept in mind that a gene tree can be different from a species tree. Tools falling into this category include general mixed Yule-Coalescent model (GMYC) (Pons et al. 2006), Automatic Barcode Gap Discovery (ABGD) (Puillandre, Lambert et al. 2012), Poisson tree process (PTP) (J. Zhang et al. 2013) and Assemble Species by Automatic Partitioning (ASAP) (Puillandre, Brouillet et al. 2021). The input for GMYC and PTP is a phylogenetic tree while the input for ABGD and ASAP is an alignment or genetic-distance between different specimens.

2.6 Ortholog inference

Inferring phylogenetic relationships among taxa relies on homologous characters, which can exhibit multiple character states. Homologous characters—such as a locus (gene), a site within an alignment, or morphological traits—are traits inherited by different lineages from a common ancestor. The terms gene(s)

and locus/loci are used interchangeably hereafter. The various forms of a homologous character are referred to as character states. For example, the forelimb in vertebrates is a homologous character, with its different states manifesting as the wings of bats and birds or the arms of primates. In traditional morphology-based cladistics, homologous characters are categorized into apomorphies, which are derived character states unique to specific taxa, and plesiomorphies, which represent ancestral or primitive character states. In molecular-based phylogenetic methods, identifying homologous loci, particularly single-copy orthologs, is crucial for accurate inference, as these loci provide a clear and consistent basis for comparison across lineages.

RNA-Seq data

There are several pipelines available for detecting single copy orthologs from transcriptome assembly, such as HaMStRad (<https://github.com/mptrsen/HaMStRad>) or Orthograph (Petersen et al. 2017). The process consists of several key steps: (1) choosing a specific taxonomic level, such as Arthropoda or Arachnida, and downloading the predefined single-copy orthologs (proteins) under this level from OrthoDB or BUSCO database. It is worth noting that selecting a lower taxonomic level closer to the organism of interest can significantly increase the number of available orthologs, providing a substantial advantage for phylogenetic inference. (2) Aligning such reference orthologs using software like MAFFT (Katoh et al. 2013; <https://mafft.cbrc.jp/alignment/software/>), and then Profile hidden Markov models (pHMMs) are constructed from the resulting alignments using HMMER (Eddy 2011; <http://hmmer.org/>). (3) Within the chosen taxonomic level, selecting a few representative species (i.e., reference species) and obtaining their annotated gene sets (i.e., all protein sequences) (e.g., from NCBI). (4) Searching the transcriptome assembly for candidate orthologs with the constructed pHMMs model and HMMER software (Eddy 2011). (5) Aligning all the candidate orthologs against the gene set of the reference species in step 3 using BLAST+ to find the best reciprocal best hits, i.e., the best match of the candidate ortholog must also be used for the construction of the corresponding pHMM model. (6) In order to further consider frameshift mutations caused by sequencing and assembly errors, the Exonerate (Slater et al. 2005) software and reference orthologs are used to predict the open reading frames of the selected transcript sequences. (7) Transcript fragments belonging to the same ortholog are connected by genomic order. (8) Due to sequencing or assembly errors, some identified orthologs may have internal stop codons, which will be replaced with “NNN”s and translated into unknown amino acid “X”s for subsequent analysis.

TCS data

The inference of orthologs from TCS assembly is less complicated, since most sequences in the assembly should belong to the predefined orthologs used for bait design. In the PHYLUCE pipeline (Faircloth 2016; <https://github.com/faircloth-lab/phylyce>), the contigs assembled by one of the assemblers Velvet (Zerbino et al. 2008; <https://github.com/dzerbino/velvet>), Trinity (Grabherr et al. 2011; <https://github.com/trinityrnaseq/trinityrnaseq>) and AbySS (Simpson et al. 2009; <https://github.com/bcgsc/abyss>) are aligned with the designed probe sequences, during which paralogs are also removed.

lcWGS data

The use of BUSCO pipeline (Manni et al. 2021) and OrthoDB database can be used to identify orthologs from lcWGS assembly. The process involves three main steps: (1) Gene prediction is performed using tools such as Augustus (Stanke et al. 2006), Metaeuk (Levy Karin et al. 2020) or Miniprot (H. Li 2023); this step quickly identifies gene structures within the assembly. (2) The predicted genes are parsed and scored against pHMM models of chosen reference orthologs using HMMER. This scoring process allows for further filtering and identification of candidate orthologs. (3) Finally, the predicted genes are classified into Complete and Single-copy (CS), Complete and Duplicated (CD), Fragmented or Missing (FM) categories. Among them, the genes of CS type will be used for subsequent analysis. Furthermore, based on my experience, it is advisable to scrutinize the orthologs predicted by ortholog inference tools. For example, perform blast search between sequences within each sample and remove any hits which could be paralogs.

In addition to the above ortholog inference tools, some recently developed methods have been introduced which can bypass these intermediate steps entirely. For example, the Read2Tree method (Dylus et al. 2024; <https://github.com/DessimozLab/read2tree>) begins with mapping clean data against reference ortholog alignments, producing consensus sequences and filtered orthologs based on sequencing coverage. If high quality genomes of samples are available, the ROADIES pipeline (Gupta et al. 2024; <https://github.com/TurakhiaLab/ROADIES>) can be used. It randomly (e.g. 500 times) chooses a sub-sequence (e.g. 500 bp in length) from one of the input genomes, and then LASTZ (Schwartz et al. 2003; <https://github.com/lastz/lastz>) is adopted to perform pairwise alignment between the sub-sequence and the genome assemblies of all other species. Next, low-quality pairwise alignments are discarded, the retained alignments are combined by locus, to be used for subsequent analysis. ROADIES eliminates the need for gene annotation and homolog inference, allowing the inclusion of both intergenic regions and protein-coding genes in phylogenetic analyses.

2.7 Multiple sequence alignment (MSA)

MSA construction

MSA is a process in which homologous sites of a locus are aligned across species to ensure proper comparison. For protein-coding genes (PCGs), it is common practice construct coding-sequence MSA (CDS-MSA) based on corresponding protein-MSA, to achieve higher accuracy. This is because codon degeneracy, where multiple codons can encode the same amino acid, makes it easier to establish reliable protein-MSAs, avoiding the introduction of spurious frameshift mutations during alignment (Suyama et al., 2006; Suyama et al. 2006). Popular alignment tools include MAFFT, ClustalW (Larkin et al. 2007; <http://www.clustal.org/clustal2/>), and MUSCLE (Edgar 2004; <https://github.com/rcedgar/muscle>); pal2nal.pl (Suyama et al. 2006) or TrimAl (Capella-Gutierrez et al. 2009; <http://trimal.cgenomics.org/>) can create CDS-MSA based on established protein-MSA.

Based on CDS-MSA, we can also create subset MSAs for different purposes. For example, a subset-MSA consisting of the first and second bases of each codon, and a subset-MSA consisting of four-fold degenerate sites (4D sites), where different nucleotides (A/T/G/C) at a certain position of a codon all encode the same amino acid. The 4D sites are thought to be under neutrally selected and meet the constant molecular clock model for divergence time estimation. In addition, a PCG often encodes one or more protein domains, which are the basic units of protein structure, function, and evolution, and is sometimes

considered to be more suitable for phylogenetic analysis (e.g., Misof et al. 2014). The Pfam (<http://pfam.xfam.org/>) and `pfam_scan.pl` (Mistry et al. 2021) and HMMER can be used to identify relevant protein domains and from which protein-MSA and CDS-MSA composed of these domains can be constructed.

Aligning homologous sites in rDNA genes, such as 18S, 28S, or tRNA genes, is challenging due to the existence of highly variable regions in the genes. Fortunately, their secondary structures are often highly conserved, which aids in the construction of their MSAs. Tools designed for this purpose include the `mafft-xinsi` algorithm in MAFFT and MXSCARNA (Tabei et al. 2008).

MSA quality control

The constructed MSAs should be subjected to checks for potential outlier sequences and poorly-aligned regions that may require realignment or pruning from the MSAs (Misof et al. 2014). Four main sources of such outliers have been identified by Jehl et al. (2015): (1) non-homologous sequences arising due to defects in ortholog inference; (2) errors in sequencing, assembly, gene annotation, and translation leading to non-homology between certain regions; (3) incomplete genes resulting from experimental treatments or insufficient sequencing coverage; and (4) highly divergent sequences among different species. Outlier sequences arising from non-homologous sequences are easily identifiable due to their significantly different gap and conserved block patterns compared to other sequences. However, detecting outliers in the second case becomes more challenging as longer non-homologous sequence parts increase. While MSA techniques have matured over time, unreliable alignment results may occur when there are widespread sequences with significant length differences or large sequence divergences; hence, extra attention must be given to this issue for accurate and reliable conclusions.

Popular tools for trimming poorly-aligned regions from MSAs include Gblocks (Castresana 2000), TrimAl (Capella-Gutierrez et al. 2009; <https://vicfero.github.io/trimal/>) and ClipKIT (Steenwyk et al. 2020; <https://github.com/JLSteenwyk/ClipKIT>). However, it is recommended to view the MSAs in UGENE (Okonechnikov et al. 2012; <https://ugene.net/>) or AliView (Larsson 2014; <https://ormbunkar.se/aliview/>) to confirm the trimmed MSAs, because, for example, some programs may retain sites with too many gaps.

A few software have also been developed for detecting outliers in MSA in recent years. For instance, PREQUAL (Whelan et al. 2018), unlike other tools, it takes unaligned homologous sequences as input, and then uses a probability model to identify and mask regions with non-homologous adjacent characters. HmmCleaner (Di Franco et al. 2019) uses hidden Markov model profiles to identify protein sequences that do not match the MSA as a whole. Divvier (Ali et al. 2019) uses a residue graph-based filtering method to partition the residues in a column of a MSA into different clusters based on pairwise posteriors. Within each cluster, the residues are highly reliable homologous to each other. Users can then choose whether to keep only the largest cluster or keep all clusters by creating a new column for each cluster. Phylopypruner (<https://gitlab.com/fethalen/phylopypruner>; last accessed 25 Dec 2024) infers sequence homology based on abnormal branch lengths in phylogenetic trees to remove possible contaminating sequences.

2.8 Systematic bias assessment of MSA

Phylogenomics introduces various models and assumptions to represent molecular evolution. Phylogenetic inference is a process to calculate phylogenetic trees based on these frameworks. It is thus advisable

to test if related assumptions are met and determine the impact of assumption violation.

Gene completeness and taxon coverage

Non-random data missing may pose challenges to phylogenetic inference, as observed in studies conducted by Dell’Ampio et al. (2014), Lemmon et al. (2009), B. T. Smith et al. (2020) and Wiens et al. (2011). AliStat (T. K. F. Wong et al. 2020; <https://github.com/thomaskf/AliStat>) can be used for evaluating the distribution of missing data in MSAs by performing pairwise comparison between samples. In conjunction with identifying rogue species (see below), we can identify samples whose phylogenetic placements are probably affected by missing data.

When inferring the phylogeny of different genes/loci (i.e., gene trees), it is recommended to use only the MSAs longer than a certain cutoff (≥ 150 nt or 50 aa), as low phylogenetic information in short MSAs can lead to incorrect gene trees. If a phylogenetic tree is to be inferred using the concatenation method (see below), especially the unpartitioned strategy, the length of each MSA is not that important. However, it should be ensured that each MSA contains a certain proportion of samples (i.e. taxon coverage), and the most important samples (representatives of all key groups) are retained.

It is also advisable to check how the characteristics of individual genes may influence phylogenetic analysis. For example, examining if MSAs with differences in the GC content, interquartile range of GC content, length, number of samples included, the number/proportion of information sites and missing data (Ns and gaps) support different topologies.

Test of the treelikeness assumption

The likelihood mapping method, as described by Strimmer et al. (1997) and implemented in IQ-TREE (Minh, Schmidt et al. 2020; <http://www.iqtree.org/>), can be used to determine if the data has enough phylogenetic information. During the process, a certain number of random quartets of taxa are drawn from the concatenated MSAs of all loci and visualized in a single triangular graph. The more quartets located at the three corners of the triangular graph, the more suitable the data is for constructing a well-resolved phylogeny; the quartets located at the middle of the triangular graph are considered uninformative (Strimmer et al. 1997).

Intra-locus recombination detection

In phylogenetic tree construction methods, both concatenation and coalescence-based approaches assume no recombination within each locus to ensure that all sites in the locus have evolved under the same tree. Unlike the concatenation method, coalescence-based methods take into account linkage equilibrium among different loci. To detect sequences with intra-locus recombination, tools such as 3SEQ (Lam et al. 2018; <https://github.com/olli0601/3SEQ>) can be employed. It is generally recommended to exclude these sequences from the analysis for more accurate results.

Test of the Stationary, Reversible and Homogeneous (SRH) violation and solutions

Most evolutionary models assume that molecular evolution meets the SRH conditions: Stationary, Reversible and Homogeneous (Ababneh, Jermini, Ma et al. 2006; Ababneh, Jermini and J. Robinson 2006; Hernandez et al. 2021; Naser-Khdour et al. 2019). Stationary refers to the constant frequency of DNA or amino acids over evolutionary time, i.e., homogeneous character composition. Reversible means that the

evolutionary process is stable and has no specific direction, that is, the substitution rate between different bases or amino acids is equal in the forward and reverse directions (Naser-Khdour et al. 2019). Homogeneous refers to the constant instantaneous rate of molecular substitution, and different lineages have the same rate. Molecular data that do not meet the SRH conditions can seriously affect the topological structure of the phylogenetic tree and lead to systematic errors (Naser-Khdour et al. 2019).

The stationarity of a sequence can be evaluated using the relative composition frequency variability (RCFV), which is a measurement of absolute deviation from the mean nucleotide (or amino acid) frequency in the dataset. The higher the RCFV values, the higher the compositional heterogeneity, which can be seen in fast-evolving clades and may have negative impact on phylogenetic inference (Cai et al. 2022). RCFV calculation has been implemented in BaCoCa (Kück and Struck 2014; <https://bonn.leibniz-lib.de/en/research/research-centres-and-groups/bacoca>) and nRCFV (Fleming et al. 2023; https://github.com/JFFleming/RCFV_Reader). Compositional homogeneity can also be checked with IQ-TREE using the chi2 test and the `basefreq` command in PAUP* (Swofford 2002; Thomas et al. 2013).

The stationary and homogeneous (SH) conditions can also be evaluated with the `--symtest-only` option in IQ-TREE. During the process, the former is checked with the maximum test of marginal symmetry (MarPval), and the latter is checked with the maximum test of internal symmetry (IntPval), small p-values for MarPval or IntPval indicate the violation of corresponding assumption. A small p-value for the third parameter, SymPval, indicates a rejection of the assumptions of stationarity, homogeneity, or both.

To mitigate the impact of compositional heterogeneity on phylogenetic tree construction, several methods can be employed. One such approach is data recoding (Roberto et al. 2017), for instance, in the RY-coding, purines (A and G) are coded as “R” while pyrimidines (C and T) are coded as “Y” (Phillips and Penny 2003). This technique reduces transition biases and normalizes base frequencies across the tree (Phillips and Penny 2003; Phillips, Delsuc et al. 2004). Available tools for this purpose include the BMGE program (Criscuolo et al. 2010; <https://bioweb.pasteur.fr/packages/pack@BMGE@1.12>) and the <https://github.com/ballesterus/Utensils/blob/master/RYplace.py> script. It should be noted that users should specify the data type parameter when apply IQ-TREE the RY-coding data. Another approach is to select only the compositionally homogenous sites from the MSAs for phylogenetic inference, which can be achieved with BMGE too. A third method is to use the evolutionary models and software with relatively loose requirements on the SRH conditions (Dutheil et al. 2008; Foster 2004; Woodhams et al. 2015; Zou et al. 2012). However, these methods usually consumes a lot of computing resources, so the scope of their applications is limited to a certain extent.

Substitutional saturation assessment and solutions

Slowly evolving loci possess little phylogenetic information due to insufficient number of occurred mutations. Conversely, excessively rapid evolution leads to saturation because a single site undergoes multiple mutations, the last mutation will cover the previous mutation signal. Furthermore, it can result in homoplasy where the same mutations occur in different lineages independently, rather than inherit from their common ancestor (homoplasy, or false synapomorphies; Philippe, Y. Zhou et al. 2005). The cumulative number of mutations (expected distance) surpasses the observed differences (p-distance) between sequences. When a locus is saturated, its phylogenetic signal for phylogenetic analysis is greatly reduced to 0, and sequences of different species of the locus will be clustered together based on whether they have similar equilibrium base frequencies (i.e., homoplasy), rather than based on sequence homology, resulting

in long-branch attraction (LBA) (Philippe, Y. Zhou et al. 2005). All tree construction algorithms (e.g., maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI)) suffer this problem. Consequently, slow-evolving loci are better suited to infer phylogenetic relationships among taxa diverged at deep nodes (e.g., organisms of different orders), whereas faster-evolving loci provide more accurate results for taxa have a more recent common ancestor (e.g., populations of the same species, or species of the same genus).

PCGs generally experience increased selection pressure to maintain their functions and thus exhibit a more gradual evolution compared to non-protein-coding regions. Within PCGs, due to varying degrees of degeneracy, different codons undergo distinct substitution rates at various positions. For instance, 65% of third position substitutions are synonymous mutations (the encoded amino acid unchanged), while only 4% of first position substitutions and all second position substitutions result are nonsense mutations (changes in amino acid) (Lemey et al. 2009). Furthermore, even within a given site, transition rates (interchanges within purines (A $\rightleftharpoons$ G), or within pyrimidines (C $\rightleftharpoons$ T)) are higher than transversion rates (interchanges between purines and pyrimidines). In addition, nucleotide-level evolution is faster than protein-level evolution at the same locus.

There are several approaches to address the problem of saturation and LBA.

Firstly, identifying slow-evolving loci and compare the topological differences between the phylogenies recovered from fast- and slow-evolving loci. One approach involves creating saturation plots for individual genes and separating loci into different groups according to their slopes. For instance, one such group may include the genes whose slopes are at least 0.4 with a correlation coefficient of $r^2 \geq 0.95$ (X-axis: corrected distance, Y-axis: p-distance; when the slope is close to 1, it indicates unsaturated replacement). Another method involves using PhyloMAd (Duchene et al. 2022) software for entropy-based saturation detection. The program compares sites in a MSA with those in a fully saturated state and performs the test. In the saturated state, the pattern of base frequencies has the maximum entropy. When the entropy value of a locus is greater than a certain threshold (critical values), the topology and branch length of a gene tree are likely to be estimated incorrectly. Before using PhyloMAd, it is best to remove loci with any base frequency exceeding 50% or with less than 30 nt of informative sites, as they will affect the accuracy of the test. An informative site is one where at least two different bases (or amino acids) are present, and each base (or amino acid) appears at least twice. Finally, IQ-TREE can also estimate the site-specific rates using its `--rate` option.

Secondly, use site-heterogeneous models, such as the CAT-GTR model (Lartillot and Philippe 2004), the PMSF model (H.-C. Wang et al. 2018), and the GTRpmix model (Banos et al. 2024). In essence, site-homogeneous models are susceptible to mutational saturation among sequences which makes it challenging to discern true phylogenetic signals from homoplasy (H.-C. Wang et al. 2018). I will discuss in more details in the model selection section later.

Thirdly, a combination of the recently proposed recoding strategies (e.g., the RY-coding of DNA data mentioned above) and site heterogeneity models (Ballesteros et al. 2022; Giacomelli et al. 2022). Data recoding can reduce the influence of saturation problem and compositional heterogeneity (i.e., different base (or amino acid) equilibrium frequencies between sites and/or lineages), during which multiple nucleotides or amino acids are collapsed into a single character state (called a bin). Susko et al. (2007) discussed two ways to construct bins for protein sequences: (1) rely on a rate-matrix (e.g., the Dayhoff empirical matrix, Dayhoff et al. 1978), and set fixed character states for each bin based on amino acids substitution rates. Amino acids belonging to the same bin have higher substitution rates and tend to have more similar chemical properties, while the substitution rates between amino acids from different bins are low. (2) Based on the properties of a specific alignment, algorithms try to find the proper groupings of amino acids

to minimize the compositional heterogeneity of the alignment (Hernandez et al. 2021; Susko et al. 2007). Similarly, translating CDS sequences protein sequences for analysis can also be seen as data recoding. Due to the degeneracy of codons, even if natural selection causes changes in the base composition of DNA sequences, the frequency of amino acids can remain relatively constant.

It is important to recognize that recoding inevitably reduces the amount of phylogenetic information in a dataset. When internal branches are short, with few accumulated changes, recoding can lower the accuracy of phylogenetic analyses. Therefore, it is generally recommended to apply recoding only to long alignments and highly divergent sequences (Foster et al. 2023; Hernandez et al. 2021). Hernandez et al. (2021) demonstrated that six-state amino acid recoding is not an effective method for addressing compositional heterogeneity or saturation in phylogenetic studies. The effectiveness of recoding depends on the specific type of compositional heterogeneity and the recoding strategy used. When possible, it is advisable to consider models that explicitly account for compositional heterogeneity, such as the CAT model for site-based heterogeneity or the Node-Discrete Compositional Heterogeneity (NDCH) model for lineage-specific heterogeneity. These approaches typically outperform homogeneous models, regardless of whether recoding is implemented (Foster et al. 2023).

Finally, TreSpEx (Struck 2014) or TreeShrink (Mai et al. 2018) can be used to detect genes and species affected by LBA and then exclude them from the analysis.

2.9 Model selection

Molecular evolution models consist of three components (Lanfear, Frandsen et al. 2017): (1) the compositional frequency (or equilibrium frequency) of different character states (either nucleotides or amino acids); (2) the relative substitution rates between different character states (i.e., exchangeabilities); (3) molecular evolution rate. Homogeneous models are those that meet the SRH criteria, assuming (1) stationary composition of character states across sites and over time, and (2) different sites evolving at the same rate (Ababneh, Jermini, Ma et al. 2006; Ababneh, Jermini and J. Robinson 2006; Hernandez et al. 2021; Naser-Khdour et al. 2019). Apart from the differences in model assumptions and available parameters, various models differ in whether the relevant model parameters are fixed as priors or estimated from the user's data, as well as the methods used for estimation (Lanfear, Frandsen et al. 2017). For instance, the JC69 assumes equal nucleotide frequencies for A/T/G/C, and equal substitution rates between them. The Generalised time reversible (GTR) model, on the other hand, infers the nucleotide frequencies and substitution rates from the data.

The character state composition as well as the substitution pattern may vary among sites (or loci) (i.e., pattern heterogeneity, or more commonly, compositional heterogeneity over site) (Lartillot, Rodrigue et al. 2013; Pagel et al. 2005). For instance, to maintain functional stability of the stem regions of tRNA genes, transition rate is higher than transversion rate; but the loop regions are not under this constraint, leading to differences in both composition and exchangeabilities between stem and loop regions (Higgs 1998; Hillis et al. 1991; Savill et al. 2001). Furthermore, the character state composition of a locus may also change over time, resulting to compositional heterogeneity across lineages (Blanquart et al. 2008; Foster 2004). It has been shown that compositional heterogeneity across loci and lineages are common in both eukaryotes (Nekrutenko et al. 2000) and prokaryote (Reichenberger et al. 2015).

Two strategies have been developed to address the compositional heterogeneity over sites (Pagel et al. 2005). The first involves data partitioning, in which different partitions (e.g., loci) of the data have their own rate matrix $\mathbf{Q}$'s which consists of two components: compositional frequency (matrix π) and

exchangeabilities (matrix $\mathbf{R}$). The second is to use mixture models., such as CAT and CAT-GTR (Lartillot and Philippe 2004), CAT-BP (Blanquart et al. 2008), LG+C60 (Si Quang et al. 2008; Z. Yang 1995), universal distribution mixture models (Schrempf et al. 2020), and GHOST (Crotty et al. 2020). Unlike the homogeneous models which takes only one rate matrix $\mathbf{Q}$ to model the evolution of a single site, mixture models directly assign multiple $\mathbf{Q}$ matrices (i.e., distinct compositional frequencies and exchangeabilities) to a single site, while each $\mathbf{Q}$ matrix has its own weight; the weights vary across matrices and sites. Finally, several models allow compositions and exchangeabilities to change at speciation events or any locations of the tree (Barry et al. 1987; Blanquart et al. 2008; Foster 2004; Heaps et al. 2014; Jayaswal et al. 2005; Williams, Heaps et al. 2015); and multiple independent composition vectors are adopted in the node-discrete compositional heterogeneity (NDCH) model (Foster 2004) to account for compositional heterogeneity across lineages (Williams, Schrempf et al. 2021).

Variation in selection pressure is thought to result in differing evolutionary rates across sites (or loci), a phenomenon known as rate heterogeneity among sites. For example, functional genes typically evolve more slowly than pseudogenes due to purifying selection acting to preserve their functions (Pagel et al. 2005). This rate heterogeneity is commonly modeled using a discrete (Z. Yang 1994) or continuous (Kalyaanamoorthy et al. 2017) gamma distribution, and/or by incorporating invariant sites within the data (Bryant et al. 2005). Conversely, homologous sites (or loci) (e.g., with differing functions) may evolve at varying rates because the functional and structural constraints acting on a specific site may shift over time (Penny et al. 2001), a phenomenon referred to as heterotachy (Lopez et al. 2002; Philippe and Lopez 2001). The GHOST model addresses the problem by using a mixture model with distinct set of branch length for each class of the mixture model, and allowing varying branch length across sites (Crotty et al. 2020).

The above discussions assume that different sites and loci evolve under the same tree (i.e., same topology and branch length). However, recombination, incomplete lineage sorting (ILS), hybridization/introgression, and horizontal gene transfer (HGT) can causes different regions of a loci and different loci have distinct evolutionary histories (see the next section for their definitions; Maddison 1997; Nichols 2001; T. Zhu et al. 2022). To address this, the mixtures across sites and trees (MAST) model (T. K. Wong et al. 2024) extends the mixture framework by positing that each site in the alignment is derived from a mixture of multiple trees. Each tree is characterized by its own weight, topology, and branch lengths, with substitution models that can be either independent or shared among the trees (T. K. Wong et al. 2024).

Model misspecification, including but not limited to the inability to accurately capture the above characteristics of the data, can lead to mislead phylogenetic trees, for example, the site-homogeneous models, where all sites share the same rate matrix $\mathbf{Q}$ may produce LBA artifacts (Nekrutenko et al. 2000; Reichenberger et al. 2015).

Popular tools for determining the best-fit model and partitioning scheme of the data (see the section **Concatenation-based tree construction** below) include PartitionFinder2 (Lanfear, Frandsen et al. 2017) and ModelFinder (Kalyaanamoorthy et al. 2017) in IQ-TREE, SMS (Smart Model Selection) (Lefort et al. 2017) in PhyML, and ModelTest-NG (Darriba et al. 2020; <https://github.com/ddarriba/modeltest>). Despite significant improvements in the computational efficiency of these programs, model selection remains time-consuming. Consequently, it has been suggested that the model selection step can be bypassed by directly using the most parameter-rich model, GTR+I+G, particularly when the primary goal of the study is to infer the topology and ancestral sequences (Abadi, Azouri et al. 2019).

2.10 Phylogenetic tree/network reconstruction

In phylogenetic analysis, stochastic molecular evolution models describe how sequences evolve over time. Maximum Likelihood (ML) and Bayesian inference are two key methods for constructing phylogenies based on such models. A phylogeny hypothesis includes the topology, branch lengths, and model parameters. ML selects the phylogeny that maximizes the likelihood of the observed data, providing a single point estimate. In contrast, Bayesian inference samples from the posterior probability distribution of phylogenetic hypotheses given the observed data using methods like Markov Chain Monte Carlo (MCMC), combining prior information with the data's likelihood. The result is often a majority-rule consensus tree summarizing the most probable relationships among species. Currently, there are two main strategies for constructing molecular phylogenetic trees.

Concatenation-based tree construction

This approach concatenates all MSAs into a super-matrix. The order of MSAs is irrelevant because molecular evolution models assume all sites are independent. Then, an appropriate evolutionary model is chosen for modeling the evolution of all sites. The method (except the newly proposed MAST, described as above) relies on the assumption that genes share identical topology as well as species trees being equal to gene trees (Shen, Steenwyk et al. 2021). However, this assumption might be invalid due to factors such as hidden paralogy (genes misassumed as orthologs because of lost or incomplete identification of family members of paralogs) (Rasmussen et al. 2012), Introgression or horizontal gene transfer (HGT; genetic information's lateral transfer between species rather than vertical transfer from parent to offspring) (Davidson et al. 2015; Soucy et al. 2015), and incomplete lineage sorting (ILS; discrepancies between gene trees and species trees caused by diversification and retention of ancestral polymorphisms) (Feng, Bai et al. 2022; L. S. Kubatko et al. 2007; Mendes et al. 2018; Scornavacca et al. 2017). The branch lengths of evolutionary trees obtained by the concatenation method are generally expressed in terms of substitutions per site.

The unpartitioned method. All genes and sites assume to follow the same rate matrix $\mathbf{Q}$ and share the identical tree (including equivalent topology and branch lengths). However, it fails to capture compositional heterogeneity across sites and over time. Additionally, recombination events within a gene or between genes during their history can result in different regions of a gene or different genes evolving under distinct phylogenetic trees. With these, this method generally lead to a poor phylogenetic inference.

The partitioned method. In the analysis, different genes or sites are grouped into different partitions, each partition has its own rate matrix $\mathbf{Q}$ to model the compositional heterogeneity over sites. For instance, GTR model for one partition, JC69 model for another, or GTR partitions for all partitions but with distinct model parameters; this method also allows for a joint analysis of molecular and morphological data whereby each has its own partition. To model the rate heterogeneity among partitions, one might require all partitions to have the same branch lengths, or the branch lengths for each partition to be calculated separately, or the branch lengths between different partitions to be proportional.

In this partitioned-based analysis, basic data blocks are firstly defined (e.g., by gene or codon position), and then different combinations (called partitioning schemes) of such basic data blocks are created using algorithms (e.g., greedy, rcluster, and hcluster). Within each partitioning scheme, (1) the log likelihood for each partition is calculated and for each candidate evolutionary model; (2) the most suitable evolutionary model for the partition is selected based on the Akaike Information Criterion (AIC), corrected Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC), or the Decision-Theory criterion

(DT), where the model with the minimum AIC/AICc/BIC/DT value is the best-fitted model; (3) the log likelihoods of the best model of each partition are summed to obtain the log likelihood value for each partition scheme. Finally, the optimal partitioning scheme with the minimum AIC/AICc/BIC/DT value is selected.

Methodological comparisons reveal that the topology and support values of phylogenetic trees differ significantly between unpartitioned and partitioned analyses. Unpartitioned analyses can yield higher or lower support values without a consistent trend and may produce incorrect topologies with misleadingly high support. However, branch lengths at inconsistent internal nodes show no significant differences compared to partitioned analyses, though unpartitioned methods may overestimate or underestimate branch lengths (Kainer et al. 2015). In partitioned analyses, conflicts in tree topologies across different schemes are typically restricted to nodes with low support or short branch lengths. As the length of the multiple sequence alignment (MSA) increases, the influence of partitioning schemes diminishes. This reduction could result from: (1) increased phylogenetic signal allowing different schemes to converge on the correct tree; or (2) rising systematic errors causing convergence on an incorrect tree due to limitations of models like SRH (Kainer et al. 2015).

Finally, the choice of partitioning scheme does not always improve phylogenetic inference with more partitions; the evolutionary model must balance accurately representing the evolutionary processes of the molecular data while avoiding overparameterization (Lanfear, Calcott et al. 2012).

The unpartitioned and partitioned ML analyses have been implemented in some popular programs, including PhyML (Guindon et al. 2010; <https://github.com/stephaneguindon/phyml>), RAXML-NG (Kozlov et al. 2019; <https://github.com/amkozlov/raxml-ng>) (the successor of RAXML, Stamatakis 2014), IQ-TREE2 (Minh, Schmidt et al. 2020; <http://www.iqtree.org/>) (the successor of IQ-TREE, L.-T. Nguyen et al. 2015). To avoid local optima in the search space, these methods typically begin by constructing multiple starting trees using the maximum parsimony method and/or the BIONJ algorithm. The topology and branch lengths of these starting trees are then iteratively adjusted to identify the tree with the highest likelihood value from each starting point. Among the trees that reach maximum likelihood values, the one with the highest likelihood is selected as the potentially globally optimal tree. Increasing the number of starting trees, tuning other specific parameters of the program and performing multiple runs enhance the likelihood of finding the global optimum. In addition, it is also a good practice to use competing topological hypotheses as initial starting trees in different runs and then summarize the congruence and incongruence among the trees (Du et al. 2024).

Coalescence-based tree construction

The process of tracing the evolutionary history (i.e., genealogical relationships) of a homologous locus sampled from multiple individuals within a population back in time until their most recent common ancestor is identified is known as coalescence (Kingman 1982a; Kingman et al. 1982; Kingman 1982b). When it is applied to a group of species instead of individuals from the same population, it is called multispecies coalescence (MSC) (L. Liu, L. Yu, L. Kubatko et al. 2009; C. Zhang, Rabiee et al. 2018). MSC and its variants have been increasingly utilized in phylogenetic studies due to their ability to mitigate biases introduced by ILS, gene duplication and loss (GDL) and gene flow (introgression/HGT/hybridization), missing data (e.g., not all gene trees have the same set of species), and other challenges like differences in sampling quality, sequencing efficiency, and assembly completeness, which are common in RRS datasets. The branch lengths of the inferred trees are often expressed in coalescent units.

Standard MSC methods assume ILS alone (rather than GDL and gene flow) to be responsible for the

heterogeneity/incongruence among gene trees (i.e., individual genealogies for different loci). They can be broadly divided into three subclasses.

Summary methods. This approach begins by constructing gene trees based on individual MSAs, respectively. These gene trees are then summarized using coalescence theory to produce a “species tree,” which represents the evolutionary relationships among species. Specifically, for every four species, unrooted quartet trees (small subsets of the species tree) are computed. Each quartet tree has three possible topologies, and the method calculates the support ratio of gene trees for each topology. The objective is to infer an unrooted species tree in which nodes maximize quartet support. This method is easily scale to a lot of species due to its computational efficiency. Summary methods are also robust to certain cases of HGT (Daskalakis et al. 2016; Davidson et al. 2015; Roch et al. 2013). However, It involves two significant limitations. First, they are highly sensitive to errors in gene tree estimation (Shen, Steenwyk et al. 2021). Such errors often stem from poor-quality gene alignments, including alignments with very few informative sites due to excessively short gene sequences or extremely slow evolutionary rates. Second, they are impacted by gene flow (e.g., introgression, HGT, hybridization) (Solís-Lemus, M. Yang et al. 2016) and non-random missing data, which can further distort results (e.g., Simmons et al. 2020; Springer et al. 2016). To address the noises in gene trees, it was recommended to contract low support branches in the gene trees, but the thresholds for contracting are arbitrary and data-dependent (C. Zhang, Rabiee et al. 2018). Later, threshold-free weighting schemes for each quartet tree were introduced where the weights are obtained from tree branch lengths, branch support values or both (C. Zhang and Mirarab 2022). Representative software includes STAR (L. Liu, L. Yu, Pearl et al. 2009), MP-EST (L. Liu, L. Yu and Edwards 2010; <http://faculty.franklin.uga.edu/lliu/mp-est>), NJst (L. Liu and L. Yu 2011; <https://code.google.com/archive/p/phybase/downloads>), ASTRAL-IV (C. Zhang, Rabiee et al. 2018; <https://github.com/smirarab/ASTRAL>), wASTRAL (C. Zhang and Mirarab 2022; <https://github.com/chaoszhang/ASTER>) and Weighted ASTRID (B. Liu et al. 2023; <https://github.com/RuneBlaze/internode>). Additionally, the study of Morel, Williams et al. (2023) shows that Asteroid (<https://github.com/BenoitMorel/Asteroid>) can achieve more accurate phylogenetic trees compared to ASTRAL-III (C. Zhang, Rabiee et al. 2018) and ASTRID (Vachaspati et al. 2015) when the proportion of missing data (up to 80%) is high.

Site-based methods. This method leverages site patterns in the multiple sequence alignment (MSA) to construct “small trees,” which can be either unrooted trees of four species or rooted trees of three species. These small trees are then summarized to infer a species tree. A notable software implementing this approach is SVDquartets (Chifman et al. 2014). It calculates a singular value decomposition (SVD) score for each possible unrooted tree of four species (three such configurations exist). The species tree is then inferred by identifying the topology that minimizes the overall SVD score (Chifman et al. 2014). One limitation of SVDquartets is that it provides only the topology of the species tree without branch length information, which can restrict some evolutionary analyses. However, this method avoids errors associated with gene tree estimation and offers a more direct approach to species tree inference but with a higher computational cost.

Co-estimation methods. This approach simultaneously estimates species trees and gene trees (Szöllősi et al. 2015) under a Bayesian framework. During the process, posterior samples of gene trees of each locus and species trees are obtained by using the Markov chain Monte Carlo (MCMC) technique, and then summarized to generate the gene trees for each locus and a species tree based on the majority rule. This approach is believed to be robust to gene tree estimation errors (Knowles et al. 2012; Leaché and Rannala 2011). The best-known programs falling into this subclass are BEST (L. Liu 2008; <https://github.com/lliu1871/BEST>), *BEAST (Heled et al. 2009), MrBayes (Ronquist, Teslenko et al.

2012; <http://nbisweden.github.io/MrBayes/>), RevBayes (Höhna et al. 2016; <https://revbayes.github.io/>), and StarBEAST2 (Ogilvie, Bouckaert et al. 2017). Unfortunately, they can only be applied to a very limited of taxa and loci due to their high computational demand. Recently, StarBEAST3 (Douglas et al. 2022) was introduced and by parallelizing gene tree inference, it achieves 1 to 2 orders of magnitude faster than *BEAST and StarBEAST2. It is also crucial to ensure different MCMC chains in Bayesian inference have converged before interpreting the results, for instance, the effective size of related parameters should be larger than 200. Bayesian methods have poor scalability due to the computational burdens, and thus are only applied to small data sets in practice.

The standard MSC framework for inferring a species tree from multiple genes, as discussed above, assumes that ILS is the only source of gene tree discordance. It also assumes that (1) Different genes have independent genealogical trees. It implies that sufficient recombination between genes results in linkage equilibrium. To meet this assumption, researchers often select genes that are spatially distant on the genome. (2) All sites within a locus share a common evolutionary history, resulting in a single gene tree for the locus. Therefore, no recombination occurring within a locus is assumed, i.e., complete linkage disequilibrium (LD) among sites, which is more likely to hold for short genes. In practice, the assumption (1) is rarely problematic, especially when analyzing a large number of genes. Even when violated, its impact on phylogenetic analyses is generally minor. However, violations of this assumption (2) can significantly affect analyses, for instance, when analyzing data from different populations within the same species or long segments of closely related species. Recombination can also produce chimeric sequences, which appear artificially divergent, creating gene trees with disproportionately long branches. Readers are encouraged to refer to the above section on **Intra-locus Recombination Detection** for guidance on identifying and excluding loci that may be affected by recombination. The standard MSC framework has been shown to be statistically consistent when the assumptions are met, indicating that the topologies of the estimated tree and the true species tree will be identical as more data (i.e., the number of gene trees) are used for the analysis (L. Liu and L. Yu 2011; Solís-Lemus, M. Yang et al. 2016; Wascher et al. 2021; C. Zhang, Rabiee et al. 2018).

The standard MSC model has been extended to account for gene duplication and loss (GDL). The ASTRAL-ONE and ASTRAL-multi algorithms (Legried et al. 2021) prove to be statistical consistency under the GDL model (no ILS) proposed by Arvestad et al. (2009). Additionally, Hill et al. (2022) and Markin et al. (2021) proved the statistical consistency of these two algorithms under the unified GDL/ILS framework, i.e., the DLCoal model developed by Rasmussen et al. (2012). Meanwhile, C. Zhang, Scornavacca et al. (2020) showed that ASTRAL-Pro (<https://github.com/chaoszhang/ASTER/blob/master/tutorial/astral-pro3.md>) is also statistically consistent under the DLCoal model. Finally, the empirical study by Yan et al. (2022) showed that the presence of paralogs does not interfere with species tree inference using the summary method, no matter whether the algorithms assume single-copy orthologs only (ASTRAL-III and NJst) or explicitly model paralogs (ASTRAL-Pro and FastMulRFS (Molloy et al. 2020; <https://github.com/ekmolloy/fastmulrfs>)). This suggests that both orthologs and paralogs can be used for phylogenetic analysis by treating them as different individuals of the same species or by randomly choosing one of them, leading to an increase in the phylogenetic signal (Yan et al. 2022). However, the branch lengths derived from this method are not accurate and have to be re-estimated with authentic single-copy orthologs (Yan et al. 2022).

On the other hand, SpeciesRax (Morel, Schade et al. 2022; <https://github.com/BenoitMorel/GeneRax>) infer species trees from gene trees within the ML framework while accounting for both GDL and HGT, but without utilizing the MSC model.

Inference of phylogenetic network

In addition to ILS and GDL, gene flow between species through introgression, HGT or hybridization can also cause incongruence between gene trees. Moreover, gene flow leads to reticulations in the species phylogeny, resulting in a phylogenetic network.

A method for inferring phylogenetic networks has been implemented within the maximum likelihood framework, but it does not explicitly use the MSNC model. Therefore, it does not consider ILS during the analysis. This approach infers a phylogenetic network that maximizes likelihood based on a given molecular evolutionary model and the data. Representative programs include NEPAL (<http://old-bioinfo.cs.rice.edu/nepal/>) and PhyloDAG (Q. Nguyen et al. 2015; <https://phyloDAG.wordpress.com/2015/04/17/phyloDAG/>), PhyLiNC (Allen-Savietta 2020) in PhyloNetworks (Solís-Lemus, Bastide et al. 2017), NetRAX (Lutteropp et al. 2022; <https://github.com/lutteropp/NetRAX>). For instance, NetRAX accepts a partitioned MSA, assuming there are no ILS nor recombination among MSA sites. Sites within each partition are assumed to have the same evolutionary history while different partitions can either share same set of branch lengths or independent set of branch lengths.

Another approach to inferring phylogenetic networks has been implemented using the multispecies network coalescent (MSNC) model within either a maximum likelihood framework or a Bayesian framework (Wen, Y. Yu and Nakhleh 2016; Y. Yu, Barnett et al. 2013; Y. Yu, Dong et al. 2014), or sometimes the maximum parsimony (MP) framework. Using the MSNC model enables them to account for ILS during the analysis and they can also be divided into three subclasses.

Summary methods. The input data in this subclass comprise either rooted or unrooted gene trees, employing either maximum likelihood (ML) or Bayesian methods for phylogenetic network inference. The `InferNetwork_ML` command (Wen, Y. Yu, J. Zhu et al. 2018; Y. Yu, Dong et al. 2014) in PhyloNet (Than et al. 2008; <https://phylogenomics.rice.edu/index.html>) is an implementation of the ML methods, while the `MCMC_GT` command (Wen, Y. Yu and Nakhleh 2016) represents a Bayesian inference implementation. The outcomes from both commands include species network(s), branch lengths (in coalescent units) and inheritance probabilities (Wen, Y. Yu, J. Zhu et al. 2018; Y. Yu, Dong et al. 2014). Depending on the settings, the branch lengths of the gene trees may or may not be used during the process. Additionally, a maximum pseudolikelihood method (the `InferNetwork_MPL` command, Y. Yu and Nakhleh 2015 in PhyloNet) has been developed to reduce the computational burden of likelihood calculations, similar to that (Solís-Lemus and Ané 2016) of PhyloNetworks (<https://github.com/JuliaPhylo/PhyloNetworks.jl>). Finally, the `InferNetwork_MP` command (Y. Yu, Barnett et al. 2013) in PhyloNet estimates a species network with a predefined number of reticulation node, employing an extended MDC (minimizing the number of deep coalescences) criterion. However, the branch lengths in the input gene trees are ignored and only the phylogenetic network topology is estimated (Wen, Y. Yu, J. Zhu et al. 2018; Y. Yu, Barnett et al. 2013).

Site-based methods. The input data for this approach are unlinked biallelic markers, which are either single nucleotide polymorphisms (SNPs) or amplified fragment length polymorphisms (AFLPs). J. Zhu et al. 2018 introduced a Bayesian framework (the `MCMC_BiMarkers` command in PhyloNet) to sample the posterior of phylogenetic networks given bi-allelic marker data using a reversible-jump MCMC technique. It estimate a species network together with its divergence times (measured in the expected number of mutations per site), population mutation rates (expressed as mutations per site), and inheritance probabilities, utilizing Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Wen, Y. Yu, J. Zhu et al.

2018; J. Zhu et al. 2018).

Co-estimation methods. This approach has been implemented in the MCMC_SEQ command (Wen and Nakhleh 2018) of PhyloNet, the SpeciesNetwork package (C. Zhang, Ogilvie et al. 2018) of BEAST2, and BPP (Flouri et al. 2020; Z. Yang 2015; <https://github.com/bpp/>). Because they are implemented in the Bayesian framework, the phylogenetic networks, gene trees, as well as their parameters are estimated from the posterior samples. Unfortunately, the heavy computation requirement of this approach limits its application to large data sets. Heuristic methods (Y. Wang and Nakhleh 2018) and constraining gene tree space searching (Y. Wang, Ogilvie et al. 2020) have been introduced to improve its efficiency.

One straightforward approach to examining a large dataset is to begin by constructing a species tree. This tree can then be used to divide the dataset into smaller subgroups, focusing on specific subtrees for further phylogenetic network analysis (Wen, Y. Yu, J. Zhu et al. 2018). Readers are also encouraged to refer to Cao et al. (2023) for a more detailed description on using PhyloNet.

Inference of gene flow

Gene flow refers to “the exchange of genes between two populations” through interbreeding (Pickrell et al. 2014). Admixture, on the other hand, is characterized by “a sudden increase in gene flow between two differentiated populations” (Pickrell et al. 2014). Introgression (or “introgressive hybridization”) involves the transfer of gene material from one population to another via hybridization between them, followed by backcrossing between the hybrids and one of the parental populations (Harrison et al. 2014). Hybridization may also lead to new species, known as hybridization speciation. Another mechanism of gene flow is horizontal gene transfer (HGT), which typically occurs among asexual organisms, such as bacteria. Admixture, introgression and hybridization have often been used interchangeably in the literature. From the perspective of data modeling, gene flow via different processes all leads to reticulation events in the phylogeny, making it difficult for current inference methods to distinguish among them (Cao et al. 2023).

Additionally, gene flow plays a critical role in adaptation and speciation by introducing new genetic variants into populations (Abbott et al. 2013; Fontaine et al. 2015; Huerta-Sánchez et al. 2014; Whitney et al. 2015). These variants (alleles) may either enhance fitness (adaptive introgression) or hinder it (maladaptive introgression). During the course of evolution, beneficial introgressed alleles are subject to positive selection, while deleterious ones are typically purged through purifying selection. The effectiveness of selection process is influenced by several factors, including the relative fitness effects of the alleles (i.e., selection coefficient), the effective population size, and the strength of recombination (Moran et al. 2021). A larger effective population size increases the efficiency of selection by mitigating the effects of genetic drift, allowing beneficial alleles to spread and deleterious ones to be eliminated more effectively (Moran et al. 2021). Conversely, low recombination rates can exacerbate linkage disequilibrium (LD), maintaining long stretches of deleterious DNA. However, under these conditions, selection can act more cohesively on linked deleterious fragments, accelerating their removal from the population (Moran et al. 2021). On the other hand, higher recombination rate may also break the linkage between beneficial allele at one locus and deleterious allele at another locus, facilitating the fixation of introgressed beneficial alleles. This dynamic interplay between selection, recombination, and population size shapes the evolutionary trajectory of introgressed variants.

Over the past two decades, numerous algorithms have been developed to detect the presence, timing, direction, and admixture proportions of both ancient and recent gene flow among species or populations. These methods typically fall into three main categories: (1) Inferring phylogenetic networks –as discussed in the previous section, which models reticulation events directly. (2) Using summary statistics –such as

identity by descent (rIBD) (Browning et al. 2013), χ^2 test (C. Zhang, Lin et al. 2019), and a variety of tests focus on site patterns among populations/species, including the D test (also known as ABBA-BABA test) (Green et al. 2010), f_4 or $f_4\text{-ratio}$ (Patterson et al. 2012; Reich et al. 2009), f_d (S. H. Martin et al. 2015), f_{dM} (Malinsky et al. 2015), D_p (Hamlin et al. 2020), D_{FOIL} (Pease et al. 2015). (3) population genetic clustering approaches, such as *STRUCTURE* and *ADMIXTURE*. Readers are encouraged to consult the relevant publications for details on these methods. Here, I will briefly introduce the D test.

Suppose three species ($H1, H2, H3$) and an outgroup species (O) form the topology $((H1, H2), H3), O$. At a homologous site, ancestral alleles are denoted as “A,” and derived alleles as “B,” with the outgroup used to identify the ancestral state. The D test compares the frequency of derived allele sharing between $H1$ and $H3$ (site pattern: BABA) and $H2$ and $H3$ (site pattern: ABBA). If ILS is the sole cause of gene tree discordance, BABA and ABBA should occur equally. Deviations suggest additional processes, such as gene flow, contributing to allele-sharing patterns.

Compared to population clustering methods like *STRUCTURE* and *ADMIXTURE*, the D test demonstrates greater robustness in various scenarios: single and multiple introgressions, combinations of ancestral and recent introgressions, as well as varying timings of introgressions scenarios (Kong et al. 2021). However, its performance reduces when the level of ILS is excessively high, and often show an unacceptably high false discovery rate (FDR) (Kong et al. 2021).

Recent studies highlighted the challenges posed by “ghost” lineages—extinct or unsampled species—in the detection and interpretation of results using the D test (Tricou et al. 2022). When ghost lineages are present, the D test fails to accurately identify the true donors and recipients of gene flow (Tricou et al. 2022). Furthermore, the use of a distant outgroup exacerbates this issue, significantly increasing the false discovery rate (FDR) (Tricou et al. 2022). Very recently, Pang et al. (2024) showed that compared to the site-pattern- or gene-tree topogoy-based methods, as implemented in HyDe (Blischak et al. 2018) and the `InferNetwork_MPL` command of PhyloNetL, the full-likelihood method BPP (Flouri et al. 2020) performs better at identifying ghost introgression by considering both gene-tree topologies and branch lengths.

Additionally, Frankel et al. 2023 revealed that model assumption violation due to evolutionary rate variation across lineages has great impact on summary tests of introgression. Recently, Koppetsch et al. 2024 proposed a new test leveraging the expected clustering of introgressed sites across the genome to address the issue.

2.11 Phylogenetic post-analysis

Rogue species identification

A rogue species refers to a species that may appear in multiple positions across a set of phylogenetic trees (e.g., bootstrap trees), making it difficult to accurately infer its evolutionary relationships with related species (Aberer et al. 2013). The presence of rogue species often leads to lower bootstrap support values for other nodes on the phylogenetic tree and can even result in polytomy (i.e., a situation in a phylogenetic tree where an ancestral node gives rise to three or more descendant lineages). The software *RogueNaRok* (Aberer et al. 2013; <https://github.com/aberer/RogueNaRok>) can identify rogue species through an exclusion method, whereby the support of the consensus tree improves after removing certain species from the bootstrap trees. Additionally, the *AliGROOVE* software (Kück, Meid et al. 2014) can analyze whether certain genes or taxa exhibit predominantly randomized sequence similarity compared to the rest of the species, which may cause the species to appear in incorrect positions on the phylogenetic

tree. M. R. Smith (2022) developed another rogue species detection method based on information theory, which is more effective than RogueNaRok in improving the reliability of the consensus tree.

Comparison of species trees

To obtain a robust phylogeny, we typically analyze multiple datasets or use various methods. These approaches often yield phylogenies with different topologies, branch lengths, and support values. It is thus necessary to identify the discrepancies and congruence among these phylogenies, and choose one or several optimal ones for subsequent analysis.

Tree support values. Maximum likelihood-based phylogenetic tree construction methods calculate a log-likelihood value for the inferred tree, with higher values indicating stronger support (Goldman et al. 2000). The posterior probability from Bayesian methods is a similar measurement. To increase the chance of obtaining a global optimal tree, multiple rounds of the analysis are usually performed with the same dataset, and then the tree with the highest likelihood value will be selected, as implemented in the parameter `--run` of IQ-TREE. Additionally, IQ-TREE and RAxML-NG can also compute the log-likelihood values, corresponding AIC, AICc, and BIC scores for user-provided trees through model optimization (`--opt-model`) and branch length optimization (`--opt-branches`). An optimal tree may be obtained by choosing the one with lowest AIC, AICc, or BIC scores. On the other hand, ASTRAL calculates the normalized quartet score for a species tree, i.e., the fraction of input gene tree quartet trees that are consistent with the species tree, where higher scores indicate stronger support. Low scores may indicate high levels of ILS or other factors causing discrepancies between gene trees and species trees.

Topological distance. The Robinson-Foulds topological distance (RF distance) (D. F. Robinson et al. 1979) can be used to measure the topological similarity between two phylogenetic trees, where smaller RF values indicate higher similarity. This value can be calculated using the `ape` package (Paradis 2019).

Topology tests. This involves testing for significant differences in support for various phylogenetic trees within a dataset. Users may conduct tree searches under different topological constraints, such as enforcing the grouping of specific species into a predefined clade, and then test for significant differences in support for various phylogenetic trees within a dataset. Popular tests include the Kishino-Hasegawa test (KH test) (Hasegawa et al. 1989; Kishino et al. 1989), the Shimodaira-Hasegawa test (SH test) (Shimodaira and Hasegawa 1999), and the Approximately Unbiased test (AU test) (Shimodaira 2002). If a tree has a P-value < 0.05, it is rejected by the data.

Assessing species tree robustness

Gene tree and site concordance factors (gCF and sCF). Studies on taxa like birds show that even with nearly 100% bootstrap support, tree topologies often differ across studies. High bootstrap values mainly reflect low sampling variance in large datasets, with values increasing as sample size grows, but gCF and sCF become insensitive to further increases in loci once a certain sample size is reached. DiscoVista (Sayyari et al., 2018) can be used for gene tree support calculation, while IQ-TREE can calculate both gCF and sCF. The gCF range is 0 to 100%, where 0% means the topology is absent from all gene trees, while 100% means the topology is present in every gene tree (Minh, Hahn et al. 2020). The sCF range is typically 33% to 100%, where a 33%, indicates that the data do not significantly support any particular quartet tree topology (Minh, Hahn et al. 2020). Low gCF value may result from either gene tree estimation error (e.g., the insufficient phylogenetic signal in short MSAs or short branch lengths due to rapid diversification)

or biological reasons (e.g. ILS, gene flow) (Minh, Hahn et al. 2020). When ILS is the only cause for gene tree incongruence, gCF and sCF values should be similar. Gene tree estimation errors can result in a significantly lower gCF compared to sCF. A combination of gene tree estimation error and high level of ILS, or missing data, or gene flow may lead to low values of both gCF and sCF. For a more detailed explanation, please refer to Minh, Hahn et al. (2020) and http://www.robertlanfear.com/blog/files/concordance_factors.html.

Testing the distribution of phylogenetic signals and systematic errors. Phylogenetic inference using maximum likelihood can be disproportionately influenced by a few highly impactful sites or genes with extreme likelihood values across different resolutions of a split (Brown et al. 2017; Shen, Hittinger et al. 2017). In contrast, maximum parsimony-based inference, which assigns equal weight to all sites without considering likelihood differences, can mitigate the influence of such sites (Minh, Hahn et al. 2020). Consequently, it is good practice to identify these highly influential genes and sites and examine whether they produce topologies that differ from those inferred using the remaining genes or sites. For instance, genes can be categorized based on their support for different topologies, the level of saturation or evolutionary rates, proportion of missing data, number of informative sites, alignment length, presence of intra-locus recombination and taxon occupancy. Similarly, sites can be categorized and analyzed based on factors such as missing data, gaps, and Shannon entropy. Shannon entropy measures uncertainty: higher entropy indicates greater uncertainty and randomness, suggesting more information is needed to resolve the event. A recent study (Ballesteros et al. 2022) has shown that genes supporting arachnid monophyly tend to have shorter alignment lengths and fewer informative sites. Additionally, sites supporting arachnid monophyly exhibit higher Shannon entropy and simpler, more random structures. Therefore, the monophyly of Arachnida was refuted (Ballesteros et al. 2022).

Gene tree and species tree discordance

Differences between species trees and gene trees can result from historical events such as homoplasy, introgression, and rapid diversification (Stiller and G. Zhang 2019). Factors leading to homoplasy include convergent evolution, parallel evolution, and evolutionary reversal. Processes like horizontal gene transfer and hybridization can cause certain genes to introgress from one species into another. Rapid and successive speciation events within a relatively short time period can result in incomplete lineage sorting (ILS). This happens because new species do not have enough time to accumulate new mutations, and ancestral genetic variation is retained across successive speciation events, which is more pronounced in ancestor species with larger effective population sizes (Stiller and G. Zhang 2019).

Molecular phylogenetic trees often differ from traditional morphology-based trees, leading to disputes when molecular results challenge long-established evolutionary views. In such cases, molecular trees are frequently dismissed as artifacts. Notable examples include debates over phylogenetic relationships at the base of clades like Metazoa, Aves, Angiospermae, and Euchelicerata. To address these conflicts, it is crucial to re-evaluate the dataset and analysis process, ensuring that potential issues like long-branch attraction (LBA) affecting certain taxa are identified and mitigated. Tools like TreeShrink Mai et al. 2018 can be used for the detection. The presence of paralogs (misidentified ortholog) may also need to be re-evaluated. If molecular and morphological trees remain incongruent after ruling out data and methodological issues, it suggests that certain morphological traits may be due to homoplasy.

Finally, the following methods can be considered to improve the reliability of phylogenetic inference: (1) Increase taxonomic sampling density. Ontano et al. (2021) showed that denser sampling at the base of clades can more effectively resolve long-branch attraction (LBA) issues in the Pseudoscorpion phylo-

genetic analysis compared to different tree-building strategies (concatenation or coalescence) and data processing methods (e.g., data completeness, evolutionary rate, model choice). (2) Choosing appropriate evolutionary models, for instance, by adopting better partitioning schemes and models accounting for data heterogeneity. (3) Testing different types of data (e.g., protein-coding genes, non-coding regions) and choose appropriate datasets with more informative sites, and different methods to minimize the impact of gene flow and other factors (Stiller and G. Zhang 2019). Please refer to previous sections for possible solutions.

2.12 Estimation of species divergence times

The molecular clock theory posits that the rate of molecular evolution remains relatively constant during species diversification. This allows for the calibration of the phylogenetic tree with fossil dates, enabling the inference of divergence times of related lineages by serving as time constraints on the phylogeny (Bromham et al. 2003; Donoghue et al. 2007). There are two ways to use fossils for species divergence time calibration: the traditional node-dating method and the tip-dating method (also known as total-evidence dating).

Node-dating

While fossils themselves do not indicate the actual timings of speciation events, in the node-dating process, fossil dates provide precise minimum constraints (“hard” lower bound) but poor maximum constraints (“soft” upper bound) of such events (Benton et al. 2007). Probability distribution models (e.g., normal, lognormal, exponential, and uniform distributions) are then used to describe the possible origin dates of a clade within the bounds of the maximum and minimum constraints. However, it is often advisable to allow a low probability (e.g., 5%) of dates extending beyond the maximum constraint to account for uncertainty of finding an older fossil belonging to the same clade in question (thus it is called “soft” upper bound) (Benton et al. 2007; Donoghue et al. 2007; Drummond et al. 2006). The minimum age constraint should correspond to the age of the clade’s oldest known fossil, while the soft maximum age constraint should be based on the age of fossils from the clade’s (immediate) outgroups or stem groups (Benton et al. 2007; Parham et al. 2012).

Tip-dating

Unlike the node-dating method, in the tip-dating process, fossil taxa are treated as sampled taxa as extant species, whose phylogenetic positions will be determined based on their morphological traits and that of extant lineages (Ronquist, Klopfstein et al. 2012). Therefore, all fossils with morphological traits available can be included in the analysis, removing the need to choose proper fossils to set the minimum and maximum age constraints (A. Luo et al. 2020; Ronquist, Klopfstein et al. 2012). The information of both fossils and extant lineages, as well the age information of the fossils is incorporated into the fossilized birth–death-like models, which have parameters including speciation rate, extinction rate, fossil recovery rate, taxonomic sampling fraction of extant lineages, origin time, root age and crown age (Gavryushkina et al. 2014; Heath et al. 2014; Ogilvie, Vaughan et al. 2018; Stadler 2010; Stadler et al. 2018; C. Zhang, Stadler et al. 2016). This method has recently been assessed by A. Luo et al. (2020) using simulated data. The authors showed that the number and maximum age of fossil occurrences significantly influence estimates of node times and tree topology, more so than the degree of among-lineage rate variation or the

number of morphological characters (A. Luo et al. 2020). This method provides accurate age estimates for both the root and crown groups; however, the precision of these estimates depends on the probability of fossil occurrence (A. Luo et al. 2020). Omitting morphological characters leads to a slight overestimation of node ages, while excluding nucleotide sequences adversely affects the inferred tree topology (A. Luo et al. 2020).

The use of fossil evidence involves uncertainties related to both taxonomy and geology, making the selection of appropriate fossil evidence a critical issue, particularly in the node-dating method. Parham et al. (2012) proposed key criteria for fossil selection: (1) Ingroup and outgroup Fossils: Consider all well-identified fossils from both the ingroup taxa (taxa of interest) and their closest outgroups. (2) Phylogenetic relationships: Summarize the phylogenetic relationships using apomorphy-based diagnostics or recent fossil studies. Apomorphies are traits newly acquired by the ingroup and inherited from their most recent common ancestor (MRCA). These traits help identify monophyletic groups, while shared traits with outgroups are termed plesiomorphies. In node dating, fossils should be assigned to internal nodes (ancestral species) based on apomorphies rather than solely on taxonomic classifications, as misidentified fossils or taxonomic inaccuracies could result in fossils being placed in the wrong groups. (3) Geological ages and calibration: Document the geological ages and stratigraphic ranges of fossils, ensuring alignment with the latest International Chronostratigraphic Chart. The reliability of molecular clock estimates improves logarithmically with an increased number of fossil calibration points (Benton et al. 2007).

Widely used software for inferring species divergence times using Bayesian methods includes MCMCtree (Z. Yang 2007), BEAST, and BEAST2. Detailed tutorials can be found in Dos Reis et al. 2019 and at <https://beast2-dev.github.io/beast-docs/beast2/DivergenceDating/DivergenceDatingTutorial.html> and <https://taming-the-beast.org/tutorials/>.

Finally, integrating information on species trees, divergence times, trait evolution, biogeographical history and environmental factors enables a comprehensive understanding of evolutionary processes.

2.13 Perspectives

Reduced representation sequencing methods, including RNA-Seq, TCS, and lcWGS, have a wide range of applications in the study of animals, plants, and other organisms. Currently, these data are predominantly generated using short-read sequencing platforms. However, short-read sequencing has inherent limitations, such as GC bias during the experimental process and the challenge of accurately aligning short sequences to repetitive regions in the reference genome (e.g., pseudogenes and multi-copy genes) (Alkan et al. 2011; Oszlak et al. 2011). Additionally, repeat regions in the genome are difficult to assemble using short reads alone. The advent of long-read sequencing technologies, such as PacBio's Single-Molecule Real-Time (SMRT) sequencing (Chin et al. 2013) and Oxford Nanopore Technologies' (ONT) nanopore sequencing (Laver et al. 2015), effectively addresses these issues. For instance, long-read sequencing enables the direct sequencing of RNA molecules (or cDNA molecules) to their full length without requiring assembly, thereby facilitating the more accurate identification of diverse transcript types (Jain et al. 2018). With ongoing technological advancements, the accuracy of data generated by these sequencing platforms continues to improve. For example, the accuracy of PacBio's HiFi reads and the recently released ONT R10.4 flow cells has reached 99.9% and 99%, respectively (Sereika et al. 2022). Additionally, the introduction of higher-throughput instruments, such as PacBio's Sequel II System and ONT's PromethION sequencer, has further reduced the cost of per-base sequencing, creating new opportunities for researchers to undertake larger-scale and higher-quality phylogenetic studies. Phylogenetic analysis based on whole

genomes enables the construction of more robust phylogenies and offers detailed insights into species' evolutionary history, including processes such as introgression, hybridization, the genetic basis of species divergence, and the evolutionary potential of species to adapt to rapid climate change (e.g., Ferraj et al. 2023; Liao et al. 2023).

However, the generation of larger datasets also introduces new challenges, particularly in data storage, transmission, and the development of algorithms and software for phylogenetic research. Therefore, the recently introduced Read2Tree (Dylus et al. 2024) and ROADIES (Gupta et al. 2024) pipelines try to speed up phylogenetic analysis by bypassing contig assembly and ortholog assignment. Moreover, CASTER (C. Zhang, Nielsen et al. 2025) allows for accurate species tree inference based on quartet site patterns and the MSC model (thus accounting for ILS) from the whole-genome alignments of several hundreds of mammals directly, which is at least 60 times faster than previous methods. Model selection, phylogenetic inference and divergence times estimation are all computationally intensive. For instance, while the Bayesian framework achieves high accuracy of phylogenetic inference by incorporating various information into the model and priors, its intensive computational cost has restricted its application to small datasets and a limited number of species. To address these issues, several machine-learning methods have been developed in recent years. For instance, ModelTeller achieves speeds 7 to 80 times faster than existing methods for selecting the optimal model, without compromising accuracy (Abadi, Avram et al. 2020; <https://modelteller.tau.ac.il/>). The Genomatnn program leverages a convolutional neural network (CNN) to detect adaptive introgression (Gower et al. 2021; <https://github.com/grahamgower/genomatnn/>), while the CoalNN program employs a CNN to estimate coalescence times for different individuals and variants (Nait Saada et al. 2023; <https://palamaralab.github.io/software/coalnn/>). Additionally, phyloGAN (M. L. Smith et al. 2023; <https://github.com/meganlsmith/phyloGAN/>) uses generative adversarial networks (GANs) to infer phylogenetic relationships among species. In conclusion, emerging sequencing technologies and advanced algorithms have the potential to significantly deepen our understanding of species' evolutionary history in the near future.

References

- Ababneh, F., L. S. Jermini, C. Ma and J. Robinson (2006),
Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences,
Bioinformatics **22** 1225.
- Ababneh, F., L. S. Jermini and J. Robinson (2006), *Generation of the exact distribution and simulation of matched nucleotide sequences on a phylogenetic tree*,
Journal of Mathematical Modelling and Algorithms **5** 291.
- Abadi, S., O. Avram, S. Rosset, T. Pupko and I. Mayrose (2020),
ModelTeller: model selection for optimal phylogenetic reconstruction using machine learning,
Mol Biol Evol **37** 3338.
- Abadi, S., D. Azouri, T. Pupko and I. Mayrose (2019),
Model selection may not be a mandatory step for phylogeny reconstruction, Nat Commun **10** 934.
- Abbott, R., D. Albach, S. Ansell et al. (2013), *Hybridization and speciation*, J Evol Biol **26** 229.
- Aberer, A. J., D. Krompass and A. Stamatakis (2013),
Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice,
Syst Biol **62** 162.

- Alexander, D. H., J. Novembre and K. Lange (2009),
Fast model-based estimation of ancestry in unrelated individuals, Genome Res **19** 1655.
- Ali, R. H., M. Bogusz and S. Whelan (2019),
Identifying clusters of high confidence homologies in multiple sequence alignments, Mol Biol Evol **36** 2340.
- Alkan, C., S. Sajjadian and E. E. Eichler (2011),
Limitations of next-generation genome sequence assembly, Nat Methods **8** 61.
- Allen-Savietta, C. (2020),
Estimating Phylogenetic Networks from Concatenated Sequence Alignments, PhD thesis: The University of Wisconsin - Madison 165.
- Amemiya, C. T., J. Alföldi, A. P. Lee et al. (2013),
The African coelacanth genome provides insights into tetrapod evolution, Nature **496** 311.
- Andrews, S. (2010), *FastQC: a quality control tool for high throughput sequence data*.
- Arakawa, K., N. Kono, A. D. Malay et al. (2022),
1000 spider silkomes: Linking sequences to silk physical properties, Sci Adv **8** eabo6043.
- Arnold, M. L. (2016), *Divergence with genetic exchange*, Oxford University Press.
- Arvestad, L., J. Lagergren and B. Sennblad (2009),
The gene evolution model and computing its associated probabilities, Journal of the ACM (JACM) **56** 1.
- Ballesteros, J. A., C. E. Santibáñez-López, C. M. Baker et al. (2022), *Comprehensive Species Sampling and Sophisticated Algorithmic Approaches Refute the Monophyly of Arachnida*, Mol Biol Evol **39** msac021.
- Banos, H., T. K. Wong, J. Daneau et al. (2024),
GTRpmix: a linked general time-reversible model for profile mixture models, Mol Biol Evol **41** msae174.
- Barry, D. and J. A. Hartigan (1987), *Statistical analysis of hominoid molecular evolution*, Statistical Science **2** 191.
- Benton, M. J. and P. C. J. Donoghue (2007), *Paleontological evidence to date the tree of life*, Mol Biol Evol **24** 26.
- Blanquart, S. and N. Lartillot (2008),
A site- and time-heterogeneous model of amino acid replacement, Mol Biol Evol **25** 842.
- Blischak, P. D., J. Chifman, A. D. Wolfe and L. S. Kubatko (2018),
HyDe: a Python package for genome-scale hybridization detection, Syst Biol **67** 821.
- Bolger, A. M., M. Lohse and B. Usadel (2014),
Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics **30** 2114.
- Bromham, L. and D. Penny (2003), *The modern molecular clock*, Nat Rev Genet **4** 216.
- Brown, J. M. and R. C. Thomson (2017), *Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses*, Syst Biol **66** 517.
- Browning, B. L. and S. R. Browning (2013),
Improving the accuracy and efficiency of identity-by-descent detection in population data, Genetics **194** 459.
- Bryant, D., N. Galtier and M.-A. Poursat (2005), “Likelihood Calculation in Molecular Phylogenetics”, *Mathematics of Evolution and Phylogeny*, Oxford University Press, ISBN: 9780198566106.

-
- Cai, C., E. Tihelka, M. Giacomelli et al. (2022),
Integrated phylogenomics and fossil data illuminate the evolution of beetles,
 R Soc Open Sci **9** 211771.
- Camacho, C., G. Coulouris, V. Avagyan et al. (2009), *BLAST+: architecture and applications*,
 BMC Bioinformatics **10** 421.
- Cao, Z., X. Liu, H. A. Ogilvie, Z. Yan and L. Nakhleh (2023),
 “CHAPTER 5 Practical Aspects of Phylogenetic Network Analysis Using PhyloNet”,
A Guide to Methods and Applications, ed. by L. Kubatko and L. L. Knowles,
 Princeton: Princeton University Press 89, ISBN: 9780691245157.
- Capella-Gutierrez, S., J. M. Silla-Martinez and T. Gabaldon (2009),
trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses,
 Bioinformatics **25** 1972.
- Castresana, J. (2000),
Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis,
 Mol Biol Evol **17** 540.
- Chaw, S.-M., Y.-C. Liu, Y.-W. Wu et al. (2019),
Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution,
 Nat Plants **5** 63.
- Chen, Y., Y. Chen, C. Shi et al. (2018), *SOAPnuke: a MapReduce acceleration-supported software
 for integrated quality control and preprocessing of high-throughput sequencing data*,
 Gigascience **7** gix120.
- Chifman, J. and L. Kubatko (2014), *Quartet inference from SNP data under the coalescent model*,
 Bioinformatics **30** 3317.
- Chin, C.-S., D. H. Alexander, P. Marks et al. (2013),
Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data,
 Nat Methods **10** 563.
- Cock, P. J. A., C. J. Fields, N. Goto, M. L. Heuer and P. M. Rice (2010), *The Sanger FASTQ file format
 for sequences with quality scores, and the Solexa/Illumina FASTQ variants*,
 Nucleic Acids Res **38** 1767.
- Consortium, i5K (2013), *The i5K Initiative: advancing arthropod genomics for knowledge, human
 health, agriculture, and the environment*, J Hered **104** 595.
- Criscuolo, A. and S. Gribaldo (2010), *BMGE (Block Mapping and Gathering with Entropy): a new
 software for selection of phylogenetic informative regions from multiple sequence alignments*,
 BMC Evol Biol **10** 210.
- Crotty, S. M., B. Q. Minh, N. G. Bean et al. (2020),
GHOST: recovering historical signal from heterotachously evolved sequence alignments,
 Syst Biol **69** 249.
- Danecek, P., J. K. Bonfield, J. Liddle et al. (2021), *Twelve years of SAMtools and BCFtools*,
 Gigascience **10** giab008.
- Darriba, D., D. Posada, A. M. Kozlov et al. (2020),
ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models,
 Mol Biol Evol **37** 291.
- Daskalakis, C. and S. Roch (2016),
 “Species trees from gene trees despite a high rate of lateral genetic transfer: A tight bound”,

- Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SIAM 1621.
- Davey, J. W., P. A. Hohenlohe, P. D. Etter et al. (2011), *Genome-wide genetic marker discovery and genotyping using next-generation sequencing*, *Nat Rev Genet* **12** 499.
- Davidson, R., P. Vachaspati, S. Mirarab and T. Warnow (2015), *Phylogenomic species tree estimation in the presence of incomplete lineage sorting and horizontal gene transfer*, *BMC Genomics* **16** 1.
- Dayhoff, M., R. Schwartz and B. Orcutt (1978), “A model of evolutionary change in proteins”, *Atlas of Protein Sequence and Structure*, ed. by M. Dayhoff, vol. 5, Washington, D. C.: National Biomedical Research Foundation 345.
- De Bruijn, N. G. (1946), *A combinatorial problem*, *Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam* **49** 758.
- Dell’Ampio, E., K. Meusemann, N. U. Szucsich et al. (2014), *Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects*, *Mol Biol Evol* **31** 239.
- Di Franco, A., R. Poujol, D. Baurain and H. Philippe (2019), *Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences*, *BMC Evol Biol* **19** 1.
- Dobin, A., C. A. Davis, F. Schlesinger et al. (2013), *STAR: ultrafast universal RNA-seq aligner*, *Bioinformatics* **29** 15.
- Donoghue, P. C. J. and M. J. Benton (2007), *Rocks and clocks: calibrating the Tree of Life using fossils and molecules*, *Trends Ecol Evol* **22** 424.
- Dos Reis, M. and Z. Yang (2019), *Bayesian molecular clock dating using genome-scale datasets*, *Evolutionary genomics: Statistical and computational methods* **1910** 309.
- Douglas, J., C. L. Jiménez-Silva and R. Bouckaert (2022), *StarBeast3: adaptive parallelized Bayesian inference under the multispecies coalescent*, *Syst Biol* **71** 901.
- Drummond, A. J., S. Y. W. Ho, M. J. Phillips and A. Rambaut (2006), *Relaxed phylogenetics and dating with confidence*, *PLoS Biol* **4** e88.
- Du, S., E. Tihelka, D. Yu et al. (2024), *Revisiting the four Hexapoda classes: Protura as the sister group to all other hexapods*, *Proc Natl Acad Sci U S A* **121** e2408775121.
- Duchene, D. A., C. Mather Niklas and Van Der Wal and S. Y. W. Ho (2022), *Excluding Loci With Substitution Saturation Improves Inferences From Phylogenomic Data*, *Syst Biol* **71** 676.
- Dutheil, J. and B. Boussau (2008), *Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs*, *BMC Evol Biol* **8** 1.
- Dylus, D., A. Altenhoff, S. Majidian, F. J. Sedlazeck and C. Dessimoz (2024), *Inference of phylogenetic trees directly from raw sequencing reads using Read2Tree*, *Nat Biotechnol* **42** 139.
- Eddy, S. R. (2011), *Accelerated profile HMM searches*, *PLoS Comput Biol* **7** e1002195.
- Edgar, R. C. (2004), *MUSCLE: multiple sequence alignment with high accuracy and high throughput*, *Nucleic Acids Res* **32** 1792.

-
- Eldridge, M. D. B., J. E. Deakin, A. J. MacDonald et al. (2020), *The Oz Mammals Genomics (OMG) initiative: developing genomic resources for mammal conservation at a continental scale*, Aust Zool **40** 505.
- Ewels, P., M. Magnusson, S. Lundin and M. Käller (2016), *MultiQC: summarize analysis results for multiple tools and samples in a single report*, Bioinformatics **32** 3047.
- Faircloth, B. C. (2016), *PHYLUCE is a software package for the analysis of conserved genomic loci*, Bioinformatics **32** 786.
- Feng, S., M. Bai, I. Rivas-González et al. (2022), *Incomplete lineage sorting and phenotypic evolution in marsupials*, Cell **185** 1646.
- Feng, S., J. Stiller, Y. Deng et al. (2020), *Dense sampling of bird diversity increases power of comparative genomics*, Nature **587** 252.
- Ferraj, A., P. A. Audano, P. Balachandran et al. (2023), *Resolution of structural variation in diverse mouse genomes reveals chromatin remodeling due to transposable elements*, Cell Genomics **3**.
- Fleming, J. F. and T. H. Struck (2023), *nRCFV: a new, dataset-size-independent metric to quantify compositional heterogeneity in nucleotide and amino acid datasets*, BMC Bioinformatics **24** 145.
- Flouri, T., X. Jiao, B. Rannala and Z. Yang (2020), *A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis*, Mol Biol Evol **37** 1211.
- Fontaine, M. C., J. B. Pease, A. Steele et al. (2015), *Extensive introgression in a malaria vector species complex revealed by phylogenomics*, Science **347** 1258524.
- Foster, P. G., D. Schrempf, G. J. Szöllősi et al. (2023), *Recoding amino acids to a reduced alphabet may increase or decrease phylogenetic accuracy*, Syst Biol **72** 723.
- Foster, P. G. (2004), *Modeling compositional heterogeneity*, Syst Biol **53** 485.
- Frankel, L. E. and C. Ané (2023), *Summary tests of introgression are highly sensitive to rate variation across lineages*, Syst Biol **72** 1357.
- Gavryushkina, A., D. Welch, T. Stadler and A. J. Drummond (2014), *Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration*, PLoS Comput Biol **10** e1003919.
- Giacomelli, M., M. E. Rossi, J. Lozano-Fernandez, R. Feuda and D. Pisani (2022), *Resolving tricky nodes in the tree of life through amino acid recoding*, Iscience **25**.
- Goldman, N., J. P. Anderson and A. G. Rodrigo (2000), *Likelihood-based tests of topologies in phylogenetics*, Syst Biol **49** 652.
- Good, I. J. (1946), *Normal recurring decimals*, J Lond Math Soc **1** 167.
- Good, J. M. (2011), *Reduced representation methods for subgenomic enrichment and next-generation sequencing*, Molecular methods for evolutionary genetics **772** 85.
- Gower, G., P. I. Picazo, M. Fumagalli and F. Racimo (2021), *Detecting adaptive introgression in human evolution using convolutional neural networks*, Elife **10** e64669.
- Grabherr, M. G., B. J. Haas, M. Yassour et al. (2011), *Full-length transcriptome assembly from RNA-Seq data without a reference genome*, Nat Biotechnol **29** 644.

- Green, R. E., J. Krause, A. W. Briggs et al. (2010), *A draft sequence of the Neandertal genome*, science **328** 710.
- Guindon, S., J.-F. Dufayard, V. Lefort et al. (2010), *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0*, Syst Biol **59** 307.
- Gupta, A., S. Mirarab and Y. Turakhia (2024), *Accurate, scalable, and fully automated inference of species trees from raw genome assemblies using ROADIES*, bioRxiv 2024.
- Hamlin, J. A. P., M. S. Hibbins and L. C. Moyle (2020), *Assessing biological factors affecting postspeciation introgression*, Evol Lett **4** 137.
- Harrison, R. G. and E. L. Larson (2014), *Hybridization, introgression, and the nature of species boundaries*, J Hered **105** 795.
- Hasegawa, M. and H. Kishino (1989), *Confidence limits of the maximum-likelihood estimate of the hominoid three from mitochondrial-DNA sequences*, Evolution **672**.
- Heaps, S. E., T. M. Nye, R. J. Boys, T. A. Williams and T. M. Embley (2014), *Bayesian modelling of compositional heterogeneity in molecular phylogenetics*, Stat Appl Genet Mol Biol **13** 589.
- Heath, T. A., J. P. Huelsenbeck and T. Stadler (2014), *The fossilized birth–death process for coherent calibration of divergence-time estimates*, Proc Natl Acad Sci U S A **111** E2957.
- Heled, J. and A. J. Drummond (2009), *Bayesian inference of species trees from multilocus data*, Mol Biol Evol **27** 570.
- Hernandez, A. M. and J. F. Ryan (2021), *Six-state amino acid recoding is not an effective strategy to offset compositional heterogeneity and saturation in phylogenetic analyses*, Syst Biol **70** 1200.
- Higgs, P. G. (1998), *Compensatory neutral mutations and the evolution of RNA*, Genetica **102** 91.
- Hill, M., B. Legried and S. Roch (2022), *Species tree estimation under joint modeling of coalescence and duplication: sample complexity of quartet methods*, The Annals of Applied Probability **32** 4681.
- Hillis, D. M. and M. T. Dixon (1991), *Ribosomal DNA: molecular evolution and phylogenetic inference*, The Quarterly review of biology **66** 411.
- Höhna, S., M. J. Landis, T. A. Heath et al. (2016), *RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language*, Syst Biol **65** 726.
- Hoyal Cuthill, J. F., N. Guttenberg and G. E. Budd (2020), *Impacts of speciation and extinction measured by an evolutionary decay clock*, Nature **588** 636.
- Huang, X., X.-G. Chen and P. A. Armbruster (2016), *Comparative performance of transcriptome assembly methods for non-model organisms*, BMC Genomics **17** 1.
- Huerta-Sánchez, E., X. Jin, Asan et al. (2014), *Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA*, Nature **512** 194.
- Jain, M., S. Koren, K. H. Miga et al. (2018), *Nanopore sequencing and assembly of a human genome with ultra-long reads*, Nat Biotechnol **36** 338.
- Jayaswal, V., L. S. Jermin and J. Robinson (2005), *Estimation of phylogeny using a general Markov model*, Evolutionary Bioinformatics **1** 62.
- Jehl, P., F. Sievers and D. G. Higgins (2015), *OD-seq: outlier detection in multiple sequence alignments*, BMC Bioinformatics **16** 1.

-
- Jensen, A., F. Swift, D. de Vries et al. (2023),
Complex evolutionary history with extensive ancestral gene flow in an African primate radiation,
Mol Biol Evol **40** msad247.
- Jiang, H., R. Lei, S.-W. Ding and S. Zhu (2014),
Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads,
BMC Bioinformatics **15** 1.
- Kainer, D. and R. Lanfear (2015), *The effects of partitioning on phylogenetic inference*,
Mol Biol Evol **32** 1611.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermiin (2017),
ModelFinder: fast model selection for accurate phylogenetic estimates, *Nat Methods* **14** 587.
- Katoh, K. and D. M. Standley (2013), *MAFFT multiple sequence alignment software version 7: improvements in performance and usability*, *Mol Biol Evol* **30** 772.
- Kim, D., J. M. Paggi, C. Park, C. Bennett and S. L. Salzberg (2019),
Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype,
Nat Biotechnol **37** 907.
- Kim, D., G. Pertea, C. Trapnell et al. (2013), *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions*, *Genome Biol* **14** 1.
- Kingman, J. F. C. (1982a), *On the genealogy of large populations*,
Journal of applied probability **19** 27.
- Kingman, J. F. C., G. Koch and F. Spizzichino (1982),
Exchangeability and the evolution of large populations,
Exchangeability in probability and statistics **91** 112.
- Kingman, J. F. C. (1982b), *The coalescent*, *Stoch Process Their Appl* **13** 235.
- Kishino, H. and M. Hasegawa (1989), *Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea*,
J Mol Evol **29** 170.
- Knowles, L. L., H. C. Lanier, P. B. Klimov and Q. He (2012), *Full modeling versus summarizing gene-tree uncertainty: method choice and species-tree accuracy*, *Mol Phylogenet Evol* **65** 501.
- Kong, S. and L. S. Kubatko (2021),
Comparative performance of popular methods for hybrid detection using genomic data,
Syst Biol **70** 891.
- Koppetsch, T., M. Malinsky and M. Matschiner (2024),
Towards reliable detection of introgression in the presence of among-species rate variation,
Syst Biol **73** 769.
- Kozlov, A. M., D. Darriba, T. Flouri, B. Morel and A. Stamatakis (2019), *RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference*, *Bioinformatics* **35** 4453.
- Kubatko, L. S. and J. H. Degnan (2007),
Inconsistency of phylogenetic estimates from concatenated data under coalescence,
Syst Biol **56** 17.
- Kück, P., S. A. Meid, C. Groß, J. W. Wägele and B. Misof (2014),
AliGROOVE—visualization of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support, *BMC Bioinformatics* **15** 1.
- Kück, P. and T. H. Struck (2014), *BaCoCa—a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions*, *Mol Phylogenet Evol* **70** 94.

- Kuznetsov, D., F. Tegenfeldt, M. Manni et al. (2023), *OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity*, *Nucleic Acids Res* **51** D445.
- Lam, H. M., O. Ratmann and M. F. Boni (2018), *Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm*, *Mol Biol Evol* **35** 247.
- Lanfear, R., B. Calcott, S. Y. Ho and S. Guindon (2012), *PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses*, *Mol Biol Evol* **29** 1695.
- Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld and B. Calcott (2017), *PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses*, *Mol Biol Evol* **34** 772.
- Larkin, M. A., G. Blackshields, N. P. Brown et al. (2007), *Clustal W and Clustal X version 2.0*, *Bioinformatics* **23** 2947.
- Larsson, A. (2014), *AliView: a fast and lightweight alignment viewer and editor for large datasets*, *Bioinformatics* **30** 3276.
- Lartillot, N. and H. Philippe (2004), *A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process*, *Mol Biol Evol* **21** 1095.
- Lartillot, N., N. Rodrigue, D. Stubbs and J. Richer (2013), *PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment*, *Syst Biol* **62** 611.
- Laver, T., J. Harrison, P. A. O'Neill et al. (2015), *Assessing the performance of the oxford nanopore technologies minion*, *Biomol Detect Quantif* **3** 1.
- Leaché, A. D., M. K. Fujita, V. N. Minin and R. R. Bouckaert (2014), *Species delimitation using genome-wide SNP data*, *Syst Biol* **63** 534.
- Leaché, A. D. and B. Rannala (2011), *The accuracy of species tree estimation under simulation: a comparison of methods*, *Syst Biol* **60** 126.
- Lefort, V., J.-E. Longueville and O. Gascuel (2017), *SMS: smart model selection in PhyML*, *Mol Biol Evol* **34** 2422.
- Legried, B., E. K. Molloy, T. Warnow and S. Roch (2021), *Polynomial-time statistical estimation of species trees under gene duplication and loss*, *J Comput Biol* **28** 452.
- Lemey, P., M. Salemi and A.-M. Vandamme (2009), *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*, Cambridge University Press 133.
- Lemmon, A. R., J. M. Brown, K. Stanger-Hall and E. M. Lemmon (2009), *The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference*, *Syst Biol* **58** 130.
- Levy Karin, E., M. Mirdita and J. Söding (2020), *MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics*, *Microbiome* **8** 1.
- Lewin, H. A., G. E. Robinson, W. J. Kress et al. (2018), *Earth BioGenome Project: Sequencing life for the future of life*, *Proc Natl Acad Sci U S A* **115** 4325.
- Li, B., N. Fillmore, Y. Bai et al. (2014), *Evaluation of de novo transcriptome assemblies from RNA-Seq data*, *Genome Biol* **15** 1.

-
- Li, H. (2023), *Protein-to-genome alignment with miniprot*, *Bioinformatics* **39** btad014.
- Li, H. and R. Durbin (2009),
Fast and accurate short read alignment with Burrows–Wheeler transform, *Bioinformatics* **25** 1754.
- Li, R., Y. Li, K. Kristiansen and J. Wang (2008), *SOAP: short oligonucleotide alignment program*, *Bioinformatics* **24** 713.
- Li, R., H. Zhu, J. Ruan et al. (2010),
De novo assembly of human genomes with massively parallel short read sequencing, *Genome Res* **20** 265.
- Li, Y., J. L. Steenwyk, Y. Chang et al. (2021), *A genome-scale phylogeny of the kingdom Fungi*, *Curr Biol* **31** 1653.
- Liao, W.-W., M. Asri, J. Ebler et al. (2023), *A draft human pangenome reference*, *Nature* **617** 312.
- Liu, B. and T. Warnow (2023),
Weighted ASTRID: fast and accurate species trees from weighted internode distances, *Algorithms Mol Biol* **18** 6.
- Liu, L. (2008), *BEST: Bayesian estimation of species trees under the coalescent model*, *Bioinformatics* **24** 2542.
- Liu, L. and L. Yu (2011), *Estimating species trees from unrooted gene trees*, *Syst Biol* **60** 661.
- Liu, L., L. Yu and S. V. Edwards (2010),
A maximum pseudo-likelihood approach for estimating species trees under the coalescent model, *BMC Evol Biol* **10** 1.
- Liu, L., L. Yu, L. Kubatko, D. K. Pearl and S. V. Edwards (2009),
Coalescent methods for estimating phylogenetic trees, *Mol Phylogenet Evol* **53** 320.
- Liu, L., L. Yu, D. K. Pearl and S. V. Edwards (2009),
Estimating species phylogenies using coalescence times among sequences, *Syst Biol* **58** 468.
- Liu, S., C. Zhou, G. Meng et al. (2022),
Evolution and diversification of Mountain voles (Rodentia: Cricetidae), *Commun Biol* **5** 1417,
DOI: [10.1038/s42003-022-04371-z](https://doi.org/10.1038/s42003-022-04371-z).
- Lopez, P., D. Casane and H. Philippe (2002), *Heterotachy, an important process of protein evolution*, *Mol Biol Evol* **19** 1.
- Luo, A., D. A. Duchêne, C. Zhang, C.-D. Zhu and S. Y. W. Ho (2020),
A simulation-based evaluation of tip-dating under the fossilized birth–death process, *Syst Biol* **69** 325.
- Luo, R., B. Liu, Y. Xie et al. (2012),
SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, *Gigascience* **1** 2047.
- Lutteropp, S., C. Scornavacca, A. M. Kozlov, B. Morel and A. Stamatakis (2022),
NetRAX: accurate and fast maximum likelihood phylogenetic network inference, *Bioinformatics* **38** 3725.
- Maddison, W. P. (1997), *Gene trees in species trees*, *Syst Biol* **46** 523.
- Mai, U. and S. Mirarab (2018), *TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees*, *BMC Genomics* **19** 23.
- Malinsky, M., R. J. Challis, A. M. Tyers et al. (2015),
Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake, *Science* **350** 1493.

- Manni, M., M. R. Berkeley, M. Seppey, F. A. Simão and E. M. Zdobnov (2021), *BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes*, *Mol Biol Evol* **38** 4647.
- Markin, A. and O. Eulenstein (2021), *Quartet-based inference is statistically consistent under the unified duplication-loss-coalescence model*, *Bioinformatics* **37** 4064.
- Martin, J., V. M. Bruno, Z. Fang et al. (2010), *Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads*, *BMC Genomics* **11** 1.
- Martin, J. A. and Z. Wang (2011), *Next-generation transcriptome assembly*, *Nat Rev Genet* **12** 671.
- Martin, M. (2011), *Cutadapt removes adapter sequences from high-throughput sequencing reads*, *EMBnet. journal* **17** 10.
- Martin, S. H., J. W. Davey and C. D. Jiggins (2015), *Evaluating the use of ABBA–BABA statistics to locate introgressed loci*, *Mol Biol Evol* **32** 244.
- Mayer, C., M. Sann, A. Donath et al. (2016), *BaitFisher: a software package for multispecies target DNA enrichment probe design*, *Mol Biol Evol* **33** 1875.
- McCormack, J. E., W. L. E. Tsai and B. C. Faircloth (2016), *Sequence capture of ultraconserved elements from bird museum specimens*, *Mol Ecol Resour* **16** 1189.
- Mendes, F. K. and M. W. Hahn (2018), *Why concatenation fails near the anomaly zone*, *Syst Biol* **67** 158.
- Meng, G., C. Zhou, S. Liu and X. Zhou (2022), “Phylogenomics based on transcriptomics and short-read sequences”, *Zoological Systematics and Evolution Protocol eBook*, ed. by G. Qiao, M. Zhu, W. Wang et al., 1st ed., Bio-protocol LLC 793, doi: [10.21769/BioProtoc.1010649](https://doi.org/10.21769/BioProtoc.1010649).
- Metsky, H. C., K. J. Siddle, A. Gladden-Young et al. (2019), *Capturing sequence diversity in metagenomes with comprehensive and scalable probe design*, *Nat Biotechnol* **37** 160.
- Minh, B. Q., M. W. Hahn and R. Lanfear (2020), *New methods to calculate concordance factors for phylogenomic datasets*, *Mol Biol Evol* **37** 2727.
- Minh, B. Q., H. A. Schmidt, O. Chernomor et al. (2020), *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era*, *Mol Biol Evol* **37** 1530.
- Mirarab, S., L. Nakhleh and T. Warnow (2021), *Multispecies coalescent: theory and applications in phylogenetics*, *Annu Rev Ecol Evol Syst* **52** 247.
- Misof, B., S. Liu, K. Meusemann et al. (2014), *Phylogenomics resolves the timing and pattern of insect evolution*, *Science* **346** 763.
- Mistry, J., S. Chuguransky, L. Williams et al. (2021), *Pfam: The protein families database in 2021*, *Nucleic Acids Res* **49** D412.
- Molloy, E. K. and T. Warnow (2020), *FastMulRFS: fast and accurate species tree estimation under generic gene duplication and loss models*, *Bioinformatics* **36** i57.
- Moran, B. M., C. Payne, Q. Langdon et al. (2021), *The genomic consequences of hybridization*, *Elife* **10** e69016.

-
- Morel, B., P. Schade, S. Lutteropp et al. (2022), *SpeciesRax: a tool for maximum likelihood species tree inference from gene family trees under duplication, transfer, and loss*, Mol Biol Evol **39** msab365.
- Morel, B., T. A. Williams and A. Stamatakis (2023), *Asteroid: a new algorithm to infer species trees from gene trees under high proportions of missing data*, Bioinformatics **39** btac832.
- Nait Saada, J., Z. Tsangalidou, M. Stricker and P. F. Palamara (2023), *Inference of coalescence times and variant ages using convolutional neural networks*, Mol Biol Evol **40** msad211.
- Naser-Khdour, S., B. Q. Minh, W. Zhang, E. A. Stone and R. Lanfear (2019), *The prevalence and impact of model violations in phylogenetic analysis*, Genome Biol Evol **11** 3341.
- Nekrutenko, A. and W.-H. Li (2000), *Assessment of compositional heterogeneity within and between eukaryotic genomes*, Genome Res **10** 1986.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler and B. Q. Minh (2015), *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies*, Mol Biol Evol **32** 268.
- Nguyen, Q. and T. Roos (2015), “Likelihood-based inference of phylogenetic networks from sequence data by phylodag”, *Algorithms for Computational Biology: Second International Conference, AlCoB 2015, Mexico City, Mexico, August 4-5, 2015, Proceedings 2*, Springer 126.
- Nichols, R. (2001), *Gene trees and species trees are not the same*, Trends Ecol Evol **16** 358.
- Ogilvie, H. A., R. R. Bouckaert and A. J. Drummond (2017), *StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates*, Mol Biol Evol **34** 2101.
- Ogilvie, H. A., T. G. Vaughan, N. J. Matzke et al. (2018), *Inferring species trees using integrative models of species evolution*, bioRxiv 242875.
- Okonechnikov, K., O. Golosova, M. Fursov and U. Team (2012), *Unipro UGENE: a unified bioinformatics toolkit*, Bioinformatics **28** 1166.
- One Thousand Plant Transcriptomes Initiative (2019), *One thousand plant transcriptomes and the phylogenomics of green plants*, Nature **574** 679.
- Ontano, A. Z., G. Gainett, S. Aharon et al. (2021), *Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions*, Mol Biol Evol **38** 2446.
- Ozsolak, F. and P. M. Milos (2011), *RNA sequencing: advances, challenges and opportunities*, Nat Rev Genet **12** 87.
- Pagel, M. and A. Meade (2005), “Mixture Models in Phylogenetic Inference”, *Mathematics of Evolution and Phylogeny*, Oxford University Press, ISBN: 9780198566106.
- Pang, X.-X. and D.-Y. Zhang (2024), *Detection of Ghost Introgression Requires Exploiting Topological and Branch Length Information*, Syst Biol **73** 207.
- Paradis Emmanuel and Schliep, K. (2019), *ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R*, Bioinformatics **35** 526.
- Parham, J. F., P. C. Donoghue, C. J. Bell et al. (2012), *Best practices for justifying fossil calibrations*, Syst Biol **61** 346.

- Patel, R. K. and M. Jain (2012),
NGS QC Toolkit: a toolkit for quality control of next generation sequencing data,
PloS One **7** e30619.
- Patterson, N., P. Moorjani, Y. Luo et al. (2012), *Ancient admixture in human history*,
Genetics **192** 1065.
- Pease, J. B. and M. W. Hahn (2015),
Detection and polarization of introgression in a five-taxon phylogeny, Syst Biol **64** 651.
- Peng, Y., H. C. M. Leung, S.-M. Yiu et al. (2013), *IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels*, Bioinformatics **29** i326.
- Peng, Y., H. C. Leung, S.-M. Yiu and F. Y. Chin (2012), *IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth*, Bioinformatics **28** 1420.
- Penny, D., B. J. McComish, M. A. Charleston and M. D. Hendy (2001),
Mathematical elegance with biochemical realism: the covarion model of molecular evolution,
J Mol Evol **53** 711.
- Pertea, M., G. M. Pertea, C. M. Antonescu et al. (2015),
StringTie enables improved reconstruction of a transcriptome from RNA-seq reads,
Nat Biotechnol **33** 290.
- Petersen, M., K. Meusemann, A. Donath et al. (2017), *Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes*, BMC Bioinformatics **18** 1.
- Philippe, H. and P. Lopez (2001), *On the conservation of protein sequences in evolution*,
Trends Biochem Sci **26** 414.
- Philippe, H., Y. Zhou, H. Brinkmann, N. Rodrigue and F. Delsuc (2005),
Heterotachy and long-branch attraction in phylogenetics, BMC Evol Biol **5** 1.
- Phillips, M. J. and D. Penny (2003),
The root of the mammalian tree inferred from whole mitochondrial genomes,
Mol Phylogenet Evol **28** 171.
- Phillips, M. J., F. Delsuc and D. Penny (2004),
Genome-scale phylogeny and the detection of systematic biases, Mol Biol Evol **21** 1455.
- Pickrell, J. K. and D. Reich (2014),
Toward a new history and geography of human genes informed by ancient DNA,
Trends Genet **30** 377.
- Pons, J., T. G. Barraclough, J. Gomez-Zurita et al. (2006),
Sequence-based species delimitation for the DNA taxonomy of undescribed insects,
Syst Biol **55** 595.
- Pritchard, J. K., M. Stephens and P. Donnelly (2000),
Inference of population structure using multilocus genotype data, Genetics **155** 945.
- Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus and A. Korobeynikov (2020),
Using SPAdes De Novo Assembler, Curr Protoc Bioinformatics **70** e102.
- Puillandre, N., S. Brouillet and G. Achaz (2021), *ASAP: assemble species by automatic partitioning*,
Mol Ecol Resour **21** 609.
- Puillandre, N., A. Lambert, S. Brouillet and G. Achaz (2012),
ABGD, Automatic Barcode Gap Discovery for primary species delimitation, Mol Ecol **21** 1864.
- Rasmussen, M. D. and M. Kellis (2012),
Unified modeling of gene duplication, loss, and coalescence using a locus tree,
Genome Res **22** 755.

-
- Reich, D., K. Thangaraj, N. Patterson, A. L. Price and L. Singh (2009), *Reconstructing Indian population history*, Nature **461** 489.
- Reichenberger, E. R., G. Rosen, U. Hershberg and R. Hershberg (2015), *Prokaryotic nucleotide composition is shaped by both phylogeny and the environment*, Genome Biol Evol **7** 1380.
- Rhie, A., S. A. McCarthy, O. Fedrigo et al. (2021), *Towards complete and error-free genome assemblies of all vertebrate species*, Nature **592** 737.
- Roberto, F., D. Martin, P. Walker et al. (2017), *Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals*, Curr Biol **27** 3864.
- Robertson, G., J. Schein, R. Chiu et al. (2010), *De novo assembly and analysis of RNA-seq data*, Nat Methods **7** 909.
- Robinson, D. F. and L. R. Foulds (1979), “Comparison of weighted labelled trees”, *Combinatorial Mathematics VI*, ed. by W. D. Horadam A. F. and Wallis, Berlin, Heidelberg: Springer Berlin Heidelberg 119.
- Roch, S. and S. Snir (2013), *Recovering the treelike trend of evolution despite extensive lateral genetic transfer: a probabilistic analysis*, J Comput Biol **20** 93.
- Ronquist, F., S. Klopfstein, L. Vilhelmsen et al. (2012), *A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera*, Syst Biol **61** 973.
- Ronquist, F., M. Teslenko, P. Van Der Mark et al. (2012), *MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space*, Syst Biol **61** 539.
- Savill, N. J., D. C. Hoyle and P. G. Higgs (2001), *RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods*, Genetics **157** 399.
- Schrempf, D., N. Lartillot and G. Szöllősi (2020), *Scalable empirical mixture models that account for across-site compositional heterogeneity*, Mol Biol Evol **37** 3616.
- Schulz, M. H., D. R. Zerbino, M. Vingron and E. Birney (2012), *Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels*, Bioinformatics **28** 1086.
- Schwartz, S., W. J. Kent, A. Smit et al. (2003), *Human–mouse alignments with BLASTZ*, Genome Res **13** 103.
- Scientists, G. I. C. of (2009), *Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species*, J Hered **100** 659.
- Scornavacca, C. and N. Galtier (2017), *Incomplete lineage sorting in mammalian phylogenomics*, Syst Biol **66** 112.
- Sereika, M., R. H. Kirkegaard, S. M. Karst et al. (2022), *Oxford Nanopore R10. 4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing*, Nat Methods **19** 823.
- Shen, X.-X., C. T. Hittinger and A. Rokas (2017), *Contentious relationships in phylogenomic studies can be driven by a handful of genes*, Nat Ecol Evol **1** 0126.
- Shen, X.-X., J. L. Steenwyk and A. Rokas (2021), *Dissecting incongruence between concatenation-and quartet-based approaches in phylogenomic data*, Syst Biol **70** 997.

- Shimodaira, H. (2002), *An approximately unbiased test of phylogenetic tree selection*, Syst Biol **51** 492.
- Shimodaira, H. and M. Hasegawa (1999), *Multiple comparisons of log-likelihoods with applications to phylogenetic inference*, Mol Biol Evol **16** 1114.
- Si Quang, L., O. Gascuel and N. Lartillot (2008), *Empirical profile mixture models for phylogenetic reconstruction*, Bioinformatics **24** 2317.
- Simmons, M. P. and J. Kessenich (2020), *Divergence and support among slightly suboptimal likelihood gene trees*, Cladistics **36** 322.
- Simpson, J. T., K. Wong, S. D. Jackman et al. (2009), *ABYSS: a parallel assembler for short read sequence data*, Genome Res **19** 1117.
- Slater, G. S. C. and E. Birney (2005), *Automated generation of heuristics for biological sequence comparison*, BMC Bioinformatics **6** 1.
- Smith, B. T., W. M. Mauck III, B. W. Benz and M. J. Andersen (2020), *Uneven missing data skew phylogenomic relationships within the lories and lorikeets*, Genome Biol Evol **12** 1131.
- Smith, M. R. (2022), *Using information theory to detect rogue taxa and improve consensus trees*, Syst Biol **71** 1088.
- Smith, M. L. and M. W. Hahn (2023), *Phylogenetic inference using generative adversarial networks*, Bioinformatics **39** btad543.
- Solís-Lemus, C. and C. Ané (2016), *Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting*, PLoS Genet **12** e1005896.
- Solís-Lemus, C., P. Bastide and C. Ané (2017), *PhyloNetworks: A Package for Phylogenetic Networks*, Mol Biol Evol **34** 3292.
- Solís-Lemus, C., M. Yang and C. Ané (2016), *Inconsistency of species tree methods under gene flow*, Syst Biol **65** 843.
- Soucy, S. M., J. Huang and J. P. Gogarten (2015), *Horizontal gene transfer: building the web of life*, Nat Rev Genet **16** 472.
- Springer, M. S. and J. Gatesy (2016), *The gene tree delusion*, Mol Phylogenet Evol **94** 1.
- Stadler, T. (2010), *Sampling-through-time in birth–death trees*, J Theor Biol **267** 396.
- Stadler, T., A. Gavryushkina, R. C. M. Warnock, A. J. Drummond and T. A. Heath (2018), *The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes*, J Theor Biol **447** 41.
- Stamatakis, A. (2014), *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies*, Bioinformatics **30** 1312.
- Stanke, M., O. Keller, I. Gunduz et al. (2006), *AUGUSTUS: ab initio prediction of alternative transcripts*, Nucleic Acids Res **34** W435.
- Steenwyk, J. L., T. J. Buida III, Y. Li, X.-X. Shen and A. Rokas (2020), *ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference*, PLoS Biol **18** e3001007.
- Stiller, J., S. Feng, A.-A. Chowdhury et al. (2024), *Complexity of avian evolution revealed by family-level genomes*, Nature **629** 851.
- Stiller, J. and G. Zhang (2019), *Comparative phylogenomics, a stepping stone for bird biodiversity studies*, Diversity **11** 115.

-
- Strimmer, K. and A. von Haeseler (1997),
Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment,
 Proc Natl Acad Sci U S A **94** 6815.
- Struck, T. H. (2014), *TreSpEx—detection of misleading signal in phylogenetic reconstructions based on tree information*, Evolutionary Bioinformatics **10** EBO.
- Surget-Groba, Y. and J. I. Montoya-Burgos (2010),
Optimization of de novo transcriptome assembly from next-generation sequencing data,
 Genome Res **20** 1432.
- Susko, E. and A. J. Roger (2007), *On reduced amino acid alphabets for phylogenetic inference*,
 Mol Biol Evol **24** 2139.
- Suyama, M., D. Torrents and P. Bork (2006), *PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments*, Nucleic Acids Res **34** W609.
- Swofford, D. L. (2002), *PAUP: phylogenetic analysis using parsimony (and other methods)*, 4.0 beta,
<http://paup.csit.fsu.edu/>.
- Szöllősi, G. J., E. Tannier, V. Daubin and B. Boussau (2015),
The inference of gene trees with species trees, Syst Biol **64** e42.
- Tabei, Y., H. Kiryu, T. Kin and K. Asai (2008),
A fast structural multiple alignment method for long RNA sequences, BMC Bioinformatics **9** 1.
- Than, C., D. Ruths and L. Nakhleh (2008), *PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships*, BMC Bioinformatics **9** 1.
- Thomas, J. A., J. W. H. Trueman, A. Rambaut and J. J. Welch (2013), *Relaxed phylogenetics and the Palaeoptera problem: resolving deep ancestral splits in the insect phylogeny*, Syst Biol **62** 285.
- Trapnell, C., D. G. Hendrickson, M. Sauvageau et al. (2013),
Differential analysis of gene regulation at transcript resolution with RNA-seq,
 Nat Biotechnol **31** 46.
- Tricou, T., E. Tannier and D. M. de Vienne (2022),
Ghost lineages highly influence the interpretation of introgression tests, Syst Biol **71** 1147.
- Vachaspati, P. and T. Warnow (2015), *ASTRID: accurate species trees from internode distances*,
 BMC Genomics **16** 1.
- Wang, H.-C., B. Q. Minh, E. Susko and A. J. Roger (2018), *Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation*, Syst Biol **67** 216.
- Wang, X., Z. He, Z. Guo et al. (2022),
Extensive gene flow in secondary sympatry after allopatric speciation, Natl Sci Rev **9** nwac280.
- Wang, Y. and L. Nakhleh (2018),
Towards an accurate and efficient heuristic for species/gene tree co-estimation,
 Bioinformatics **34** i697.
- Wang, Y., H. A. Ogilvie and L. Nakhleh (2020), *Practical speedup of Bayesian inference of species phylogenies by restricting the space of gene trees*, Mol Biol Evol **37** 1809.
- Wascher, M. and L. Kubatko (2021), *Consistency of SVDQuartets and maximum likelihood for coalescent-based species tree estimation*, Syst Biol **70** 33.
- Wen, D. and L. Nakhleh (2018),
Coestimating reticulate phylogenies and gene trees from multilocus sequence data,
 Syst Biol **67** 439.

- Wen, D., Y. Yu and L. Nakhleh (2016), *Bayesian inference of reticulate phylogenies under the multispecies network coalescent*, PLoS Genet **12** e1006006.
- Wen, D., Y. Yu, J. Zhu and L. Nakhleh (2018), *Inferring phylogenetic networks using PhyloNet*, Syst Biol **67** 735.
- Whelan, S., I. Irisarri and F. Burki (2018), *PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences*, Bioinformatics **34** 3929.
- Whitney, K. D., K. W. Broman, N. C. Kane et al. (2015), *Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower*, Mol Ecol **24** 2194.
- Wiens, J. J. and M. C. Morrill (2011), *Missing data in phylogenetic analysis: reconciling results from simulations and empirical data*, Syst Biol **60** 719.
- Williams, T. A., S. E. Heaps, S. Cherlin et al. (2015), *New substitution models for rooting phylogenetic trees*, Philosophical Transactions of the Royal Society B: Biological Sciences **370** 20140336.
- Williams, T. A., D. Schrempf, G. J. Szöllősi et al. (2021), *Inferring the deep past from molecular data*, Genome Biol Evol **13** evab067.
- Wong, T. K. F., S. Kalyaanamoorthy, K. Meusemann et al. (2020), *A minimum reporting standard for multiple sequence alignments*, NAR Genom Bioinform **2** lqaa024.
- Wong, T. K., C. Cherryh, A. G. Rodrigo et al. (2024), *MAST: phylogenetic inference with mixtures across sites and trees*, Systematic biology **73** 375.
- Woodhams, M. D., J. Fernández-Sánchez and J. G. Sumner (2015), *A new hierarchy of phylogenetic models consistent with heterogeneous substitution rates*, Syst Biol **64** 638.
- Wright, E. S. and K. H. Vetsigian (2016), *Quality filtering of Illumina index reads mitigates sample cross-talk*, BMC Genomics **17** 1.
- Xie, Y., G. Wu, J. Tang et al. (2014), *SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads*, Bioinformatics **30** 1660.
- Yakimowski, S. B. and L. H. Rieseberg (2014), *The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology*, Am J Bot **101** 1247.
- Yan, Z., M. L. Smith, P. Du, M. W. Hahn and L. Nakhleh (2022), *Species tree inference methods intended to deal with incomplete lineage sorting are robust to the presence of paralogs*, Syst Biol **71** 367.
- Yang, Z. (1994), *Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods*, J Mol Evol **39** 306.
- Yang, Z. (1995), *A space-time process model for the evolution of DNA sequences*, Genetics **139** 993.
- Yang, Z. (2007), *PAML 4: phylogenetic analysis by maximum likelihood*, Mol Biol Evol **24** 1586.
- Yang, Z. (2015), *The BPP program for species tree estimation and species delimitation*, Curr Zool **61** 854.

-
- Yang, Z. and B. Rannala (2014),
Unguided species delimitation using DNA sequence data from multiple loci,
Mol Biol Evol **31** 3125.
- Yu, Y., R. M. Barnett and L. Nakhleh (2013),
Parsimonious inference of hybridization in the presence of incomplete lineage sorting,
Syst Biol **62** 738.
- Yu, Y., J. Dong, K. J. Liu and L. Nakhleh (2014),
Maximum likelihood inference of reticulate evolutionary histories,
Proc Natl Acad Sci U S A **111** 16448.
- Yu, Y. and L. Nakhleh (2015), *A maximum pseudo-likelihood approach for phylogenetic networks*,
BMC Genomics **16** 1.
- Zerbino, D. R. and E. Birney (2008),
Velvet: algorithms for de novo short read assembly using de Bruijn graphs, Genome Res **18** 821.
- Zhang, C. and S. Mirarab (2022),
Weighting by Gene Tree Uncertainty Improves Accuracy of Quartet-based Species Trees,
Mol Biol Evol **39** msac215.
- Zhang, C., R. Nielsen and S. Mirarab (2025),
CASTER: Direct species tree inference from whole-genome alignments, Science **387** eadk9688.
- Zhang, C., M. Rabiee, E. Sayyari and S. Mirarab (2018),
ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees,
BMC Bioinformatics **19** 153.
- Zhang, C., C. Scornavacca, E. K. Molloy and S. Mirarab (2020),
ASTRAL-Pro: quartet-based species-tree inference despite paralogy, Mol Biol Evol **37** 3292.
- Zhang, C., H. A. Ogilvie, A. J. Drummond and T. Stadler (2018),
Bayesian inference of species networks from multilocus sequence data, Mol Biol Evol **35** 504.
- Zhang, C., T. Stadler, S. Klopstein, T. A. Heath and F. Ronquist (2016),
Total-evidence dating under the fossilized birth–death process, Syst Biol **65** 228.
- Zhang, C., D. Lin, Y. Wang et al. (2019),
Widespread introgression in Chinese indigenous chicken breeds from commercial broiler,
Evol Appl **12** 610.
- Zhang, G. (2015), *Bird sequencing project takes off*, Nature **522** 34.
- Zhang, J., P. Kapli, P. Pavlidis and A. Stamatakis (2013),
A general species delimitation method with applications to phylogenetic placements,
Bioinformatics **29** 2869.
- Zhou, C., G. Meng and S. Liu (2020), “From Reads to Genes”,
Transcriptomics in Entomological Research, ed. by M. Shelomi, CABI 24.
- Zhu, J., D. Wen, Y. Yu, H. M. Meudt and L. Nakhleh (2018),
Bayesian inference of phylogenetic networks from bi-allelic genetic markers,
PLoS Comput Biol **14** e1005932.
- Zhu, T., T. Flouri and Z. Yang (2022), *A simulation study to examine the impact of recombination on phylogenomic inferences under the multispecies coalescent model*, Mol Ecol **31** 2814.
- Zou, L., E. Susko, C. Field and A. J. Roger (2012), *Fitting nonstationary general-time-reversible models to obtain edge-lengths and frequencies for the Barry–Hartigan model*, Syst Biol **61** 927.

Unraveling the Phylogeny of the Pholcidae Family Using Targeted Capture Sequencing

Guanliang Meng^{1,*}, Lars Podsiadlowski¹, Dimitar Dimitrov^{2,*}, and Bernhard A. Huber¹

¹ Zoological Research Museum Alexander Koenig, LIB, Bonn, Germany

² University Museum of Bergen, University of Bergen, Bergen, Norway

* Corresponding author: G.Meng@leibniz-lib.de; dimitard.gwu@gmail.com

3.1 Summary

This chapter and Chapter 4 have been submitted as:

G. Meng, L. Podsiadlowski, D. Dimitrov and B. A. Huber (2025), *The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders*, submitted (currently under review by *Systematic Biology*)

Summary

The daddy long-legs spiders, Pholcidae, currently consist of almost 2,000 described species (WSC 2025) and five formally recognized subfamilies (Arteminae, Smeringopinae, Pholcinae, Modisiminae, and Ninetinae) (Huber 2011), however, the subfamily-level phylogenetic relationships within the family have long been debated (Dimitrov, Astrin et al. 2013; Eberle et al. 2018; Huber, Eberle et al. 2018), preventing researchers from addressing questions such as the origin and evolution of many morphological traits and patterns of pholcid diversity. With the emergence of high-throughput sequencing (HTS) and the decreasing cost associated with it, researchers can now address such fundamental problems with a great amount of molecular data.

In this study, we collected a total of 96 pholcid species from all the five recognized subfamilies of the family. Previous studies showed that the sister-group relationships between Pholcinae and Smeringopinae, as well as Modisiminae and “other Arteminae” (excluding *Artema*), are supported by strong evidence (Eberle et al. 2018; Huber, Eberle et al. 2018). On the other hand, according to Huber, Eberle et al. (2018), the positions of Ninetinae, *Artema*, and *Priscula* were considered the three most notable gaps in pholcid phylogeny. Our sampling thus focused on the latter three groups. We applied the target capture sequencing method and the spider-specific probe set for ultra-conserved elements (UCE) developed by

Kulkarni, Wood et al. (2020) to our 96 pholcids samples. For phylogenetic analysis, we included *Scytodes* sp. (Scytodidae), *Maoriata* sp. (Orsolobidae), *Hypochilus gertschi* (Hypochilidae), and *Loxosceles reclusa* (Sicariidae) as outgroups (Kulkarni, Kallal et al. 2021), whose data were downloaded from NCBI.

We created six different matrices using various alignment trimming strategies, taxon coverage, and locus lengths. To mitigate the impact of saturation that may lead to long branch attraction, we removed any saturated loci from our analysis. Additionally, we employed RY-coding, selected compositionally homogeneous sites, and adopted a compositional heterogeneity-aware model (CAT-GTR+G4) in order to reduce the interference of compositional heterogeneity for phylogenetic inference. In total, we analyzed ten data matrices. During the tree construction process, we utilized both partitioned and unpartitioned maximum likelihood inference with popular tools such as IQ-TREE (Minh, Schmidt et al. 2020) and RAxML-NG (Kozlov et al. 2019). We also tested summary-based multi-species coalescence methods like ASTRAL (Zhang et al. 2018) and SVDquartets (Chifman et al. 2014), as well as Bayesian MSC-based analysis using PhyloBayes-MPI (Lartillot, Rodrigue et al. 2013).

Our analyses revealed that approximately 36.9% of the loci were affected by substitutional saturation, while only 10.7% of the loci violated the assumptions of stationarity and homogeneity in molecular evolutionary models. Likelihood mapping analysis also supported the high phylogenetic signal present in our data. Furthermore, compositional homogeneity tests conducted using IQ-TREE and RCFV scores from BaCoCa indicated a low level of compositional heterogeneity. Finally, only 5.5% of the total sequences across all loci were found to be affected by intra-locus recombination. Overall, these analyses demonstrated that our datasets are suitable for phylogenetic inference.

Our analyses consistently recovered the same subfamily phylogenetic relationships, independent of data matrices (except for two matrices with very limited sites) and tree construction methods used. Our results lend further support to the sister group relationship between Smeringopinae and Pholcinae as well as “other Arteminae” and Modisiminae, similar to what has been previously reported by the studies based on a few genetic markers and morphological characters (Bruvo-Madaric et al. 2005; Dimitrov, Astrin et al. 2013; Eberle et al. 2018; Huber 2011; Huber, Eberle et al. 2018).

Our results revealed that the conventional concept of the subfamily Arteminae is not monophyletic. Instead, we found that the genera *Artema* and *Priscula* form a sister group, which is united with the subfamily Ninetinae. The remaining genera traditionally included in Arteminae are provisionally classified as “other Arteminae”. The (Ninetinae, (*Artema*, *Priscula*)) clade forms a sister group to (“other Arteminae,” Modisiminae), while previous studies suggested Ninetinae is sister to all other pholcids. Further analysis revealed that approximately 43% of the genes and 61% of the sites supported a sister group relationship of *Artema* and *Priscula*. Additionally, about 65% of the genes supported the monophyly of (Ninetinae, (*Artema*, *Priscula*)). However, this topology presents difficulties in reconciling with morphological evidence, especially concerning sperm ultrastructure and genital mechanics. The closest relatives of Pholcidae feature a sperm transfer form called synspermia, which is also seen in Ninetinae. The derived form of the sperm transfer is called cleistospermia. According to our reconstructed phylogeny, synspermia evolved in Ninetinae independently, or the transition from synspermia to cleistospermia occurred in Pholcidae at least four times. Our results also suggest that the male intromittent organ, known as the procursus, present in both *Artema* and “other Arteminae,” may have evolved independently in each of the two groups or been lost in other groups repeatedly.

Finally, our analysis indicates that the most recent common ancestor (MRCA) of Pholcidae can be traced to about 117 million years ago. This coincides with a recent study on related spider families (Magalhaes et al. 2020) but remarkably later than proposed for Pholcidae by a previous study (Dimitrov, Astrin et al. 2013).

Personal contribution

This project was a collaborative effort. Bernhard A. Huber initiated the project and secured funding from the German Research Foundation (DFG; project HU980/12-1), while Lars Podsiadlowski and Dimitar Dimitrov were also involved in the funding acquisition. Bernhard A. Huber and Guanliang Meng further received financial support from the Alexander Koenig Stiftung, while Dimitar Dimitrov received additional support from the Norwegian Metacenter for Computational Science (NOTUR; project NN9601K). Bernhard A. Huber collected specimens and conducted initial taxonomic assignments based on morphological characteristics. Lars Podsiadlowski coordinated the molecular work, while Guanliang Meng and Sandra Kukowka performed molecular experiments including DNA isolation, library construction, and target-locus capture. Guanliang Meng analyzed the molecular data with assistance from Lars Podsiadlowski, Dimitar Dimitrov, and Bernhard A. Huber. The draft version of the manuscript was written by Guanliang Meng. All authors contributed to the revision and improvement of the manuscript.

3.2 Introduction

The spiders (order: Araneae) is one of the most species-rich animal groups, consisting of 134 families and 52,179 nominated species around the world, with around 1,000 new species being formally described per year in the past 10 years (<https://wsc.nmbe.ch/statistics/>; accessed date: June 17th, 2024). They have colonized all continents except the Antarctic and adapted to a wide range of habitats (Dimitrov and Hormiga 2021). The daddy-longlegs-spiders, Pholcidae, is one of the largest spider families worldwide and currently consists of five recognized subfamilies (Arteminae, Modisiminae, Ninetinae, Smeringopinae and Pholcinae) (Huber 2011; Huber, Eberle et al. 2018) and over 2000 nominated species (<https://wsc.nmbe.ch/statistics/>; access date: Jan 1st, 2025). Pholcids exhibit numerous notable traits and evolutionary characteristics, such as long legs associated with whirling as a response to visual predators, multiple convergence of microhabitat shifts even among closely related species, multiple convergent evolutionary losses of certain characters, a large number of sexual dimorphisms (Eberle et al. 2018; Huber 2021; Huber, Eberle et al. 2018). Some pholcids display distinctive features that are rare among spiders, like asymmetric genitalia and modifications in the ocular region, construct highly structured webs, and utilize sticky silk to wrap their prey (Deeleman-Reinhold 1986; Huber and Nuñez 2015; Huber, Nuñez and Leh Moi Ung 2016; Japyassú et al. 2004). These characteristics demonstrate the high ecological diversity of pholcids in terms of both morphology and behavior. However, challenges in evolutionary interpretation of these characteristics arise due to a lack of well-supported phylogeny (Huber, Eberle et al. 2018).

A few studies based on morphology and several genetic markers and extensive sampling have shed some important lights on our understanding of Pholcidae evolution. For instance, both the evidence from four genetic markers (*12S rDNA*, *16S rDNA*, *cytochrome c oxidase subunit I (COI)*, *28S rDNA*) and forty-five morphological traits (Bruvo-Madaric et al. 2005), and the evidence from 6-7 genetic markers (*wingless*, *Histone H3* and *18S rDNA*, plus the four markers just mentioned) (Dimitrov, Astrin et al. 2013; Eberle et al. 2018) consistently support a sister relationship between Arteminae and Modisiminae. Moreover, the sister group relationship between Pholcinae and Smeringopinae as well as the monophyly of Ninetinae, are both confirmed by Dimitrov, Astrin et al. (2013) and Eberle et al. (2018). However, the phylogenetic positions of Ninetinae and of the genera *Artema* Walckenaer, 1837, and *Priscula* Simon, 1893 are still under debated (Huber, Eberle et al. 2018). For example, Ninetinae was placed with Pholcinae by Bruvo-Madaric et al. (2005), but later it was suggested to be sister to all other pholcids by Dimitrov,

Astrin et al. (2013), or the group (Ninetinae, *Artema*) was sister of all other pholcids (Eberle et al. 2018). The latter grouping was also in contrast with Dimitrov, Astrin et al. (2013) who supported the grouping of *Artema* with ‘other Arteminae,’ supporting the traditional concept of Arteminae.

With the advancement and rapidly decreasing cost of sequencing technologies, researchers are now able to access the genomic resources for a myriad of clades, and therefore improve our understanding of biodiversity and address some controversial phylogenetic problems and test previously intractable evolutionary problems. Compared to other fauna (e.g., insects, vertebrates) (Eldridge et al. 2020; Misof et al. 2014; Stiller et al. 2024), there are very few studies on spiders conducted based on large-scale genomic sampling (Arakawa et al. 2022; Kulkarni, Kallal et al. 2021; Kulkarni, Wood et al. 2020). In this study, we performed the first genomic-scale experiment on the Pholcidae family, constructing a phylogeny based on target-captured ultra-conserved elements (UCEs). Specifically, we wanted to solve the phylogenetic placements of Ninetinae, *Artema* and *Priscula*, paving the way for studying the evolution of the remarkable traits and biogeographic history of pholcids.

3.3 Materials and methods

3.3.1 Taxonomic Sampling and Target Capture Sequencing

For the phylogenomic analyses, we selected 96 pholcid species (Table A.1) representing all five formally recognized subfamilies (Arteminae, Smeringopinae, Pholcinae, Modisiminae and Ninetinae) (Huber 2011), with a strong emphasis on those groups with unclear or dubious placement in the latest and largest molecular phylogeny of the family (Eberle et al. 2018). This concerned Ninetinae, *Artema* and *Priscula*, which were identified as the three “most obvious gaps” in pholcid phylogeny (Huber, Eberle et al. 2018). The sister-group relationships Pholcinae + Smeringopinae, and Modisiminae + ‘other Arteminae’ (i.e. the traditional Arteminae without *Artema*), appear beyond doubt (Eberle et al. 2018; Huber, Eberle et al. 2018); therefore, we have a weak representation of these four groups. We use an operational division of Pholcidae into seven subfamily-level groups: Pholcinae, Smeringopinae, Modisiminae, Ninetinae, *Artema*, ‘other Arteminae’, and *Priscula*. For the sake of simplicity, we will use ‘subfamilies’ as a shorthand for ‘subfamily-level groups’ throughout the manuscript; we use ‘Arteminae’ as a shorthand for ‘other Arteminae’, and ‘traditional Arteminae’ to address the original concept (*Artema* + ‘other Arteminae’). Vouchers of all analyzed species are deposited in ZFMK. Additionally, we chose the four families most closely related to Pholcidae as outgroups, following the topology of Kulkarni, Kallal et al. (2021). This includes the fastq data for three species, i.e., *Scytodes* sp. (Scytodidae), *Maoriata* sp. (Orsolobidae), *Hypochilus gertschi* (Hypochilidae), and the genome assembly of *Loxosceles reclusa* (Sicariidae), from NCBI, respectively. They were analyzed together with our 96 newly sequenced samples.

DNA was extracted with the Dneasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer’s instruction. Hybrid enrichment generally followed the approach and methods described in Sann et al. (2018). Indexed gDNA libraries for Illumina sequencing were constructed using the SureSelect XT HS2 Library Preparation Kit (Cat # 5500-0147, Lot # 0006583499, Agilent Technologies, USA), which were then hybridized to the spider-specific probe set for ultra-conserved elements (UCE) (Kulkarni, Wood et al. 2020) and amplified using the SureSelect XT HS2 Target Enrichment Kit (Cat # 5191-6690, Lot # 0006599172, Agilent Technologies, USA) following the manufacturer’s instruction. Finally, sequencing of captured gDNA libraries was performed by MacroGen Europe (Amsterdam, NL) using pair-end (PE) 150 bp strategy. Each specimen was sequenced with ca. 1.5 Gb of data. More details on the experiment are available in the Supplementary Text A.

3.3.2 Data Matrix Construction

To study the subfamily-level phylogenetic relationships of Pholcidae, we constructed 10 different matrices based on the UCE data (Figure 3.1).

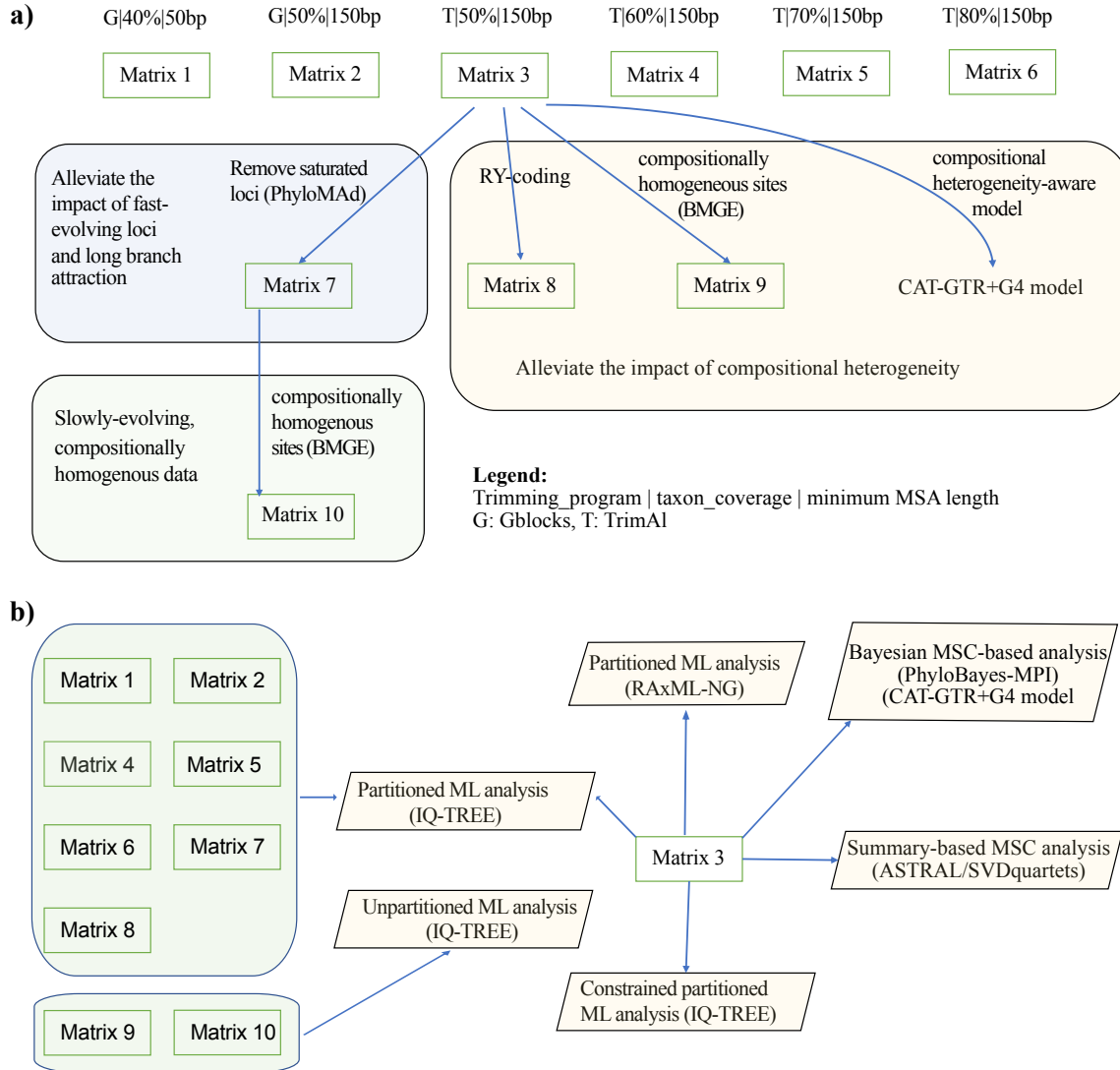


Figure 3.1. Analytical workflow of target capture sequencing data. (a) The creation of different data matrices based on different taxon coverage of genes. The impact of compositional heterogeneity and varying evolutionary rates are also considered (matrices 7 to 10). (b) Phylogenetic tree construction with different data matrices and strategies. MSA, multiple sequence alignment. ML, maximum likelihood. Text in parenthesis are program names.

Data quality control and sequence assembly. The raw data of all 100 samples except *L. reclusa* were filtered with fastp (v0.23.1) (Chen et al. 2018). Reads were filtered out if: (1) one read contains more than ten Ns; (2) More than 40% of bases are unqualified; (3) reads are duplicates; (4) reads are shorter

than 40 bp. Clean reads were assembled into contigs using SPAdes (v3.14.1) (Prjibelski et al. 2020) in Phyluce (v1.7.1) (Faircloth 2016). To detect potential cross-contaminations among species, “all-blast-all” analysis of the assembled contigs was performed with megablast in Blast+ (v2.9.0) (Camacho et al. 2009) under the default parameters. The blast hits with similarity $\geq 99\%$ and alignment length ≥ 100 bp were kept and further checked for biological sources by blasting their sequences against a local NT database. We used the following criteria to determine which sequences to keep or remove from the analysis: if the abundance difference (as output by the SPAdes program) between the query and subject sequences was less than 10X, both sequences were marked as “removed”. If both sequences had abundances $\geq 20X$ and their abundance difference exceeded 10X, the sequences with lower abundance was marked as “removed”. Any sequences previously marked as “removed” were excluded from the assembly.

Multiple-sequence-alignment (MSA). The UCE sequences of all samples (including *L. reclusa*) were extracted with Phyluce, aligned with MAFFT (v7.475) (Katoh et al. 2013) and trimmed with Gblocks (v0.91b) (Castresana 2000) (both implemented in Phyluce) and TrimAl (v1.4.rev15) (Capella-Gutierrez et al. 2009) (-automated 1), respectively.

Finally, six data matrices (matrices 1 to 6, see Figure 3.1) were generated based on different taxonomic occupancy (40%, 50%, 60%, 70%, and 80%) and locus length (≥ 50 bp vs. ≥ 150 bp). Another four data matrices (matrices 7 to 10) were generated based on matrix 3 to test if there are systematic biases in our data (see the Systematic bias assessment section and Figure 3.1).

3.3.3 Systematic Bias Assessment

Information content and substitutional saturation assessment. To detect systematic biases in our UCE datasets, the trimmed MSAs were evaluated for their phylogenetic information. For each locus, we calculated its average GC content, the interquartile range of GC content, the number and proportion of missing data (i.e., gaps and Ns), the number and proportion of parsimonious informative sites, the length of the MSA, the taxon number of the MSA, and a saturation plot using custom scripts. The density plots of these features as well as the missing data distribution plot were created using ggplot2 (v3.3.6) (Ginestet 2011) and R (v4.2.0) (R Core Team 2022). Substitutional saturation was also assessed by PhyloMAad (v1.2) (Duchene et al. 2022) based on measures of entropy. Loci with high risks of saturation were excluded from matrix 3 and thus matrix 7 was created (Figure 3.1).

Test of stationarity, homogeneity and treelikeness evolution. Many evolutionary models hold the assumptions of stationarity (constant nucleotide/amino-acid frequencies over time), reversibility (substitution rates in both directions are equal), homogeneity (constant substitution rates over time), and treelikeness (different sites of an MSA share the same evolutionary history). IQ-TREE (v2.2.0) (Minh, Schmidt et al. 2020) was used to test the assumptions of stationarity and homogeneity (–symtest-only). The two assumptions were rejected when $p - \text{values}(\text{SymPval}) < 0.01$ were found. The likelihood mapping method (Strimmer and Haeseler 1997) in IQ-TREE was used and 100,000 random quartets of taxa were drawn from the concatenated MSA of all loci and visualized in a single triangular graph. The more quartets there are located at the three corners of the triangular graph, the more suitable the data is to construct a well-resolved phylogeny (Strimmer and Haeseler 1997).

Test of base compositional heterogeneity. We calculated the compositional heterogeneity (relative composition frequency variability (RCFV)) with BaCoCa (v1.109) (Kück et al. 2014). RCFV is a measurement of absolute deviation from the mean nucleotide (or amino acid) frequency in the dataset. The higher the RCFV values, the higher the compositional heterogeneity, which can be seen in fast-evolving clades and may have negative impact on phylogenetic inference (Cai et al. 2022). Compositional homogeneity was

also checked with IQ-TREE using the chi2 test, in which a sequence is labeled as “failed” if its character composition differs significantly ($p - value < 0.05$) from the average composition of the alignment. To alleviate the impact of compositional heterogeneity on phylogenetic tree construction, we created a RY-coding dataset (matrix 8) from matrix 3 using the script at <https://raw.githubusercontent.com/ballesterus/Utensils/master/RYplace.py>. Additionally, we selected compositionally homogenous sites from the concatenated MSA of matrix 3 and created matrix 9 using the BMGE program (v1.12) (Criscuolo et al. 2010). Using the same method, we also created matrix 10 from matrix 7 (Figure 3.1).

Intra-locus recombination detection. Coalescence-based phylogenetic tree construction methods generally assume there is no recombination within each locus, while different loci are in linkage equilibrium (Zhu et al. 2022). We employed the program 3SEQ (v1.8.0) (Lam et al. 2018) to detect the significant presence of intra-locus recombination ($p - value < 0.05$) using a non-parametric test for mosaicism (Boni et al. 2007). Before the analysis, a $3000 * 3000 * 3000$ p-value table was pre-computed.

3.3.4 Phylogenetic Inference

To build a robust phylogenetic hypothesis, we used both concatenation- and coalescence-based methods using IQ-TREE (v2.2.0), RAXML-NG (v1.1) (Kozlov et al. 2019), ASTRAL (v5.7.8) (Zhang et al. 2018), SVDquartets (Chifman et al. 2014) implemented in PAUP* (v4a168) (Swofford 2002), and PhyloBayes-MPI (v1.8) (Lartillot, Rodrigue et al. 2013) (Figure 3.1).

Maximum likelihood (ML) and Bayesian inference (BI). We applied IQ-TREE to all data matrices except matrices 9 and 10 using a partitioned strategy (-p). For matrices 9 and 10, we applied the unpartitioned strategy due to their short lengths (8.6 Kbp). Best-fit models of molecular evolution were determined by ModelFinder (Kalyaanamoorthy et al. 2017) in IQ-TREE automatically (-m MFP+MERGE). To overcome local optima during heuristics, we performed ten independent runs on each dataset, with a smaller perturbation strength (-pers 0.2) and a larger number of stop iterations (-nstop 500) (Liu et al. 2024; L.-T. Nguyen et al. 2015). Branch supports were evaluated with 2000 ultrafast bootstrap (UFBoot) (Minh, M. A. T. Nguyen et al. 2013) considering potential model violations (-B 2000 -bnni -sampling GENESITE). The SH-aLRT branch test (Guindon et al. 2010) was performed using 2,000 bootstrap replicates.

For comparison, we also used both RAXML-NG (two runs) and PhyloBayes-MPI with matrix 3, which balances the number of loci and taxon coverage. Compared to IQ-TREE, RAXML-NG is less sensitive to different starting trees (Liu et al. 2024). During each RAXML-NG run, the GTR+R4+I+FO model, and 25 parsimony starting trees + 25 random starting trees were used to explore the tree space more thoroughly, and the number of bootstrap replicates was automatically determined (-bs-trees autoMRE) (Pattengale et al. 2010). The infinite mixture model, i.e., the CAT-GTR+G4 model (Lartillot and Philippe 2004), was used in our PhyloBayes-MPI analysis. It can model both rate-heterogeneity across sites and pattern-heterogeneity across sites (i.e., across-site compositional heterogeneity), and is thus less sensitive to systematic errors than classical empirical models (Lartillot, Rodrigue et al. 2013). It is increasingly used to resolve phylogenetic controversies despite of its intensive computational demands (e.g., Ballesteros et al. 2022; Cai et al. 2022). Two independent PhyloBayes-MPI runs were performed, and convergences of the continuous parameters of the model as well as the tree spaces were evaluated using the tracecomp and bpcomp commands in PhyloBayes-MPI. One-third of samplings of each run was used as burn-in, and a majority consensus tree was generated by combining the trees of the two runs with the bpcomp command.

Coalescence-based inference. We used ASTRAL and SVDquartets with matrix 3. In the ASTRAL analysis, individual gene trees were firstly estimated by IQ-TREE as described above. To reduce noise,

internal nodes of gene trees with low branch supports (≤ 30) were collapsed (Zhang et al. 2018) and ASTRAL was run with default parameters. ASTRAL is sensitive to gene tree estimation errors (Mirarab et al. 2021); therefore, we also used SVDquartets, bypassing the gene tree construction step. In the SVDquartets analysis, we evaluated all quartets (evalQuartets=all) under the multispecies coalescent model (treemodel=mscoalescent) which allows for gene tree discordance due to incomplete lineage sorting. The “Erik+2” normalization was used (plus2=yes) to consider the heterogeneity among sites and lineages (Fernandez-Sanchez et al. 2016). The QFM algorithm (Reaz et al. 2014) was used to search for an optimal tree under the maximum quartet compatibility algorithm (treeInf=QFM), and 200 multilocus bootstrap replicates were sampled (bootstrap=multilocus nreps=200).

Constrained tree search and topology test. Using matrix 3 and the same parameter settings of IQ-TREE as described above, we performed two constrained tree searches. Based on previous studies (Bruvo-Madaric et al. 2005; Dimitrov, Astrin et al. 2013; Eberle et al. 2018; Huber 2011; Huber, Eberle et al. 2018), we tested two hypotheses: (1) the traditional Arteminae (i.e., ‘other Arteminae’+ *Artema*) is a monophyletic group. (2) Ninetinae is sister to all other pholcids.

Finally, a topology test with the bootstrap proportion using the REL method (bp-REL) (Kishino et al. 1990) and Expected Likelihood Weight (c-ELW) (Strimmer and Rambaut 2002) were performed across all unconstrained and constrained trees using IQ-TREE. DiscoVista (Sayyari et al. 2018) was used to visualize the relative gene frequency support for different topologies, while the likelihood mapping method (Strimmer and Haeseler 1997) in IQ-TREE was applied to visualize the site support. All trees were visualized with iTOL (v6.5.8; <https://itol.embl.de/>) (Letunic et al. 2021).

Divergence time calibration. To date the tree we used eight calibration points based on fossil data from the PaleoDB database (<https://paleobiodb.org>) and publications (García-Villafuerte et al. 2020; Magalhaes et al. 2020; Penney 2006; Serrano-Sánchez et al. 2015; Wunderlich 1988) (Table A.2). We followed Magalhaes et al. 2020 for the justifications of fossil assignments; details can be found in Table A.2. The MCMCTREE program in PAML (v4.10.6) (Yang 2007) and node-dating strategy were applied to calibrate the divergence times based on fossil records following the recommendations in Parham et al. (2012). Five independent MCMCTree runs were performed and the convergence of MCMC (Markov Chain Monte Carlo) chains was checked.

3.4 Results

A total of 145 Gbp raw data (on average 1.5 Gbp per sample) was generated from the target capture sequencing of the 96 samples and has been submitted to the NCBI SRA database (Table A.1). From these, ten data matrices have been generated for phylogenetic analysis (Figure 3.1). The “all-blast-all” filtering removed 13.2 Mbp from the 100 samples (including four from NCBI). It is worth noting that blast hits do not necessarily mean real cross-contaminations; they can, for example, represent microbial DNA such as *Wolbachia*, *Rickettsia* and *Spiroplasma*, which are known endosymbionts of spiders (Goodacre et al. 2006), or highly conserved genomic regions, such as ribosomal DNA (rDNA) genes. Most samples had little missing data across loci (Figure A.1 and A.2)

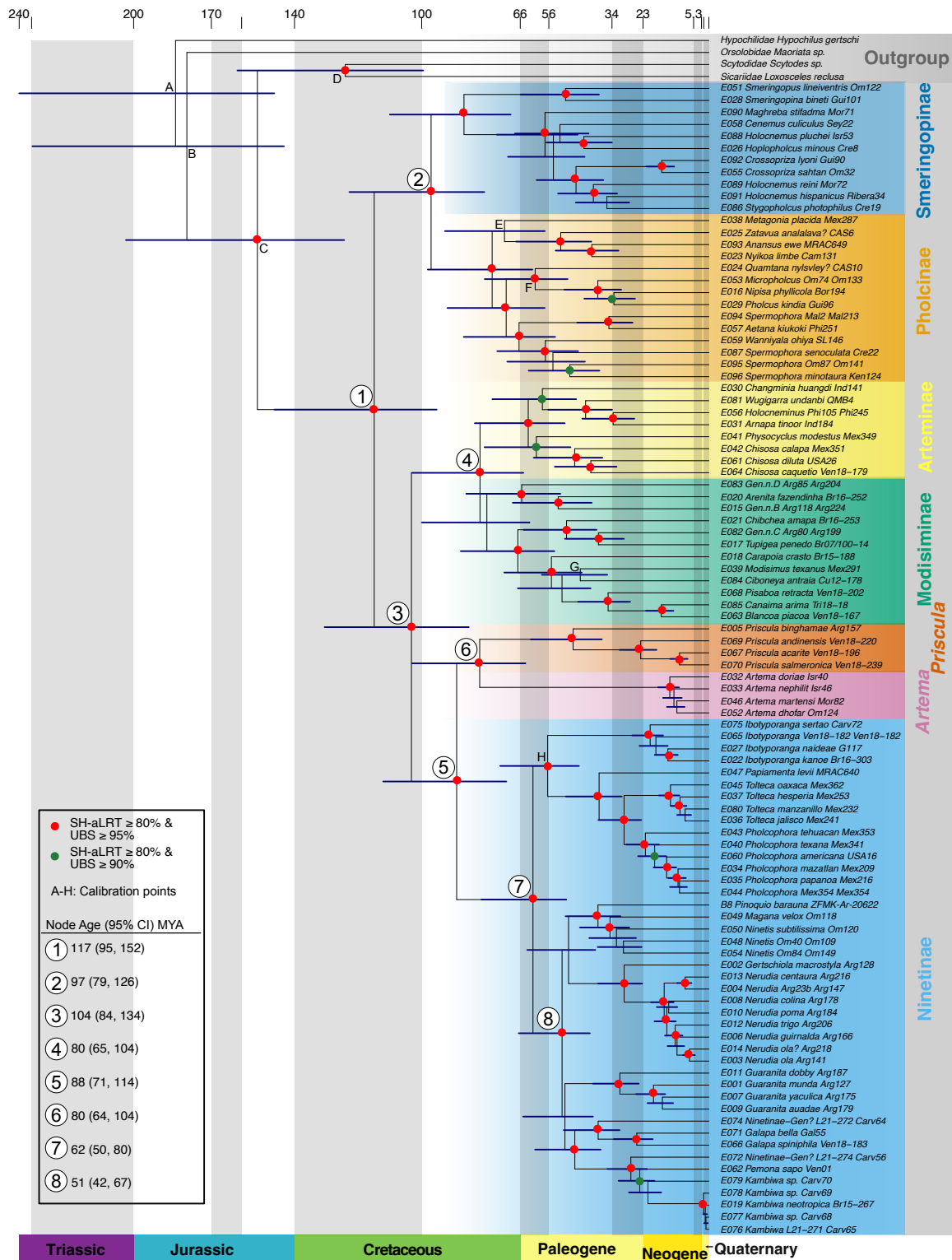


Figure 3.2. Maximum likelihood phylogenetic tree and divergence times of Pholcidae. The tree was constructed with IQ-TREE and data matrix 3 using a partitioned model in which all loci share the same set of branch lengths

(Continued) while each locus has its own evolution rate. Branch supports were evaluated with 2,000 ultrafast bootstrap (UBS) and SH-aLRT branch test based on 2,000 bootstrap replicates. MCMCTree program in PAML and node-dating strategy were applied to calibrate the divergence times based on fossil records.

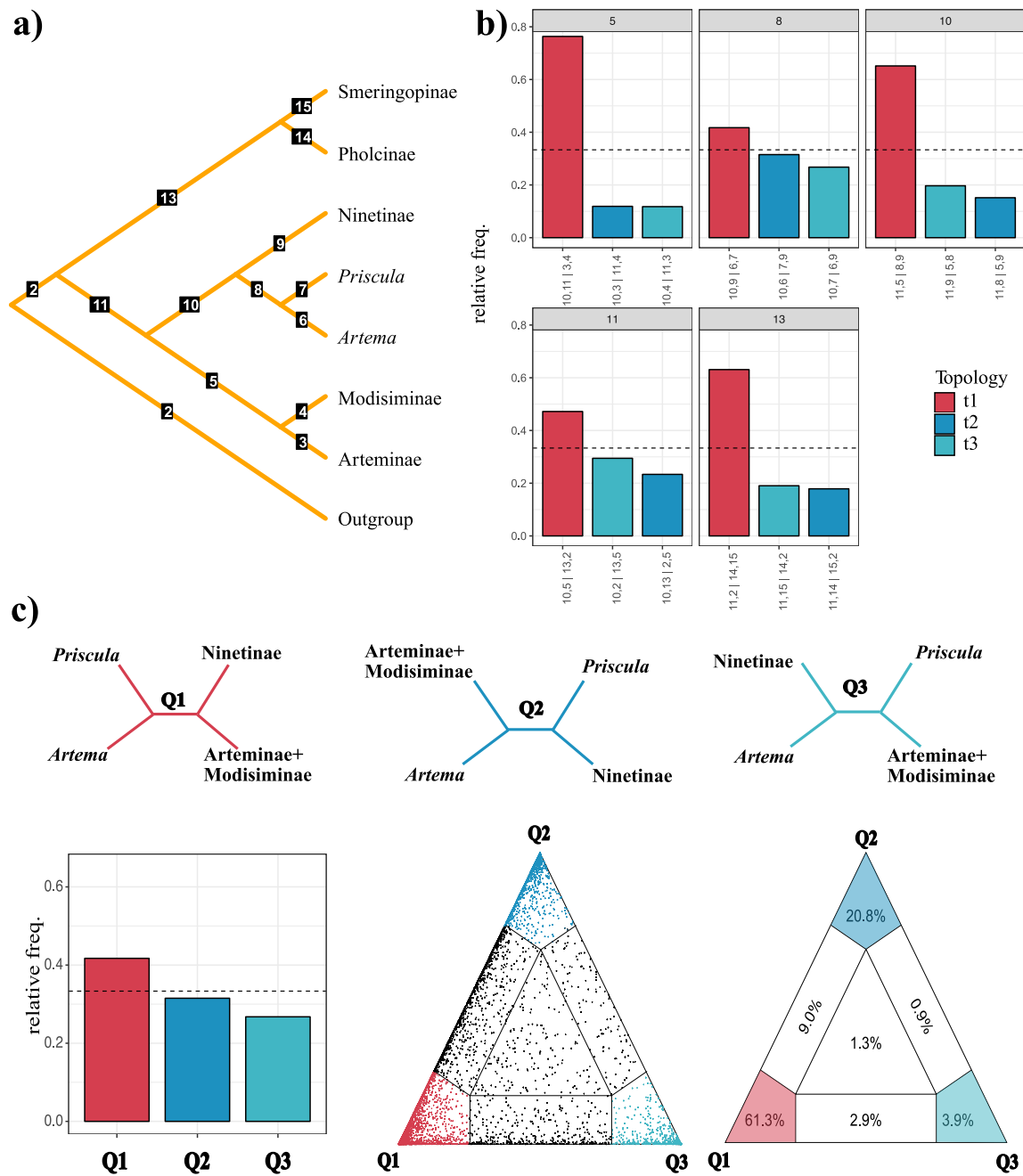
There were 122 loci in matrix 3 (total sites: 32,100 bp; total sequences: 8,704), and the length of the loci ranged approximately from 200 to 500 bp. The saturation analysis shows that 45 loci (36.9%) were saturated (Table A.3). The assumptions of stationarity and homogeneity for 13 loci (10.7%) were rejected ($p - \text{values}(\text{SymPval}) < 0.01$) (Table A.4). The likelihood mapping plot (Figure A.3) shows that there are equal numbers of quartets (32.7%) located at the three corners of the triangular graph, indicating the data is suitable to construct a well-resolved phylogeny (Strimmer and Haeseler 1997). The compositional homogeneity test by IQ-TREE shows that there were only 61 sequences (0.7%) whose character composition significantly deviates from the average composition of the alignment (Figure A.4). Similarly, RCFV scores from BaCoCa ranges from 0 to 0.0081 (average: 0.00052), which also indicates a low level of compositional heterogeneity (Kocot et al. 2017) (Figure A.5). However, taxon-wise, the RCFV scores (0.0168 - 0.0695) of all taxa are quite high, indicating a high level of compositional heterogeneity (Figure A.6). The intra-locus recombination analysis ($p - \text{value} < 0.05$) identified 483 recombinant sequences (5.5% of total sequences) involving 52 loci (42.6%); however, there were only 36 recombinants involving 7 loci (5.7%) with $p - \text{value} < 0.01$ and longer than 100 bp (Figure A.7).

All analyses except those based on matrices 9 and 10 recovered the same topology in terms of subfamily-level phylogenetic relationships, and very similar intergeneric and interspecific relationships. An overview of tree topology consistency across all analyses and matrices can be found in Figure A.8, while Figure A.9, A.10, A.11 and A.12 show the trees (based on matrix 3) from RAXML-NG, PhyloBayes-MPI, AS-TRAL and SVDquartets, respectively. We chose the tree from IQ-TREE partitioned analysis based on matrix 3 as our preferred tree (Figure 3.2), whose branches all show high gene frequency supports ($\geq 43\%$) (Figure 3.3a-b). The analyses based on matrices 9 and 10 (i.e., MSAs consisting of compositionally homogenous sites only) failed to recover monophyletic subfamilies due to too few sites left (not shown).

Our phylogeny (Figure 3.2) concurs with previous studies (Bruvo-Madaric et al. 2005; Dimitrov, Astrin et al. 2013; Eberle et al. 2018; Huber 2011; Huber, Eberle et al. 2018) that were based on a few genes or a combination with morphological characters in terms of the following subfamily-level relationships: (1) Smeringopinae is sister to Pholcinae (gene frequency support: 63%), and (2) Modisiminae is sister to Arteminae (gene frequency support: 79%).

However, our study suggests that (1) the traditional Arteminae is not monophyletic but contains two independent lineages, *Artema* and ‘other Arteminae’, which is also supported by the topology test analysis (Table A.5). (2) *Priscula* is not part of Modisiminae but sister to *Artema* (gene frequency support: 43%). (3) The subfamily Ninetinae that contains some of the tiniest Pholcidae, is not sister to all other pholcids but to *Artema* + *Priscula*, two genera that contain some of the biggest Pholcidae (gene frequency support: 65%), as evidenced by the topology test analysis (Table A.5). There are 61.3% of quartets supporting the topology ((Ninetinae, (*Artema*, *Priscula*)), (Arteminae, Modisiminae)) in the likelihood mapping analysis (Figure 3.3c). Ninetinae has been an enigmatic subfamily. Our large sample (48 species) provides not only a new hypothesis about their position among Pholcidae but also a first hypothesis on intergeneric relationships. Three major groups are recognized, group 1: (*Ibotyporanga*, (*Papiamenta*, (*Tolteca*, *Pholcophora*))); group 2: ((*Pinoquio*, (*Magana*, *Ninetis*)), (*Gertschiola*, *Nerudia*)); and group 3: (*Guaranita*, (*Galapa*, (*Pemona*, *Kambiwa*))). Group 1 is sister to (group 2, group 3) (Figure 3.2).

The most recent common ancestor (MRCA) of pholcids dates back to 117 million years ago (Ma) (95%



confidence interval: (95, 152)) (Figure 3.2), and then it diverged into two major groups. The first group consists of Pholcinae and Smeringopinae, which diverged from each other 97 Ma (79, 126). The second group includes all the remaining pholcids. Arteminae is sister to Modisiminae, and they diverged 80 Ma (65, 104). The (Ninetinae, (*Artema*, *Priscula*)) group split from (Arteminae, Modisiminae) 104 Ma (84, 134). The (*Artema*, *Priscula*) group split from Ninetinae 88 Ma (71, 114), while *Artema* and *Priscula* diverged from each other 80 Ma (64, 104).

3.5 Discussion

Based on a sample of 96 taxa covering all subfamilies of Pholcidae, and using UCEs target capture we obtained the largest number of loci (122) to date for pholcids. Since our phylogeny conflicts with previous molecular and morphological data (see below), we did an extensive search for possible errors. The outcome of phylogenetic analysis can be affected by various factors, such as contamination, model misspecification, missing data, incomplete lineage sorting (ILS), and introgression (Tiley et al. 2023; Zhang et al. 2018). We scrutinized cross-contamination by filtering out potential cross-contaminant sequences using the “all-blast-all” method with sequence abundance, similarity ($\geq 99\%$) and alignment length (≥ 100 bp) being considered. We also tried blast similarity cutoffs ($\geq 98\%$ or $\geq 97\%$) (not shown), but the results of phylogenetic analyses (i.e., the subfamily-level relationships) remained the same, indicating a good quality of our dataset. Our dataset appears unaffected by compositional heterogeneity. Only 61 sequences (0.7%) show character composition deviates from the average composition of the alignment, as indicated by the compositional homogeneity test (Figure A.4). Additionally, the RCFV measurement values are very low (0 - 0.0081, average 0.000516174) (Kocot et al. 2017) (Figure A.5). Indeed, both the RY-coding strategy-based phylogeny (matrix 8; Figure A.8), which aims to reduce to the impact of compositional heterogeneity on phylogenetic inference (Criscuolo et al. 2010), and the compositional heterogeneity-aware model CAT-GTR+G4 in PhyloBayes-MPI (matrix 3; Figure A.10) yielded topologies similar to those in Figure 3.2. The intra-locus recombination analysis identified only 36 recombinants involving seven loci (5.7%) with p -value < 0.01 and longer than 100 bp (Figure A.7), suggesting intra-locus recombination should not be a severe problem either, and our coalescent analyses (i.e., ASTRAL) gave similar result (Figure A.11). Most quartets (98.1%) were distributed at the three corners of the triangular graph in the likelihood mapping analysis, indicating a high level of phylogenetic signal (Figure A.3). We also assessed the influences of different MSA trimming programs (Gblocks vs. TrimAl) (Figure 3.1); however, all data matrices resulted in the same subfamily-level relationships for Pholcidae. Overall, our UCE dataset is of high quality and appropriate for phylogenetic analysis. All partitioned ML analyses (IQ-TREE (Figure 3.1) vs. RAxML-NG) (Figure A.9), Bayesian multi-species coalescence (MSC) analysis (PhyloBayes-MPI; Figure A.10), summary-based MSC analyses (ASTRAL (Figure A.11) and SVDquartets (Figure A.12)) consistently recovered the same subfamily-level topology, and very similar genus- and species-level relationships (Figure A.8), indicating our phylogeny is robust.

Our phylogeny supports numerous previously hypothesized relationships, resolves dubious nodes, and solves several open questions. At subfamily level, the sister-group relationships Pholcinae + Smeringopinae and Modisiminae + Arteminae have been proposed before (Dimitrov, Astrin et al. 2013; Eberle et al. 2018) and receive additional strong support (gene frequency support: 63% and 79%, respectively; Figure 3.3a-b). The separation of *Artema* from ‘other Arteminae’ was first proposed by Eberle et al. (2018) and is also strongly supported by our new data (gene frequency support: 79%; Figure 3.3a-b). The most fundamental changes to previous ideas concern three taxa that were identified as the “most obvious gaps”

in understanding pholcid phylogeny (Huber, Eberle et al. 2018): Ninetinae, *Artema*, and *Priscula*. Our data strongly suggest that these three taxa together constitute a monophyletic group: (Ninetinae, (*Artema*, *Priscula*)) (gene frequency support: 65%; Figure 3.3a-b), while the sister group relationship of *Artema* and *Priscula* is supported by 43% of the genes and by 61% of the sites (Figure 3.3c).

This topology is difficult to reconcile with morphological evidence, particularly in relation to sperm ultrastructure and genital mechanics. Pholcidae are part of a large clade of spiders characterized by a specific sperm transfer form, where sperm are united in packages (synspermia) (Michalik et al. 2014). However, most Pholcidae have switched to another sperm transfer form: they transfer sperm individually (cleistospermia) (Dederichs et al. 2022). Synspermia in Pholcidae have only been found in certain Ninetinae, which has been interpreted as a retention of the ancestral state, supporting a ‘basal’ position of Ninetinae, as sister to all other Pholcidae (Dederichs et al. 2022). Our phylogeny suggests that synspermia have either been regained in Ninetinae or transformed to cleistospermia four times independently within Pholcidae. Details of genital mechanics, on the other hand, have been considered as strong evidence for the monophyly of traditional Arteminae, i.e. *Artema* + ‘other Arteminae’. All traditional Arteminae share a pair of unique structures on each male intromittent organ (procursus) that allows for an asymmetric insertion of right and left procursi during copulation (Aharon et al. 2017; Huber and Eberhard 1997). Our phylogeny suggests that this complex and seemingly identical mechanism has either evolved independently in *Artema* and ‘other Arteminae’ or that it has been lost at least three times independently. Interestingly, the monophyly of *Artema* + ‘other Arteminae’ has previously been supported by molecular data (12S/16S/18S/28S/H3/CO1) (Dimitrov, Astrin et al. 2013). Finally, we note that we are not aware of any specific morphological similarity between *Priscula*, *Artema*, and Ninetinae that is likely to constitute a synapomorphy.

Our estimates of divergence times (Figure 3.2) are consistently younger than those of a previous study (Dimitrov, Astrin et al. 2013). For instance, we recovered the most recent common ancestor (MRCA) of Pholcidae to be 117 (95, 152) Ma, while Dimitrov, Astrin et al. (2013) recovered 207 (185, 228) Ma. A more recent study by Magalhaes et al. (2020) highlighted that over half (54%) of fossil calibrations in previous spider studies were problematic. By critically curating key fossil records and integrating transcriptomic data, they produced a revised dated spider tree. Their findings suggest that Pholcidae originated no earlier than 145 Ma, with a range of 80 to 200 Ma, which is congruent with our estimate of 117 Ma for the divergence of Pholcidae.

3.6 Conclusion

In this study, we conducted the largest phylogenomic analysis of the Pholcidae family to date, utilizing target capture sequencing of UCEs. Our results provide new insights into subfamily-level phylogenetic relationships within Pholcidae and reject the monophyly of the subfamily Arteminae. This updated phylogenetic framework also offers novel interpretations of trait evolution within the family. Future studies incorporating additional data types (e.g., transcriptomes, protein-coding genes), as well as broader genomic and taxon sampling, will help confirm and further refine our findings.

3.7 Supplementary material

The raw sequencing data has been deposited in the NCBI SRA database under BioProject ID PRJNA851037. Supplementary information, data and scripts are available in the Dryad Digital Repository and and Ap-

pendix [A](#).

3.8 Funding

This project was financially supported by the German Research Foundation (DFG; project HU980/12-1) and the Alexander Koenig Stiftung. DD received additional support by the Norwegian Metacenter for Computational Science (NOTUR; project NN9601K).

3.9 Acknowledgments

BAH is particularly grateful to Leonardo S. Carvalho, Jimmy Cabra García, Diego Javier Inclán Luna, Matias Izquierdo, Alejandro Valdez-Mondragón, Osvaldo Villarreal, and Andrea Acurio for their help with permits and logistics in Brazil, Colombia, Ecuador, Argentina, Mexico, and Galápagos, respectively. GLM is deeply grateful to Jonas Eberle for his generous sharing of knowledge and experiences from previous projects. We thank the HPC unit of Leibniz Institute for the Analysis of Biodiversity Change (LIB) and the molecular lab, especially Sandra Kukowka for hands-on introduction to the laboratory experiments.

References

- Aharon, S., B. A. Huber and E. Gavish-Regev (2017), *Daddy-long-leg giants: revision of the spider genus *Artema* Walckenaer, 1837 (Araneae, Pholcidae)*, Eur J Taxon **0**.
- Arakawa, K., N. Kono, A. D. Malay et al. (2022), *1000 spider silkomes: Linking sequences to silk physical properties*, Sci Adv **8** eabo6043.
- Ballesteros, J. A., C. E. Santibáñez-López, C. M. Baker et al. (2022), *Comprehensive Species Sampling and Sophisticated Algorithmic Approaches Refute the Monophyly of Arachnida*, Mol Biol Evol **39** msac021.
- Boni, M. F., D. Posada and M. W. Feldman (2007), *An exact nonparametric method for inferring mosaic structure in sequence triplets*, Genetics **176** 1035.
- Bruvo-Madaric, B., B. A. Huber, A. Steinacher and G. Pass (2005), *Phylogeny of pholcid spiders (Araneae: Pholcidae): combined analysis using morphology and molecules*, Mol Phylogenet Evol **37** 661.
- Cai, C., E. Tihelka, M. Giacomelli et al. (2022), *Integrated phylogenomics and fossil data illuminate the evolution of beetles*, R Soc Open Sci **9** 211771.
- Camacho, C., G. Coulouris, V. Avagyan et al. (2009), *BLAST+: architecture and applications*, BMC Bioinformatics **10** 421.
- Capella-Gutierrez, S., J. M. Silla-Martinez and T. Gabaldon (2009), *trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses*, Bioinformatics **25** 1972.
- Castresana, J. (2000), *Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis*, Mol Biol Evol **17** 540.

- Chen, S., Y. Zhou, J. Chen and J. Gu (2018), *fastp: an ultra-fast all-in-one FASTQ preprocessor*, *Bioinformatics* **34** i884.
- Chifman, J. and L. Kubatko (2014), *Quartet inference from SNP data under the coalescent model*, *Bioinformatics* **30** 3317.
- Criscuolo, A. and S. Gribaldo (2010), *BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments*, *BMC Evol Biol* **10** 210.
- Dederichs, T. M., B. A. Huber and P. Michalik (2022), *Evolutionary morphology of sperm in pholcid spiders (Pholcidae, Synspermiata)*, *BMC Zoology* **7** 52.
- Deeleman-Reinhold, C. L. (1986), "Leaf-dwelling Pholcidae in Indo-Australian rain forests", *Proceedings of the Ninth International Congress of Arachnology, Panama*, ed. by W. Eberhard, Y. Lubin and B. Robinson, Washington: Smithsonian Institution Press 45.
- Dimitrov, D., J. J. Astrin and B. A. Huber (2013), *Pholcid spider molecular systematics revisited, with new insights into the biogeography and the evolution of the group*, *Cladistics* **29** 132.
- Dimitrov, D. and G. Hormiga (2021), *Spider diversification through space and time*, *Annu. Rev. Entomol.* **66** 225.
- Duchene, D. A., C. Mather Niklas and Van Der Wal and S. Y. W. Ho (2022), *Excluding Loci With Substitution Saturation Improves Inferences From Phylogenomic Data*, *Syst Biol* **71** 676.
- Eberle, J., D. Dimitrov, A. Valdez-Mondragon and B. A. Huber (2018), *Microhabitat change drives diversification in pholcid spiders*, *BMC Evol Biol* **18** 141.
- Eldridge, M. D. B., J. E. Deakin, A. J. MacDonald et al. (2020), *The Oz Mammals Genomics (OMG) initiative: developing genomic resources for mammal conservation at a continental scale*, *Aust Zool* **40** 505.
- Faircloth, B. C. (2016), *PHYLUCE is a software package for the analysis of conserved genomic loci*, *Bioinformatics* **32** 786.
- Fernandez-Sanchez, J. and M. Casanellas (2016), *Invariant Versus Classical Quartet Inference When Evolution is Heterogeneous Across Sites and Lineages*, *Syst Biol* **65** 280.
- García-Villafuerte, M. Á. and A. Valdez-Mondragón (2020), *The oldest fossils of the spider subfamily Modisiminae from the Americas: Description of a new species of the genus Modisimus Simon (Araneae: Pholcidae) from the amber of Mexico with a checklist of the extant Mexican species*, *J South Am Earth Sci* **103** 102702.
- Ginestet, C. (2011), *ggplot2: Elegant Graphics for Data Analysis*, *J R Stat Soc Ser A Stat Soc* **174** 245.
- Goodacre, S. L., O. Y. Martin, C. F. G. Thomas and G. M. Hewitt (2006), *Wolbachia and other endosymbiont infections in spiders*, *Mol Ecol* **15** 517.
- Guindon, S., J.-F. Dufayard, V. Lefort et al. (2010), *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0*, *Syst Biol* **59** 307.
- Huber, B. A. (2011), *Phylogeny and classification of Pholcidae (Araneae): an update*, *J Arachnol* **39** 211.
- Huber, B. A. (2021), *Beyond size: sexual dimorphisms in pholcid spiders*, *Arachnology* **18** 656.
- Huber, B. A. and W. G. Eberhard (1997), *Courtship, copulation, and genital mechanics in Physocyclus globosus (Araneae, Pholcidae)*, *Can J Zool* **75** 905.

-
- Huber, B. A., J. Eberle and D. Dimitrov (2018), *The phylogeny of pholcid spiders: a critical evaluation of relationships suggested by molecular data (Araneae, Pholcidae)*, Zookeys **51**.
- Huber, B. A. and O. M. Nuñez (2015), *Evolution of genital asymmetry, exaggerated eye stalks, and extreme palpal elongation in Panjange spiders (Araneae: Pholcidae)*, Eur J Taxon.
- Huber, B. A., O. M. Nuñez and C. Leh Moi Ung (2016), *The Philippine hair wax spiders and their relatives: revision of the Pholcus bicornutus species group (Araneae, Pholcidae)*, Eur J Taxon **225** 1.
- Japyassú, H. F. and C. R. Macagnan (2004), *Fishing for prey: the evolution of a new predatory tactic among spiders (Araneae, Pholcidae)*, Revista de Etologia **6** 79.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermiin (2017), *ModelFinder: fast model selection for accurate phylogenetic estimates*, Nat Methods **14** 587.
- Katoh, K. and D. M. Standley (2013), *MAFFT multiple sequence alignment software version 7: improvements in performance and usability*, Mol Biol Evol **30** 772.
- Kishino, H., T. Miyata and M. Hasegawa (1990), *Maximum likelihood inference of protein phylogeny and the origin of chloroplasts*, J Mol Evol **31** 151.
- Kocot, K. M., T. H. Struck, J. Merkel et al. (2017), *Phylogenomics of Lophotrochozoa with Consideration of Systematic Error*, Syst Biol **66** 256.
- Kozlov, A. M., D. Darriba, T. Flouri, B. Morel and A. Stamatakis (2019), *RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference*, Bioinformatics **35** 4453.
- Kück, P. and T. H. Struck (2014), *BaCoCa—a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions*, Mol Phylogenet Evol **70** 94.
- Kulkarni, S., R. J. Kallal, H. Wood et al. (2021), *Interrogating Genomic-Scale Data to Resolve Recalcitrant Nodes in the Spider Tree of Life*, Mol Biol Evol **38** 891.
- Kulkarni, S., H. Wood, M. Lloyd and G. Hormiga (2020), *Spider-specific probe set for ultraconserved elements offers new perspectives on the evolutionary history of spiders (Arachnida, Araneae)*, Mol Ecol Resour **20** 185.
- Lam, H. M., O. Ratmann and M. F. Boni (2018), *Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm*, Mol Biol Evol **35** 247.
- Lartillot, N. and H. Philippe (2004), *A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process*, Mol Biol Evol **21** 1095.
- Lartillot, N., N. Rodrigue, D. Stubbs and J. Richer (2013), *PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment*, Syst Biol **62** 611.
- Letunic, I. and P. Bork (2021), *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation*, Nucleic Acids Res **49** W293.
- Liu, C., X. Zhou, Y. Li et al. (2024), *The influence of the number of tree searches on maximum likelihood inference in phylogenomics*, Syst Biol **73** 807.

- Magalhaes, I. L. F., G. H. Azevedo, P. Michalik and M. J. Ramírez (2020), *The fossil record of spiders revisited: implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic*, *Biological Reviews* **95** 184.
- Meng, G., L. Podsiadlowski, D. Dimitrov and B. A. Huber (2025), *The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders*, submitted.
- Michalik, P. and M. J. Ramírez (2014), *Evolutionary morphology of the male reproductive system, spermatozoa and seminal fluid of spiders (Araneae, Arachnida)—Current knowledge and future directions*, *Arthropod Struct Dev* **43** 291.
- Minh, B. Q., M. A. T. Nguyen and A. von Haeseler (2013), *Ultrafast approximation for phylogenetic bootstrap*, *Mol Biol Evol* **30** 1188.
- Minh, B. Q., H. A. Schmidt, O. Chernomor et al. (2020), *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era*, *Mol Biol Evol* **37** 1530.
- Mirarab, S., L. Nakhleh and T. Warnow (2021), *Multispecies coalescent: theory and applications in phylogenetics*, *Annu Rev Ecol Evol Syst* **52** 247.
- Misof, B., S. Liu, K. Meusemann et al. (2014), *Phylogenomics resolves the timing and pattern of insect evolution*, *Science* **346** 763.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler and B. Q. Minh (2015), *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies*, *Mol Biol Evol* **32** 268.
- Parham, J. F., P. C. Donoghue, C. J. Bell et al. (2012), *Best practices for justifying fossil calibrations*, *Syst Biol* **61** 346.
- Pattengale, N. D., M. Alipour, O. R. P. Bininda-Emonds, B. M. Moret and A. Stamatakis (2010), *How many bootstrap replicates are necessary?*, *J Comput Biol* **17** 337.
- Penney, D. (2006), *The oldest fossil pholcid and selenopid spiders (Araneae) in lowermost eocene amber from the Paris Basin, France*, *J Arachnol* **34** 592.
- Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus and A. Korobeynikov (2020), *Using SPAdes De Novo Assembler*, *Curr Protoc Bioinformatics* **70** e102.
- R Core Team, R. (2022), *R: A language and environment for statistical computing*.
- Reaz, R., M. S. Bayzid and M. S. Rahman (2014), *Accurate phylogenetic tree reconstruction from quartets: a heuristic approach*, *PloS One* **9** e104008.
- Sann, M., O. Niehuis, R. S. Peters et al. (2018), *Phylogenomic analysis of Apoidea sheds new light on the sister group of bees*, *BMC Evol Biol* **18** 71.
- Sayyari, E., J. B. Whitfield and S. Mirarab (2018), *DiscoVista: Interpretable visualizations of gene tree discordance*, *Mol Phylogenet Evol* **122** 110.
- Serrano-Sánchez, M. d. L., T. A. Hegna, P. Schaaf et al. (2015), *The aquatic and semiaquatic biota in Miocene amber from the Campo LA Granja mine (Chiapas, Mexico): Paleoenvironmental implications*, *J South Am Earth Sci* **62** 243.
- Stiller, J., S. Feng, A.-A. Chowdhury et al. (2024), *Complexity of avian evolution revealed by family-level genomes*, *Nature* **629** 851.

-
- Strimmer, K. and A. von Haeseler (1997),
Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment,
Proc Natl Acad Sci U S A **94** 6815.
- Strimmer, K. and A. Rambaut (2002), *Inferring confidence sets of possibly misspecified gene trees*,
Proc Biol Sci **269** 137.
- Swofford, D. L. (2002), *PAUP: phylogenetic analysis using parsimony (and other methods)*, 4.0 beta,
<http://paup.csit.fsu.edu/>.
- Tiley, G. P., T. Flouri, X. Jiao et al. (2023),
Estimation of species divergence times in presence of cross-species gene flow, Syst Biol **72** 820.
- WSC (2025), *World Spider Catalog*, doi: [10.24436/2](https://doi.org/10.24436/2), (visited on 20/01/2025).
- Wunderlich, J. (1988), *Die fossilen Spinnen im dominikanischen Bernstein*, Beit. Araneol. **2** 1.
- Yang, Z. (2007), *PAML 4: phylogenetic analysis by maximum likelihood*, Mol Biol Evol **24** 1586.
- Zhang, C., M. Rabiee, E. Sayyari and S. Mirarab (2018),
ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees,
BMC Bioinformatics **19** 153.
- Zhu, T., T. Flouri and Z. Yang (2022), *A simulation study to examine the impact of recombination on
phylogenomic inferences under the multispecies coalescent model*, Mol Ecol **31** 2814.

Evolutionary Flexibility and Diversification of Pholcidae

Guanliang Meng^{1,*}, Lars Podsiadlowski¹, Dimitar Dimitrov^{2,*}, and Bernhard A. Huber¹

¹ Zoological Research Museum Alexander Koenig, LIB, Bonn, Germany

² University Museum of Bergen, University of Bergen, Bergen, Norway

* Corresponding author: G.Meng@leibniz-lib.de; dimitard.gwu@gmail.com

4.1 Summary

This chapter and Chapter 3 have been submitted as:

G. Meng, L. Podsiadlowski, D. Dimitrov and B. A. Huber (2025), *The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders*, submitted (currently under review by *Systematic Biology*)

Summary

Understanding the mechanisms of how morphological traits drive speciation and contribute to species richness is pivotal in evolutionary biology. In this context, the evolutionary flexibility of morphological traits may play a significant role. Body size has a profound impact on a species' interactions with its environment and its fitness, influencing various aspects such as survival, reproduction, dispersal, and physiology. Larger body sizes are often correlated with increased fitness through enhanced survival, higher reproductive success, and greater dispersal ability. However, the mechanisms driving body size evolution are complex and involve both neutral and selective processes. Cope's rule posits that body size tends to increase over time due to competition, climatic cooling, and selection, but has received mixed support and remains a topic of debate. In contrast, the size-grain hypothesis (SGH) (Kaspari and M. D. Weiser 1999; Pearce-Duvet et al. 2011) suggests that smaller body sizes enable greater specialization as organisms perceive their environment as more complex. Furthermore, research also indicates that reduced body size disparity can be linked to increased extinction risk, which has been observed in crocodilians (Solórzano et al. 2020).

In a similar way, legs or appendages are for most animals essential for adhesion and locomotion. Leg length thus has a significant impact on an organism's fitness, influencing stride length, walking speed, and movement ability. The SGH posits that smaller body sizes are often correlated with shorter legs, which enable access to small spaces providing sustenance and shelter (Kaspari and M. D. Weiser 1999;

Pearce-Duvet et al. 2011). While some studies support this hypothesis (e.g., Kaspari and M. Weiser 2007; Pearce-Duvet et al. 2011), others contradict it (e.g., Parr et al. 2003; Teuscher et al. 2009), yielding a nuanced picture of the evidence. Furthermore, leg length also affects an organism's locomotion cost, dispersal ability, and speciation through dispersal. Notably, terrestrial animals with smaller body sizes tend to expend more energy (relative to body mass) when traveling the same distance as their larger counterparts (Pontzer 2007). Additionally, dispersal ability is affected by body size and leg length, influencing speciation rates (Cadotte 2006; Claramunt et al. 2012; Czekanski-Moir et al. 2019; Venail et al. 2008).

The connection between the measurement of one trait (e.g., leg length) and another reference trait (e.g., body size) is referred to as morphological scaling or allometric scaling (West et al. 1997). Despite its importance in understanding morphological evolution, the relationship between morphological scaling, phenotypic divergence and speciation rate remains unclear. For instance, do the allometric parameters (including slope and intercept) differ between species-rich and species-poor sister clades?

Pholcidae spiders are highly variable in body size, leg length and microhabitat, making them an ideal system to explore the interplay between speciation rate, trait evolution rate, microhabitat shift rate, species richness, interspecific variability of body size, leg length, relative leg length, and leg proportions. Based on the subfamily-level phylogenetic relationships of Pholcidae discussed in Chapter 3, in this study, we constructed the largest phylogeny of any spider family, consisting of 991 species (including 35 outgroup species) with both newly sequenced and public genetic markers. We performed a series of comparative analyses to assess the dynamics of diversification and evolutionary flexibility (measured as either the standard variance or disparity of traits), as well as their interactions.

We found that increased evolutionary flexibility is correlated with accelerated rates of speciation and trait evolution, and with higher species richness. We also observed near-isometry of leg length and body size in the species-rich lineages, suggesting stabilizing selection, but positive allometry in the species-poor lineages, which indicates directional selection. Additionally, we found a positive correlation between trait evolution rates and microhabitat shifts. We argue that environmental heterogeneity and frequent microhabitat shifts may contribute to the origin and maintenance of evolutionary flexibility, which in turn influences the organisms' ability to exploit new resources and habitats. On the other hand, our study suggests that a lack of evolutionary flexibility, e.g. that observed in Pholcidae taxa tending towards dwarfism (Ninetinae) or gigantism (*Artema*, *Priscula*), could be an evolutionary dead end. Our findings also support the idea that taxa do not always follow Cope's rule, which postulates that body size increases with lineage age. This study enhances our understanding of the mechanisms of diversification, demonstrating that evolutionary flexibility of morphological traits may vary among closely related taxa and is likely to have contributed to the uneven distribution of biodiversity across the tree of life.

Personal contribution

This project was a collaborative effort. Bernhard A. Huber initiated the project and secured funding from the German Research Foundation (DFG; project HU980/12-1), while Lars Podsiadlowski and Dimitar Dimitrov were also involved in the funding acquisition. Bernhard A. Huber and Guanliang Meng further received financial support from the Alexander Koenig Stiftung, while Dimitar Dimitrov received additional support from the Norwegian Metacenter for Computational Science (NOTUR; project NN9601K). Bernhard A. Huber collected specimens and conducted initial taxonomic assignments based on morphological characteristics. Lars Podsiadlowski coordinated the molecular work, including DNA isolation, PCR, and Sanger sequencing. Guanliang Meng analyzed the molecular data with assistance from Dimitar Dimitrov, Lars Podsiadlowski and Bernhard A. Huber. The draft version of the manuscript was written by

Guanliang Meng. All authors contributed to the revision and improvement of the manuscript.

4.2 Introduction

The mechanisms underlying the uneven distribution of biodiversity among clades represent one of the core questions in evolutionary biology. Organisms with higher levels of evolutionary flexibility in terms of niches, morphological traits, life-history traits, etc., are able to adapt to different and changing environments more effectively, which may contribute to their diversification (Aiello et al. 2021; Mowery et al. 2022; Pfennig et al. 2010; Ricklefs et al. 1994; Wright et al. 2016). On the other hand, taxa exhibiting evolutionary stasis experience slow rates of morphological evolution and have lower species diversity (Brownstein et al. 2024). While some studies suggest that speciation rates can be decoupled from phenotypic evolution (e.g., Crouch et al. 2019; Rabosky 2013), others have found a correlation between speciation rate and the evolution of phenotypic traits (e.g., Cardillo et al. 2005; Cerezer et al. 2023; Cooney et al. 2021; Etienne et al. 2012; Mahler et al. 2010). However, whether this is a causal relationship remains to be determined—taxonomic diversification can follow, coincide with, or even precede increases in morphological disparity (Folk et al. 2019; Jablonski 2017). Many authors have tried to identify key innovations and ecological opportunities that might have played important roles in the radiation of different groups (Gillet et al. 2019; Lee et al. 2014; Siqueira et al. 2020), or have focused on the correlation between trait values and speciation/extinction rates (Cardillo et al. 2005; Etienne et al. 2012). At the same time recent development of more sophisticated models (e.g., Rabosky 2014; Venditti et al. 2011), has allowed to assess the relationship between the rates of trait evolution and speciation more precisely, particularly in vertebrates (Beltran et al. 2021; Cerezer et al. 2023; Cooney et al. 2021; Rabosky 2013). Few studies have focused on invertebrates, e.g. arthropods, which include more than 50% of all known animals (Scheffers et al. 2012).

Body size has fundamental impacts on the way species interact with the environment (P. A. Arnold et al. 2017; Shao et al. 2023), affecting their fitness, ecology, physiology and life history (LaBarbera 1989; A. N. Smith et al. 2018). Previous studies show that larger body size can lead to higher fitness by: (1) enhanced survival, e.g. by facilitating escape from predators, contributing to foraging efficiency or improving resistance to poor environmental conditions; (2) higher fecundity or mating success during male-male competition and intraspecific competition; (3) greater dispersal ability and thus easier access to new resources (Avaria-Llautureo et al. 2021; Hone et al. 2005; Kingsolver et al. 2004; Marshall et al. 1994). Large body size can result from both neutral and selective processes. In a neutral process, the increase in body size arises from the increase of trait variance over time following a Brownian motion model, without the influence of unidirectional selection (P. H. Harvey et al. 1991). On the other hand, Cope's rule describes an active process, hypothesizing that body size tends to increase over the course of evolution (Stanley 1973), and it has been suggested that competition, climatic cooling and selection are among the drivers (Benson et al. 2014; Hunt et al. 2006; Kingsolver et al. 2004). Cope's rule has received mixed support and remains controversial. Evidence supporting the rule includes studies by Baker et al. (2015), Benson et al. (2014) and Heim et al. (2015). Evidence against the rule has been presented by Dale Guthrie (2003), Gotanda et al. (2015), Jablonski (1997), Knouft et al. (2003), Monroe et al. (2010) and Waller et al. (2017). On the other side, the size-grain hypothesis (SGH) (Kaspari and M. D. Weiser 1999) suggests that, as body size decreases, organisms perceive the Earth's surface as more rugged compared to larger organisms. The relatively increased complexity may provide opportunities for further specialization. In addition, an earlier study indicates that reduced body size disparity among crocodylians is linked to a higher risk of extinction (Solórzano et al. 2020).

Legs or appendages in animals are mostly responsible for adhesion and locomotion (Griethuijsen et al. 2014; Holmes et al. 2006; Labonte et al. 2013; Silva et al. 2021). Absolute and relative leg lengths have important implications. For instance, leg length controls stride length, affecting stride frequency and walking speed (Kaspari and M. D. Weiser 1999; Taylor et al. 1980). In birds, it has been shown that leg length in relation to body mass has a significant impact on gait dynamics (Daley et al. 2018); in the red flour beetle *Tribolium castaneum* (Coleoptera), movement ability is strongly positively associated with intermediate values of relative leg length (P. A. Arnold et al. 2017). Additionally, the SGH suggests that smaller body size comes with shorter legs, because they facilitate access to small interstices that can provide food and shelter (Kaspari and M. D. Weiser 1999). Therefore, leg length and relative leg length affect habitat selection (Kaspari and M. D. Weiser 1999; C. Li et al. 2011). Similar to Cope's rule, the SGH is supported by some studies (e.g., Kaspari and M. Weiser 2007; Pearce-Duvet et al. 2011) but not by others (e.g., Parr et al. 2003; Teuscher et al. 2009). Like body size, leg length is also correlated with locomotion cost (Grossi et al. 2016; Pontzer 2007) and dispersal ability (Corcobado et al. 2012; Phillips et al. 2006). Terrestrial animals with smaller body sizes and shorter legs are found to use more energy (relative to body mass) than larger animals when traveling the same distance (Pontzer 2007). Leg length is indirectly linked with speciation via dispersal ability (Cadotte 2006; Claramunt et al. 2012; Czekanski-Moir et al. 2019; Venail et al. 2008).

Different traits are often correlated both developmentally and genetically (a phenomenon known as integration, as opposed to modularity) (Armbruster et al. 2014; Conner et al. 2014; Evans et al. 2021) and so are their evolutionary flexibilities. The relationship between the measure of a trait (e.g., leg length) and another reference trait (e.g., body size) is called morphological scaling or allometric scaling, and can be formulated as: $\log y = \log b + a \log x$, where y and x are the values for the two traits, respectively; b is the intercept and a is the allometric scaling factor. Such allometric scaling relationships are frequently observed in morphological, physiological and life-history traits (e.g., Andresen et al. 2002; Bolstad et al. 2015; Chong et al. 2024; West et al. 1997; White et al. 2019). When $a = 1$, it is called isometry. For $a < 1$ or $a > 1$, it is referred to as negative and positive allometry, respectively. Allometric relationships between traits may be ontogenetic, static, or evolutionary, depending on whether the relationship occurs among different developmental stages of the same individual, different individuals at the same developmental stage within the same population, or across different species (Cheverud 1982; A. G. Cock 1966; Gould 1966). Understanding whether morphological scaling parameters are under evolutionary constraints, and the nature of such constraints, may provide insights into the formation of phenotypic and species diversity. If variations in genetics and developmental processes that regulate the growth of related traits exist, natural selection may act on these variations, potentially driving the evolution of allometry—unless strong negative selection or stabilizing selection prevents such changes (Houle et al. 2019). The allometric-constraint hypothesis proposes that factors such as functional and physiological mechanisms can strongly restrict the evolution of morphological scaling parameters and the evolution of related morphological traits (Gould 1966; Pélabon et al. 2014; Voje, Bell et al. 2022; Voje, Hansen et al. 2014). On the other hand, natural selection could also drive the evolution of the slope and intercept of scaling relationships and pleiotropy is thought to be responsible for the rapid reversal to the initial slope once selection is suspended (Bolstad et al. 2015; Frankino et al. 2005; Houle et al. 2019). However, it has been shown that the slope evolves more slowly than the intercept (Bolstad et al. 2015; Egset et al. 2012; Pélabon et al. 2014; Voje, Bell et al. 2022; Voje, Hansen et al. 2014), suggesting the slope is subject to stronger evolutionary constraints. Additionally, natural selection may also break trait-trait correlations, resulting in phenotypic modularity and high variability in each trait (Voje, Bell et al. 2022). However, it is unclear how morphological scaling is related to phenotypic divergence and species diversification. For example, are there any differences in

the allometric parameters (i.e., slope and intercept) between species-rich and species-poor sister clades? Are allometric parameters in different parameter spaces under the same direction and strength of natural selection, and how does this correlate with evolutionary flexibility?

Spiders (Araneae), with their diverse range of body sizes and leg lengths (Shao et al. 2023; Wolff, Wierucka et al. 2022), make an ideal group for studying the evolution of these traits. Spider body size and leg length have been attributed to various factors and shown to affect spider biology in many ways. Mean body size increases from cool/moist to warm/dry environments (Entling et al. 2010). Larger spiders tend to win intraspecific male-male fights (Wells 1988). Larger-bodied spiders typically undergo more molts and longer intermolt intervals to reach sexual maturity compared to smaller-bodied spiders (Head 1995), and the size of a spider also significantly influences the diversification of both web architectures and reproductive strategies (Dong et al. 2023; Sensenig et al. 2011). Relative leg length varies a lot among different groups of organisms and has different implications. Some spiders (e.g., jumping spiders, Salticidae) have very short relative leg lengths, while others (e.g., funnel-web spiders, Agelenidae; cellar spiders, Pholcidae) generally have very long legs (Pechmann et al. 2010). Spiders that live in silk-lined burrows tend to have larger bodies with relatively shorter legs. Conversely, spiders that build aerial webs generally have smaller bodies and relatively longer legs. Spiders that hunt exhibit an intermediate body size and morphology, falling between the two extremes (Shao et al. 2023). We hypothesize that the evolutionary flexibility of morphological traits—measured by trait disparity and/or variance—rather than their absolute size, is what truly matters in terms of impact on species diversification.

Diversification in spiders has previously been attributed to: (1) silk innovations and web design/loss, by which spiders were able to explore new niches due to new silk types and web architectures (Bond et al. 1998); (2) the evolution of venom to quickly subdue prey (King et al. 2013); (3) the evolution of adhesive setae to help in prey capture (Wolff, Nentwig et al. 2013); (4) the transformation of the second pairs of book lungs into tracheae to meet high energy consumption requirements in locomotion (Schmitz 2016); (5) the coevolution with insect prey (J. Liu et al. 2016); (6) geological processes and climate changes (Xu et al. 2015); (7) sexual selection (Gage et al. 2002); and (8) microhabitat shifts (Eberle et al. 2018). The impact of body size and leg length evolution, as well as their allometric pattern on diversification in spiders have rarely been studied (Kuntner and Coddington 2020; Kuntner, Hamilton et al. 2019; Shao et al. 2023; Wolff, Wierucka et al. 2022). Here, we investigate the relationships between evolutionary flexibility, diversification and allometry using the daddy long-legs spiders Pholcidae as a study system. Pholcidae is one of the largest and ecologically most diverse spider families, including over 2,000 described species (WSC 2025). The family has a worldwide distribution, predominantly in tropical and subtropical regions (Huber, Eberle et al. 2018). Pholcids occupy a wide range of microhabitats, which can be classified into three main types: “space,” “leaf,” and “ground” (Eberle et al. 2018). They are highly variable in body size, shape, and color; and, despite their vernacular name, include both species with very short legs (1–2 mm) and species with leg-spans of up to 180 mm.

4.3 Materials and methods

4.3.1 Molecular methods

To better understand the evolutionary dynamics of pholcids and to assess the impact of trait evolution on pholcid diversification, we constructed the most comprehensive pholcid phylogeny so far, using six genes (12S rDNA/16S rDNA/18S rDNA/28S rDNA/Histone 3 (H3)/Cytochrome c oxidase I (CO1)) combining newly generated and public data from the following sources: (1) the target capture sequencing of ultra-

conserved elements (UCE) from Chapter 3 (citmeng2025complex); (2) 101 newly sequenced specimens (84 CO1, 2 H3, 1 12S, 3 18S, 11 28S) with the same protocols as in Eberle et al. (2018); (3) the dataset of Eberle et al. (2018) (almost 600 species); (4) all other pholcid sequences available in GenBank (access date: October 26th, 2022). In total, we were able to obtain data for 991 species (including 35 outgroup species) and 424 12S, 521 16S, 462 18S, 450 28S, 426 H3, 873 CO1 sequences (Table B.1).

Screening for the 12S/16S/18S/28S/H3/CO1 sequences. The 18S and 28S genes in the previous SPAdes assemblies were identified with RNAmmer (v1.2) (Lagesen et al. 2007), which uses hidden Markov models trained on rRNA data for different super kingdoms (Archaea, Bacteria, and Eukaryotes). The deep-learning-based software, Tiara (v1.0.2) (Karlicki et al. 2022), was employed to screen for mitochondrial sequences (`-to_fasta mit pla pro -prob_cutoff 0.65 0.60 -probabilities -min_len 500 -t 8`). The mitochondrial sequences were annotated by MitoZ (v3.4) (Meng, Y. Li et al. 2019) for 12S and 16S genes. A method combining tBlastn (v2.2.26) (Camacho et al. 2009) and Genewise (v2.4.1) (Birney et al. 2004) is implemented in MitoZ and was used to search for H3 and CO1 genes. In all cases, candidate gene sequences not overlapping with the dataset of Eberle et al. (2018) were identified with Blast+ and filtered out. Next, to exclude potential false positives, the remaining sequences were aligned with the corresponding genes of Eberle et al. (2018) using MAFFT (v7.487) (Katoh et al. 2013) with either the mafft-linsi algorithm (for CO1 and H3) or the secondary-structure-based mafft-xinsi algorithm (for rDNA genes). The alignments were checked by eye in AliView (v1.28) (Larsson 2014). Whenever there were multiple copies of the same gene for one sample, only the candidate copies with the highest abundances (from the Phyluce pipeline (v1.7.1), Faircloth (2016)) were kept, since copies with lower abundances were assumed to be artifacts (e.g., assembly errors, contaminations, etc.). Finally, all retained sequences were blasted against the NT database to further confirm their biological sources.

Multiple sequence alignment (MSA). For the protein-coding genes CO1 and H3, DNA sequences were translated into protein using BioPython (v1.78) (P. J. Cock et al. 2009) with invertebrate mitochondrial genetic code and standard genetic code, respectively. Next, protein-MSAs were constructed using the mafft-linsi algorithm of MAFFT, which then assisted the construction of nucleotide level MSAs with pal2nal.pl (Suyama et al. 2006). This helps avoid the introduction of biologically meaningless frameshifts to the alignments (Suyama et al. 2006). The alignments of rDNA genes (12S, 16S, 18S, and 28S) were constructed based on secondary structure information using the mafft-xinsi algorithm in MAFFT (v7.487) and MXSCARNA (Tabei et al. 2008). Poorly aligned regions in the MSAs were trimmed with Gblocks (-b5=h) (v0.91b) (Castresana 2000), TrimAl (-automated 1) (v1.4.rev15) (Capella-Gutierrez et al. 2009) and ClipKIT (v1.1.3) (Steenwyk et al. 2020), respectively. For the ClipKIT program, we tested eight different trimming strategies (`-modes gappy, kpi, kplic, kplic-gappy, kplic-smart-gap, kpi-gappy, kpi-smart-gap, and smart-gap`).

Phylogenetic Inference. The trimmed six-gene matrices were concatenated with the FASconCAT-G.pl (v1.05) program (Kück et al. 2014) and constrained tree search analyses were performed with IQ-TREE (v2.2.0) (Minh et al. 2020) using the partitioned strategy and the same parameter settings as in the UCE-tree analysis. As a constraint we used the tree from the IQ-TREE maximum likelihood (ML) analysis of matrix 3 from Chapter 3 (Meng, Podsiadlowski et al. (2025)). To account for topological uncertainty, we chose three trees for subsequent analyses: a primary tree based on the ClipKIT smart-gap trimmed-MSA; and two alternative trees based on the ClipKIT gappy and kplic-smart-gap trimmed-MSAs. These trees were chosen based on two criteria: minimum number of species obviously misplaced at subfamily level; and monophyly of uncontested genera. Next, the divergence times of the three trees were estimated using TreePL (v1.0) (S. A. Smith et al. 2012) and the R function `congruify.phylo()` in `geiger` (v2.0.11) (Pennell et al. 2014) using our dated UCE tree as a reference tree (see the section *Divergence time calibration of*

Chapter 3, Meng, Podsiadlowski et al. (2025)).

4.3.2 Comparative analysis

We focus on four traits: (1) body size, using male carapace width as a proxy; (2) leg length, using the combined length of male tibia 1 + metatarsus 1 as a proxy, because these can be measured easily and precisely and they are available from the taxonomic literature; (3) relative leg length, which is the ratio of leg length to body size and (4) leg proportion, which is the ratio of male metatarsus 1 to tibia 1; it can be used as a proxy for microhabitat preference (i.e., “space,” “leaf,” and “ground”), in which leaf and space living pholcids tend to have larger leg proportion values (Eberle et al. 2018). The analyses were applied for each of the three selected six-gene trees. Body size and leg length were log-transformed before the analyses (base 2).

Speciation and extinction dynamics. We modeled the complex dynamics of speciation and extinction of pholcids using BAMM (v2.5.0) (Rabosky 2014), because model-based approaches have higher or equivalent accuracy (Cerezer et al. 2023; Jablonski 2017; Title et al. 2019) compared to other methods including the inverse terminal branch lengths (Steel et al. 2010), node density metric (Freckleton et al. 2008), and DR statistic (Jetz et al. 2012). Moore et al. (2016) criticized BAMM by raising concerns about the correctness of its likelihood function, sensitivity to priors, and reliability of diversification rate estimates. However, subsequent works have shown that these points are generally incorrect or unjustified, as they were based on flawed assessments or inappropriate parameterizations (Rabosky et al. 2017). To assess the robustness of our inference, both genus- and subfamily-specific sampling fractions were tested, in which the total species number of each group was estimated using the same method as described in Eberle et al. (2018). Additionally, for the analysis using genus-specific sampling fractions, we further tested it against the currently known species richness (including all described species and undescribed species available in collections and known to BAH). The prior on the number of shifts in diversification was 50, while the priors for the λ InitPrior, λ ShiftPrior and μ InitPrior parameters were estimated with BAMMtools (v2.1.10) (Rabosky 2014). For each tree, two independent runs were applied, and for each run, we used 50 million generations for MCMC simulation. The convergences of MCMC chains were confirmed by plotting the log-likelihood traces of MCMC output files, and by using the R package coda (v0.19-4) (Plummer et al. 2006) to estimate the effective sizes of all runs, where most runs have the effective sizes varying from 1000 to 6000 (minimum 440). We used 20% of the generations as burn-in. The single best shift configuration was identified with the function `getBestShiftConfiguration()` in BAMMtools. Group-specific (i.e., genus and subfamily) speciation rates, extinction rates and diversification rates were calculated and we assessed if different groups have different rates using the Kruskal-Wallis H-test in the package pingouin (v0.5.3) (Vallat 2018). We also conducted a rate-through-time analysis using the `plotRateThroughTime()` function and performed a macroevolutionary cohort analysis with the `cohorts()` function in BAMMtools.

Phylogenetic signal and model fitting of trait evolution. Blomberg’s K (Blomberg et al. 2003) and Pagel’s λ (Pagel 1999) of the four traits were assessed using the R package phytools (v1.9-16) (Revell 2012). A series of trait evolution models (BM, OU, EB, rate_trend, lambda, delta and white) were fitted using the `fitContinuous()` function in the package geiger (v2.0.11) (Pennell et al. 2014). BM (Brownian Motion) is a “random walk” model without constraints on rate evolution (Felsenstein 1973); OU (Ornstein–Uhlenbeck) is a BM model with a trend evolving to an optimal trait value (Butler et al. 2004); rate_trend is a diffusion model with linear trends in rates through time; EB is BM with an exponential acceleration or deceleration of the evolutionary rate through time (Harmon et al. 2010); delta is a time-dependent model

of trait evolution (Pagel 1999). Lambda measures the extent to which the phylogeny predicts covariance among trait values for species, with $\lambda = 0$ corresponding to a star-like phylogeny while $\lambda = 1$ corresponds to the BM model (Pagel 1999); finally, white is a null-model assuming trait data is independent from each other. To find the optimal solution, we applied 30,000 random starting points and all available optimization methods in the function. The optimal models were chosen based on AICc (the sample-size corrected Akaike information criterion).

Ancestral state reconstruction. We reconstructed the ancestral states of the four traits using the `ace()`, `anc.ML()`, `anc.trend()`, and `anc.Bayes()` functions in the packages `ape` (v5.7-1) (Paradis 2019) and `phytools` based on the available BM, EB, OU, and the `mean_trend` models, respectively.

Trait evolution rate estimation. The evolutionary rates of the four traits were estimated with BMM whose performance is similar to other tools such as BayesTraits (Cerezer et al. 2023; Venditti et al. 2011). The `betaInitPrior` and `betaShiftPrior` priors were estimated with BMMtools and `expectedNumberOfShifts` was set to 50. Similar to the speciation/extinction analysis described above, we ran two independent runs for each tree using 50 million generations for MCMC simulation. The convergences and effective sizes were confirmed with the R package `coda`. To assess the impact of incomplete taxonomic sampling, we randomly selected 650 species from the 991 species of the original tree and reran the analysis. Subsequently, subfamily-specific evolutionary rate and evolutionary rate through time plots of the traits were calculated with BMMtools.

Correlation analysis between traits. The phylogenetic generalized least squares (PGLS) analyses were applied to estimate the correlations between different traits using the `gls()` function in the R package `nlme` (v3.1-163) (Pinheiro 2023).

Trait disparity and diversification. Disparity through time (DTT) for each trait was estimated using the `dtm()` function in the R package `geiger` for each sub-family. Trait evolutionary flexibility was also assessed with standard variance of trait values within a group (O'Meara et al. 2006), and we further evaluated the relationships between trait evolutionary flexibility, speciation rate, extinction rate, diversification rate, trait evolution rate and estimated species richness using PGLS.

Trait-dependent diversification. We tested the correlation between the four traits and pholcid diversification using two methods. (1) The QuaSSE (Quantitative State Speciation and Extinction) model in the R package `diversitree` (v0.9-20) (FitzJohn 2010), which simultaneously analyzes the relationship between changes in continuous traits and diversification and can suffer from model misspecification and phylogenetic pseudoreplication (M. G. Harvey et al. 2018). A global sampling fraction was applied, since the model does not support clade-specific sampling fractions. We fitted the speciation and extinction rates as constant, linear, sigmoidal, and hump-shaped functions of the traits, respectively. We also tested the models without the drift parameter (i.e., assuming a constant mean of trait values through time). Finally, we performed a likelihood ratio test to compare the model fits. (2) The tip-rate correlation ES-sim test, which avoids the requirement for a fully parameterized model of diversification and trait evolution and directly examines the correlation between trait variation across the tips of a phylogeny and tip-specific speciation rate estimates and is thus believed to have rare false inferences (M. G. Harvey et al. 2018). We used 20,000 simulations to build the null distribution of trait-speciation associations for significance testing.

4.4 Results

The phylogenetic trees based on 12S/16S/18S/28S/H3/CO1 genes using constrained searches are shown in Figure B.1, B.2 and B.3. The differences between them include: (1) the number of specimens placed in wrong subfamilies; monophyly of uncontested genera recovered or not; (2) morphologically supported sister-group relationships of genera either recovered or not, and (3) individual species misplaced in wrong genera. These differences come from different alignment trimming strategies. We focus on the question about whether the main results from our comparative analyses are stable using these different topologies, rather than on the possible causes for variation among topologies and on judging which topology is the “best”.

4.4.1 Speciation and Extinction Rates

BAMM diversification analyses based on the three selected six-gene trees reveal similar patterns, with only four speciation rate shifts (Figure B.4), suggesting that speciation rate shifts are unlikely to be the primary factor contributing to the uneven levels of species richness among subfamily-level taxa. In the analyses based on estimated species-richness (Figure B.4a), two speciation rate shifts within the subfamily Pholcinae were identified. The first rate shift involves a clade containing 60% of described Pholcinae species, and the second rate shift occurs within that clade. A third rate shift occurs within the genus *Smeringopina* of the subfamily Smeringopinae and a fourth rate shift occurs within the clade which contains 91% of described Modisiminae species. The BAMM diversification analyses based on currently known species-richness show similar patterns, except that there are two rate shifts for the genus *Pholcus*, and no rate shift in the subfamily Smeringopinae (Figure B.4b). At the subfamily-level, Pholcinae, Modisiminae and Smeringopinae consistently show higher speciation rates (Pholcinae > Modisiminae > Smeringopinae) than Arteminae, Ninetinae, *Artema* and *Priscula* (p – value << 0.001; Figure 4.1b, B.5 and Table B.2; B.2). Speciation rate generally increases through time (Figure 4.1a). Pholcinae, Modisiminae and Smeringopinae also show higher extinction rates (Pholcinae > Smeringopinae > Modisiminae) than the other groups (p – value << 0.001; Figure B.6 and Table B.2). The same pattern also applies to the net diversification rates (i.e., speciation - extinction) (Modisiminae > Pholcinae > Smeringopinae) (p – value << 0.001; Figure B.7 and Table B.2).

4.4.2 Trait Evolution

Both Blomberg’s K ($0.64 \leq K \leq 1.5$) (p –values = 0.001) and Pagel’s λ ($0.95 \leq \lambda \leq 0.99$) (p –values << 0.001) show high phylogenetic signal for the four analyzed traits (body size, leg length, relative leg length and leg proportion), indicating that taxa which are more closely related tend to exhibit similar traits as a result of shared ancestry (Figure B.8). Both measures show that leg length has the highest phylogenetic signal, indicating that leg length is a conservative trait. Body size and relative leg length are much more variable.

Our model-fitting analyses for the four traits consistently support the idea that these morphological traits evolve following a Brownian Motion-like dynamics. The best-fitted model is rate_trend, followed by EB, then BM or delta, with OU, lambda, and finally white as the least fitted (Table B.3). Both rate_trend and EB models are extended BM models which are characterized by an increasing or decreasing rate through time (linear or exponential). For all traits, both the parameter *slope* of the rate_trend model and

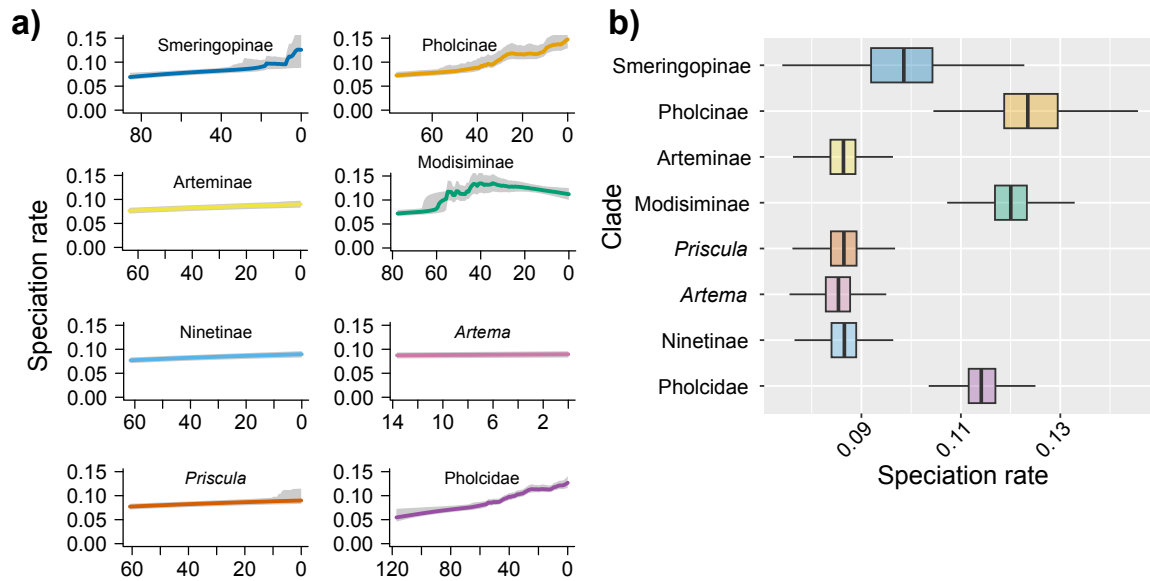


Figure 4.1. Speciation rate dynamics. Group-specific speciation rates were extracted from the BAMM analysis, based on genus-specific sampling fraction, estimated species richness and the ClipKIT smart-gap tree. **(a)** Speciation rate through time. Smeringopinae, Pholcinae and Modisiminae show a rapid increase of speciation rates after their emergence, compared to Arteminae, Ninetinae, *Artema* and *Priscula*. **(b)** Boxplot of speciation rates. Pholcinae and Modisiminae show the highest average speciation rates.

the parameter α of the EB model are negative, and the parameter δ of the delta model is < 1 , indicating the trait evolution rates decrease through time. On the other hand, for all fitted models, the parameters σ^2 (i.e., the diffusion rate of trait changes) are very small (e.g., $\sigma^2 \leq 0.01$ for the BM model) except for the trait relative leg length ($\sigma^2 = 0.51$ for the BM model), indicating that relative leg length shows substantial change over time, leading to greater variability among species, whereas the other three traits exhibited relatively limited variability throughout the course of evolution.

On the whole, the reconstructed ancestral states of the four traits from different models (BM and OU; but the EB model failed to run even though we had tried many different combinations of phytools versions and platforms) and methods (ML vs. Bayesian) show heterogeneous patterns of trait values for all groups (except for Ninetinae, *Artema* and *Priscula*). The pholcid ancestors had intermediate values for the traits, and then the trait values fluctuated in different lineages (Figure 4.2, B.9, B.10 and B.11). Evolution in Ninetinae is skewed towards small values, while *Artema* and *Priscula* evolved towards large values for both body size and leg length; other groups generally evolved towards a much wider range of current values (Figure 4.3b, B.12, B.13 and B.14).

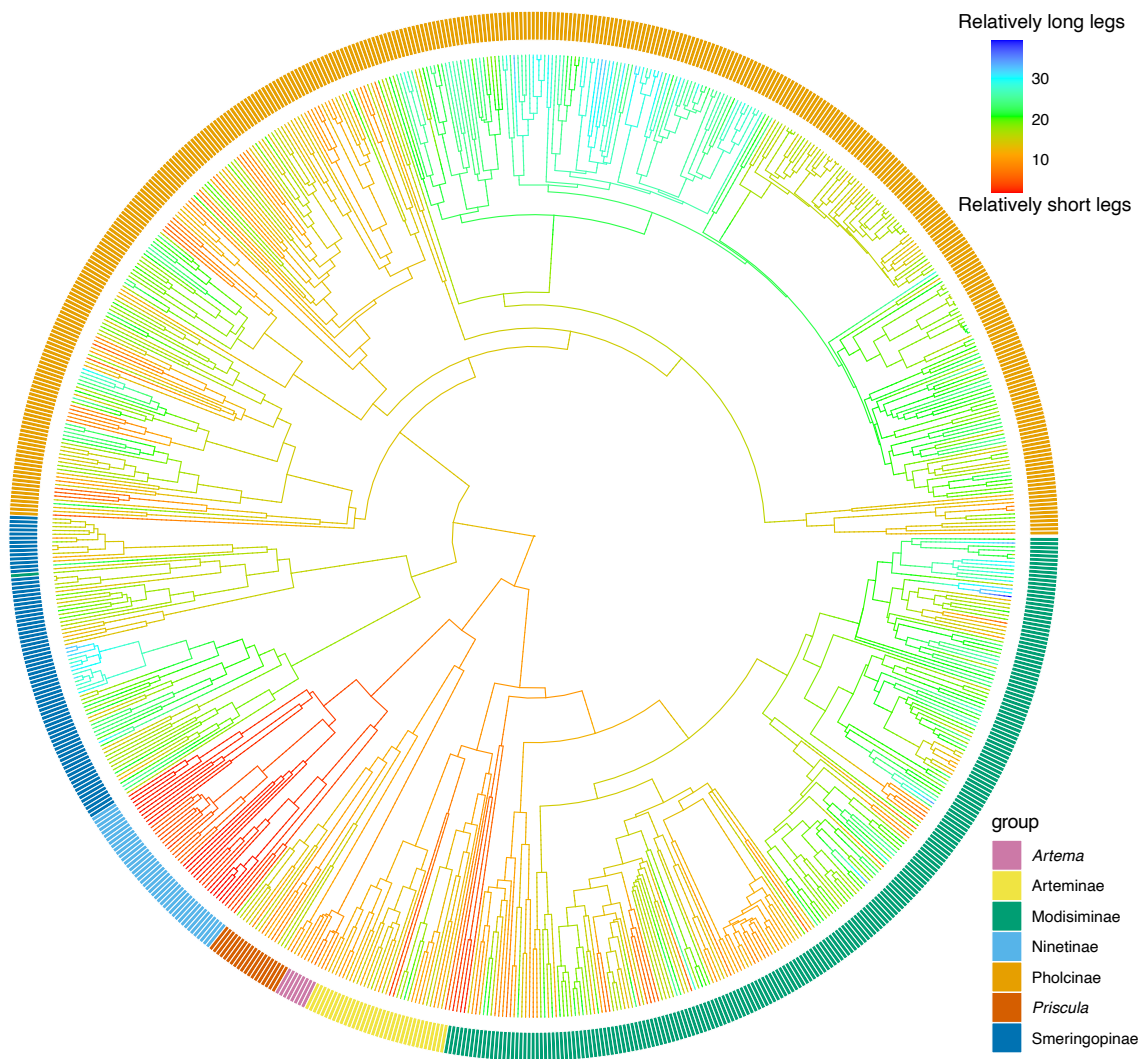


Figure 4.2. Ancestral state reconstruction. The data on relative leg length is presented; the ClipKIT smart-gap tree was used for analysis. The `anc.ML()` function in phytools was used to estimate ancestral states (the inner tree), with cooler (blue) color indicating relatively long legs and warmer (red) color indicating relatively short legs. Smeringopinae, Pholcinae and Modisiminae show higher level of heterogeneity than Arteminae, Ninetinae, *Artema* and *Priscula* in terms of relative leg length.

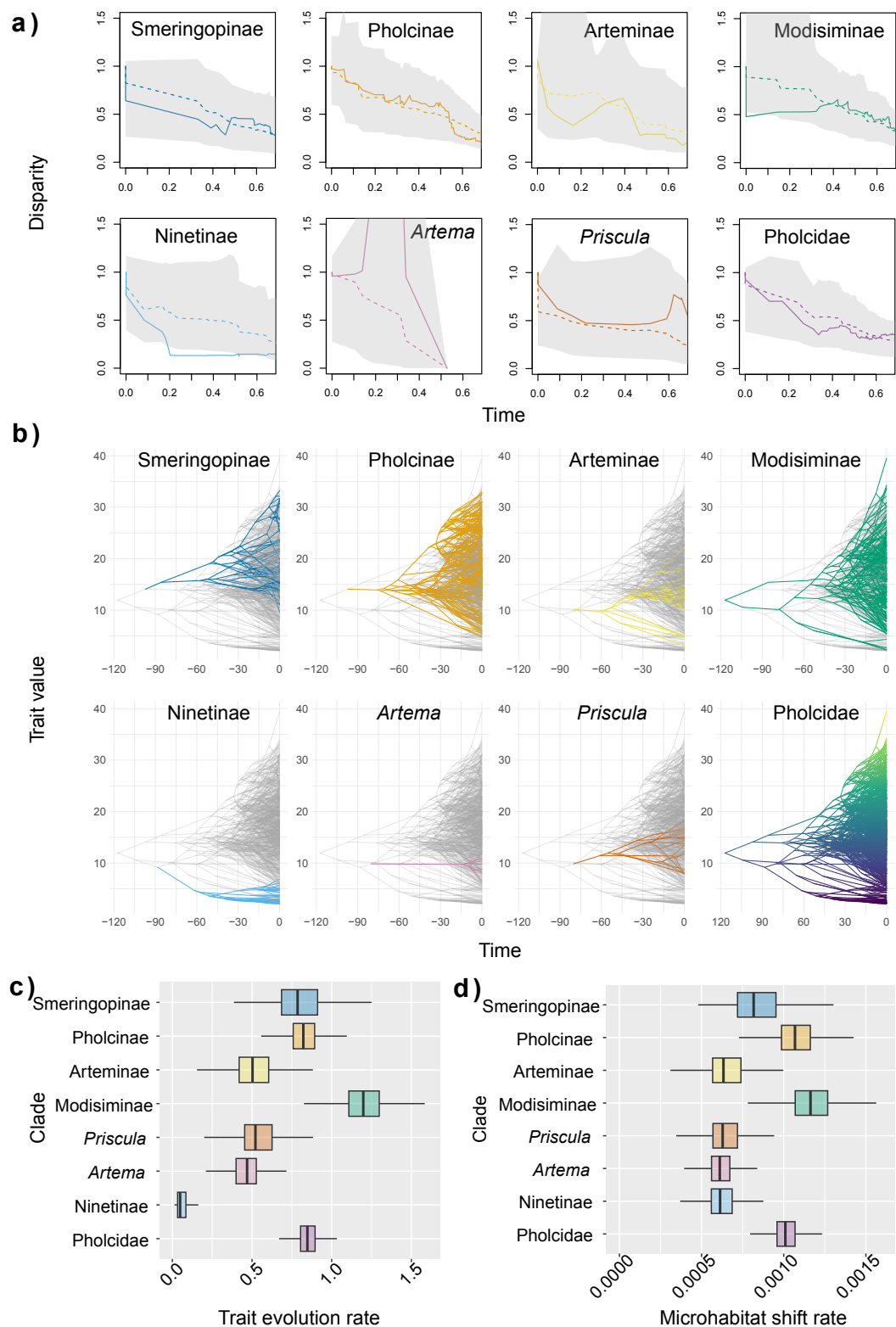


Figure 4.3. Trait evolution dynamics. The data on relative leg length (**a**, **b**, **c**) and leg proportion (**d**) are shown,

(Continued) and the ClipKIT smart-gap tree was used for analysis. (a) Disparity dynamics, estimated by the `dti()` function in `geiger`. Smeringopinae, Pholcinae and Modisiminae show higher disparity, compared to Arteminae, Ninetinae and *Priscula*. The high disparity in *Artema* could be an artifact due to its small sample size ($n=9$). (b) Phenograms for different groups, based on the BM model and the `anc.ML()` function in `phytools`. Smeringopinae, Pholcinae and Modisiminae show much wider ranges of trait values than Arteminae, Ninetinae, *Artema* and *Priscula*. (c and d) Group-specific trait evolution rates, extracted from the BAMM trait analysis, which was based on genus-specific sampling fraction and estimated species richness. Smeringopinae, Pholcinae and Modisiminae show higher evolution rates for relative leg length and leg proportion than Arteminae, Ninetinae, *Artema* and *Priscula*. Leg proportion in d is a proxy of microhabitat preference (Eberle et al. (2018)); the evolution rate of leg proportion can thus be interpreted as a microhabitat shift rate.

The subfamily Ninetinae is noteworthy for having the smallest and most homogeneous values for all four traits among pholcids, especially for relative leg length (Figure 4.2, 4.3a-b and B.15). Pholcinae, Modisiminae and Smeringopinae occupy a broad morphospace and show large disparity (standard variance: 5.79, 5.82 and 5.42, respectively, for relative leg length), while the morphospace of Arteminae, Ninetinae, *Artema* and *Priscula* is more limited (standard variance: 3.17, 1.15, 1.24 and 2.71, respectively) (Figure 4.2, 4.3a-b, and B.15).

Evolutionary rates of the four traits calculated using BAMM based on different trees and different taxonomic sampling were largely consistent (Figure B.16), indicating that estimates of rates were not significantly impacted by the uncertainty in tree topology or sampling fractions. Specifically, for both relative leg length and leg proportion, Smeringopinae, Pholcinae and Modisiminae consistently show highest rates, while Ninetinae shows the lowest rate for relative leg length, and Arteminae, Ninetinae, *Artema* and *Priscula* show similar low rates for leg proportion (Figure 4.3c-d).

Similar to Eberle et al. (2018), we found that space and leaf dwelling spiders tend to have larger leg proportion values (Figure B.17a). Space and ground dwelling spiders tend to have larger body size than leaf dwellers, while space and leaf dwellers tend to have longer legs than ground spiders (Figure B.17b-c). On the other side, all pholcids, regardless of habitat, have a wide range for relative leg length (Figure B.17d).

PGLS analysis based on the three six-gene trees consistently found a linear correlation between body size and leg length for most groups (Figure 4.4). Overall, Pholcidae species show a slightly positive allometry in terms of leg length ($slope \pm$ standard error (SE): 1.108 ± 0.044 , p - value $<< 0.001$). Interestingly, the species-diverse groups Pholcinae, Modisiminae and Smeringopinae all show a trend similar to isometry with slopes of 1.012 ± 0.059 , 1.123 ± 0.096 and 0.979 ± 0.119 (p - values $<< 0.001$), respectively, while the species-poor groups Arteminae, Ninetinae, and *Artema* show a strong positive allometry with slopes of 1.522 ± 0.175 , 1.567 ± 0.127 and 1.612 ± 0.307 (p - values for Arteminae and Ninetinae $<< 0.001$, for *Artema*: 0.002), respectively. *Priscula* shows a non-significant negative allometry ($slope$: 0.254 ± 0.343 , p - value : 0.469). Regarding intercepts, Pholcinae, Modisiminae, Smeringopinae, Arteminae and *Priscula* have similar, large values (3.648 ± 0.164 , 3.256 ± 0.572 , 4.117 ± 0.208 , 3.03 ± 0.205 , and 4.414 ± 0.399 , p - values $<< 0.001$), while Ninetinae and *Artema* have similar but smaller values (1.94 ± 0.143 , 2.079 ± 0.513 , p - values $<< 0.001$ and 0.007, respectively) (Figure 4.4). Body size, leg length or relative leg length all show a linear correlation with leg proportion (Figure B.18).

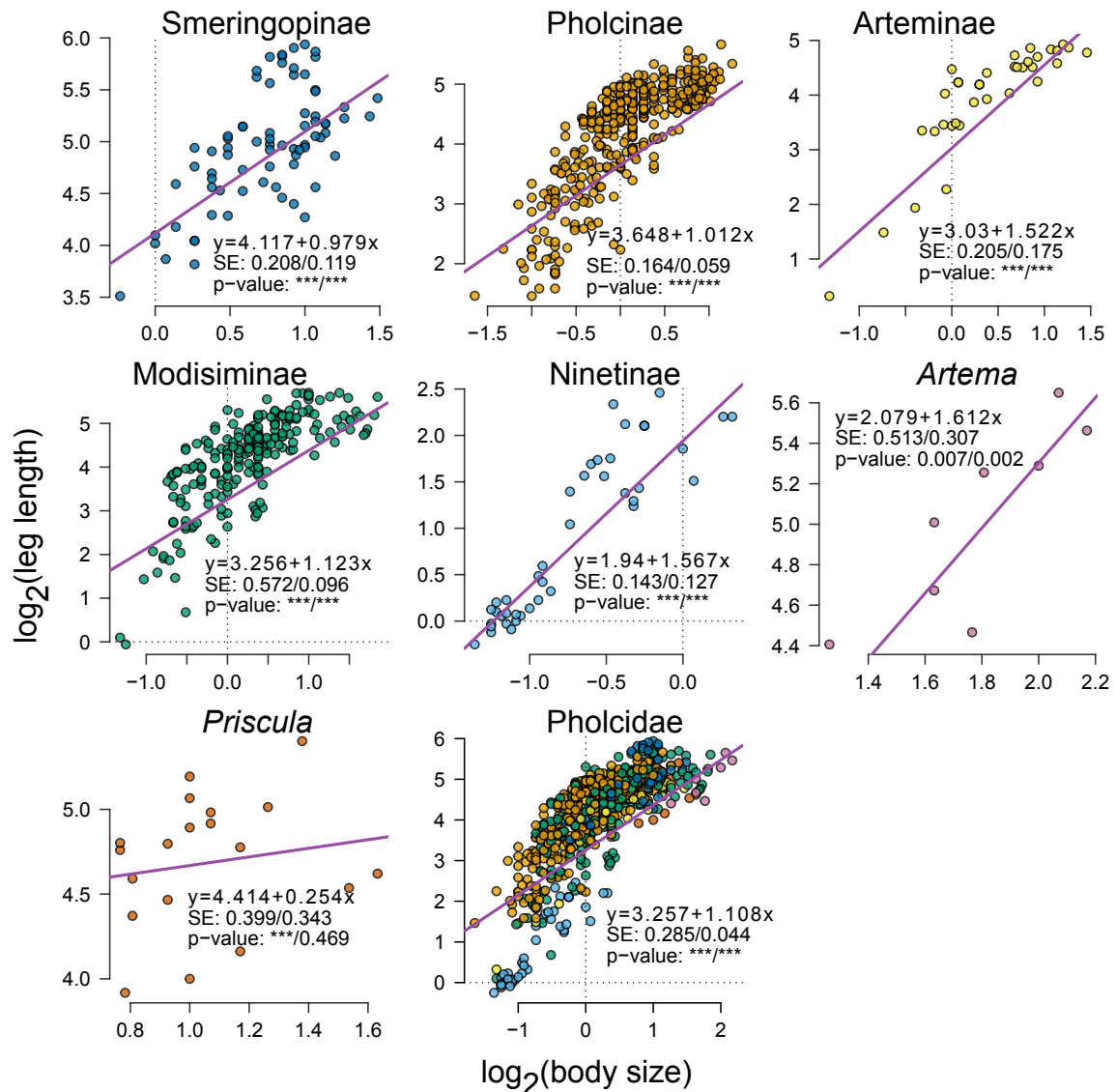


Figure 4.4. Evolutionary allometry of pholcids. Data points for each group are shown in different colors. Relationships between $\log_2(\text{leg length})$ and $\log_2(\text{body size})$ were determined by Phylogenetic Generalized Least Squares regression. SE: Standard Error. ***: $p - \text{value} < 0.001$. The two items separated by “/” are SE and $p - \text{values}$ for the intercept and slope, respectively. Leg length generally shows a positive allometry to body size in all groups but with varied slopes and intercepts.

4.4.3 Relationships between Trait Variation and Diversification

The tip-rate correlation test (Table B.4) shows that body size and leg length are positively related to net diversification rates (Pearson's correlation coefficient $\rho = 0.34$ and $p\text{-value} < 0.05$), while the correlation with relative leg length or leg proportion is not significant (Pearson's correlation coefficient $\rho = 0.24$ and $p\text{-value} > 0.05$). The QuaSSE analysis, on the other hand, shows that speciation and extinction rates are a sigmoidal or hump function of the traits (Table B.5).

The PGLS analysis shows that the amount of disparity (standard variance) of traits is significantly positively correlated with species richness and speciation rate (Figure 4.5a-b and B.19). For example, relative leg length has a correlation of slope with species richness = 0.871 ($p\text{-value} < 0.001$) (Figure 4.5a) while it has a correlation of slope with speciation rate = 0.663 ($p\text{-value} < 0.001$) (Figure 4.5b). Pholcinae, Modisiminae and Smeringopinae have the greatest variance and highest speciation rates while Arteminae, Ninetinae, *Artema* and *Priscula* have the opposite. Pholcinae and Modisiminae are the pholcids with the largest net diversification rates ($\text{speciation} - \text{extinction}$) and biggest species richness (Figure 4.5c). Finally, the trait evolution rate of relative leg length is positively correlated with disparity, speciation rate and the rate of microhabitat shift ($\text{slopes} \pm \text{SE}$: 1.045 ± 0.1563 , 0.954 ± 0.0002 , 1.088 ± 0.0002 , $p\text{-values} : < 0.001$, < 0.001 , and 0.001 , respectively) (Figure 4.5d-f).

4.5 Discussion

4.5.1 Subfamily Species Diversity and Morphological Trait Evolution in Pholcids

Pholcids exhibit variable morphological traits, with body sizes ranging from 1 to 10 mm, and leg spans from 3 to 180 mm. Within Pholcidae, Pholcinae and Modisiminae (excl. *Priscula*) are the two largest subfamilies, with currently (Oct. 2024) 1,122 and 510 described species, respectively (B.A. Huber, unpubl. data). Their diversity is largely concentrated in humid tropical and subtropical regions. Smeringopinae, the third-largest subfamily, includes 159 described species. It is often found in semi-arid regions, except for the tropical African genus *Smeringopina* Kraus, 1957 (Huber, Eberle et al. 2018). The less diverse groups are Arteminae (107), Ninetinae (77), and *Artema* + *Priscula* (40). Among the species-poor groups, Ninetinae, Arteminae, and *Artema* are mostly or strictly distributed in rather dry regions (Huber 2021; Huber, Eberle et al. 2018), while *Priscula* is largely restricted to high elevation Andean forests (Huber, Meng, Dupérré et al. 2023).

What are the drivers behind the high species diversity, speciation and net-diversification rates in Pholcinae and Modisiminae, as shown in Figure 4.1b and 4.5c? One indirect but intuitive explanation refers to the complexity of the environments these spiders inhabit, combined with their ability to adapt to various microhabitats (Eberle et al. 2018). Eberle et al. (2018) demonstrated frequent microhabitat shifts throughout pholcid evolution, with many back-and-forth transitions between near-ground habitats and leaves, particularly in the two largest subfamilies, Pholcinae and Modisiminae. Additionally, leaf- and space-dwelling species were found to have higher diversification rates. Thus, high evolutionary flexibility in microhabitat use may have played a key role in driving pholcid diversification. Our analysis further supports this idea. Leg proportion can be seen as a proxy for different types of microhabitats (i.e., "space," "leaf" and "ground"), and space and leaf dwellers tend to have larger values of leg proportion (Eberle et al. 2018). We found that Modisiminae, Pholcinae and Smeringopinae have higher trait evolution rates for leg proportion, indicating these groups have experienced more frequent microhabitat transitions

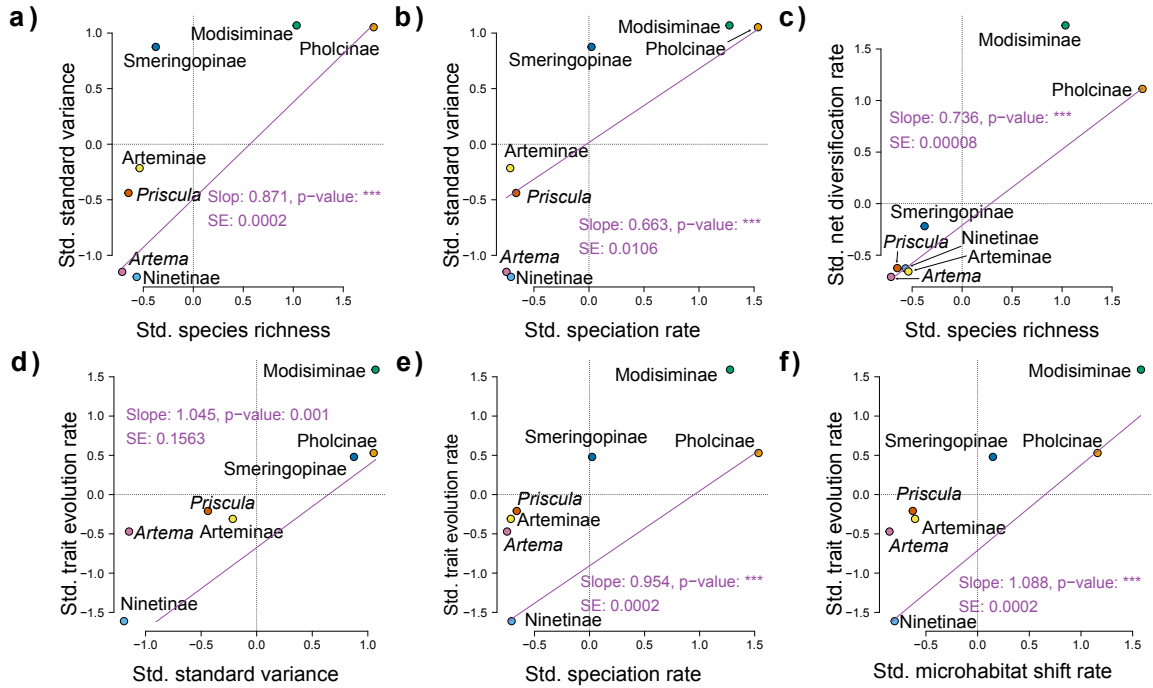


Figure 4.5. Relationships between disparity, species richness, trait evolution, diversification and microhabitat shift. Relative leg length and the ClipKIT smart-gap tree were used for analysis. Disparity was quantified as the Standardized Standard Variance of the trait (the y-axis in **a** and **b**), which was calculated by combining the trait values of extant taxa (tips) and the ancestral trait states reconstructed using the Brownian motion model. Species richness (**a** and **c**) was estimated at the genus level based on the number of currently described species and their distribution (either temperate or tropical region), as outlined by Eberle et al. (2018). Trait evolution rates (the y-axis in **d**, **e** and **f**) were estimated by BAMM. Leg proportion is a proxy of microhabitat preference (Eberle et al. 2018), the evolution rate of leg proportion (estimated by BAMM) can thus be seen as a microhabitat shift rate (the x-axis in **f**). The relationships between these variables were determined by Phylogenetic Generalized Least Squares regression. Std.: Standardized. SE: Standard Error of the slope. ***: $p - value < 0.001$.

(Figure 4.3d), and indeed, the extent of flexibility in microhabitat use is positively correlated with species richness (Figure B.19e).

Understanding the factors that shape phenotypic differences among lineages is also crucial for explaining the variation in species richness across different lineages. Overall, the four analyzed traits show a high level of phylogenetic signal ($0.64 \leq \text{Blomberg's } K \leq 1.5$, $0.95 \leq \text{Pagel's } \lambda \leq 0.99$), indicating closely related species are prone to resemble each other for these traits. Leg length is the most conserved of the four traits, as indicated by its high values for both Blomberg's K (1.5) and Pagel's λ (0.99), reflecting minimal deviation from evolutionary expectations. In contrast, relative leg length is the most variable trait, with lower values for Blomberg's K (0.76) and Pagel's λ (0.95), suggesting greater evolutionary lability. This coincides with our model-fitting analysis. For all fitted models, the diffusion rates, σ^2 , were found to be remarkably small. Most traits have a $\sigma^2 \leq 0.01$; only the trait relative leg length has a higher $\sigma^2 = 0.51$. This indicates that most traits evolve much slower than relative leg length.

Our analysis also shows that the four analyzed traits are consistently best-fitted by Brownian Motion-like models including rate_trend and EB, then BM or delta (Table B.3) and estimates of related parameters

(both *slope* and $\alpha < 0$, $\delta < 1$) of these models suggest that trait evolution rates are decreasing through time. A possible explanation is that there were few competitors when pholcids entered new adaptive zones, leading to diversification and increased phenotypic diversity. However, as the available ecological space filled up and competition increased, the rate of morphological evolution naturally slowed down. The need for further rapid changes diminished because the organisms had already adapted to their environments, leading to a deceleration in the pace of trait evolution over time (Harmon et al. 2010; Losos et al. 2002; Simpson 1984).

Body size.—Wolff, Wierucka et al. (2022) suggested that most morphological traits in spiders, including body size, do not significantly vary across ecological guilds. Our analysis adds a temporal perspective (Figure B.12). While the body size of all pholcids evolves in a Brownian-like motion, distinct subgroups exhibit unique patterns: Smeringopinae, *Artema*, and *Priscula* show a trend toward larger body sizes, supporting Cope’s rule, whereas Ninetinae exhibit a shift toward smaller sizes, contradicting Cope’s rule. Meanwhile, Modisiminae, Pholcinae, and Arteminae continue to evolve in line with the Brownian motion model. It has been suggested that web-building spiders tend to have smaller body sizes, while hunting spiders are generally larger (Shao et al. 2023; Wolff, Wierucka et al. 2022). Our study provides more detail on body size variation within web-building spiders. Benefits of evolving towards large body size for *Priscula*, *Artema* and Smeringopinae may include: (1) enhanced protection against predators and improved prey capture due to increased strength; (2) increased mating success through sexual selection; (3) higher fecundity (Kuntner and Coddington 2020; Wells 1988). On the other side, adult and juvenile viability selection (Blanckenhorn 2000) could be responsible for Ninetinae evolving towards small body size. Thus, gigantism and dwarfism may represent two distinct strategies employed within Pholcidae to reduce competition with close relatives with intermediate body size (Figure B.9 and B.12). On the other hand, multidirectional selection governs the body size evolution of Modisiminae, Pholcinae, and Arteminae.

(Relative) leg length.—The evolution of leg length in pholcids mirrors the patterns observed in body size; however, relative leg length shows different trends. While the evolution of relative leg length in most groups follows a Brownian motion-like pattern, Smeringopinae exhibit a slight shift toward longer relative leg lengths. The most intriguing group, however, is Ninetinae, which evolves toward an extremely small relative leg length. This aligns with their ground-dwelling habits and fast-running behavior (Huber, Meng, Dederichs et al. 2024; Huber, Meng, garcía et al. 2024; Huber, Meng, Král et al. 2023; Huber, Meng, Valdez-Mondragón et al. 2023). Similar patterns have also been observed among spiders in general. For instance, aerial web builders generally have longer legs compared to substrate-bound web builders and ground runners (Wolff, Wierucka et al. 2022).

Phenotypic flexibility.—While the direction of trait evolution (e.g., Cope’s rule) may play a role in species diversification, the significance of variation in trait values across time (referred to here as evolutionary flexibility) has received less attention. The reconstructed ancestral states of body size and (relative) leg length (Figure 4.2, B.9 and B.10) clearly demonstrate the high degree of flexibility in these traits, as evidenced by the significant heterogeneity in trait values, especially in relative leg length (Figure 4.2). Our analyses also clearly indicate that different subfamilies exhibit varying degrees of phenotypic flexibility. The species-rich subfamilies (Modisiminae, Pholcinae, and Smeringopinae) consistently show higher levels of phenotypic flexibility compared to the species-poorer subfamilies (Arteminae, Ninetinae, *Artema*, and *Priscula*) (Figure 4.2, B.9 and B.10). Since body size and relative leg length are key traits influencing species’ interactions with their biotic and abiotic environments, the observed differences in flexibility and species richness among subfamilies suggest that phenotypic flexibility may play a critical role in pholcid diversification. This flexibility could enable species to explore a broader morphospace, facilitating adaptation to diverse environmental conditions and potentially driving evolutionary diversification.

ation (see next section).

4.5.2 Evolutionary Flexibility is Positively Correlated with Trait Evolution Rate and Speciation Rate

To quantify phenotypic flexibility, we measured disparity and standard variance of traits (O’Meara et al. 2006). Our results show that the species-rich subfamilies Pholcinae, Modisiminae, and Smeringopinae exhibit larger standard variances for relative leg length (5.79, 5.82, and 5.42, respectively), compared to the species-poorer subfamilies Arteminae, Ninetinae, *Artema*, and *Priscula*, which have much lower standard variances (3.17, 1.15, 1.24, and 2.71, respectively) (Figure 4.3a-b). Similar results were obtained in the analyses based on all the three six-gene trees, trait evolutionary models (BM, OU, trend) and body size and leg length (Figure B.12 and B.13), suggesting that our estimate of flexibility is robust.

Our PGLS analysis shows that the standard variance of the studied traits is positively correlated with species richness and speciation rates (slopes: 0.871 ± 0.0002 , 0.663 ± 0.0106 , p – value : < 0.001 , respectively) (Figure 4.5a-b). The higher species diversities and speciation rates in Modisiminae, Pholcinae and Smeringopinae are accompanied with higher phenotypic flexibility and larger net-diversification rates (Figure 4.5a-c). Further analysis shows that phenotypic flexibility is positively correlated with trait evolution rate (Figure 4.5d), while the latter is also positively correlated with speciation rate (Figure 4.5e). For instance, Modisiminae, Pholcinae and Smeringopinae show higher evolution rates for relative leg length than Arteminae, Ninetinae, *Artema*, and *Priscula* (Figure 4.5e).

Two key aspects of our results should be emphasized, regarding sister groups with the same evolutionary age: (1) a higher speciation rate and greater net diversification rate lead to increased species diversity (Figure 4.5a and 4.5c), and (2) a higher rate of trait evolution results in greater phenotypic diversity (Figure 4.5d). Pholcinae + Smeringopinae and Modisiminae + Arteminae form two sister groups (Eberle et al. 2018; Huber, Eberle et al. 2018; Meng, Podsiadlowski et al. 2025). Both Pholcinae and Modisiminae exhibit greater species richness, higher speciation rates, and increased phenotypic diversity than their sister taxa Smeringopinae and Arteminae, respectively. This pattern of convergent evolution suggests that increased phenotypic diversity and accelerated trait evolution rate have likely played a significant role in the diversification of pholcids. Indeed, positive correlation between trait evolution rate and speciation rate has also been observed in several recent studies of vertebrates, including hummingbirds (plumage color evolution) (Beltran et al. 2021) and fishes (body size evolution) (Cerezer et al. 2023; Rabosky 2013). This positive correlation generally holds true at broad macroevolutionary scales for body size evolution, which has been tested in the five major vertebrate clades (Cooney et al. 2021): amphibians, birds, mammals, ray-finned fish, and squamate reptiles; but negative correlation has also been found in clades including *Bufo* toads (amphibian) and lacertid lizards (squamates). Furthermore, non-significant relationships are also observed in many subgroups (Cooney et al. 2021). It should be noted, however, that phenotypic diversification can occur after, alongside, or even before species diversification (Folk et al. 2019; Jablonski 2017). As such, additional evidence from biogeography, ecology, genetics and other aspects is required to establish causality.

The higher trait evolution rate in Modisiminae, Pholcinae, and Smeringopinae indicates they exhibit a higher evolutionary potential to produce diverse phenotypes, which in turn are adaptations to corresponding environments. Evolutionary constraints may limit the evolutionary potential (S. J. Arnold 1992). In the phenograms of the four traits (Figure 4.3b, B.12, B.13 and B.14), these three subfamilies occupy much wider morphospace, while the other four species-poor subfamilies (Arteminae, Ninetinae, *Artema*, and *Priscula*) occupy much narrower morphospace. Additional support for this finding comes from the fact

that the majority of troglomorphic pholcids are found in two subfamilies, Modisiminae and Pholcinae, with only a few slightly troglomorphic species in Smeringopinae, *Artema*, Arteminae, and *Priscula*, and none in Ninetinae (Huber 2018). Furthermore, our data supports the habitat heterogeneity hypothesis which claims that increased environmental heterogeneity promotes greater species diversity by offering more ecological niches (Allouche et al. 2012). Indeed, we found that Modisiminae, Pholcinae and Smeringopinae have greater microhabitat shift rates (Figure 4.5f), aligning with many microhabitat shifts found in these subfamilies in a previous study (Eberle et al. 2018). Modisiminae and Pholcinae are mostly found in humid tropical and subtropical regions, and their varied traits are thought to be adapted to the wide range of forest microhabitats (Huber, Eberle et al. 2018). The subfamily Smeringopinae is intermediate in many respects, with most genera confined to dry environments, but with one genus (*Smeringopina*) in tropical African forests (Huber 2013). Ninetinae exemplify the other extreme: they occupy the ground in semi-arid environments that offer a limited number of different microhabitats. Lastly, in addition to evolutionary potential, inter-specific competition may also favor representatives of species-rich groups. For instance, Huber and Acurio (2022) observed that while the introduced species *Modisimus culicinus* (Modisiminae) is becoming more prevalent on Galápagos, the abundance of the native *Galapa bella* (Ninetinae) that occupies the same microhabitat is declining.

In summary, we found that the species-rich subfamilies (Modisiminae, Pholcinae, and Smeringopinae) have higher phenotypic flexibility compared to the species-poor subfamilies (Arteminae, Ninetinae, *Artema*, and *Priscula*), and that trait evolution rate positively correlates with the rates of speciation and microhabitat shift. However, the factors driving the disparity in phenotypic flexibility between species-rich and species-poor groups remain poorly understood. Future research should investigate the roles of genetic and developmental constraints, as well as selective regimes, to better address this knowledge gap.

4.5.3 The Interplay between Allometry and Evolutionary Flexibility

The allometric scaling relationship between traits is often highly conserved across lineages (Bolstad et al. 2015), and it has been suggested that allometry may limit morphological diversification (Pélabon et al. 2014; Voje, Hansen et al. 2014). However, empirical evidence shows that the allometric scaling slope and intercept can be modified by natural selection (e.g., Chong et al. 2024; Emberts et al. 2023; Pélabon et al. 2014) and sexual selection (reviewed by W. Eberhard et al. (2009)). Thus, by overcoming the existing constraints in allometric scaling, organisms are able to explore new ecological opportunities (Chong et al. 2024; DeLong et al. 2010).

We found that pholcids generally exhibit a pattern of positive allometry between leg length and body size (scaling factor: 1.108) (Figure 4.4), indicating that relatively long legs are positively selected. At the subfamily level, however, different subfamilies display distinct allometric slopes and intercepts, suggesting they have undergone varying selection pressures. Interestingly, while Pholcinae, Modisiminae and Smeringopinae have higher species diversity, speciation rate, trait evolution rate and phenotypic flexibility, they all show an allometric pattern similar to isometry, with scaling *slopes* close to 1 (i.e., 1.012, 1.123, and 0.979) (Figure 4.4). One plausible explanation is that stabilizing selection may be acting on the slopes, as suggested by the declining rates of trait evolution revealed in our model-fitting analysis (Table B.3). This process likely facilitates the specialization of related species into distinct ecological niches (Maher et al. 2022). More specifically, based on the ecological roles of leg length and body size in spiders (W. G. Eberhard 1992; W. G. Eberhard and Briceno 1984; Huber 1998; Shao et al. 2023), such specializations may be related to locomotion mode, male-male fighting, web construction, courtship and copulation. For instance, relatively long legs resulting from near-isometry allow spiders to reduce locomotor

costs (Pontzer 2007) and bridge relatively large gaps and move efficiently across uneven surfaces (W. G. Eberhard 1992), as also predicted by the size-grain hypothesis (Kaspari and M. D. Weiser 1999; Pearce-Duvet et al. 2011). The latter adaptation is particularly important as many Pholcinae, Modisiminae, and some Smeringopinae species inhabit complex forest environments with high levels of three-dimensionality (Huber 2011). Indeed, we observed larger intercepts in the species-rich groups (Pholcinae, Modisiminae, and Smeringopinae) (3.648 ± 0.164 , 3.256 ± 0.672 , and 4.117 ± 0.208 , all with p – values $<< 0.001$) compared to the species-poor groups (Arteminae, Ninetinae, *Artema*, but not *Priscula*) (3.03 ± 0.205 , 1.94 ± 0.143 , 2.079 ± 0.513 , and 4.414 ± 0.399 , all with p – values $<< 0.001$, except *Artema* with p – value = 0.007) (Figure 4.4). This suggests that the species-rich groups generally maintain relatively longer legs than the species-poor groups, as illustrated in Figure 4.3b and B.15. On the other hand, how can high phenotypic flexibility in the species-rich groups evolve under stabilizing selection? Stabilizing selection is generally believed to reduce both genetic and phenotypic diversity (Wirthlin et al. 2018; Wolff, Wierucka et al. 2022). We propose three possible explanations. Firstly, isometry and stabilizing selection do not necessarily lead to low phenotypic flexibility. Organisms can still explore a large proportion of morphospace (i.e., larger or smaller trait values) while maintaining slope and intercept unchanged, as exemplified by Pholcinae, Modisiminae and Smeringopinae (Figure 4.3b, B.12, B.13 and B.14). Secondly, environmental heterogeneity favors the maintenance of genetic and morphological variation by favoring divergent adaptive peaks and locally well-adapted phenotypes (Boussange et al. 2022; Byers 2005; Kalan et al. 2020; McDonald et al. 2018; Rainey et al. 1998). This coincides with the habitat heterogeneity hypothesis, which suggests that increased habitat diversity supports a larger number of species by offering a wider range of niches compared to more homogeneous habitats (D. Chen et al. 2020; MacArthur et al. 1961; Tews et al. 2004). For instance, a study by Nooten et al. (2019) indicates that vegetation structural complexity can function as an environmental filter, shaping ant assemblages by influencing both species diversity and functional traits. Moreover, interactions among key species can enhance environmental heterogeneity, such as facilitating cascades initiated by foundation species (Thomsen et al. 2022). Lastly, frequent environmental changes or habitat shifts (as revealed by Eberle et al. (2018)) can drive a change in optimal trait values, leading to a shift in the adaptive peak and increased rates of trait evolution (Coyne et al. 1997; Marques et al. 2022; Van Tienderen 1991), as well as increased phenotypic variance as suggested by theoretical research (Norberg et al. 2001). For instance, experimental research revealed that greater habitat heterogeneity led to rapid shifts in the feeding behavior of generalist predators including spiders (Staudacher et al. 2018).

On the other side, we found strong positive allometry in the species-poor groups Arteminae, Ninetinae, and *Artema*, with scaling slopes of 1.522, 1.567, and 1.612, respectively, suggesting strong directional selection (W. Eberhard et al. 2009) on leg length in these groups. *Priscula* had a non-significant negative allometry, suggesting either a decoupling of leg length and body size or an artifact. The steeper slopes in the species-poor groups mean that their leg length increases more rapidly than in the species-rich groups when a change occurs in body size. This suggests that body size has a stronger influence on leg length in the species-poor groups, which could be due to stronger evolutionary constraints. Relatively long legs can effectively reduce locomotor cost and increase stride length in terrestrial animals (Pontzer 2007), such as tetrapods (Maher et al. 2022), Theraphosidae spiders (Grossi et al. 2016), and Lycosidae spiders (Framenau 2005). However, what factors contribute to the low evolutionary flexibility in these species-poor groups? Like stabilizing selection, directional selection also tends to reduce genetic and phenotypic diversity (Assis et al. 2016). Furthermore, these subfamilies typically occupy relatively homogeneous environments, which contribute to their low rates of trait evolution, microhabitat shifts (Figure 4.3c-d, B.16), speciation, and overall species richness (Figure 4.5). With these similar characteristics,

Ninetinae exhibits dwarfism while *Artema* and *Priscula* show gigantism (Figure B.15). Both may lead to an evolutionary dead end, as indicated by their low net diversification rates and species richness (Figure 4.5c). However, further research is needed to determine causalities from the correlations identified in this study, such as between trait evolution rate and microhabitat shift rate.

4.6 Conclusion

Speciation is a complex process which is influenced by a number of intrinsic and extrinsic factors (Czekanski-Moir et al. 2019; Nosil et al. 2009; Rundell et al. 2009). In this study, we discovered that the speciation rate in Pholcidae is positively correlated with the rates of trait evolution, microhabitat shifts, and phenotypic flexibility. Our findings indicate that species-rich groups (Pholcinae, Modisiminae, and Smeringopinae) that exhibit allometric scaling close to isometry show the greatest phenotypic flexibility, whereas species-poor groups (Arteminae, Ninetinae, *Artema*, and *Priscula*) with positive allometry are least flexible. Notably, Ninetinae exhibits a trend toward dwarfism, while *Artema* and *Priscula* display gigantism. Species-rich groups are likely influenced by stabilizing selection, whereas species-poor groups experience directional selection; the observed differences in phenotypic flexibility may stem from variations in environmental heterogeneity. Further research is needed to elucidate the causal relationships underlying the correlations observed in this study.

4.7 Supplementary material

The raw sequencing data has been deposited in the NCBI SRA database under BioProject ID PRJNA851037. Supplementary information, data and scripts are available in the Dryad Digital Repository and Appendix B.

Funding

This project was financially supported by the German Research Foundation (DFG: project HU980/12-1) and the Alexander Koenig Stiftung. DD received additional support by the Norwegian Metacenter for Computational Science (NOTUR; project NN9601K).

Acknowledgments

BAH is particularly grateful to Leonardo S. Carvalho, Jimmy Cabra García, Diego Javier Inclán Luna, Matias Izquierdo, Alejandro Valdez-Mondragón, Osvaldo Villarreal, and Andrea Acurio for their help with permits and logistics in Brazil, Colombia, Ecuador, Argentina, Mexico, and Galápagos, respectively. GLM is deeply grateful to Jonas Eberle for his generous sharing of knowledge and experiences from previous projects. We thank the HPC unit of Leibniz Institute for the Analysis of Biodiversity Change (LIB) and the molecular lab, especially Sandra Kukowka for hands-on introduction to the laboratory experiments.

References

- Aiello, B. R., M. Tan, U. Bin Sikandar et al. (2021), *Adaptive shifts underlie the divergence in wing morphology in bombycoid moths*, *Proc Biol Sci* **288** 20210677.
- Allouche, O., M. Kalyuzhny, G. Moreno-Rueda, M. Pizarro and R. Kadmon (2012), *Area-heterogeneity tradeoff and the diversity of ecological communities*, *Proc Natl Acad Sci U S A* **109** 17495.
- Andresen, B., J. S. Shiner and D. E. Uehlinger (2002), *Allometric scaling and maximum efficiency in physiological eigen time*, *Proc Natl Acad Sci U S A* **99** 5822.
- Armbruster, W. S., C. Pelabon, G. H. Bolstad and T. F. Hansen (2014), *Integrated phenotypes: understanding trait covariation in plants and animals*, *Philos Trans R Soc Lond B Biol Sci* **369** 20130245.
- Arnold, P. A., P. Cassey and C. R. White (2017), *Functional traits in red flour beetles: the dispersal phenotype is associated with leg length but not body size nor metabolic rate*, *Functional Ecology* **31** 653.
- Arnold, S. J. (1992), *Constraints on phenotypic evolution*, *Am Nat* **140** S85.
- Assis, A. P., J. L. Patton, A. Hubbe and G. Marroig (2016), *Directional selection effects on patterns of phenotypic (co)variation in wild populations*, *Proc Biol Sci* **283**.
- Avaria-Llautureo, J., C. Venditti, M. M. Rivadeneira et al. (2021), *Historical warming consistently decreased size, dispersal and speciation rate of fish*, *Nat Clim Chang* **11** 787.
- Baker, J., A. Meade, M. Pagel and C. Venditti (2015), *Adaptive evolution toward larger size in mammals*, *Proc Natl Acad Sci U S A* **112** 5093.
- Beltran, D. F., A. J. Shultz and J. L. Parra (2021), *Speciation rates are positively correlated with the rate of plumage color evolution in hummingbirds*, *Evolution* **75** 1665.
- Benson, R. B. J., R. A. Frigot, A. Goswami, B. Andres and R. J. B. Butler (2014), *Competition and constraint drove Cope's rule in the evolution of giant flying reptiles*, *Nat Commun* **5** 3567.
- Birney, E., M. Clamp and R. Durbin (2004), *GeneWise and Genomewise*, *Genome Res* **14** 988.
- Blanckenhorn, W. U. (2000), *The evolution of body size: what keeps organisms small?*, *Q Rev Biol* **75** 385.
- Blomberg, S. P., T. J. Garland and A. R. Ives (2003), *Testing for phylogenetic signal in comparative data: behavioral traits are more labile*, *Evolution* **57** 717.
- Bolstad, G. H., J. A. Cassara, E. Marquez et al. (2015), *Complex constraints on allometry revealed by artificial selection on the wing of *Drosophila melanogaster**, *Proc Natl Acad Sci U S A* **112** 13284.
- Bond, J. E. and B. D. Opell (1998), *Testing adaptive radiation and key innovation hypotheses in spiders*, *Evolution* **52** 403.
- Boussange, V. and L. Pellissier (2022), *Eco-evolutionary model on spatial graphs reveals how habitat structure affects phenotypic differentiation*, *Commun Biol* **5** 668.
- Brownstein, C. D., D. J. MacGuigan, D. Kim et al. (2024), *The genomic signatures of evolutionary stasis*, *Evolution* **78** 821.

- Butler, M. A. and A. A. King (2004),
Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution,
Am Nat **164** 683.
- Byers, D. L. (2005), *Evolution in heterogeneous environments and the potential of maintenance of genetic variation in traits of adaptive significance*, Genetica **123** 107.
- Cadotte, M. W. (2006), *Dispersal and species diversity: a meta-analysis*, Am Nat **167** 913.
- Camacho, C., G. Coulouris, V. Avagyan et al. (2009), *BLAST+: architecture and applications*, BMC Bioinformatics **10** 1.
- Capella-Gutierrez, S., J. M. Silla-Martinez and T. Gabaldon (2009),
trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses,
Bioinformatics **25** 1972.
- Cardillo, M., G. M. Mace, K. E. Jones et al. (2005),
Multiple causes of high extinction risk in large mammal species, Science **309** 1239.
- Castresana, J. (2000),
Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis,
Mol Biol Evol **17** 540.
- Cerezer, F. O., C. S. Dambros, M. T. P. Coelho et al. (2023), *Accelerated body size evolution in upland environments is correlated with recent speciation in South American freshwater fishes*,
Nat Commun **14** 6070.
- Chen, D., J. Liao, D. Bearup and Z. Li (2020),
Habitat heterogeneity mediates effects of individual variation on spatial species coexistence,
Proc Biol Sci **287** 20192436.
- Cheverud, J. M. (1982), *Relationships among ontogenetic, static, and evolutionary allometry*,
Am J Phys Anthropol **59** 139.
- Chong, K. L., A. Grahn, C. D. Perl and L. Sumner-Rooney (2024),
Allometry and ecology shape eye size evolution in spiders, Curr Biol **34** 3178.
- Claramunt, S., E. P. Derryberry, J. Remsen J. V. and R. T. Brumfield (2012),
High dispersal ability inhibits speciation in a continental radiation of passerine birds,
Proc Biol Sci **279** 1567.
- Cock, A. G. (1966), *Genetical aspects of metrical growth and form in animals*, Q Rev Biol **41** 131.
- Cock, P. J., T. Antao, J. T. Chang et al. (2009),
Biopython: freely available Python tools for computational molecular biology and bioinformatics,
Bioinformatics **25** 1422.
- Conner, J. K., I. A. Cooper, R. J. La Rosa, S. G. Pérez and A. M. Royer (2014),
Patterns of phenotypic correlations among morphological traits across plants and animals,
Philosophical Transactions of the Royal Society B: Biological Sciences **369** 20130246.
- Cooney, C. R. and G. H. Thomas (2021), *Heterogeneous relationships between rates of speciation and body size evolution across vertebrate clades*, Nat Ecol Evol **5** 101.
- Corcobado, G., M. A. Rodríguez-Gironés, J. Moya-Laraño and L. Avilés (2012),
Sociality level correlates with dispersal ability in spiders, Functional Ecology **26** 794.
- Coyne, J. A., N. H. Barton and M. Turelli (1997),
Perspective: a critique of Sewall Wright's shifting balance theory of evolution, Evolution **51** 643.
- Crouch, N. M. A. and R. E. Ricklefs (2019),
Speciation Rate Is Independent of the Rate of Evolution of Morphological Size, Shape, and Absolute Morphological Specialization in a Large Clade of Birds, Am Nat **193** E78.

-
- Czekanski-Moir, J. E. and R. J. Rundell (2019),
The Ecology of Nonecological Speciation and Nonadaptive Radiations, Trends Ecol Evol **34** 400.
- Dale Guthrie, R. (2003), *Rapid body size decline in Alaskan Pleistocene horses before extinction*, Nature **426** 169.
- Daley, M. A. and A. Birn-Jeffery (2018), *Scaling of avian bipedal locomotion reveals independent effects of body mass and leg posture on gait*, J Exp Biol **221** jeb152538.
- DeLong, J. P., J. G. Okie, M. E. Moses, R. M. Sibly and J. H. Brown (2010),
Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life, Proc Natl Acad Sci U S A **107** 12941.
- Dong, Y., J. A. Harvey, R. Steegh, R. Gols and M. Rowe (2023),
The role of male body size in mating success and male–male competition in a false widow spider, Anim Behav **206** 53.
- Eberhard, W., R. L. Rodriguez and M. Polihronakis (2009),
Pitfalls in understanding the functional significance of genital allometry, J Evol Biol **22** 435.
- Eberhard, W. G. (1992), *Web construction by Modisimus sp. (Araneae, Pholcidae)*, J Arachnol **20** 25.
- Eberhard, W. G. and R. D. Briceno (1984),
Behavior and ecology of four species of Modisimus and Blechroscelis (Araneae, Pholcidae), Revue arachnologique **6** 29.
- Eberle, J., D. Dimitrov, A. Valdez-Mondragon and B. A. Huber (2018),
Microhabitat change drives diversification in pholcid spiders, BMC Evol Biol **18** 141.
- Egset, C. K., T. F. Hansen, L. E. R. A et al. (2012),
Artificial selection on allometry: change in elevation but not slope, J Evol Biol **25** 938.
- Emberts, Z., U. Somjee and J. J. Wiens (2023), *Selection on weapon allometry in the wild*, Evolution **77** 2090.
- Entling, W., M. H. Schmidt-Entling, S. Bacher, R. Brandl and W. Nentwig (2010),
Body size–climate relationships of European spiders, J Biogeogr **37** 477.
- Etienne, R. S., S. N. de Visser, T. Janzen et al. (2012),
Can clade age alone explain the relationship between body size and diversity?, Interface Focus **2** 170.
- Evans, K. M., O. Larouche, S.-J. Watson et al. (2021),
Integration drives rapid phenotypic evolution in flatfishes, Proc Natl Acad Sci U S A **118**.
- Faircloth, B. C. (2016), *PHYLUCE is a software package for the analysis of conserved genomic loci*, Bioinformatics **32** 786.
- Felsenstein, J. (1973),
Maximum-likelihood estimation of evolutionary trees from continuous characters, Am J Hum Genet **25** 471.
- FitzJohn, R. G. (2010), *Quantitative traits and diversification*, Syst Biol **59** 619.
- Folk, R. A., R. L. Stubbs, M. E. Mort et al. (2019),
Rates of niche and phenotype evolution lag behind diversification in a temperate radiation, Proc Natl Acad Sci U S A **116** 10874.
- Framenau, V. W. (2005), *Gender specific differences in activity and home range reflect morphological dimorphism in wolf spiders (Araneae, Lycosidae)*, J Arachnol **33** 334.
- Frankino, W. A., B. J. Zwaan, D. L. Stern and P. M. Brakefield (2005),
Natural selection and developmental constraints in the evolution of allometries, Science **307** 718.

- Freckleton, R. P., A. B. Phillimore and M. Pagel (2008), *Relating traits to diversification: a simple test*, *Am Nat* **172** 102.
- Gage, M. J., G. A. Parker, S. Nylin and C. Wiklund (2002), *Sexual selection and speciation in mammals, butterflies and spiders*, *Proc Biol Sci* **269** 2309.
- Gillet, A., B. Frederich and E. Parmentier (2019), *Divergent evolutionary morphology of the axial skeleton as a potential key innovation in modern cetaceans*, *Proc Biol Sci* **286** 20191771.
- Gotanda, K. M., C. Correa, M. M. Turcotte, G. Rolshausen and A. P. Hendry (2015), *Linking macrorends and microrates: Re-evaluating microevolutionary support for Cope's rule*, *Evolution* **69** 1345.
- Gould, S. J. (1966), *Allometry and size in ontogeny and phylogeny*, *Biol Rev Camb Philos Soc* **41** 587.
- Griethuijsen, L. I. van and B. A. Trimmer (2014), *Locomotion in caterpillars*, *Biol Rev Camb Philos Soc* **89** 656.
- Grossi, B., C. Veloso, A. Taucare-Ríos and M. Canals (2016), *Allometry of locomotor organs and sexual size dimorphism in the mygalomorph spider *Grammostola rosea* (Walckenaer, 1837)(Araneae, Theraphosidae)*, *J Arachnol* **44** 99.
- Harmon, L. J., J. B. Losos, T. Jonathan Davies et al. (2010), *Early bursts of body size and shape evolution are rare in comparative data*, *Evolution* **64** 2385.
- Harvey, M. G. and D. L. Rabosky (2018), *Continuous traits and speciation rates: Alternatives to state-dependent diversification models*, *Methods Ecol Evol* **9** 984.
- Harvey, P. H. and M. D. Pagel (1991), *The comparative method in evolutionary biology*, Oxford university press, ISBN: 1383027404.
- Head, G. (1995), *Selection on Fecundity and Variation in the Degree of Sexual Size Dimorphism Among Spider Species (Class Araneae)*, *Evolution* **49** 776.
- Heim, N. A., M. L. Knope, E. K. Schaal, S. C. Wang and J. L. Payne (2015), *Cope's rule in the evolution of marine animals*, *Science* **347** 867.
- Holmes, P., R. J. Full, D. Koditschek and J. Guckenheimer (2006), *The Dynamics of Legged Locomotion: Models, Analyses, and Challenges*, *SIAM Review* **48** 207.
- Hone, D. W. and M. J. Benton (2005), *The evolution of large size: how does Cope's Rule work?*, *Trends Ecol Evol* **20** 4.
- Houle, D., L. T. Jones, R. Fortune and J. L. Sztepanacz (2019), *Why does allometry evolve so slowly?*, *Integr Comp Biol* **59** 1429.
- Huber, B. A. (1998), *Genital mechanics in some neotropical pholcid spiders (Araneae: Pholcidae), with implications for systematics*, *J Zool* (1987) **244** 587.
- Huber, B. A. (2011), *Phylogeny and classification of Pholcidae (Araneae): an update*, *J Arachnol* **39** 211.
- Huber, B. A. (2013), *Revision and cladistic analysis of the Guineo-Congolian spider genus *Smeringopina* Kraus (Araneae, Pholcidae)*, *Zootaxa* **3713** 1.
- Huber, B. A. (2018), *Cave-dwelling pholcid spiders (Araneae, Pholcidae): a review*, *Subterr Biol* **26** 1.
- Huber, B. A. (2021), *First northwest African species of the spider genus *Artema*, from caves in Morocco, with notes on body size in pholcid spiders (Araneae: Pholcidae)*, *Zootaxa* **4984** 324334.
- Huber, B. A. and A. E. Acurio (2022), *Are introduced spiders displacing native species on Galápagos? Observations on pholcid spiders (Araneae: Pholcidae)*, *Arachnology* **19** 191.

-
- Huber, B. A., J. Eberle and D. Dimitrov (2018), *The phylogeny of pholcid spiders: a critical evaluation of relationships suggested by molecular data (Araneae, Pholcidae)*, Zookeys 51.
- Huber, B. A., **G. Meng**, T. M. Dederichs et al. (2024), *Castaways: the Leeward Antilles endemic spider genus Papiamenta (Araneae: Pholcidae)*, Invertebr Syst **38** IS23052, DOI: [10.1071/IS23052](https://doi.org/10.1071/IS23052).
- Huber, B. A., **G. Meng**, N. Dupérré, J. Astrin and M. Herrera (2023), *Andean giants: Priscula spiders from Ecuador, with notes on species groups and egg-sac troglomorphism (Araneae: Pholcidae)*, Eur J Taxon **909** 1, DOI: [10.5852/ejt.2023.909.2351](https://doi.org/10.5852/ejt.2023.909.2351).
- Huber, B. A., **G. Meng**, J. C. garcía and L. S. Carvalho (2024), *Thriving in dry conditions: on the Neotropical spider genus Galapa (Araneae: Pholcidae)*, Zootaxa **5419** 301, DOI: [10.11646/zootaxa.5419.3.1](https://doi.org/10.11646/zootaxa.5419.3.1).
- Huber, B. A., **G. Meng**, J. Král et al. (2023), *High and dry: integrative taxonomy of the Andean spider genus Nerudia (Araneae: Pholcidae)*, Zool J Linn Soc **198** 534, DOI: [10.1093/zoolinnean/zlac100](https://doi.org/10.1093/zoolinnean/zlac100).
- Huber, B. A., **G. Meng**, A. Valdez-Mondragón et al. (2023), *Short-legged daddy-long-leg spiders in North America: the genera Pholcophora and Tolteca (Araneae, Pholcidae)*, Eur J Taxon **880** 1, DOI: [10.5852/ejt.2023.880.2173](https://doi.org/10.5852/ejt.2023.880.2173).
- Hunt, G. and K. Roy (2006), *Climate change, body size evolution, and Cope's Rule in deep-sea ostracodes*, Proc Natl Acad Sci U S A **103** 1347.
- Jablonski, D. (1997), *Body-size evolution in Cretaceous molluscs and the status of Cope's rule*, Nature **385** 250.
- Jablonski, D. (2017), *Approaches to Macroevolution: 2. Sorting of Variation, Some Overarching Issues, and General Conclusions*, Evol Biol **44** 451.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann and A. O. Mooers (2012), *The global diversity of birds in space and time*, Nature **491** 444.
- Kalan, A. K., L. Kulik, M. Arandjelovic et al. (2020), *Environmental variability supports chimpanzee behavioural diversity*, Nat Commun **11** 4451.
- Karlicki, M., S. Antonowicz and A. Karnkowska (2022), *Tiara: deep learning-based classification system for eukaryotic sequences*, Bioinformatics **38** 344.
- Kaspari, M. and M. Weiser (2007), *The size-grain hypothesis: do macroarthropods see a fractal world?*, Ecol Entomol **32** 279.
- Kaspari, M. and M. D. Weiser (1999), *The size-grain hypothesis and interspecific scaling in ants*, Functional ecology **13** 530.
- Katoh, K. and D. M. Standley (2013), *MAFFT multiple sequence alignment software version 7: improvements in performance and usability*, Mol Biol Evol **30** 772.
- King, G. F. and M. C. Hardy (2013), *Spider-venom peptides: structure, pharmacology, and potential for control of insect pests*, Annu Rev Entomol **58** 475.
- Kingsolver, J. G. and D. W. Pfennig (2004), *Individual-level selection as a cause of Cope's rule of phyletic size increase*, Evolution **58** 1608.
- Knouft, J. H. and L. M. Page (2003), *The evolution of body size in extant groups of North American freshwater fishes: speciation, size distributions, and Cope's rule*, Am Nat **161** 413.
- Kück, P. and G. C. Longo (2014), *FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies*, Frontiers in zoology **11** 1.

- Kuntner, M. and J. A. Coddington (2020), *Sexual Size Dimorphism: Evolution and Perils of Extreme Phenotypes in Spiders*, *Annu Rev Entomol* **65** 57.
- Kuntner, M., C. A. Hamilton, R.-C. Cheng et al. (2019), *Golden Orbweavers Ignore Biological Rules: Phylogenomic and Comparative Analyses Unravel a Complex Evolution of Sexual Size Dimorphism*, *Syst Biol* **68** 555.
- LaBarbera, M. (1989), *Analyzing body size as a factor in ecology and evolution*, *Annu Rev Ecol Syst* **20** 97.
- Labonte, D. and W. Federle (2013), *Functionally different pads on the same foot allow control of attachment: stick insects have load-sensitive "heel" pads for friction and shear-sensitive "toe" pads for adhesion*, *PloS One* **8** e81943.
- Lagesen, K., P. Hallin, E. A. Rødland et al. (2007), *RNAmmber: consistent and rapid annotation of ribosomal RNA genes*, *Nucleic Acids Res* **35** 3100.
- Larsson, A. (2014), *AliView: a fast and lightweight alignment viewer and editor for large datasets*, *Bioinformatics* **30** 3276.
- Lee, M. S. Y., A. Cau, D. Naish and G. J. Dyke (2014), *Dinosaur evolution. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds*, *Science* **345** 562.
- Li, C., X. Lian, J. Bi et al. (2011), *Effects of sand grain size and morphological traits on running speed of toad-headed lizard Phrynocephalus frontalis*, *J Arid Environ* **75** 1038.
- Liu, J., L. J. May-Collado, S. Pekar and I. Agnarsson (2016), *A revised and dated phylogeny of cobweb spiders (Araneae, Araneoidea, Theridiidae): A predatory Cretaceous lineage diversifying in the era of the ants (Hymenoptera, Formicidae)*, *Mol Phylogenet Evol* **94** 658.
- Losos, J. B. and D. B. Miles (2002), *Testing the hypothesis that a clade has adaptively radiated: iguanid lizard clades as a case study*, *Am Nat* **160** 147.
- MacArthur, R. H. and J. W. MacArthur (1961), *On bird species diversity*, *Ecology* **42** 594.
- Maher, A. E., G. Burin, P. G. Cox et al. (2022), *Body size, shape and ecology in tetrapods*, *Nat Commun* **13** 4340.
- Mahler, D. L., L. J. Revell, R. E. Glor and J. B. Losos (2010), *Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles*, *Evolution* **64** 2731.
- Marques, D. A., F. C. Jones, F. Di Palma, D. M. Kingsley and T. E. Reimchen (2022), *Genomic changes underlying repeated niche shifts in an adaptive radiation*, *Evolution* **76** 1301.
- Marshall, S. D. and J. L. Gittleman (1994), *Clutch size in spiders: is more better?*, *Functional Ecology* **118**.
- McDonald, T. K. and S. Yeaman (2018), *Effect of migration and environmental heterogeneity on the maintenance of quantitative genetic variation: a simulation study*, *J Evol Biol* **31** 1386.
- Meng, G., Y. Li, C. Yang and S. Liu (2019), *MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization*, *Nucleic Acids Res* **47** e63.
- Meng, G., L. Podsiadlowski, D. Dimitrov and B. A. Huber (2025), *The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders*, submitted.

-
- Minh, B. Q., H. A. Schmidt, O. Chernomor et al. (2020), *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era*, *Mol Biol Evol* **37** 1530.
- Monroe, M. J. and F. Bokma (2010), *Little evidence for Cope's rule from Bayesian phylogenetic analysis of extant mammals*, *J Evol Biol* **23** 2017.
- Moore, B. R., S. Hohna, M. R. May, B. Rannala and J. P. Huelsenbeck (2016), *Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures*, *Proc Natl Acad Sci U S A* **113** 9569.
- Mowery, M. A., Y. Lubin, A. Harari, A. C. Mason and M. C. B. Andrade (2022), *Dispersal and life history of brown widow spiders in dated invasive populations on two continents*, *Anim Behav* **186** 207.
- Nooten, S. S., P. Schultheiss, R. C. Rowe, S. L. Facey and J. M. Cook (2019), *Habitat complexity affects functional traits and diversity of ant assemblages in urban green spaces (Hymenoptera: Formicidae)*, *Myrmecological News* **29** 67.
- Norberg, J., D. P. Swaney, J. Dushoff et al. (2001), *Phenotypic diversity and ecosystem functioning in changing environments: a theoretical framework*, *Proc Natl Acad Sci U S A* **98** 11376.
- Nosil, P., L. J. Harmon and O. Seehausen (2009), *Ecological explanations for (incomplete) speciation*, *Trends Ecol Evol* **24** 145.
- O'Meara, B. C., C. Ane, M. J. Sanderson and P. C. Wainwright (2006), *Testing for different rates of continuous trait evolution using likelihood*, *Evolution* **60** 922.
- Pagel, M. (1999), *Inferring the historical patterns of biological evolution*, *Nature* **401** 877.
- Paradis Emmanuel and Schliep, K. (2019), *ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R*, *Bioinformatics* **35** 526.
- Parr, Z., C. L. Parr and S. L. Chown (2003), *The size-grain hypothesis: a phylogenetic and field test*, *Ecol Entomol* **28** 475.
- Pearce-Duvet, J. M. C., C. P. H. Elemans and D. H. Feener Jr (2011), *Walking the line: search behavior and foraging success in ant species*, *Behav Ecol* **22** 501.
- Pechmann, M., S. Khadjeh, F. Sprenger and N.-M. Prpic (2010), *Patterning mechanisms and morphological diversity of spider appendages and their importance for spider evolution*, *Arthropod Struct Dev* **39** 453.
- Pélabon, C., C. Firmat, G. H. Bolstad et al. (2014), *Evolution of morphological allometry*, *Ann N Y Acad Sci* **1320** 58.
- Pennell, M. W., J. M. Eastman, G. J. Slater et al. (2014), *geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees*, *Bioinformatics* **30** 2216.
- Pfennig, D. W., M. A. Wund, E. C. Snell-Rood et al. (2010), *Phenotypic plasticity's impacts on diversification and speciation*, *Trends Ecol Evol* **25** 459.
- Phillips, B. L., G. P. Brown, J. K. Webb and R. Shine (2006), *Invasion and the evolution of speed in toads*, *Nature* **439** 803.
- Pinheiro, J. (2023), *nlme: linear and nonlinear mixed effects models. R package version 3.1-163*, <http://cran.r-project.org/web/packages/nlme/>.
- Plummer, M., N. Best, K. Cowles and K. Vines (2006), *CODA: convergence diagnosis and output analysis for MCMC*, *R News* **6** 7.

- Pontzer, H. (2007), *Effective limb length and the scaling of locomotor cost in terrestrial animals*, J Exp Biol **210** 1752.
- Rabosky, D. L. (2013), *Diversity-Dependence, Ecological Speciation, and the Role of Competition in Macroevolution*, Annu Rev Ecol Evol Syst **44** 481.
- Rabosky, D. L. (2014), *Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees*, PloS One **9** e89543.
- Rabosky, D. L., J. S. Mitchell and J. Chang (2017), *Is BAMM Flawed? Theoretical and Practical Concerns in the Analysis of Multi-Rate Diversification Models*, Syst Biol **66** 477.
- Rainey, P. B. and M. Travisano (1998), *Adaptive radiation in a heterogeneous environment*, Nature **394** 69.
- Revell, L. J. (2012), *phytools: an R package for phylogenetic comparative biology (and other things)*, Methods Ecol Evol **3** 217.
- Ricklefs, R. E. and S. S. Renner (1994), *Species Richness within Families of Flowering Plants*, Evolution **48** 1619.
- Rundell, R. J. and T. D. Price (2009), *Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation*, Trends Ecol Evol **24** 394.
- Scheffers, B. R., L. N. Joppa, L. L. Pimm and W. F. Laurance (2012), *What we know and don't know about Earth's missing biodiversity*, Trends Ecol Evol **27** 501.
- Schmitz, A. (2016), *Respiration in spiders (Araneae)*, J Comp Physiol B **186** 403.
- Sensenig, A., I. Agnarsson and T. A. Blackledge (2011), *Adult spiders use tougher silk: Ontogenetic changes in web architecture and silk biomechanics in the orb-weaver spider*, J Zool (1987) **285** 28.
- Shao, L., Z. Zhao and S. Li (2023), *Is Phenotypic Evolution Affected by Spiders' Construction Behaviors?*, Syst Biol **72** 319.
- Silva, V., C. Biancardi, C. Perafan et al. (2021), *Giant steps: adhesion and locomotion in theraphosid tarantulas*, J Comp Physiol A Neuroethol Sens Neural Behav Physiol **207** 179.
- Simpson, G. G. (1984), *Tempo and mode in evolution*, Columbia University Press, ISBN: 0231891725.
- Siqueira, A. C., R. A. Morais, D. R. Bellwood and P. F. Cowman (2020), *Trophic innovations fuel reef fish diversification*, Nat Commun **11** 2669.
- Smith, A. N. and M. C. Belk (2018), *Does body size affect fitness the same way in males and females? A test of multiple fitness components*, Biol J Linn Soc Lond **124** 47.
- Smith, S. A. and B. C. O'Meara (2012), *treePL: divergence time estimation using penalized likelihood for large phylogenies*, Bioinformatics **28** 2689.
- Solórzano, A., M. Núñez-Flores, O. Inostroza-Michael and C. E. Hernández (2020), *Biotic and abiotic factors driving the diversification dynamics of Crocodylia*, Palaeontology **63** 415.
- Stanley, S. M. (1973), *An explanation for Cope's rule*, Evolution **27** 1.
- Staudacher, K., O. Rennstam Rubbmark, K. Birkhofer et al. (2018), *Habitat heterogeneity induces rapid changes in the feeding behaviour of generalist arthropod predators*, Funct Ecol **32** 809.
- Steel, M. and A. Mooers (2010), *The expected length of pendant and interior edges of a Yule tree*, Appl Math Lett **23** 1315.

-
- Steenwyk, J. L., T. J. Buida III, Y. Li, X.-X. Shen and A. Rokas (2020),
ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference, PLoS Biol **18** e3001007.
- Suyama, M., D. Torrents and P. Bork (2006), *PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments*, Nucleic Acids Res **34** W609.
- Tabai, Y., H. Kiryu, T. Kin and K. Asai (2008),
A fast structural multiple alignment method for long RNA sequences, BMC Bioinformatics **9** 1.
- Taylor, C. R., N. C. Heglund, T. A. McMahon and T. R. Looney (1980), *Energetic Cost of Generating Muscular Force During Running: A Comparison of Large and Small Animals*, J Exp Biol **86** 9.
- Teuscher, M., M. Brändle, V. Traxel and R. Brandl (2009),
Allometry between leg and body length of insects: lack of support for the size–grain hypothesis, Ecol Entomol **34** 718.
- Tews, J., U. Brose, V. Grimm et al. (2004), *Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures*, J Biogeogr **31** 79.
- Thomsen, M. S., A. H. Altieri, C. Angelini et al. (2022),
Heterogeneity within and among co-occurring foundation species increases biodiversity, Nat Commun **13** 581.
- Title, P. O. and D. L. Rabosky (2019), *Tip rates, phylogenies and diversification: What are we estimating, and how good are the estimates?*, Methods Ecol Evol **10** 821.
- Vallat, R. (2018), *Pingouin: statistics in Python*, J. Open Source Softw. **3** 1026.
- Van Tienderen, P. H. (1991),
Evolution of generalists and specialists in spatially heterogeneous environments, Evolution **45** 1317.
- Venail, P. A., R. C. MacLean, T. Bouvier et al. (2008),
Diversity and productivity peak at intermediate dispersal rate in evolving metacommunities, Nature **452** 210.
- Venditti, C., A. Meade and M. Pagel (2011), *Multiple routes to mammalian diversity*, Nature **479** 393.
- Voje, K. L., M. A. Bell and Y. E. Stuart (2022),
Evolution of static allometry and constraint on evolutionary allometry in a fossil stickleback, J Evol Biol **35** 423.
- Voje, K. L., T. F. Hansen, C. K. Egset, G. H. Bolstad and C. Pelabon (2014),
Allometric constraints and the evolution of allometry, Evolution **68** 866.
- Waller, J. T. and E. I. Svensson (2017), *Body size evolution in an old insect order: No evidence for Cope’s Rule in spite of fitness benefits of large size*, Evolution **71** 2178.
- Wells, M. S. (1988),
Effects of body size and resource value on fighting behaviour in a jumping spider, Anim Behav **36** 321.
- West, G. B., J. H. Brown and B. J. Enquist (1997),
A general model for the origin of allometric scaling laws in biology, Science **276** 122.
- White, C. R., D. J. Marshall, L. A. Alton et al. (2019),
The origin and maintenance of metabolic allometry in animals, Nat Ecol Evol **3** 598.
- Wirthlin, M., N. C. B. Lima, R. L. M. Guedes et al. (2018),
Parrot Genomes and the Evolution of Heightened Longevity and Cognition, Curr Biol **28** 4001.

- Wolff, J. O., W. Nentwig and S. N. Gorb (2013),
The great silk alternative: multiple co-evolution of web loss and sticky hairs in spiders,
PloS One **8** e62682.
- Wolff, J. O., K. Wierucka, G. B. Paterno et al. (2022),
Stabilized Morphological Evolution of Spiders Despite Mosaic Changes in Foraging Ecology,
Syst Biol **71** 1487.
- Wright, P. A. and A. J. Turko (2016), *Amphibious fishes: evolution and phenotypic plasticity*,
J Exp Biol **219** 2245.
- WSC (2025), *World Spider Catalog*, DOI: [10.24436/2](https://doi.org/10.24436/2), (visited on 20/01/2025).
- Xu, X., F. Liu, R.-C. Cheng et al. (2015),
Extant primitively segmented spiders have recently diversified from an ancient lineage,
Proceedings of the Royal Society B: Biological Sciences **282** 20142486.

A New Subfamily of Pholcidae from Brazilian Caatinga

Guanliang Meng¹, Leonardo S. Carvalho², Lars Podsiadlowski¹ and Bernhard A. Huber^{1,*}

¹ Zoological Research Museum Alexander Koenig, LIB, Bonn, Germany

² Campus Amílcar Ferreira Sobral, Universidade Federal do Piauí, Floriano

* Corresponding author: b.huber@leibniz-lib.de

NOTE. The taxonomic names will be made available in a journal publication, and I will not apply the formal taxonomic names in this thesis in order to avoid any possible confusion about validity and availability; for the same reason, new subfamily names below are written in a way that clearly shows that they are not made available by this thesis.

NOTE TO READERS: please ignore informal collection codes such as “[Br22-27]”; they will be deleted in the final version accepted by a scientific journal.

5.1 Summary

This chapter is in preparation to be submitted as:

G. Meng, L. Podsiadlowski and B. A. Huber (2025b), *Low coverage genome sequencing reveals a new subfamily of daddy long-legs spiders from Brazilian Caatinga (Araneae, Pholcidae)*, in preparation

Summary

Pholcid spiders have long been classified into five subfamilies, including Ninetinae Simon 1890, Arteminae Simon 1893, Modisiminae Simon 1893, Smeringopinae Simon 1893, and Pholcinae C.L. Koch 1850. This framework ultimately dates back to Eugène Simon’s non-phylogenetic system of 1893 (Simon 1893). While subfamily relationships and compositions have been updated extensively over the last decades, no new subfamily had to be erected for any of the hundreds of new species newly described since Simon.

Here we report two new species from semi-arid Brazilian Caatinga: GENUS-1 sp1, and GENUS-1 sp2. Morphologically, GENUS-1 is similar to the Chilean genus *Aucana* Huber, 2000, therefore, we hypothesized that they are each other’s closest relatives. To determine their phylogenetic placements, we applied low-coverage whole genome sequencing (lcWGS) to the two new species (three specimens) and to four representatives of *Aucana*. Then we combined these lcWGS data with the ultra-conserved elements

(UCE) of 66 Pholcidae species and four outgroup taxa from **Meng**, Podsiadlowski, Dimitrov et al. (2025) (Chapters 3 and 4 above) for phylogenetic analysis.

Contig assembly from the lcWGS data was performed with the Shovill pipeline (Seemann 2017). Following **Meng**, Podsiadlowski, Dimitrov et al. (2025), potential contaminations between samples were filtered out and the UCE sequences were extracted with the Phyluce pipeline (Faircloth 2016) from the assembled contigs. Multiple-sequence alignment (MSA) and quality control were performed with MAFFT (Katoh et al. 2013) and TrimAl (Capella-Gutierrez et al. 2009). Based on different minimum MSA length (50 vs 150 bp) and taxon coverage (50-80%), we created five different matrices. We then tested if there were systematic bias in the matrices. To achieve a robust phylogenetic placement of the new species, we applied both concatenation- and coalescence-based strategies using six different tree construction algorithms, including IQ-TREE (Minh et al. 2020) with the best-fit model determined by ModelFinder (Kalyaanamoorthy et al. 2017), RAxML-NG (model: GTR+R4+I+FO) (Kozlov et al. 2019), unweighted and weighted ASTRAL-IV (Zhang and Mirarab 2022), SVDquartets (Chifman et al. 2014) from PAUP* (Swofford 2002), and PhyloBayes-MPI (model: CAT-GTR) (Lartillot et al. 2013).

Our datasets were robust, showing high phylogenetic signal, with only a few loci (7.2%) violating the assumptions of stationarity and homogeneity, and intra-locus recombination present in no more 0.6% of the total sequences. Additionally, by adopting the multiple species coalescent (MSC) theory, ASTRAL-IV is statistically consistent when incomplete lineage sorting (ILS) is present (Zhang and Mirarab 2022; Zhang, Rabiee et al. 2018). Moreover, The models GTR+R4+I+FO and CAT-GTR are able to handle rate and compositional heterogeneity, mitigating the influences of model violation (Kozlov et al. 2019; Lartillot et al. 2013).

With respect to previously known subfamilies, our analyses consistently recovered the same subfamily-level phylogenetic relationships, which were identical to these of **Meng**, Podsiadlowski, Dimitrov et al. (2025). Furthermore, the sister-group relationship between the two new species were also strongly supported; we thus joined them conservatively in a single genus, GENUS-1, even though they show some remarkable morphological differences. This genus is sister to a large clade including all pholcid subfamilies except Pholcinae and Smeringopinae, which necessitates the erection of a new subfamily: Cai***inae. In addition, we formalized the separation of the genus *Artema* from “other Arteminae”. This had previously been suggested by **Meng**, Podsiadlowski, Dimitrov et al. (2025), and is strongly supported by our new genomic data. Arteminae is newly circumscribed to include only *Artema* and *Priscula*, and the name Phy***inae is proposed for “other Arteminae”. Pholcidae is thus divided into seven subfamilies, with the following relationships suggested by genomic data: (Pholcinae + Smeringopinae) + (Cai***inae + ((Arteminae + Ninetinae) + (Phy***inae + Modisiminae))). Finally, we tested the hypothesis that the Chilean genus *Aucana* is the closest relative of the new Brazilian species. This is strongly rejected; *Aucana* is resolved as the only known South American representative of Phy***inae.

Personal contribution

This project was a collaborative effort. Bernhard A. Huber initiated the project, collected specimens, and made taxonomic descriptions based on morphological characteristics. Leonardo S. Carvalho applied for sampling permits and collected specimens. Lars Podsiadlowski coordinated the molecular work. Guanliang Meng, Anja Bodenheimer and Sandra Kukowka performed molecular methods including DNA isolation and library construction. Lars Podsiadlowski performed contig assembly. Guanliang Meng curated and analyzed the molecular data. Bernhard A. Huber, Guanliang Meng and Leonardo S. Carvalho wrote the initial version of the manuscript. All authors contributed to the revision and improvement of the manu-

script.

Our work exemplifies that extensive taxon and gene sampling can greatly uncover hidden biodiversity. In conjunction with the work presented here, I also contributed to **14** studies on Pholcidae taxonomy, systematics and biodiversity (See **List of Publications and Contributions** at the beginning of the thesis), which in total described **73 new species in 16 genera** and **revised 5 genera**. In these publications, I was in charge of molecular analyses and writings of the corresponding parts. The molecular analyses included sequence assembly, examination of contamination, tree inference, computation of genetic distances, and delimiting species based on molecular data.

5.2 Introduction

Pholcidae are ubiquitous spiders in tropical and subtropical regions around the world. They occupy a wide range of different microhabitats, with corresponding variation in body size, body shape, leg length, and coloration (Eberle et al. 2018). More than 1,100 new species have been added to the family within the last 20 years, and assignments of new species to higher taxonomic ranks (subfamilies and genera) have usually been unproblematic. This is particularly true at the level of subfamilies, even though subfamily-level relationships and exact compositions of subfamilies have proven difficult to resolve in some cases. For example, relationships of Ninetinae to other subfamilies and the positions of the genera *Artema* Walckenaer, 1873 and *Priscula* Simon, 1893 at subfamily level have long been dubious (Dimitrov et al. 2013; Eberle et al. 2018; Huber 2000; Huber, Eberle et al. 2018) and could only be resolved recently with genomic data (Meng, Podsiadlowski, Dimitrov et al. 2025).

Thus, the current subfamily division of Pholcidae is essentially an updated version of Simon's (1893) classification (Simon 1893), with numerous corrections and changes but without any new subfamily-level names (Huber 2011). For more than a century, no new species have been discovered that would defy assignment to one of the subfamily-level taxa of Simon (1893) or their current emended versions. Here, we report on two newly discovered species for which genomic data suggest they are neither included into nor sister of an existing subfamily; thus, they require the creation of a new subfamily.

In addition, we formalize here the separation of *Artema* from "other Arteminae," which was first suggested in Eberle et al. (2018) and which is strongly supported by new genomic data (Meng, Podsiadlowski, Dimitrov et al. 2025). We thus redefine Arteminae to include only *Artema* and *Priscula* and propose a new subfamily name for the "other Arteminae" sensu Eberle et al. (2018) and Huber, Eberle et al. (2018). Finally, we provide first molecular data for the Chilean genus *Aucana* Huber, because we initially thought (based on superficial similarity) that this genus might be the closest relative of the new Brazilian species.

5.3 Materials and methods

5.3.1 Taxonomy and Morphology

The morphological part of this study is based on specimens deposited in Coleção de História Natural da Universidade Federal do Piauí, Floriano (CHNUFPI); Centro de Coleções Taxonômicas da Universidade Federal de Minas Gerais, Belo Horizonte (UFMG); and Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany (ZFMK). Taxonomic descriptions follow the style of recent publications on Pholcidae (e.g., Huber, Meng, Král, I. M. Á. Herrera et al. 2024; Huber, Meng and Valdez-Mondragn 2024; based on Huber 2000). Measurements were done on a dissecting microscope with an ocular grid and are in

mm unless otherwise noted; eye measurements are $\pm 5 \mu\text{m}$. Photos were made with a Nikon Coolpix 995 digital camera (2048 x 1536 pixels) mounted on a Nikon SMZ 18 stereo microscope or a Leitz Dialux 20 compound microscope. CombineZP (<https://combinezp.software.informer.com/>) was used for stacking photos. Drawings are partly based on photos that were traced on a light table and later improved under a dissecting microscope, or they were directly drawn with a Leitz Dialux 20 compound microscope using a drawing tube. Cleared epigyna were stained with chlorazol black. For scanning electron microscope (SEM) photos, specimens were dried in hexamethyldisilazane (HMDS) (Brown 1993), sputter-coated with gold, and photographed with a Zeiss Sigma 300 VP scanning electron microscope. The distribution map was generated with ArcMap 10.0.

5.3.2 Abbreviations

ALE - anterior lateral eye(s); **ALS** - anterior lateral spinneret(s); **AME** - anterior median eye(s); **a.s.l.** - above sea level; **L/d** - length/diameter; **PME** - posterior median eye(s); **PMS** - posterior median spinneret(s). Abbreviations used in figures only are explained in the figure legends.

5.3.3 Molecular phylogeny

Taxon sampling

To study the phylogenetic placement of the new Brazilian taxa, we sequenced both known species (three specimens) with a low coverage whole genome sequencing strategy. Since our original hypothesis was that these species might be related to the Chilean genus *Aucana* Huber, 2000 (for which no molecular data had been available), we analyzed four representatives of *Aucana* with the same method, and combined these new data with the ultra-conserved elements (UCEs) of 66 Pholcidae species from Meng, Podsiadlowski, Dimitrov et al. (2025) (see Chapter 3.1). As outgroups, we used the same four taxa as in Meng, Podsiadlowski, Dimitrov et al. (2025) (Table C.1).

Low coverage whole genome sequencing

Following the manufacturer's instruction, DNA was isolated with the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) including the optional RNase digest. DNA concentration and fragment size were determined using a QuantusTM Fluorometer (Cat # E6150, Promega Corporation, USA) and a Fragment Analyzer (Advanced Analytical Technologies, Inc), respectively. Adapting parameters to the determined fragment size, DNA was fragmented by a Bioruptor[®] Pico sonication device (Diagenode S.A.). For each specimen, a total mass of 100 ng fragmented DNA was used for library construction with the NEBNext Ultra II FS DNA Library Prep Kit (NEB #7805), following the manufacturer's instruction. During this process, the fragmented DNA was subject to end-repair, dA-tailing and ligation of molecular-barcoded adaptors, and purified with AMPure[®] XP Beads (Beckman Coulter, Inc. #A63881). The indexed libraries were then further amplified and purified. Finally, concentration and fragment size distributions of the DNA libraries were analyzed again and sequenced by a commercial company (Macrogen Europe) using pair-end (PE) 150 bp strategy. Each specimen was sequenced with ca. 11 Gb of data.

Sequence assembly and UCE extraction

We used the Shovill pipeline (ver. 1.1.0; Seemann 2017) with default parameters for raw data filtering and sequence assembly. Specifically, Trimmomatic (ver. 0.39; Bolger et al. 2014) was employed for quality

trimming and adapter clipping (`-trim`), followed by contig assembly with SPAdes (ver. 4.0.0; Prjibelski et al. 2020). Following Meng, Podsiadlowski, Dimitrov et al. (2025), we applied the “all-blast-all” method to filter out potential cross-contaminations among samples (blast similarity $\geq 99\%$ and alignment length ≥ 100 bp were subject to check) and extracted UCE sequences from the remained assembly with Phyluce (ver. 1.7.1; Faircloth 2016). The extracted UCE sequences were aligned with MAFFT (ver. 7.475; Katoh et al. 2013) (`-adjustdirection -maxiterate 1000`), and resulting alignments were trimmed using TrimAl (ver. 1.4.rev15; Capella-Gutierrez et al. 2009) (`-automated 1`). Finally, five data matrices (matrices 1–5) were created based on different taxon coverage (50, 60, 70 and 80%) and minimum alignment length (50 vs 150 bp) (Table C.2).

Systematic bias assessment

To assess potential systematic biases in the datasets, we conducted a likelihood mapping analysis (Strimmer and von Haeseler 1997) using IQ-TREE (ver. 2.2.0; Minh et al. 2020) and matrix 1 to evaluate the treelikeness of 100,000 randomly sampled quartets. We also tested the assumptions of stationarity and homogeneity, which are integral to most molecular evolutionary models, by performing symmetry tests (Naser-Khdour et al. 2019) in IQ-TREE (`-symtest-only`). To investigate potential violations of the assumption of no intra-locus recombination, we applied 3SEQ (ver. 1.8.0; Boni et al. 2007; Lam et al. 2018), with a pre-computed $3000 \times 3000 \times 3000$ p-value table for the analysis. PhyloMAd (ver. 1.2; Duchene et al. 2022) was used to assess substitutional saturation. Additionally, for each locus, we computed key features such as average GC content, GC content interquartile range, the number and proportion of missing data (gaps and Ns), the number and proportion of parsimony-informative sites, the length of the multiple sequence alignment (MSA), the number of taxa in the MSA, and created a saturation plot using custom scripts. Finally, we visualized the distributions of these features using density plots, along with a missing data distribution plot, generated with ggplot2 (ver. 3.3.6; Ginestet 2011) and R (ver. 4.2.0; R Core Team 2022).

Phylogenetic inference

To each data matrix, we applied both concatenation- and coalescence-based methods using IQ-TREE (the best-fit model was determined by ModelFinder; Kalyaanamoorthy et al. 2017), RAxML-NG (ver. 1.1; Kozlov et al. 2019) (model: GTR+R4+I+FO), unweighted and weighted ASTRAL-IV (ver. 1.19.4.5; Zhang and Mirarab 2022), SVDquartets (Chifman et al. 2014) as implemented in PAUP* (ver. 4a168; Swofford 2002), and PhyloBayes-MPI (ver. 1.8; Lartillot et al. 2013) (model: CAT-GTR), following the methods in Meng, Podsiadlowski, Dimitrov et al. (2025). By down-weighting quartets with low support, long terminal branches, or both, the weighted version of ASTRAL offers stronger theoretical guarantees and improved empirical performance compared to the unweighted version (Zhang and Mirarab 2022). The resulting trees were visualized with Figtree (ver. 1.44; <https://github.com/rambaut/figtree>). Additionally, we used DiscoVista (ver. 1.0; Sayyari et al. 2018) to calculate the relative gene frequency support for different topologies. For each tree, we compared all the other trees against it, calculating the total number of congruent and incongruent branches. The tree with the highest number of supported branches was selected as the preferred topology and used as reference for comparisons and for further discussion.

5.4 Results

5.4.1 Molecular data and phylogeny

A total of 66.28 Gbp raw data was generated from low coverage genome sequencing, with a range of 7.80 to 13.48 Gbp for each sample (Table C.1). Combining our new data with those from Meng, Podsiadlowski, Dimitrov et al. (2025) and NCBI, we created five data matrices (matrices 1–5), with total locus numbers ranging from 40 to 139, total sites ranging from 9.5 to 30.5 Kbp and numbers of contigs ranging from 2,632 to 7,573 (Table C.2). Most sequences in matrix 1 have a length of about 200 bp (ranging from 50 to 550 bp), and missing data account for less than 2.5% of the alignment length (Figure C.1 and C.2). Likelihood mapping analyses show that about 32% of the quartets are located at each of the three corners of a triangular graph, suggesting that our dataset has high information content with respect to phylogenetic analysis (Figure C.3). For only 10 out of 139 loci (7.2%), the assumptions of stationarity or homogeneity, or both, were rejected (SymPval < 0.001) (Table C.3). The substitution saturation analysis showed that all loci are under low risk of saturation (Table C.4). Intralocus recombination analysis identified 45 recombinant sequences (out of 7,573 sequences, 0.6%) (p-value < 0.05), involving 40 out of 139 loci (28.8%), but only 15 sequences (0.2%) have recombinant segments longer than 100 bp (Figure C.4).

The main phylogeny is shown in Fig. 5.1. Analyses from different data matrices and tree construction methods mostly reconstructed the same subfamily-level topology as in Meng, Podsiadlowski, Dimitrov et al. (2025) and similar genus-level relationships, with the new subfamily Cai***inae always being sister to ((Phy***inae, Modisiminae), (Arteminae, Ninetinae)) (Fig. 5.2a). The subfamily-level relationships have high gene frequency supports ($\geq 41\%$) (Fig. 5.2b). In addition, the molecular data strongly reject our original hypothesis that the Chilean genus *Aucana* might be the closest relative of the new Brazilian species. Instead, *Aucana* is resolved as the only South American genus of Phy***inae.

5.4.2 Taxonomy

Genus *GENUS-1* Huber gen. n. (Family Pholcidae C.L. Koch, 1850)

<https://zoobank.org/532DFFE5-EF60-4E09-88E7-C5CB36B226CF>

Type species. *GENUS-1 sp1* Huber sp. n.

Diagnosis. Small Pholcidae (body size <1.5 mm) with relatively short legs (tibia 1 L/d: 13–15) and globular abdomen (Fig. 5.3). Distinguished from representatives of superficially similar South American genera (e.g., *Arenita* Huber & Carvalho, 2019; *Aucana* Huber, 2000, *Blancoa* Huber 2000, *Canaima* Huber, 2000; *Kairona* Huber & Carvalho, 2019; *Kambiwa* Huber, 2000; *Nerudia* Huber, 2000; *Tupigea* Huber, 2000) by combination of: (1) carapace with pair of dark marks (Fig. 5.3); superficially similar Ninetinae lack such paired marks; (2) strong process ventrally on male palpal tarsus, at basis of procurus (arrows in Figs. 5.6b, 5.16b); (3) prominent transversal ridge on male genital bulb, on prolateral-dorsal side (arrows in Figs. 5.6d–e, 5.16d–e); (4) pair of pointed male cheliceral apophyses and corresponding pair of female epigynal pockets (Figs. 5.7a, c, 5.17b, d); (5) absence of pseudosegmentation (Fig. 5.13e); (6) very low ratio of tibia 1 / tibia 2 (~1.05–1.15); higher in most other pholcids except some Ninetinae; Fig. 5.20). For further discussion of individual diagnostic traits, see Discussion.

Description. Total body length ~1.1–1.4; carapace width 0.50–0.60. AME either present (*GENUS-1 sp1*) or strongly reduced to absent (*GENUS-1 sp2*). Legs relatively short: male tibia 1 length 0.8–1.1; female tibia 1 length: 0.7–1.0; male tibia 1 L/d: 13–15. Color (in ethanol) mostly whitish to pale ochre-yellow, with pair of darker ochre marks on carapace; sternum monochromous whitish; legs either

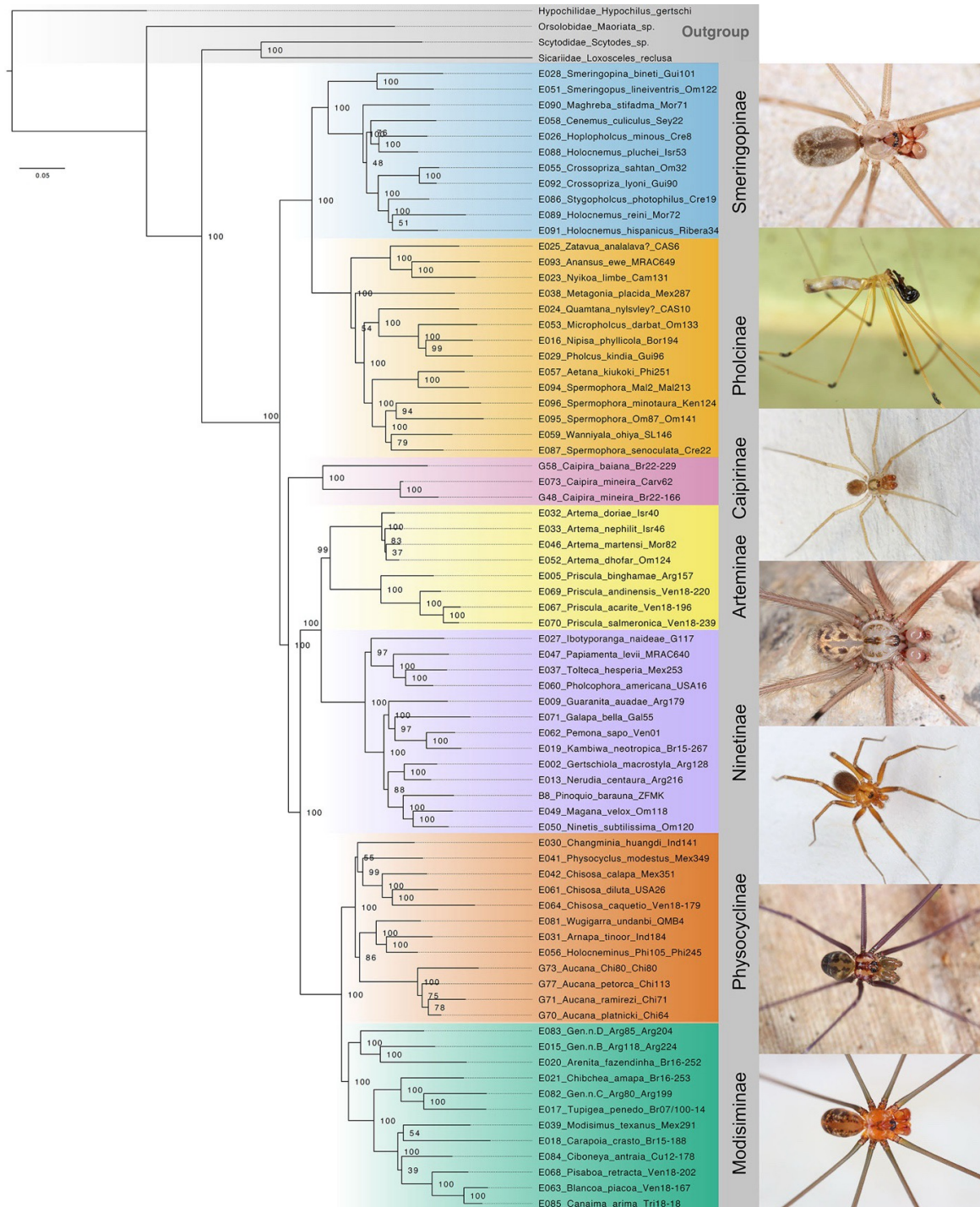


Figure 5.1. Maximum likelihood tree based on UCE data. The tree was constructed with RAXML-NG and matrix 1 using a partitioned strategy and the GTR+R4+I+FO model. Numbers on the branches are bootstrap supports. Representatives of each subfamily on the right, from top: *Hoplopholcus konya*, *Apokayana kapit*, *GENUS-1 sp1*, *Artema martensi*, *Ibotyporanga walekuru*, *Arnapa arfak*, *Mesabolivar eberhardi*; photos BAH.

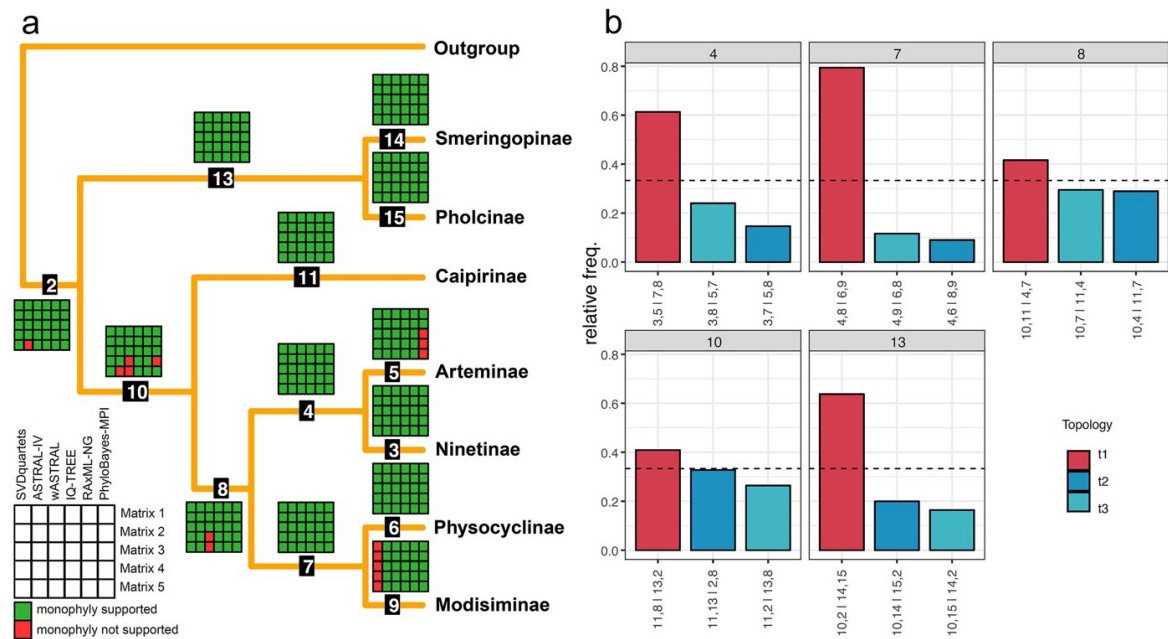


Figure 5.2. Subfamily-level phylogeny and supporting evidence. **a** Phylogenetic relationships among subfamilies. Branch panels display congruence (green) and incongruence (red) across different analyses based on data matrices (matrices 1–5) and various tree construction methods. The primary topology is derived from Fig. 5.1. **b** Relative gene frequency for alternative topologies. Subfigure titles and x-axis labels correspond to the branch labels in **a**. The subfamily-level phylogeny is supported by most analyses and the new subfamily Cai***inae is sister of ((Phy***inae, Modisiminae), (Arteminae, Ninetinae)), with high gene frequency supports (41%).

without darker rings (*GENUS-1 sp1*) or with rings on femora and tibiae (*GENUS-1 sp2*); abdomen either monochromatic (*GENUS-1 sp1*) or with distinct white marks (*GENUS-1 sp2*). Ocular area barely raised. Thoracic groove shallow (Fig. 5.9a, b). Male clypeus either unmodified (*GENUS-1 sp1*) or with pair of short rounded processes near margin and many strong hairs directed upwards (*GENUS-1 sp2*; Fig. 5.17a). Sternum wider than long, unmodified. Abdomen globular. ALS either with eight spigots each (*GENUS-1 sp1*; Fig. 5.10b–e) or with only two spigots each (*GENUS-1 sp2*; not studied with SEM; based only on cleared female abdomen observed in compound microscope). Gonopore in *GENUS-1 sp1* with four epiandrous spigots arranged in two pairs (Fig. 5.9f); not studied in *GENUS-1 sp2*. Male chelicerae with pair of distal pointed apophyses directed towards median, and pair of proximal lateral sclerotized humps (Figs. 5.7a–b, 5.17b–c); without stridulatory ridges (Fig. 5.9d). Male palps either with strongly widened femur (*GENUS-1 sp1*; Fig. 5.5) or with much smaller femur with distinct prolateral-ventral sclerotized apophysis (*GENUS-1 sp2*; Fig. 5.15); tibia-tarsus joints either shifted towards prolateral (*GENUS-1 sp1*) or towards retrolateral (*GENUS-1 sp2*); tarsus with prolateral-ventral process at basis of procursus (arrows in Figs. 5.6b, 5.16b); procursus distally with heavily sclerotized element with pointed tips (Figs. 5.6a–c, 5.16a–c); genital bulb slightly elongated and medially constricted, with prolateral-dorsal protruding flap or ridge (arrows in Figs. 5.6d–e, 5.16d–e), sperm duct opening not seen. Legs without spines, without curved hairs. With sexually dimorphic short ‘vertical’ hairs on tibia 1 only (Fig. 5.13a). Ventrally on metatarsi 3 and 4 with slender hairs (Fig. 5.12a–d) as described recently in Ninetinae and “other Arteminae” (now Phy***inae) (Huber, Meng, Král, Ávila Herrera and Izquierdo 2023; Huber, Meng and

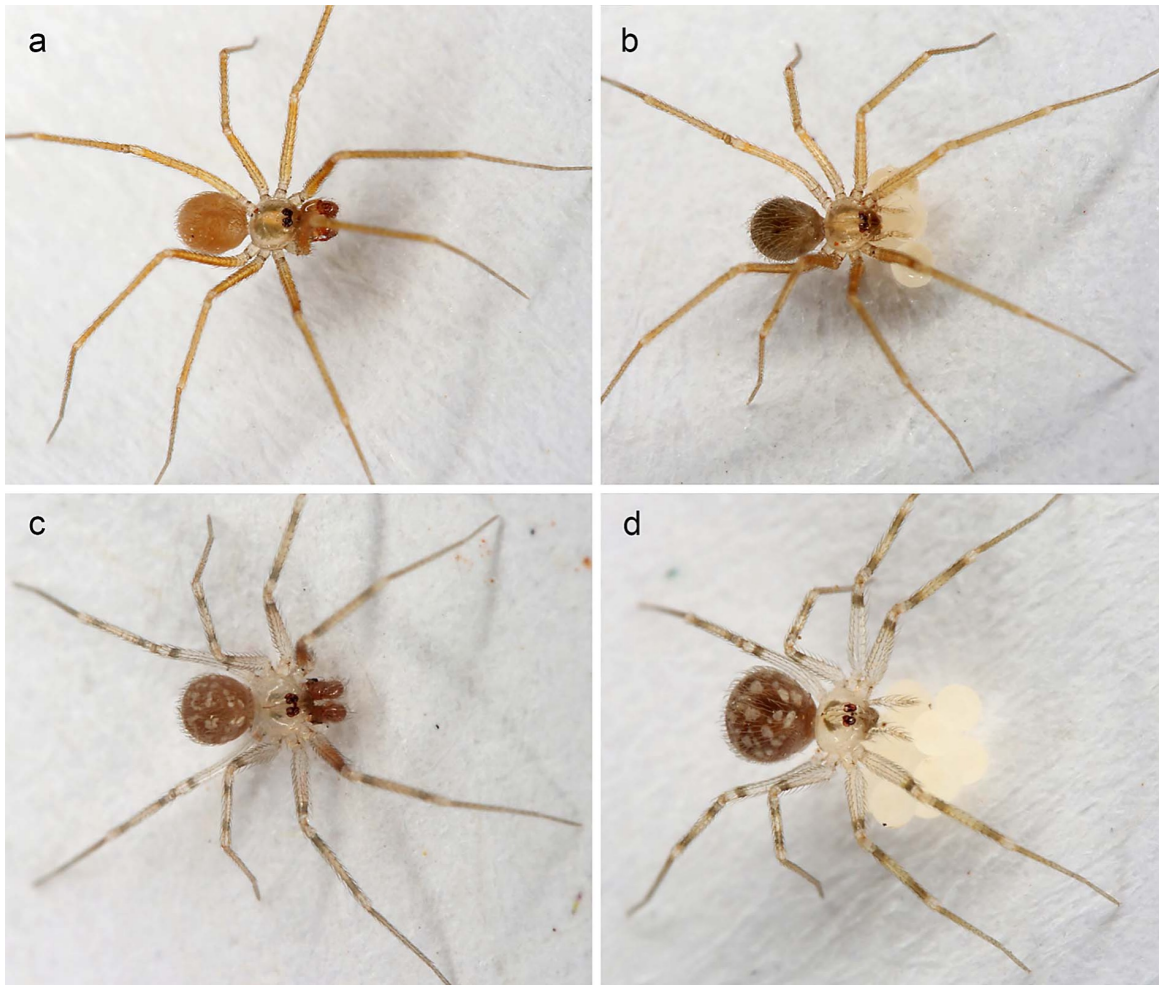


Figure 5.3. Live specimens. a–b *GENUS-1 sp1* Huber sp. n. from Gruta do Janelão; male and female with egg-sac. **c–d** *GENUS-1 sp2* Huber sp. n. from W of Queimada Nova; male and female with egg-sac.

Valdez-Mondragn 2024). Retrolateral trichobothria on leg tibiae in very distal position (male tibia 1: at ~50–60% of tibia length); prolateral trichobothrium absent on tibia 1, present on tibiae 2–4. No tarsal pseudosegmentation visible (also in SEM and in compound microscope). Tarsus 4 distally with two comb-hairs on prolateral side (arrows in Fig. 5.13f; studied in *GENUS-1 sp1* only). Main tarsal and median tarsal claws as typical for family.

Etymology. The genus name is derived from the traditional rural culture of Brazil; gender feminine.

Relationships. See results of molecular analysis above.

Distribution. Known from limestone outcrop terrains in the Irecê and Peruaçu biogeographic districts (sensu Moro et al. 2024) of the Caatinga domain, in the Brazilian states of Minas Gerais and Bahia (Fig. 5.4).

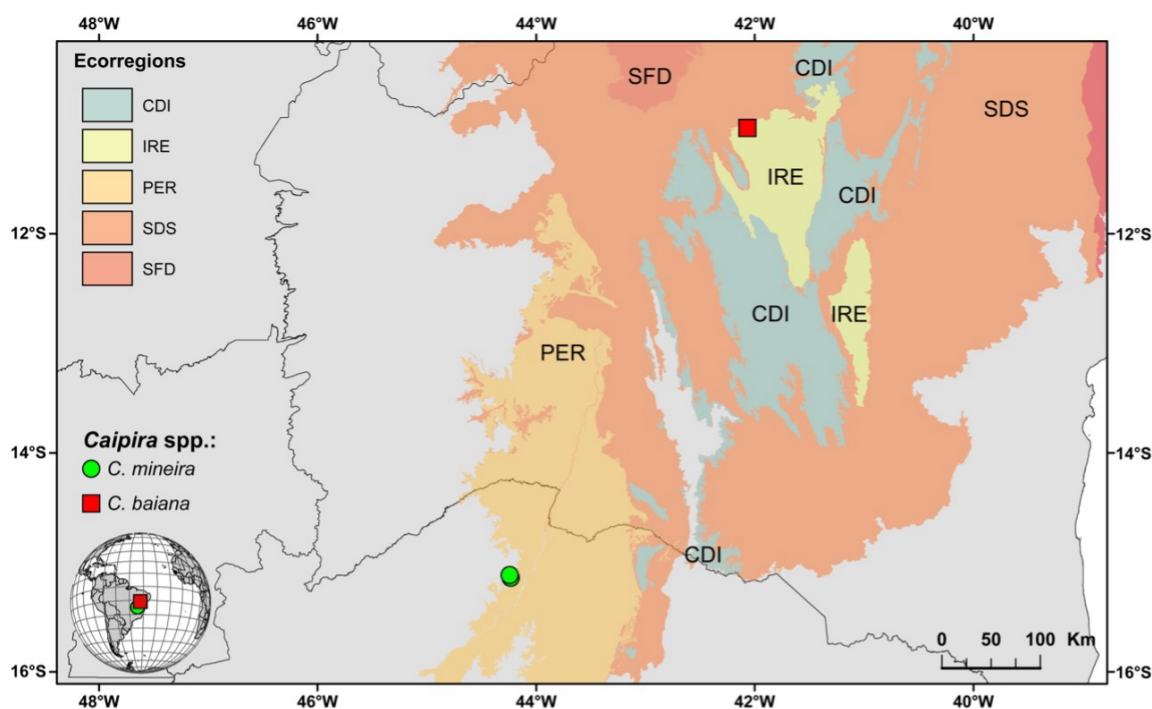


Figure 5.4. Known geographic distribution of *GENUS-1* gen. n., including *GENUS-1 sp2* sp. n. (red square) and *GENUS-1 sp1* sp. n. (green circles). Colored areas represent the main districts of the Caatinga domain: CDI –Chapada Diamantina; IRE –Irecê; PER –Peruaçu; SDS –Southern Depressão Sertaneja; and SFD –São Francisco Dunes. Other biomes are not shown. Note: Peruaçu and Irecê districts are the only limestone outcrops terrains in this region, separated by the crystalline terrains of the Southern Depressão Sertaneja and the Chapada Diamantina.

GENUS-1 sp1 Huber sp. n.

<https://zoobank.org/6A1020F5-61CE-4769-BB6B-9F75AF0033B8>

Figs. 5.3a–b, 5.5–5.14

Diagnosis. Easily distinguished from all other known Pholcidae by shape of procursus (Figs. 5.6a–c, 5.11; slender, proximally with ventral notch and ventral process, distally with dorsal rounded sclerite and ventral membranous hooked element); also by shape of genital bulb (Figs. 5.6d–f, 5.11a–e; dorsal

protruding flap; distal pointed sclerite), by armature of male chelicerae (Figs. 5.7a–b, 5.9c–d; pair of lateral sclerotized humps and pair of frontal pointed apophyses directed towards median), and by female epigynum and internal genitalia (Figs. 5.7c–d, 5.8; anterior plate with pair of pockets close together; posterior plate anteriorly with pair of dark marks; internal genitalia with complex and large median rounded sclerotized structure); from most Pholcidae also by combination of large male palpal femur (Fig. 5.5; much larger than tibia) and small size (body length ~1 mm; prosoma width ~0.5 mm).

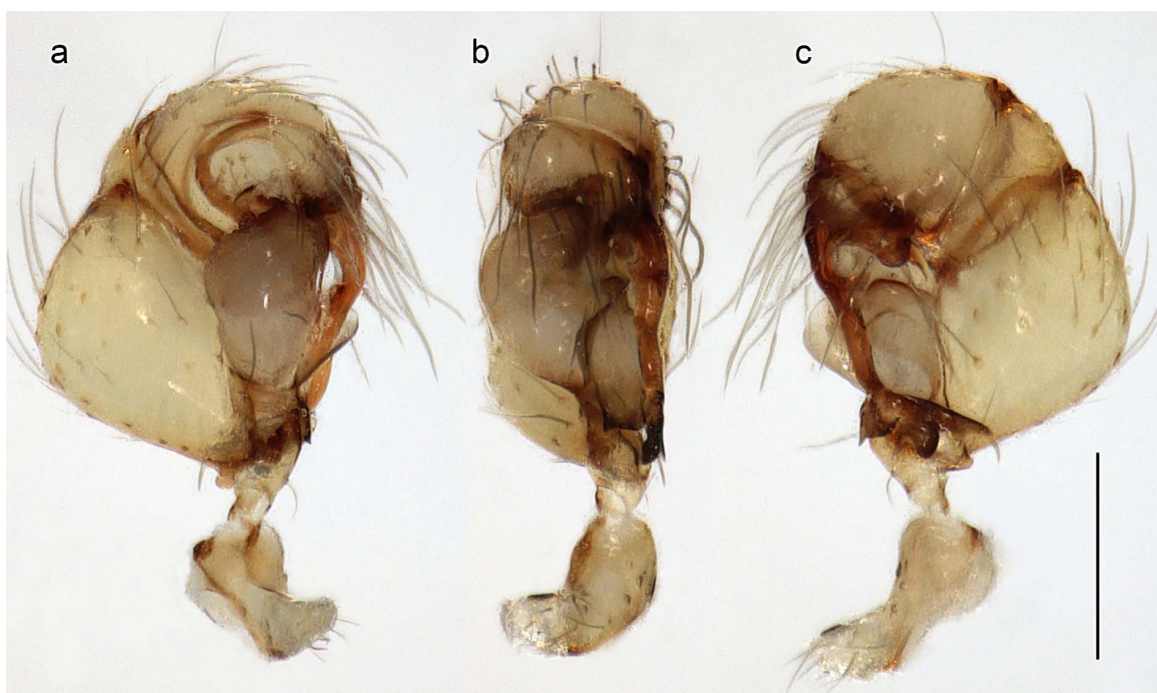


Figure 5.5. *GENUS-1 sp1* Huber sp. n.; male from Gruta do Janelão; ZFMK Ar 24704. Left genital palp, prolateral (a), dorsal (b), and retrolateral (c) views. Scale bar: 0.2 mm.

Holotype. BRAZIL - Minas Gerais • ♂; Parque Nacional Cavernas do Peruaçu, Gruta do Janelão, ~100–300 m from cave entrance; ~15.125°S, 44.240°W (coordinates: cave entrance); ~600 m a.s.l.; 14 Nov. 2022; B.A. Huber, L.S. Carvalho & R.A. Torres leg.; CHNUFPI 9069 [sep Br22-27].

Paratypes. BRAZIL - Minas Gerais • 2 ♂♂, 2 ♀♀; same collection data as for holotype; CHNUFPI 9070 [Br22-27, part1] • 1 ♂, 1 ♀; same collection data as for holotype; UFMG 33242 [Br22-27, part3] • 4 ♂♂, 4 ♀♀, and one female abdomen cleared and transferred from ZFMK Br22-166); same collection data as for holotype; CHNUFPI 9071 [deposited in ZFMK Ar 24704] [Br22-27, part2] • 1 ♀; same collection data as for holotype; CHNUFPI 9072 [Br22-29] • 2 ♂♂, 2 ♀♀, 3 juvs.; same locality as for holotype, “PNCPI5”; 29 Feb. 2020; A.J. Santos et al. leg.; CHNUFPI 4193 [Carv123].

Other material examined. BRAZIL - Minas Gerais • 1 ♂, 5 ♀♀, in pure ethanol; same collection data as for holotype; CHNUFPI 9073 [deposited in ZFMK Br22-166] (vouchers of UH485, G48; 1 ♂ and 1 ♀~used for SEM; one female abdomen cleared and transferred to ZFMK Ar 24704) • 1 ♂, in pure ethanol; same locality as holotype, “PNCPI6”; 1 Mar. 2020; A.J. Santos et al. leg.; CHNUFPI 3505 (voucher of E073) [Carv62] • 2 ♂♂, 1 ♀; Parque Nacional Cavernas do Peruaçu, Lapa do Rezar (karst cave, at cave entrance); 15.1433°S, 44.2349°W; 610 m a.s.l.; 14 Nov. 2022; B.A. Huber, L.S. Carvalho & R.A. Torres

leg.; CHNUFPI 9074 [deposited in ZFMK Ar 24705] [Br22-32] • 3 juvs, in pure ethanol; same collection data as for preceding; CHNUFPI 9075 [deposited in ZFMK Br22-169] (voucher of UH057).

Etymology. The species name is an adjective, describing someone originating from Minas Gerais.

Description. Male (holotype) –MEASUREMENTS. Total length 1.20, carapace width 0.50. Distance PME–PME 30 μ m; diameter PME 45 μ m; distance PME–ALE 10 μ m; diameter AME 20 μ m; distance AME–AME 10 μ m. Leg 1: 3.23 (0.93 + 0.17 + 0.90 + 0.83 + 0.40), tibia 2: 0.80, tibia 3: 0.70, tibia 4: 0.98; tibia 1 L/d: 15. –COLOR (in ethanol). Carapace whitish with pair of lateral submarginal ochre bands, clypeus ochre; sternum white; legs pale ochre-yellow, without dark rings; abdomen monochromatic ochre gray. –BODY. Habitus as in Fig. 5.3a. Ocular area barely raised. Thoracic groove shallow (Fig. 5.9a). Clypeus unmodified. Sternum wider than long (0.40/0.30), unmodified. Abdomen globular. ALS with one strongly widened spigot, one long pointed spigot, and apparently six cylindrical spigots (Fig. 5.10b). PMS with two slender spigots (cf. Fig. 5.10f). Gonopore with four epiandrous spigots arranged in two pairs (Fig. 5.9f). –CHELICERAE. As in Fig. 5.7a–b, with pair of frontal pointed apophyses directed towards median, without modified hairs (Fig. 5.9c–d), and pair of lateral sclerotized humps with series of frontal ridges (Fig. 5.9d), without lateral stridulatory ridges. –PALPS. As in Figs. 5.5 and 5.11; coxa unmodified; trochanter with short ventral apophysis (~25 μ m long); femur large and strongly widened, dorsally bulging, without stridulatory pick; femur-patella joints shifted toward prolateral side; tibia with two trichobothria in very proximal position; tibia-tarsus joints slightly shifted toward prolateral side; tarsus with strong process and deep notch ventrally at basis of procursus (Fig. 5.6b); tarsal organ exposed (Fig. 5.14a), diameter ~10 μ m; procursus (Fig. 5.6a–c) slender, proximally with ventral process, distally with dorsal rounded sclerite and ventral membranous hooked element (Fig. 5.11f); genital bulb (Fig. 5.6d–f) slightly elongated, with distinctive dorsal protruding flap and distal pointed sclerite, sperm duct opening not seen. –LEGS. Without spines, without curved hairs. Tibiae, metatarsi, and tarsi with round cuticular ‘plates’ (Figs. 5.12a, 5.13c) at regular intervals, diameter ~4–5 μ m. Without rimmed pores. With sexually dimorphic short ‘vertical’ hairs (Fig. 5.13a) on tibia 1 only, diameter proximally 1.4 μ m, length ~15–20 μ m; putative chemoreceptors on all legs, especially distally, i.e. on metatarsi and tarsi, similar to sexually dimorphic short ‘vertical’ hairs but with several side branches (Fig. 5.13b) and slightly thicker (diameter proximally 1.6 μ m). Ventrally on metatarsi 3 and 4 with slender hairs (Fig. 5.12a–d); diameters proximally 1.6 μ m, versus 2.8 μ m in neighboring mechanoreceptors; shape of hair basis similar to neighboring mechanoreceptors (Fig. 5.12c), but shaft similar to trichobothria (cf. Fig. 5.12e–f); metatarsus 3 with only three such hairs, metatarsus 4 with ~14; distally very thin and long, with distinctively widened tips (Fig. 5.12d; ~1.4 μ m diameter, versus 0.6 μ m subdistally). Bases of trichobothria capsulate, unmodified (Fig. 5.12e–f); trichobothria feathered (Fig. 5.12e–f); retrolateral trichobothrium of tibia 1 at 48%; prolateral trichobothrium absent on tibia 1. No tarsal pseudosegmentation visible in SEM (Fig. 5.13e) and in compound microscope. Leg tarsal organs exposed (Fig. 5.14f), diameters ~3–6 μ m. Main tarsal and median tarsal claws as typical for family; main claws with 9–10 tines (Fig. 5.13d–f). Tarsus 4 distally with two comb-hairs on prolateral side (Fig. 5.13f).

Variation (male). Tibia 1 in 13 males (incl. holotype): 0.90–1.07 (mean 0.97).

Female. In general very similar to male, including size and carapace pattern (Fig. 5.3b). Chelicerae also without stridulatory ridges (Fig. 5.9e). Tibia 1 in 14 females: 0.83–0.93 (mean 0.89). With short chemoreceptive hairs like in male but without sexually dimorphic short ‘vertical’ hairs. Spinnerets as in male (Fig. 5.10c–f). Palpal tibia with only one trichobothrium (Fig. 5.12e). Epigynum (Fig. 5.8a–b) anterior plate light brown, strongly indented posteriorly, with pair of pockets close together near posterior border (Fig. 5.7c); posterior epigynal plate large but simple, with pair of darker marks anteriorly (muscle attachment sites?). Internal genitalia partly visible in uncleared specimens; with sclerotized roundish

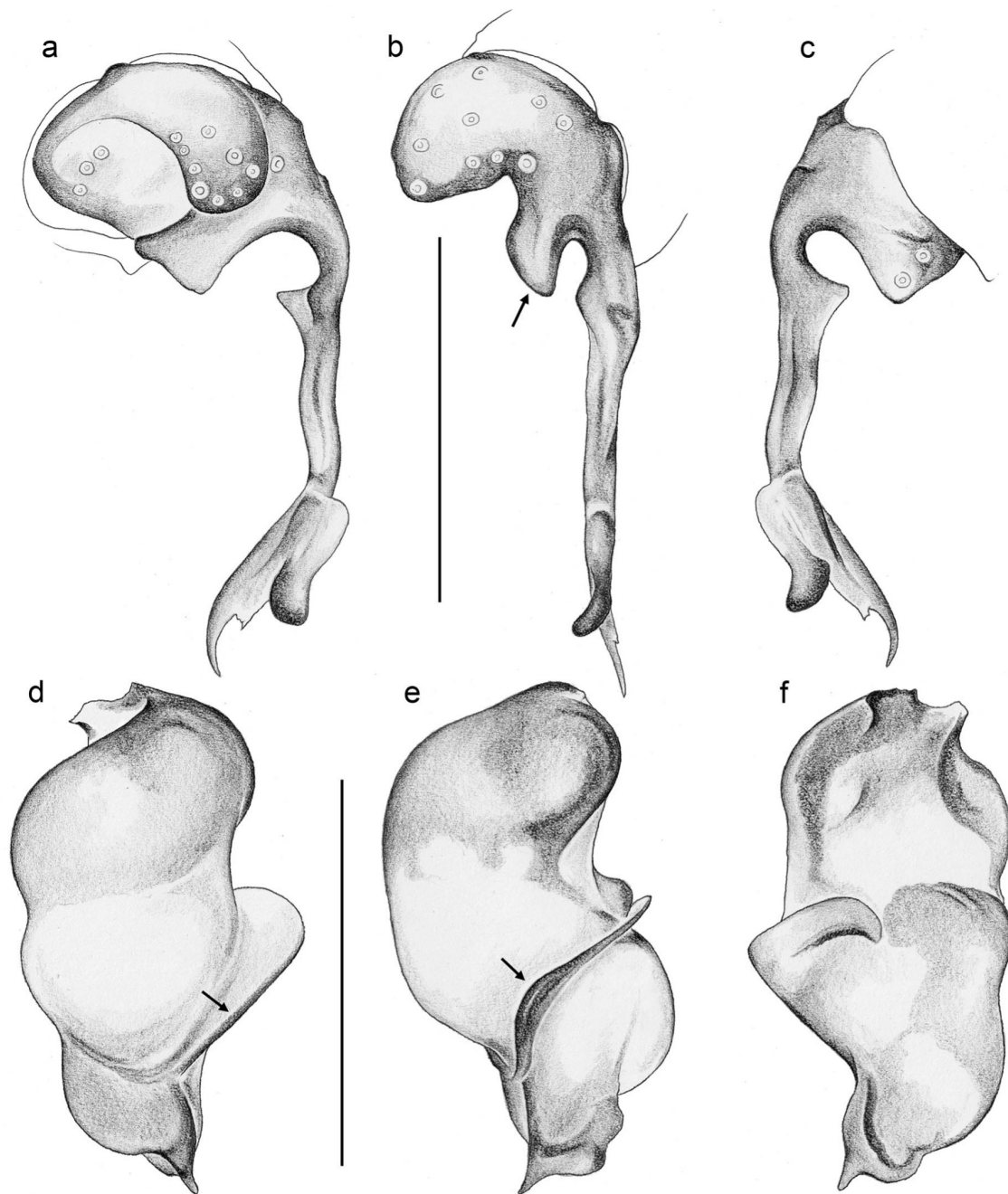


Figure 5.6. *GENUS-1 sp1* Huber sp. n.; male from Gruta do Janelão; CHNUFPI 3505. **a–c** Left procursus, prolateral, dorsal, and retrolateral views. **d–f** Left genital bulb, prolateral, dorsal, and retrolateral views. Scale bars: 0.2 mm.

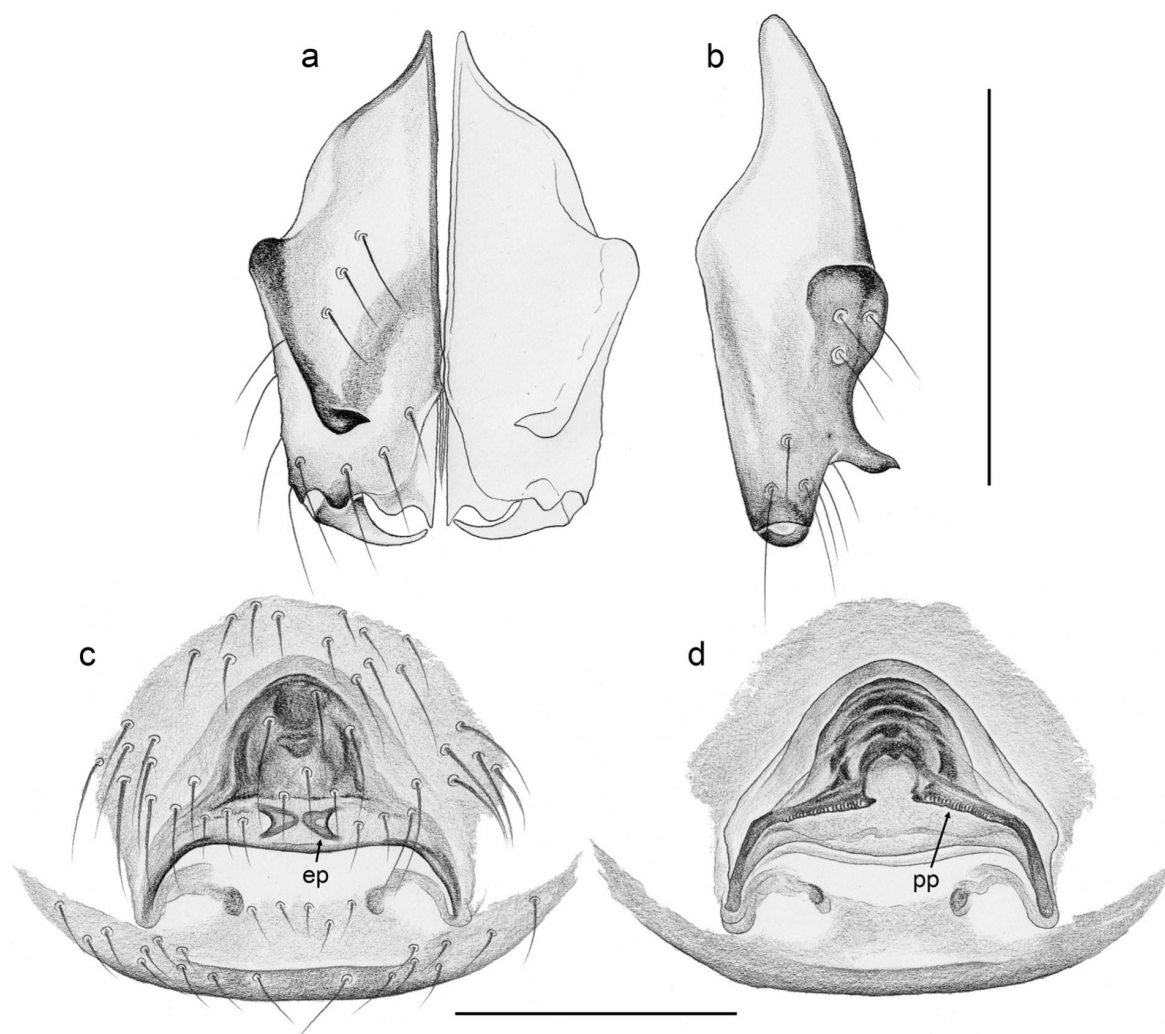


Figure 5.7. *GENUS-1 sp1* Huber sp. n.; male and female from Gruta do Janelão; CHNUFPI 3505 (male), ZFMK Ar 24704 (female). **a–b** Male chelicerae, frontal and lateral views. **c–d** Cleared female genitalia in ventral and dorsal views. Abbreviations: ep, epigynal pocket; pp, pore plate. Scale bars: 0.2 mm.

median structure from which pair of sclerotized lateral sclerites originate; pore plates poorly visible, apparently at median origin of lateral sclerites (Fig. 5.7d); with large anterior membranous arc.

Natural history. Both sites are karstic caves, located in the Peruaçu district of the Caatinga domain (sensu Moro et al. 2024). At the type locality, Gruta do Janelão (Fig. 5.19a), all specimens were found in the aphotic zone, ~100–300 m from the cave entrance. Measurements of temperature and humidity at approximately the same section of the cave: 23.7°C, 90.5% (A.J. Santos, 1 Mar. 2020). Searches carried out in Feb. 2020 and in Nov. 2022 explored about 700 of the 4,800 m extension of the cave, but no specimens were found in the twilight section and deeper in the cave. At the second locality, Lapa do Rezar, the specimens were found at the cave entrance. At both localities, the spiders were found under stones or among small limestone pebbles on the floor. Two egg sacs contained six and seven eggs, respectively, with an egg diameter of 0.40.

Distribution. Known from two caves (~2km from each other) in Parque Nacional Cavernas do Peruaçu, Minas Gerais, Brazil (Fig. 5.4).

GENUS-1 sp2 Huber sp. nov.

<https://zoobank.org/59FE9223-CA80-4281-9B73-A5F43E67E7DE>

Figs. 5.3c–d, 5.15–5.18

Diagnosis. Easily distinguished from all other known Pholcidae by shape of procurus (Fig. 5.16a–c; slender, distally with heavily sclerotized element with pointed tip); also by strong apophysis prolateral-ventrally on male palpal femur (Fig. 5.15a), by shape of genital bulb (Fig. 5.16d–f; prolateral-dorsal protruding flap; dorsal ridges), by armature of male chelicerae (Fig. 5.17a–c; pair of lateral sclerotized humps and pair of distal lateral pointed apophyses directed towards median), by pair of rounded processes on male clypeus (Fig. 5.17a), and by female epigynum and internal genitalia (Figs. 5.17d–e, 5.18; main anterior plate with pair of pockets and additional pair of frontal lateral sclerotized areas; internal genitalia with arched element and pair of small elongated pore plates).

Holotype. BRAZIL - Bahia • ♂; W of Queimada Nova; 11.0343°S, 42.0682°W; 580 m a.s.l.; 25 Nov. 2022; B.A. Huber & A.S. Michelotto leg.; CHNUFPI 9076 [sep Br22-107].

Paratypes. BRAZIL - Bahia • 1 ♂, 1 ♀; same collection data as for holotype; CHNUFPI 9077 [Br22-107, part1] • 1 ♂ (together with one female abdomen, dissected and epigynum cleared, transferred from ZFMK Br22-229; voucher for G58); same collection data as for holotype; CHNUFPI 9078 [deposited in ZFMK Ar 24706] [Br22-107, part2].

Other material examined. BRAZIL - Bahia • 1 ♂, 3 ♀♀ (one female prosoma used for molecular work; abdomen transferred to ZFMK Ar 24706), in pure ethanol; same collection data as holotype; CHNUFPI 9079 [deposited in ZFMK Br22-229; voucher of UH501].

Etymology. The species name is an adjective, describing someone originating from Bahia.

Description. Male (holotype). –MEASUREMENTS. Total length 1.30, carapace width 0.58. Distance PME–PME 40 µm; diameter PME 60 µm; distance PME–ALE 15 µm; AME absent (no lenses, only black mark). Leg 1: 3.22 (0.88 + 0.20 + 0.88 + 0.88 + 0.38), tibia 2: 0.84, tibia 3: 0.64, tibia 4: 0.84; tibia 1 L/d: 13. –COLOR (in ethanol). Prosoma and legs mostly whitish to pale ochre-yellow, carapace with pair of ochre marks beside ocular area, clypeus also slightly darkened; sternum monochromous whitish; legs with darker rings on femora (subdistally) and tibiae (proximally and subdistally); abdomen ochre gray, with distinct white marks except on ventral side. –BODY. Habitus as in Fig. 5.3c. Ocular area barely raised. Thoracic groove shallow. Clypeus with pair of short rounded processes near margin and many strong hairs directed upwards (Fig. 5.17a). Sternum wider than long (0.38/0.34), unmodified. Abdomen

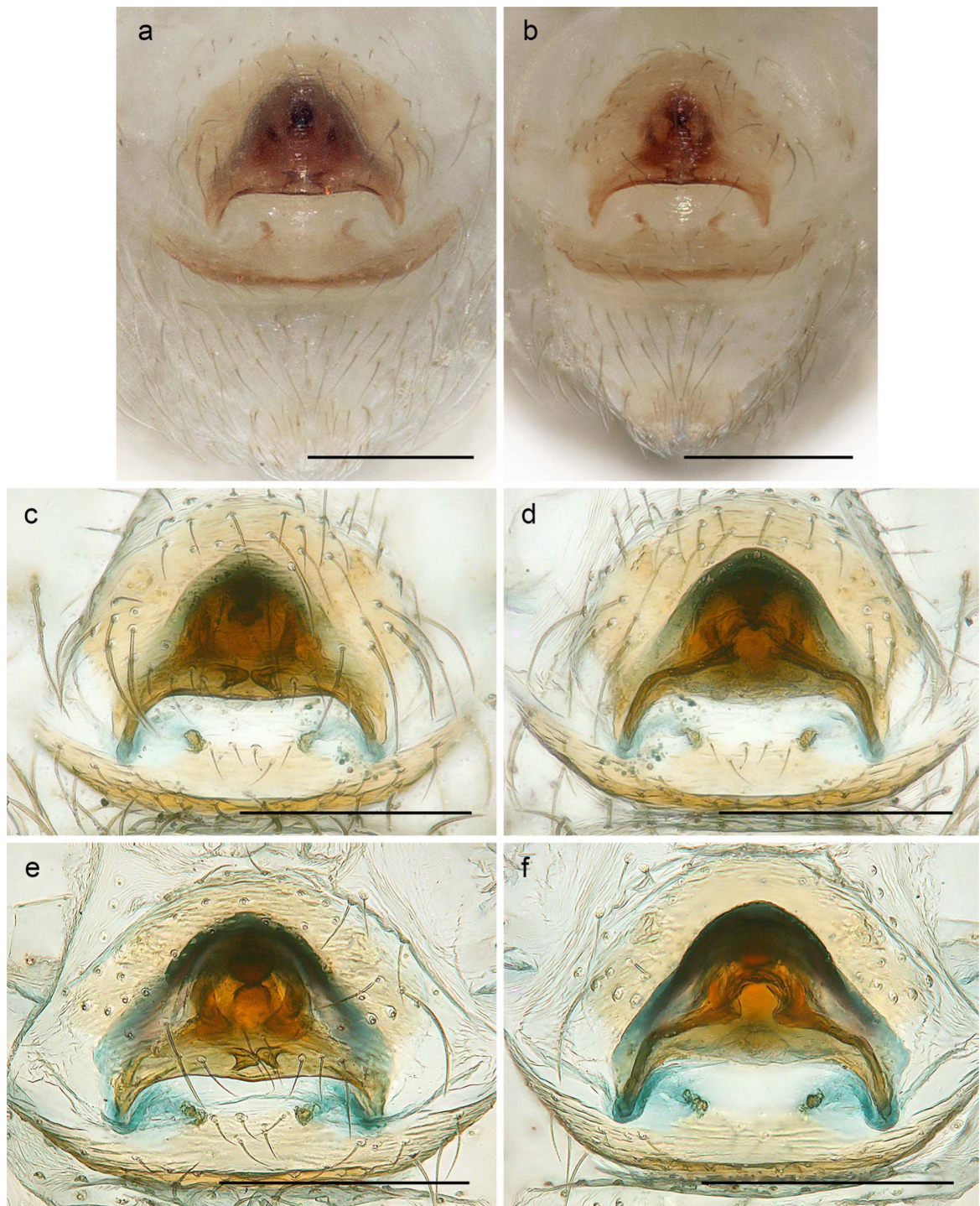


Figure 5.8. *GENUS-1 sp1* Huber sp. n.; two females from Gruta do Janelão, ZFMK Ar 24704. **a–b** Epigyna, ventral views. **c–d** Cleared female genitalia of female shown in A, ventral and dorsal views. **e–f** Cleared female genitalia of female shown in B, ventral and dorsal views. Scale bars: 0.2 mm.

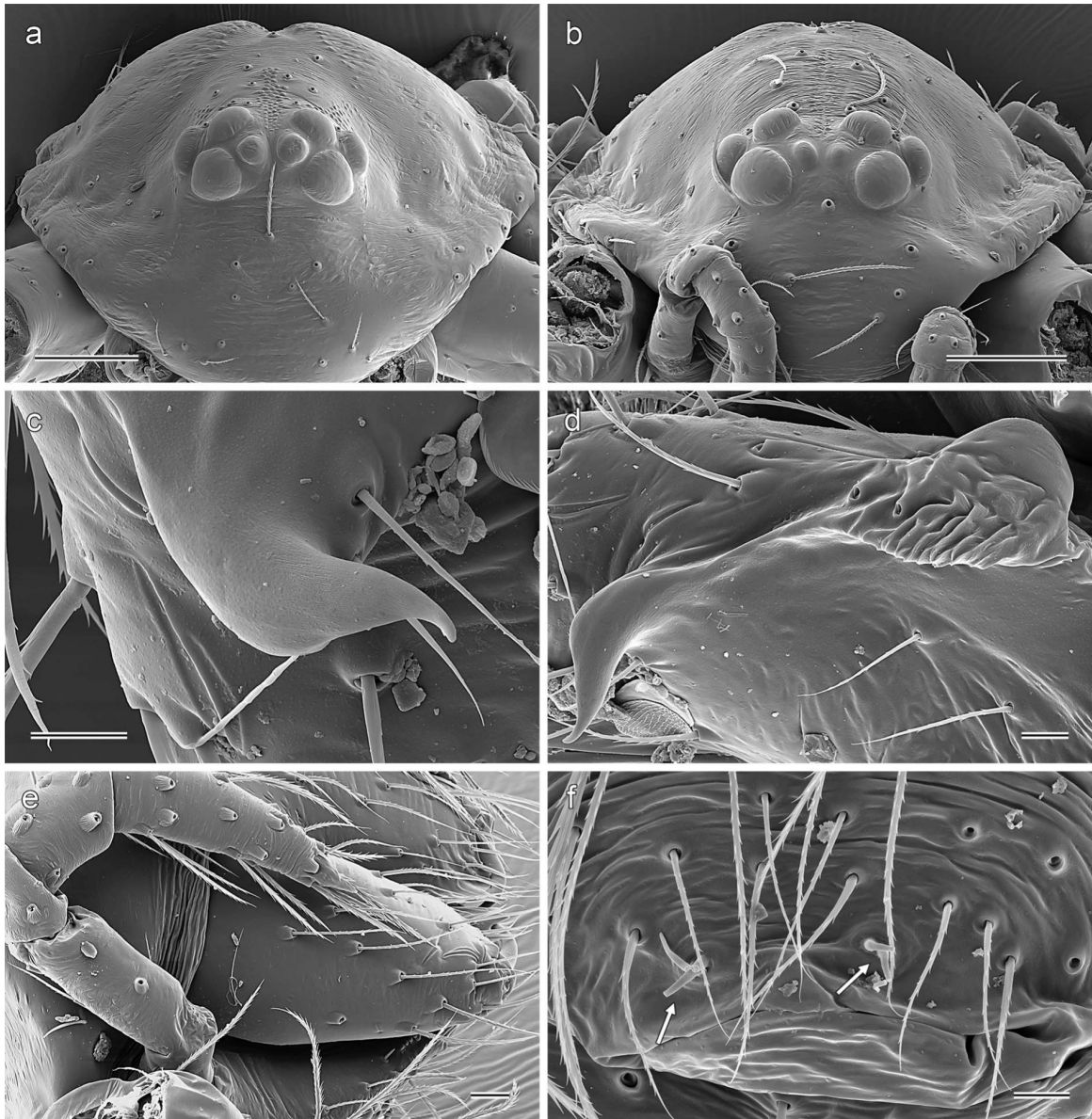


Figure 5.9. *GENUS-1 sp1* Huber sp. n.; male and female from Gruta do Janelão, ZFMK Br22-166. **a** Male prosoma, frontal view. **b** Female prosoma, frontal view. **c** Right male frontal cheliceral apophysis, frontal view. **d** Right male chelicera, frontal-lateral view. **e** Right female palp and chelicera, frontal-lateral view. **f** Male gonopore and epiandrous spigots (arrows), ventral view. Scale bars: 100 µm (a–b), 10 µm (c–d, f), 20 µm (e).

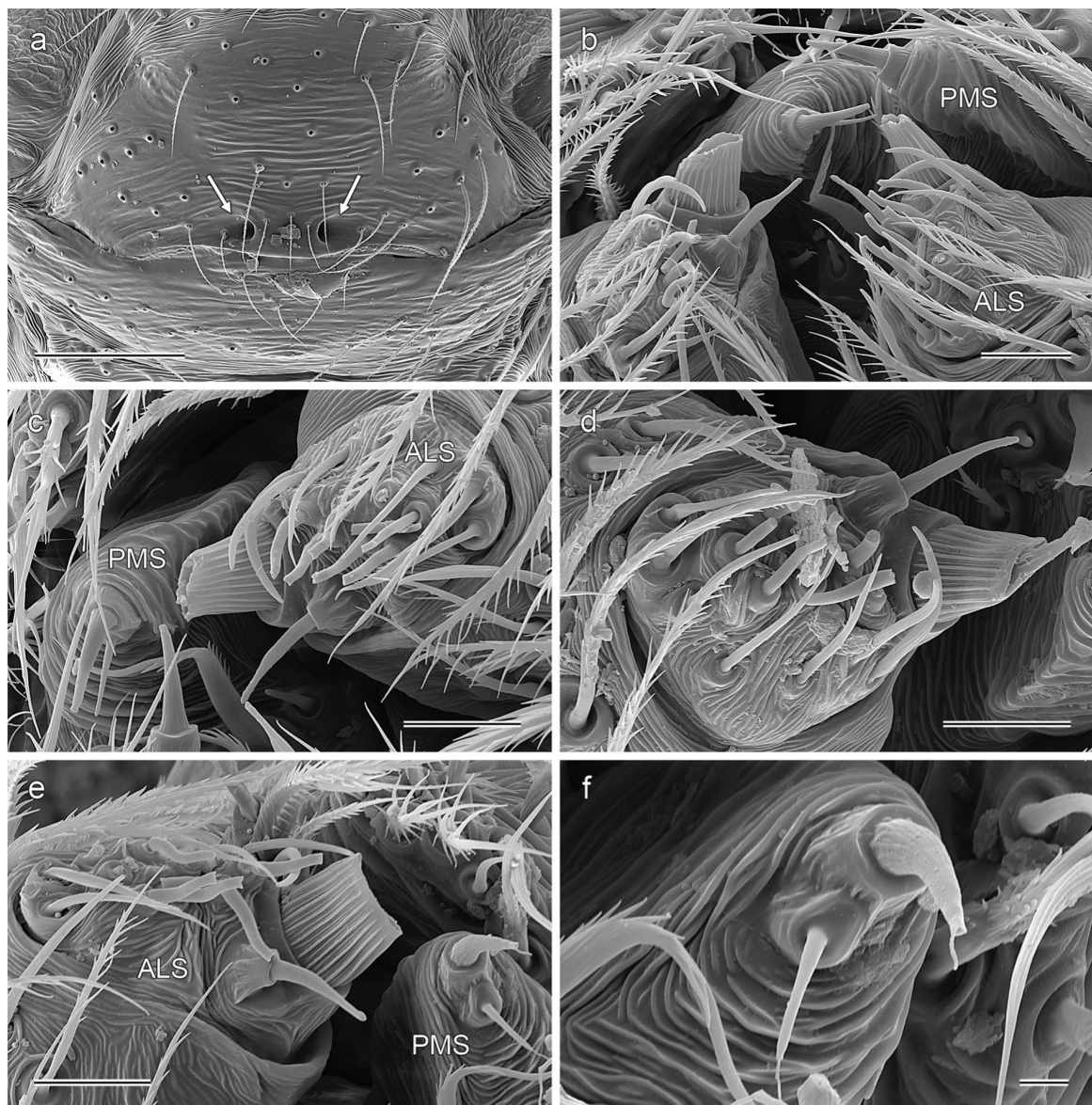


Figure 5.10. *GENUS-1 sp1* Huber sp. n.; male and female from Gruta do Janelão, ZFMK Br22-166. **a** Epigynum, ventral view, showing pockets (arrows). **b** Male anterior lateral spinnerets (ALS) and posterior median spinnerets (PMS). **c–e** Female ALS and PMS. **f** Female PMS. Scale bars: 100 μ m (a), 10 μ m (b–e), 2 μ m (f).

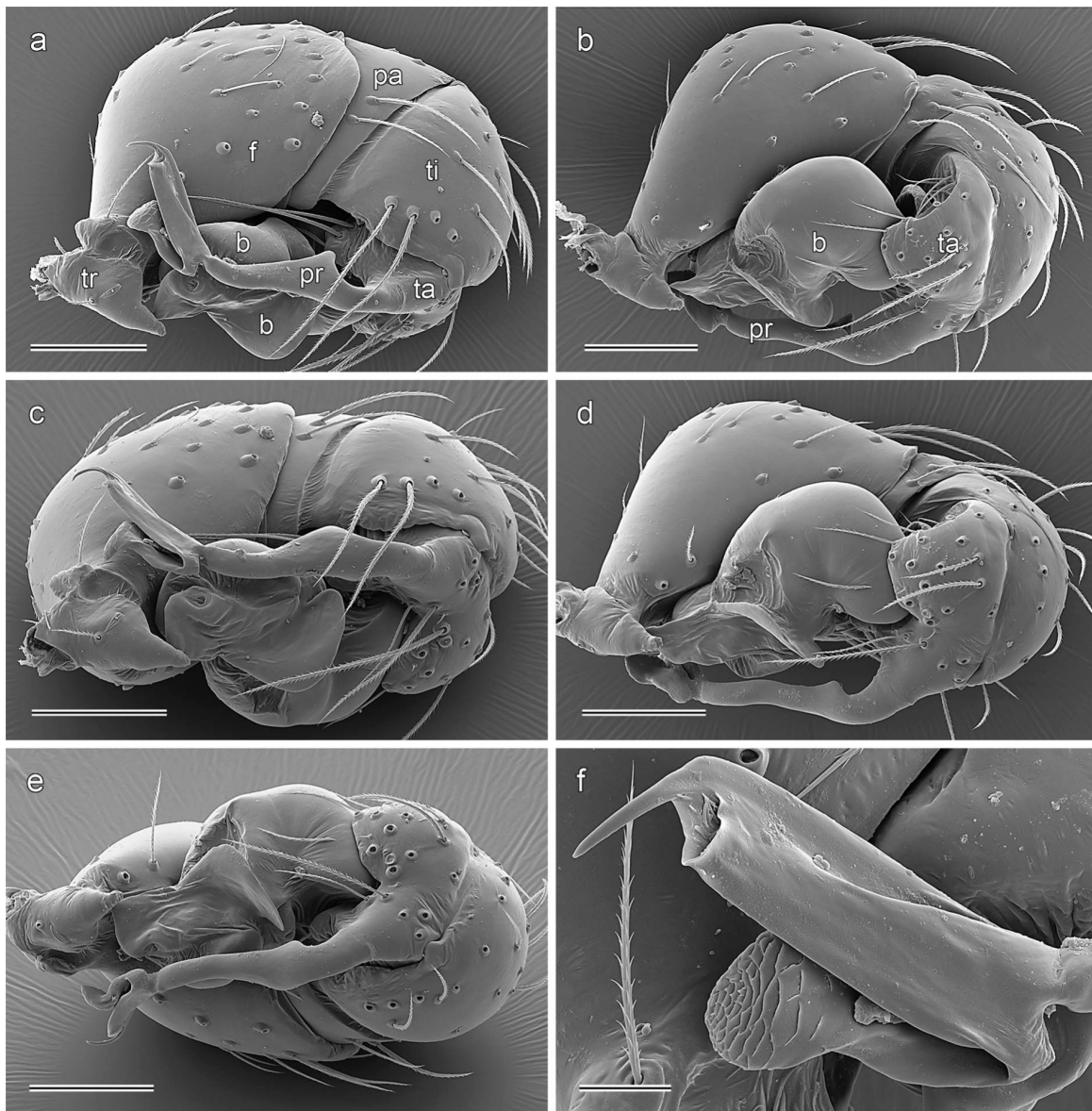


Figure 5.11. *GENUS-1 sp1* Huber sp. n.; male from Gruta do Janelão, ZFMK Br22-166. **a** Right palp, retrolateral view. **b** Left palp prolateral view. **c** Right palp, retrolateral-dorsal view. **d** Left palp, prolateral-dorsal view. **e** Left palp, dorsal view. **f** Tip of right procursus, retrolateral view. Abbreviations: b, genital bulb; f, femur; pa, patella; pr, procursus; ta, tarsus; ti, tibia; tr, trochanter. Scale bars: 100 μ m (a–e), 20 μ m (f).

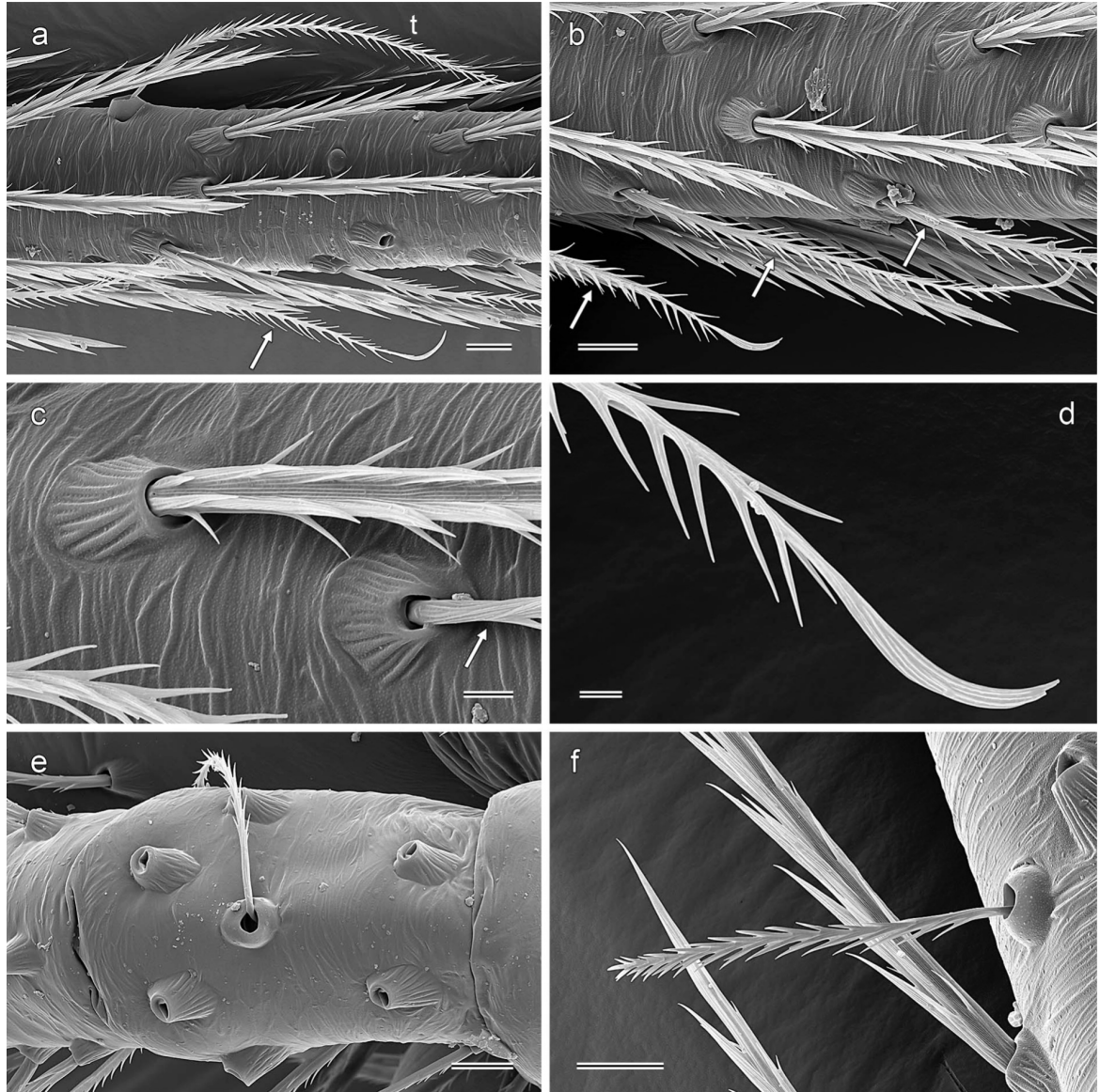


Figure 5.12. *GENUS-1 sp1* Huber sp. n.; male and female from Gruta do Janelão, ZFMK Br22-166. **a** Right female metatarsus 4, showing similarity of trichobothrium (t) and slender metatarsal hair (arrow). **b** Left male metatarsus 3, prolateral view, showing three thin metatarsal hairs (arrows). **c** Right female metatarsus 4, bases of regular mechanoreceptor and of slender metatarsal hair (arrow). **d** Tip of slender metatarsal hair on left male metatarsus 3. **e** Female palpal tibia, showing single trichobothrium. **f** Prolateral trichobothrium on left male tibia 3. Scale bars: 10 µm (a–b, e–f), 3 µm (c), 2 µm (d).

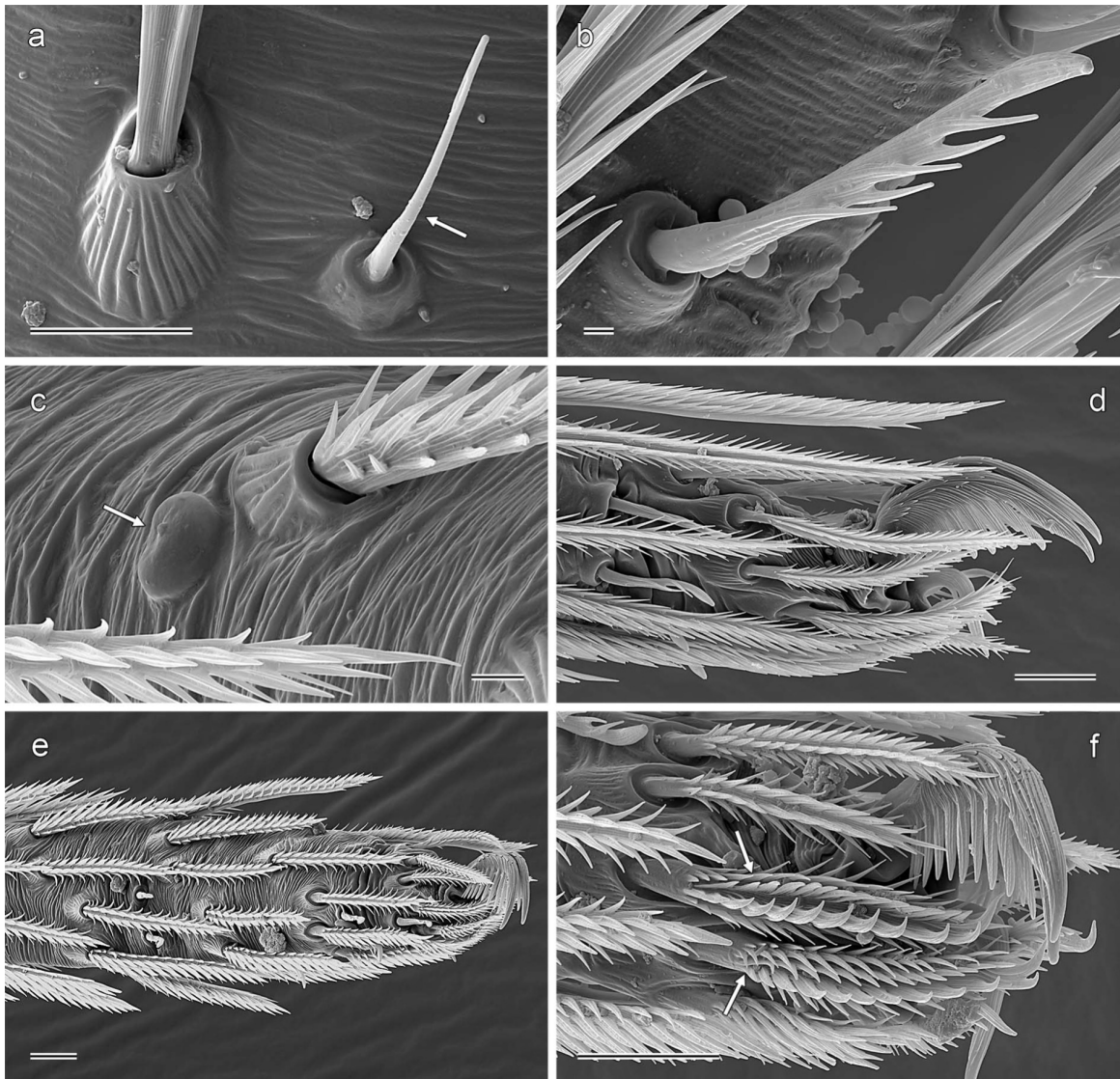


Figure 5.13. *GENUS-1 sp1* Huber sp. n.; male from Gruta do Janelão, ZFMK Br22-166. **a** Sexually dimorphic short 'vertical' hair (arrow) on left tibia 1. **b** Putative chemoreceptor on left metatarsus 1. **c** Round cuticular plate (arrow) on right tarsus 4. **d** Left tarsus 1, prolateral view. **e** Left tarsus 3, prolateral-distal view. **f** Left tarsus 4, prolateral view, showing two comb hairs (arrows). Scale bars: 10 μm (a, d–f), 1 μm (b), 2 μm (c).

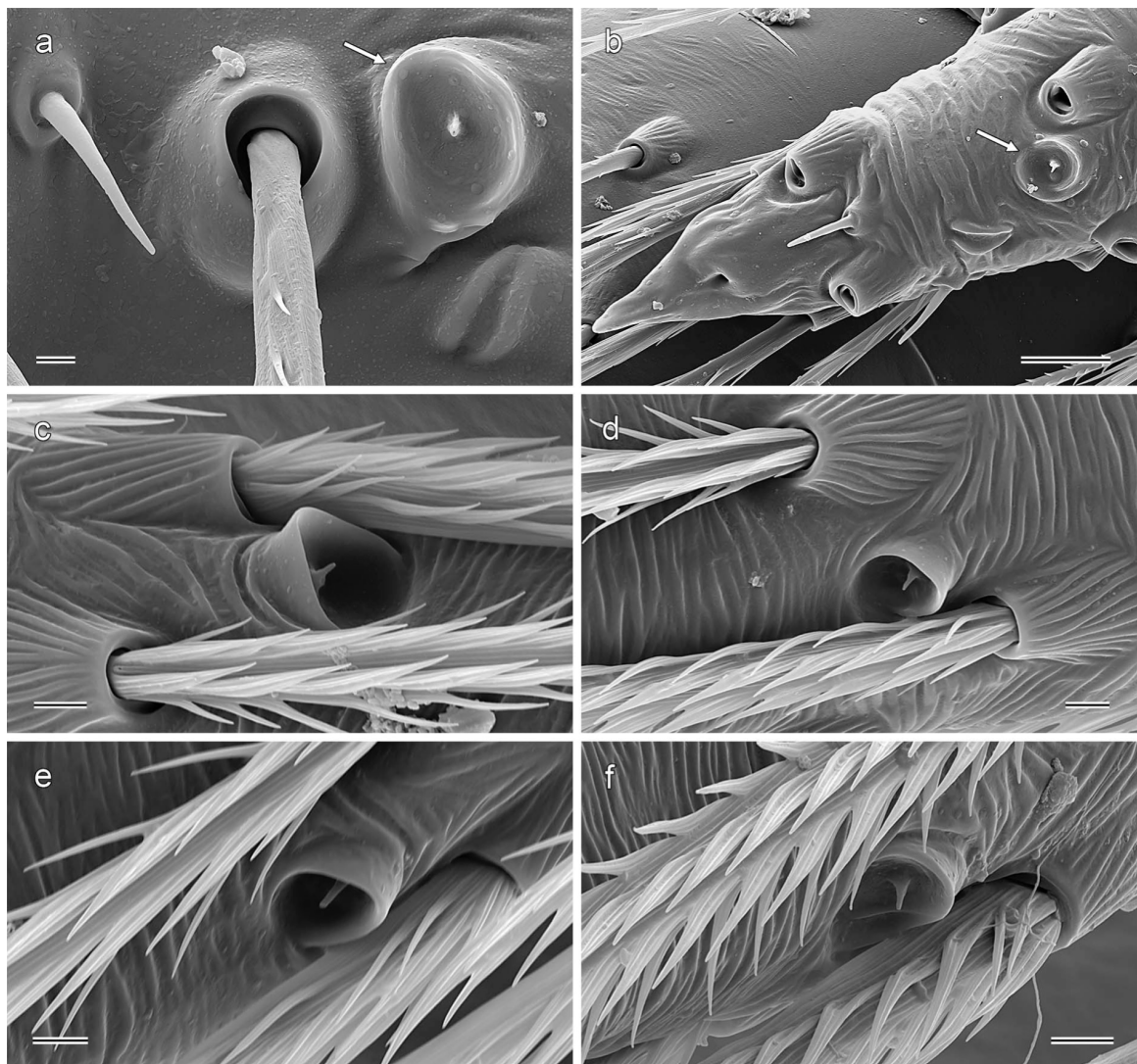


Figure 5.14. *GENUS-1 sp1* Huber sp. n.; male and female from Gruta do Janelão, ZFMK Br22-166. **a** Tarsal organ (arrow) and neighboring sensilla on right male palp. **b** Tip of right female palp, dorsal view, showing tarsal organ (arrow). **c–f** Tarsal organs on right female tarsus 1 (c), right female tarsus 2 (e), right female tarsus 3 (e), and right male tarsus 4 (f). Scale bars: 2 µm (a, c–f), 10 µm (b).

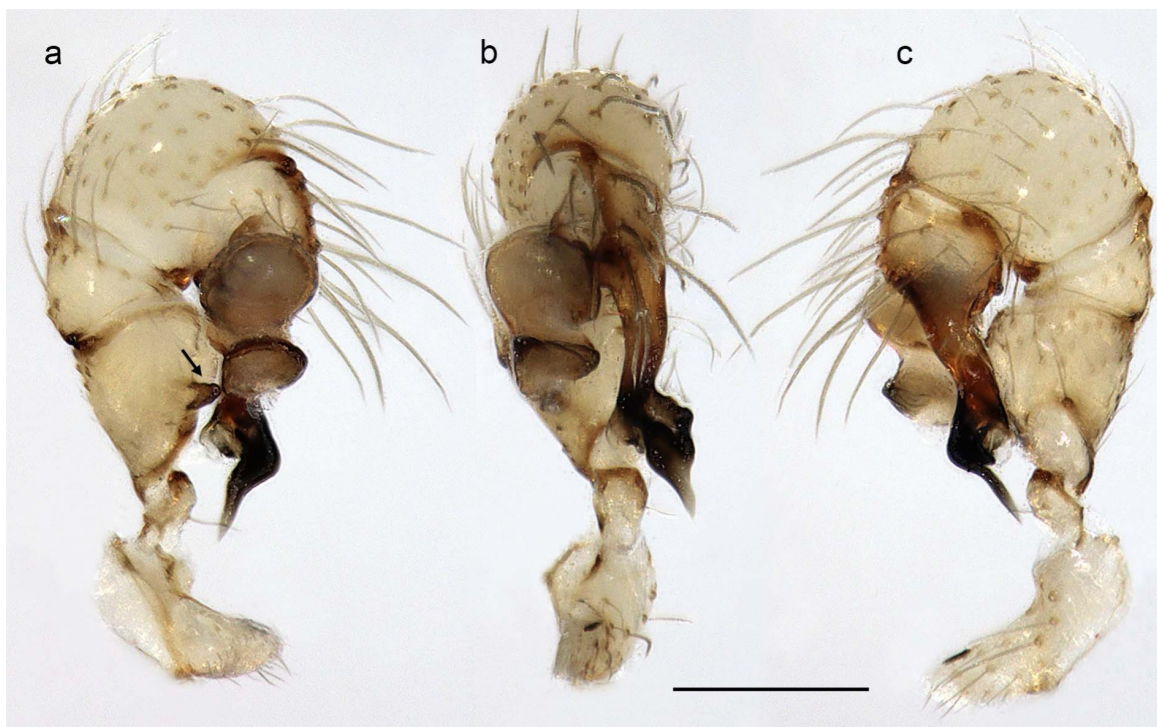


Figure 5.15. *GENUS-1 sp2* Huber sp. n.; male from W of Queimada Nova; ZFMK Ar 24706. Left genital palp, prolateral (a), dorsal (b), and retrolateral (c) views. Scale bar: 0.2 mm.

globular. –CHELICERAE. As in Fig. 5.17a–c, with pair of distal lateral pointed apophyses directed towards median, and pair of proximal lateral sclerotized humps; without stridulatory ridges. –PALPS. As in Fig. 5.15; coxa with low ventral hump; trochanter unmodified; femur relatively small (smaller than tibia), with strong prolateral-ventral sclerotized apophysis (arrow in Fig. 5.15a), distally strongly widened; femur-patella joints not shifted to one side; tibia with two trichobothria; tibia-tarsus joints strongly shifted toward retrolateral side; tarsus with prolateral-ventral process at basis of procursus; procursus (Fig. 5.16a–c) proximally simple, distally with heavily sclerotized element with pointed tip; genital bulb (Fig. 5.16d–f) slightly elongated and medially constricted, with distinctive prolateral-dorsal protruding flap and dorsal ridges, sperm duct opening not seen (presumably on distal membranous part). –LEGS. Without spines, without curved hairs. With sexually dimorphic short ‘vertical’ hairs on tibia 1 only (confirmed in compound microscope; not visible in dissecting microscope). Ventrally on metatarsi 3 and 4 with slender hairs (confirmed in compound microscope; not visible in dissecting microscope). Retrolateral trichobothrium of tibia 1 at 61%; prolateral trichobothrium absent on tibia 1. No tarsal pseudosegmentation visible (also in compound microscope). Main tarsal claws and median tarsal claw as typical for family.

Variation (male). Tibia 1 in three other males: 0.82, 0.82, 0.84.

Female. In general, very similar to male (size, body shape, color pattern; Fig. 5.3d), but clypeus unmodified and legs without sexually dimorphic short ‘vertical’ hairs. One female with tiny AME lenses (diameter ~ 5 µm). Palpal tibia with two trichobothria. Tibia 1 in three females: 0.70, 0.72, 0.82. Epigynum (Fig. 5.18a–b) anterior plate wide and short, with pair of pockets and additional pair of frontal lateral sclerotized areas; posterior epigynal plate very short, indistinct. Internal genitalia (Figs. 5.17e,

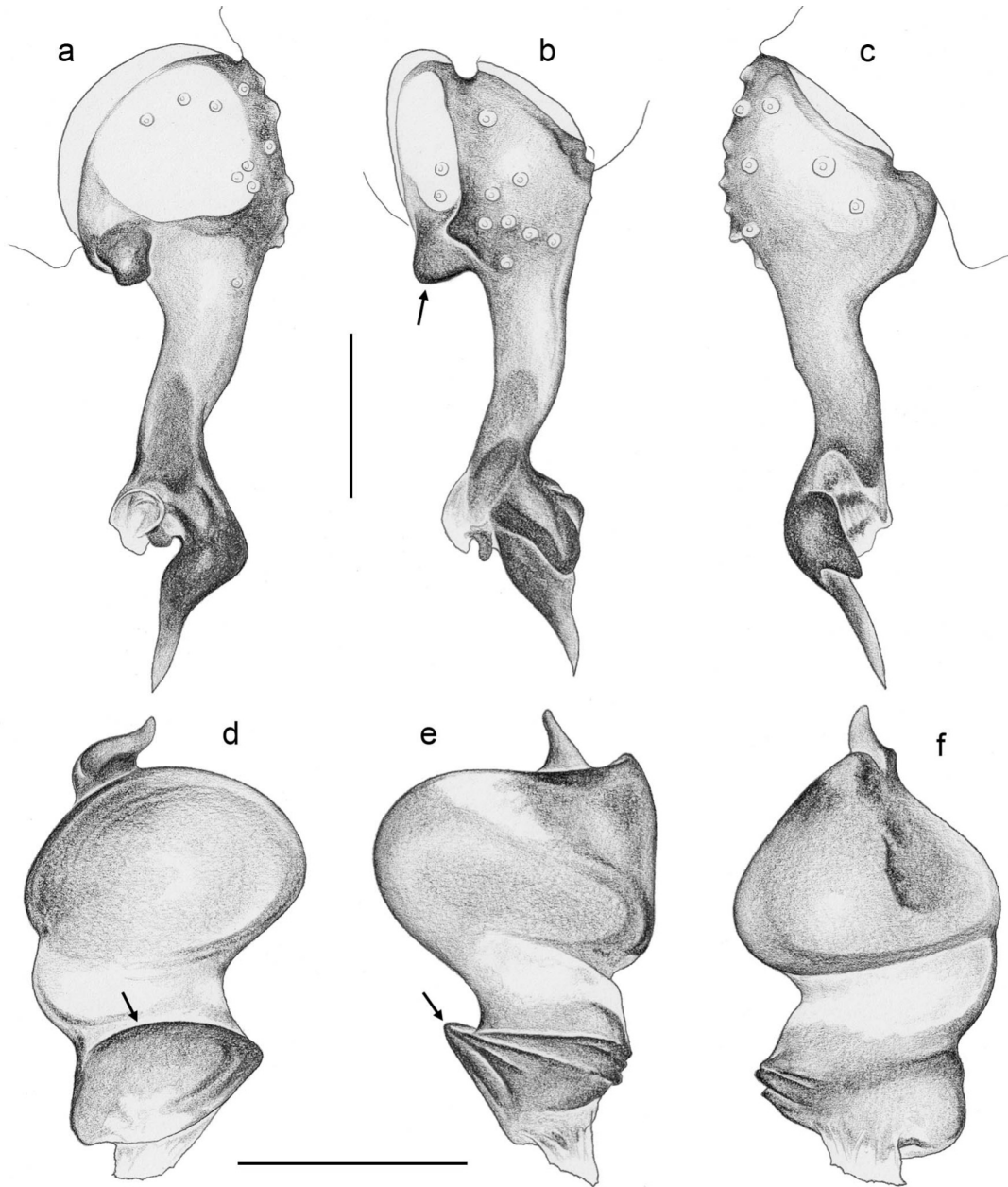


Figure 5.16. *GENUS-1 sp2* Huber sp. n.; male from W of Queimada Nova; ZFMK Ar 24706. **a–c** Left tarsus and procursus, prolateral, dorsal, and retrolateral views; arrow: process at basis of procursus. **d–f** Left genital bulb, prolateral, dorsal, and retrolateral views; arrows: prolateral-dorsal ridge. Scale bars: 0.1 mm.

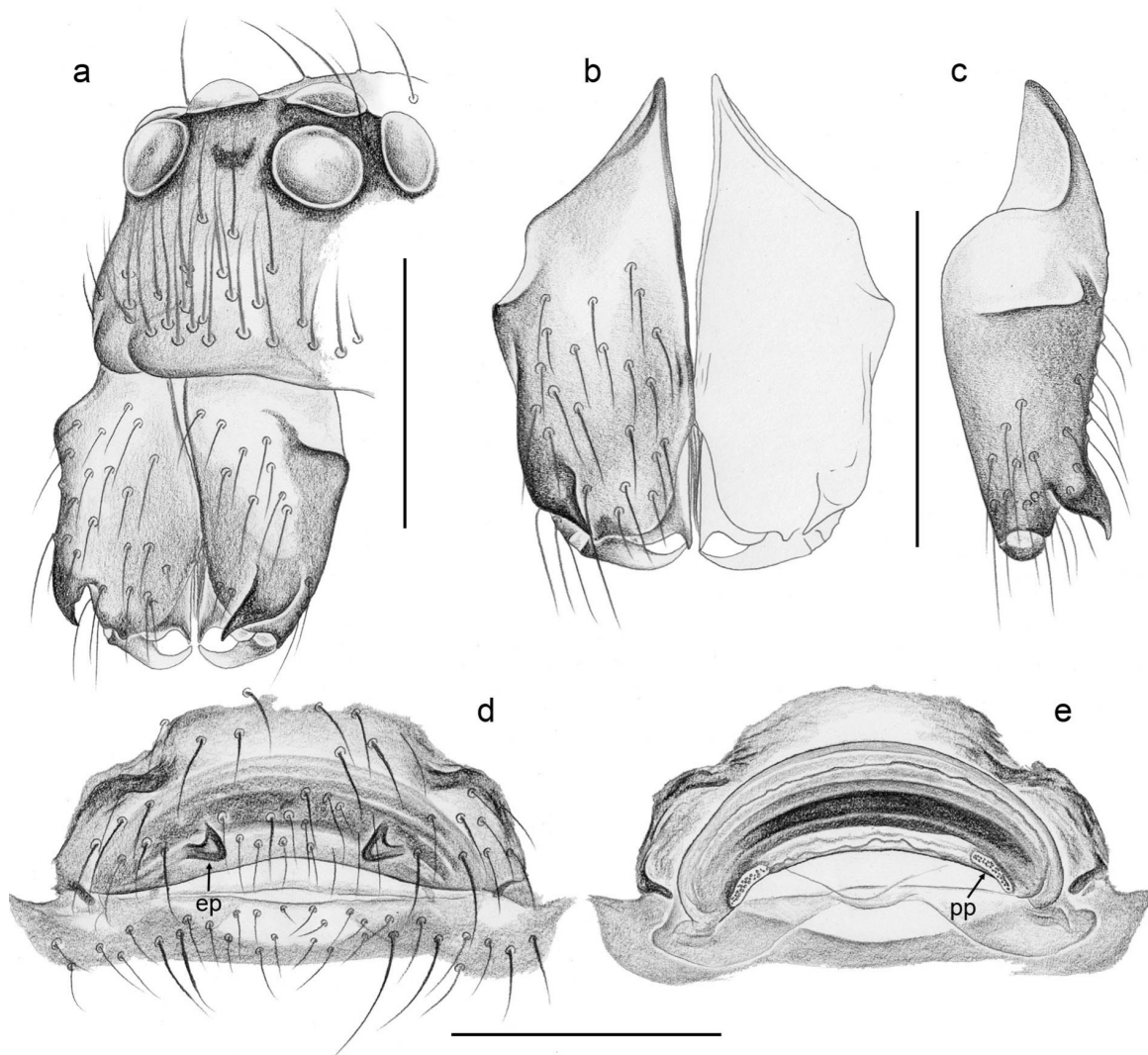


Figure 5.17. *GENUS-1 sp2* Huber sp. n.; male and female from W of Queimada Nova; ZFMK Ar 24706. **a** Male ocular area, clypeus, and chelicerae, oblique frontal view. **b–c** Male chelicerae, frontal and lateral views. **d–e** Cleared female genitalia, ventral and dorsal views. Abbreviations: ep, epigynal pocket; pp, pore plate. Scale bars: 0.2 mm.

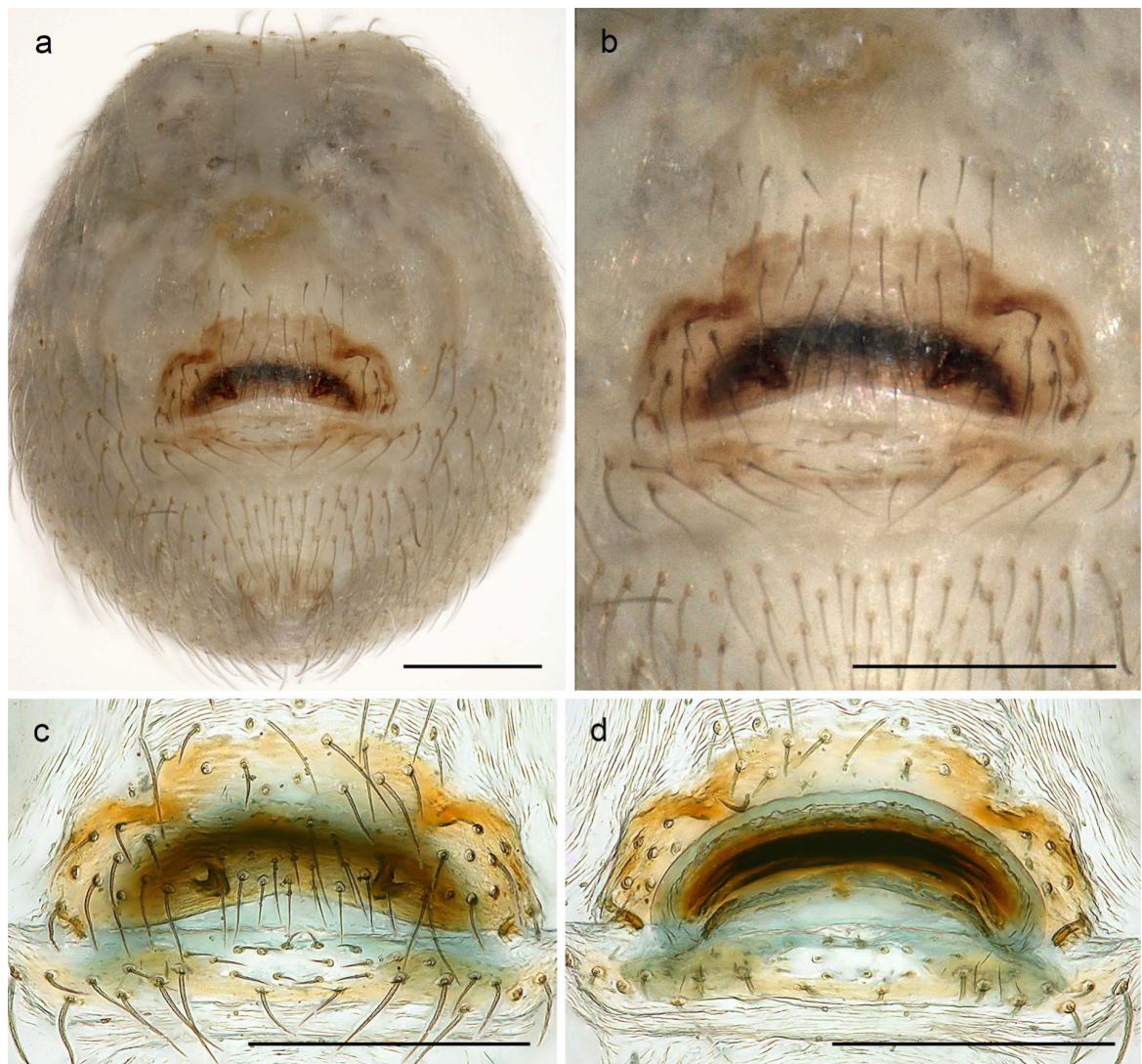


Figure 5.18. *GENUS-1 sp2* Huber sp. n.; female from W of Queimada Nova; ZFMK Ar 24706. **a–b** Abdomen ventral view, and epigynum at higher magnification. **c–d** Cleared female genitalia, ventral and dorsal views. Scale bars: 0.2 mm.

5.18c-d) with strong arched element (clearly visible in uncleared specimens) and pair of small, elongated pore plates.

Natural history. The type-locality is in the Irecê district of the Caatinga domain (sensu Moro et al. 2024). All specimens were found in an old collapsed low stone fence in thorny woodland (Fig. 5.19b). They shared this microhabitat with three further species of Pholcidae: *Ibotyporanga diroa* Huber & Brescovit, 2003 (Brescovit et al. 2003); an undescribed species of *Mesabolivar* González-Sponga, 1998; and an undescribed species tentatively assigned to *Tupigea* Huber, 2000. One egg sac contained eleven eggs; egg diameter 0.38–0.39.

Distribution. Known from type locality only, in Bahia, Brazil (Fig. 5.4).

Subfamily Phy***inae subfam. n.

Remarks. Recent multi-locus genetic data (Eberle et al. 2018) suggested that Arteminae as previously defined (Huber 2011) is polyphyletic: *Artema* and “other Arteminae” are not sister taxa. This result is strongly supported by new genomic data (Meng, Podsiadlowski, Dimitrov et al. 2025; and herein), emphasizing the need for a new name for “other Arteminae”. Since no subfamily-level name for this group is available in the literature, we chose *Physocyclus* Simon, 1893 as the type genus. This subfamily includes all Arteminae as listed in the recent literature (Huber 2011; Huber, Eberle et al. 2018; Huber, Meng and Valdez-Mondragn 2024) except for *Artema*. We also include here *Aucana* Huber, 2000 because our new genomic data strongly supports this assignment.

Composition. The subfamily as circumscribed here currently includes 107 nominal species in the following 11 genera (species numbers in parentheses): *Arnapa* Huber, 2019 (6); *Aucana* Huber, 2000 (5); *Changminia* Yao & Li, 2022 (2); *Chisosa* Huber, 2000 (4); *Holocneminus* Berland, 1942 (3); *Nita* Huber & El Hennawy, 2007 (1); *Pholcitrichocyclus* Ceccolini & Cianferoni, 2022 (= *Trichocyclus* Simon, 1908) (23); *Physocyclus* Simon 1893 (39); *Serratochorus* Wunderlich, 1988 (1); *Tibetia* Zhang, Zhu & Song, 2006 (1); *Wugigarra* Huber, 2001 (22).

5.5 Discussion

5.5.1 How reliable is the molecular phylogeny?

In our likelihood mapping analysis, approximately 32% of quartets are positioned at each of the three corners of the triangular graph, indicating a high level of phylogenetic signal in the dataset and the potential to resolve a robust phylogeny (Strimmer et al. 1997) (Figure C.3). Only 7.2% of the loci violate the assumptions of stationarity, homogeneity, or both (Table C.3), and less than 0.6% of the contigs are affected by intra-locus recombination (Figure C.4). This suggests that the dataset is well-suited for phylogenetic inference. Furthermore, the coalescence-based methods (ASTRAL-IV and wASTRAL) enable us to mitigate the impact of incomplete lineage sorting (ILS) (Zhang and Mirarab 2022). The models GTR+R4+I+FO used in RaxML-NG and CAT-GTR in PhyloBayes-MPI are designed for handling rate and compositional heterogeneity (Kozlov et al. 2019; Lartillot et al. 2013), which alleviate potential influences of model violation.

Our phylogenetic inferences from 30 different analyses (five matrices x six tree construction methods) are largely congruent in terms of subfamily monophylies and inter-subfamily relationships (375 green squares / (30 x 13 panels) = 96%; see Fig. 5.2a). The monophylies of Pholcinae, Smeringopinae, Modisiminae, Ninetinae, Phy***inae, and Arteminae appear beyond doubt. The monophyly of the new subfamily

Cai***inae is supported by all analytical methods in our study (Fig. 5.2a) with 100% bootstrap support (Fig. 5.1). Non-monophyly of subfamilies was found in only a few analyses: (1) Arteminae, when analyzed with PhyloBayes-MPI using data matrices 3–5 (3 out of 30 analyses), which may be due to an insufficient amount of data (total sites ranging from 9 to 20 Kbp), and (2) Modisiminae, when analyzed with SVDquartets using matrices 1–5. The latter case involves the uncertain placement of three genera (*Arenita* and two unnamed genera). In most analyses (25 out of 30 analyses), they were sister to all other Modisiminae (Fig. 5.1), while in a few cases (5 out of 30 analyses) they were sister to Phy***inae (not shown), suggesting that further analysis of this poorly known group is necessary. The inconsistency between SVDquartets and other programs could stem from the fact that finding the species tree with maximum quartet support from a set of quartet trees is an NP-hard problem (Jiang et al. 2001). As SVDquartets uses a heuristic search strategy, it does not guarantee to find the optimal solution (Vachaspati et al. 2018).

Since a large percentage of our dataset is identical to that analyzed in Meng, Podsiadlowski, Dimitrov et al. (2025), it is unsurprising that the subfamily relationships recovered herein align well with those reported in Meng, Podsiadlowski, Dimitrov et al. (2025). Only a small fraction of the analyses (7 out of 90) did not support the monophyly of certain higher-level groups (Fig. 5.2a): (1) the clade comprising Modisiminae + Phy***inae and Ninetinae + Arteminae; (2) its sister-group relationship with the newly proposed subfamily Cai***inae; and (3) the entire family Pholcidae. Interestingly, all of these discrepancies arose from analyses based on the smaller datasets of matrices 4 and 5 (total sites: 15 Kbp and 9 Kbp, respectively). This suggests that the incongruences are likely due to the limited data in these specific matrices.

5.5.2 Morphological aspects

The sister-group relationship between the two newly described species is strongly supported by molecular data but not immediately obvious in their morphology. The combined similarities listed in the diagnosis of the genus appear suitable for the particular purpose of differentiation from other genera, but rather weak in the context of phylogeny. Individually, most of these characters or traits do not seem to be unique among Pholcidae, i.e. they are unlikely candidates to be non-homoplastic synapomorphies. A possible exception is the transversal ridge on the male genital bulb (Figs. 5.6d-e, 5.16d-e). For all other items listed in the diagnosis, similar characters and traits occur in other, often distantly related, genera. (1) Pointed male cheliceral apophyses and a corresponding pair of female epigynal pockets also occur in other genera like *Belisana* Thorell, 1898; *Crossopriza* Simon, 1893; *Quamtana* Huber, 2003; and *Smeringopus* Simon, 1890 (Huber 2003, 2012, 2022; Huber and Schaller 2005) but are absent in the sister group of *GENUS-1*, i.e. in Phy***inae, Modisiminae, Arteminae, and Ninetinae. (2) A ventral or prolateral process on the male palpal tarsus, at the basis of procursus, exists in many Pholcidae, even though usually less distinct, as for example in *Crossopriza*, *Smeringopina* Kraus, 1957 and *Wanniyala* Huber & Benjamin, 2005 (Huber 2013, 2022; Huber and Carvalho 2019). (3) Dark marks on the carapace are ubiquitous in Pholcidae and only useful as a quick distinguishing trait in comparison to superficially similar Brazilian Ninetinae. (4) A low ratio of male tibia 1 / tibia 2 (~1.05–1.15) is very rare among Pholcidae (Fig. 5.20), but similar values also occur in Ninetinae. (5) The absence of leg tarsal pseudosegmentation is difficult to evaluate since the pseudosegmentation of some Pholcidae is very indistinct, or the individual pseudosegments are “broken” into many seemingly irregular small platelets (in some *Smeringopinae* and some Phy***inae; Huber 2000, 2022).

On the other hand, the two species show some remarkable differences that are unusual among closely related and congeneric Pholcidae. The male palpal femur is particularly different: strongly enlarged in



Figure 5.19. Type localities of newly described species. **a** Gruta do Janelão, entrance area; *GENUS-1 sp1* Huber sp. n. was found ~100–300 m from the cave entrance, in the aphotic zone. **b** Thorny woodland W of Queimada Nova; *GENUS-1 sp2* Huber sp. n. was found in the collapsed rock wall visible in the left central part of the photo. Photos BAH.

GENUS-1 sp1 (very similar to most *Phy***inae*, also *Artema*, some *Priscula*, some *Mesabolivar*; e.g. Aharon et al. (2017), Huber (2018) and Huber and Villarreal (2020)) but much smaller and with strong ventral process in *GENUS-1 sp2* (ventral apophyses are common in some *Modisiminae* genera and usually relatively conserved among close relatives; e.g. Huber 2000). Noteworthy, but less surprising, is the different number of ALS spigots: *GENUS-1 sp1* has the plesiomorphic set of eight spigots; *GENUS-1 sp2* has the reduced set of only two spigots. A reduction of spigots has occurred repeatedly in *Pholcidae*, even though rarely within genera, e.g., in *Belisana*; *Metagonia* Simon, 1893; *Pholcus* Walckenaer, 1805; and *Smeringopus* (Huber 2000, 2011, 2012; Huber and Schaller 2005). Male clypeus modifications have also evolved repeatedly in *Pholcidae*, sometimes within genera (Huber 2021), and are present in *GENUS-1 sp2* but absent in *GENUS-1 sp1*. Finally, the AME have been reduced repeatedly in *Pholcidae*, sometimes even within genera, as in *Leptopholcus* Simon, 1893; *Mecolaesthus* Simon, 1893; *Modisimus* Simon, 1893; *Panjange* Deeleman-Reinhold & Deeleman, 1983; *Pholcus*, *Priscula*, and *Quamtana* (Huber 2003, 2011; Huber and Villarreal 2020); their presence in *GENUS-1 sp1* but absence in *GENUS-1 sp2* just adds one more reduction to the list. In sum, the discovery of further species may well justify a future split of *GENUS-1* into two genera, but currently we see no need and no justification to do so.

5.5.3 Notes on the biogeography of *Cai***inae*

The distribution of *Cai***inae* is restricted to limestone outcrop terrains in northeastern Brazil (Fig. 5.4). Each described species is recorded from a single biogeographic district of the Caatinga domain, separated by over 500 km. While *GENUS-1 sp1* was collected in the Peruaçu district, *GENUS-1 sp2* was collected in the Irecê district (sensu Moro et al. 2024). These karstic districts are separated by the crystalline terrains of the Southern Depressão Sertaneja district (SDS in Fig. 5.4) and the Chapada Diamantina subprovince

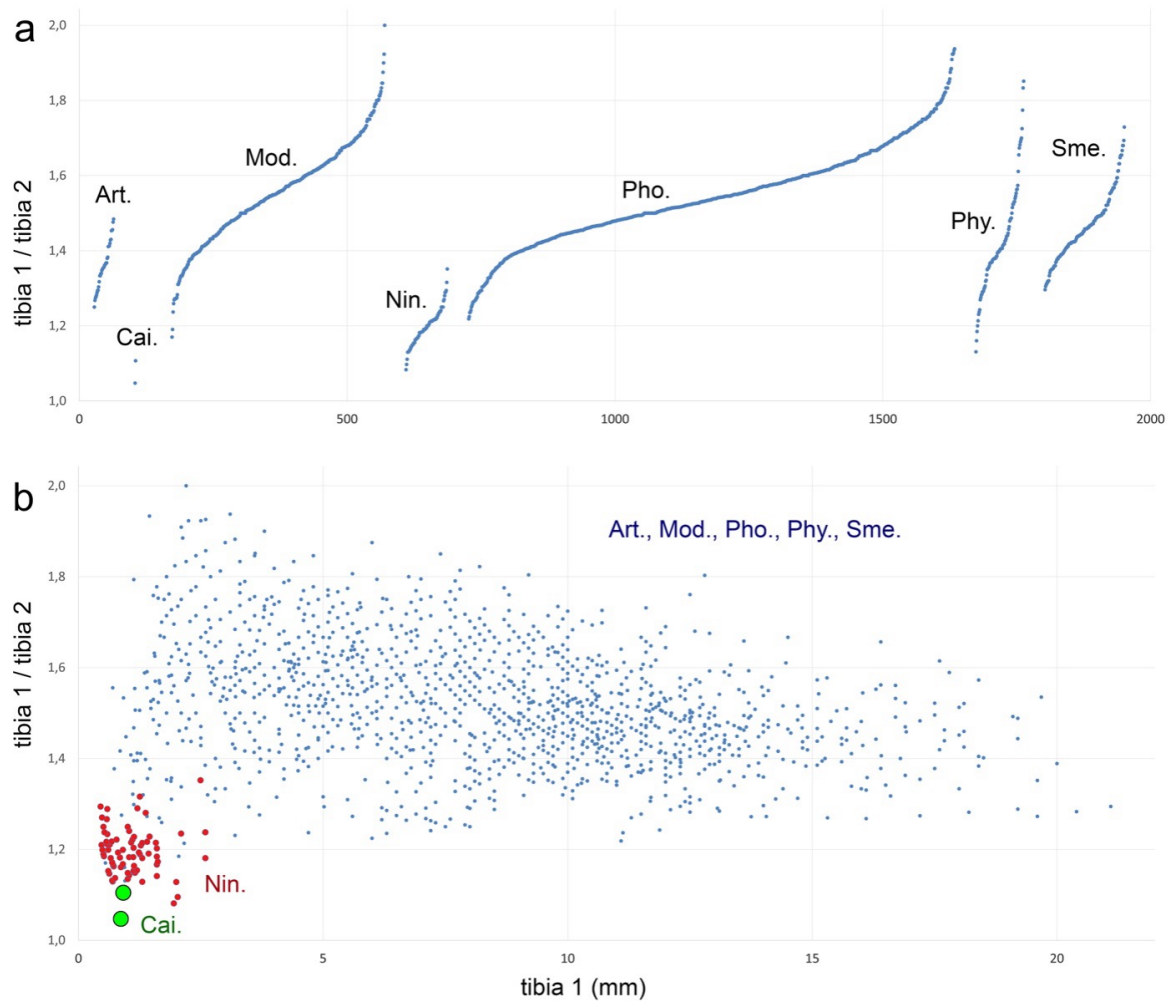


Figure 5.20. a Ratio of male tibia 1 / male tibia 2 in 1662 species of Pholcidae, taken from the taxonomic literature and sorted in each subfamily from smallest to largest values (Art. = Arteminae; Cai. = Cai***inae; Mod. = Modisiminae; Nin. = Ninetinae; Pho. = Pholcinae; Phy. = Physocyxclinae; Sme. = Smeringopinae). Note that the values for the two new species described herein are among the lowest known in Pholcidae. **b** Scatter plot of the ratio male tibia 1 / male tibia 2 on a proxy of male leg length (tibia 1 length). Note that the lower left corner of the plot is almost entirely filled by Ninetinae and Cai***inae. Erroneous and dubious measurements reported in the literature were excluded (see Suppl. material, Table C.5)

(CPD in Fig. 5.4). Pholcidae sampling in this region is sparse (see Fig. C.5) and it is thus likely that other species or additional occurrences of the described species will be found in future focused campaigns. Likewise, the existence of a third limestone outcrop district in the Caatinga domain (i.e., the Potiguar district; see Moro et al. 2024) raises the possibility of additional Cai***inae taxa further north. It is worth noting that targeted pholcid sampling is fundamental to reduce this sampling bias, as previously demonstrated for Ninetinae taxa (see Huber, Meng, Král, Ávila Herrera, Izquierdo and Carvalho 2023; Huber, Meng, Král, I. M. Á. Herrera et al. 2024).

As sister-group to a large lineage of Pholcidae hypothesized to have originated during the Cretaceous period (Meng, Podsiadlowski, Dimitrov et al. 2025), Cai***inae appears to be a relict taxon, as previously reported for other organisms in the Peruaçu district. A striking example is *Relictopiolus galadriel* Pérez-González, Monte and Bichuette, 2017, a troglobitic Kimulidae harvestman from another cave of the Parque Nacional Cavernas do Peruaçu (Perez-Gonzalez et al. 2017), the type-locality of *GENUS-1 sp2*. *Relictopiolus galadriel* diverged from its epigean sister-taxon *Tegipiolus pachypus* Roewer, 1949 about 40 Mya, during the Eocene, corroborating the hypothesis of an ancient divergence between relictual taxa from this region and their sister-taxa (Perez-Gonzalez et al. 2017). These two species are the only currently known kimulids from eastern Brazil, with the remaining species recorded from northern South America and Central America (Perez-Gonzalez et al. 2017). The Peruaçu district also harbors several other troglobites, including other harvestmen, whip-spiders, pseudoscorpions, isopods, amphipods, crickets, heteropterans, centipedes, and fish (Ázara et al. 2020; Cardoso, Bastos-Pereira, L. A. Souza et al. 2021; Monte et al. 2020; Perez-Gonzalez et al. 2017; Polhemus et al. 2018). Furthermore, the richest hotspot of subterranean biodiversity in South America (i.e., the Águas Claras cave system; see Ferreira, Berbert-Born et al. 2023), is in the Peruaçu district.

The floristic composition of the Peruaçu district is well-studied and distinct from the Caatinga sensu stricto and sandy Caatinga of other regions in the Caatinga domains, thus forming a particular sub-type of Caatinga, called ‘arboreal Caatinga’ (Moro et al. 2024). Conversely, the flora of the Irecê district is the least studied karstic region in the Caatinga domain (Moro et al. 2024). Both districts share the high abundance of caves, with several troglobites described for the Irecê district too, including other pholcids (Machado et al. 2011), as well as scorpions (Esposito et al. 2017), isopods (Cardoso, Bastos-Pereira and Ferreira 2022), amphipods (Bueno et al. 2022), beetles (Gnaspini et al. 1998) and catfishes (Bockmann et al. 2010). The Pholcidae species *Metagonia diamantina* Machado, Ferreira & Brescovit, 2011, described from the Irecê district (Ferreira, M. F. V. Souza et al. 2011), also represents a relict taxon. This species belongs to the *potiguar* species-group whose representatives are typically found in low primary productivity environments such as caves and high-altitude regions (Huber, Meng, Acurio et al. 2022). This distribution suggests that species of the *potiguar* group of *Metagonia* may have been replaced by modern taxa in more favorable environments (Huber, Meng, Acurio et al. 2022).

These abundant records of endemic and cave-associated taxa within both limestone outcrop districts underscore the importance of allopatric speciation in shaping species composition in this region. This pattern is corroborated by studies on other taxa, such as spiny orb-web spiders (Magalhaes et al. 2024), pholcid spiders (Huber, Meng, Acurio et al. 2022) and frogs (Teixeira Jr et al. 2012).

5.6 Conclusion

Two new species of Pholcidae from semiarid Brazilian Caatinga environments are described that do not fit into any of the previously available subfamilies. Genomic data suggest that together they are the sister

taxon of a large clade of Pholcidae that includes several subfamilies. We conservatively join these two species in a single new genus *GENUS-I* Huber gen. n. in the new subfamily Cai***inae subfam. n., even though there are some morphological differences between them that are uncommon within Pholcidae genera. Based of superficial similarities, our initial hypothesis was that the new Brazilian species might be related to the Chilean genus *Aucana* Huber, 2000. We thus also generated first molecular data for *Aucana*. These data reject our initial hypothesis and place *Aucana* in a clade that was previously informally called “other Arteminae” (Huber, Eberle et al. 2018) and for which we propose the new formal name Phy***inae subfam. n.; the revised subfamily-level relationships in Pholcidae are thus: (Pholcinae + Smeringopinae) + (Cai***inae + ((Arteminae + Ninetinae) + (Phy***inae + Modisiminae))).

5.7 Acknowledgements

We thank Anja Bodenheimer and Sandra Kukowka (LIB, Bonn) for their help with molecular lab work, and Richard A. Torres Contreras and Alexandre S. Michelotto for help with field work in Brazil. Specimens were collected under sampling permits issued by the Instituto Chico Mendes de Conservação da Biodiversidade through the Sistema de Autorização e Informação em Biodiversidade (SISBio; #59280-5). Field expeditions were supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG (PPM-00605-17 to Adalberto J. Santos), Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (311843/2022-0 to LSC), Fundação de Amparo à Pesquisa do Estado do Piauí – FAPEPI (Termo de Aceitação e Outorga Nº 001/2022 to LSC), and the German Research Foundation (DFG, project HU980/12-1 to BAH). The DFG is also acknowledged for funding molecular work and the PhD position of GM (project HU980/12-1). This publication is registered in compliance with Brazilian regulations at the Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SISGen; #A4CC17C).

References

- Aharon, S., B. A. Huber and E. Gavish-Regev (2017), *Daddy-long-leg giants: revision of the spider genus Artema Walckenaer, 1837 (Araneae, Pholcidae)*, Eur J Taxon **0**.
- Ázara, L. N. de, M. R. Hara and R. L. Ferreira (2020), *Cladistic analysis of the Brazilian troglitic harvestmen genus Iandumoema Pinto-da-Rocha (Opiliones: Gonyleptidae) with the description of three new species: a brief exercise over the use of troglomorphisms in cladistic analysis*, Invertebr Syst **34** 474.
- Bockmann, F. A. and R. Castro (2010), *The blind catfish from the caves of Chapada Diamantina, Bahia, Brazil (Siluriformes: Heptapteridae): description, anatomy, phylogenetic relationships, natural history, and biogeography*, Neotrop Ichthyol **8** 673.
- Bolger, A. M., M. Lohse and B. Usadel (2014), *Trimmomatic: a flexible trimmer for Illumina sequence data*, Bioinformatics **30** 2114.
- Boni, M. F., D. Posada and M. W. Feldman (2007), *An exact nonparametric method for inferring mosaic structure in sequence triplets*, Genetics **176** 1035.
- Brescovit, A. and B. A. Huber (2003), *Ibotoporanga Mello-Leitão: tropical spiders in Brazilian semi-arid habitats (Araneae: Pholcidae)*, Insect Syst Evol **34** 15.

- Brown, B. V. (1993),
A further chemical alternative to critical-point-drying for preparing small (or large) flies,
 Fly times **11** 10.
- Bueno, A. A. d. P., M. E. Bichuette, T. Zepon and L. R. Penoni (2022),
A new species of the troglobitic genus Spelaeogammarus da Silva Brum, 1975 (Amphipoda: Artesiidae) from a cave in the Brazilian semi-arid region, with new records of its congener, Spelaeogammarus spinilacertus Koenemann and Holsinger, 2000, Nauplius **30** e2022020.
- Capella-Gutierrez, S., J. M. Silla-Martinez and T. Gabaldon (2009),
trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses,
 Bioinformatics **25** 1972.
- Cardoso, G. M., R. Bastos-Pereira and R. L. Ferreira (2022), *Two new troglobitic species of Iansaoniscus from Brazilian caves (Crustacea, Isopoda, Pudeoniscidae)*, Subterr Biol **43** 127.
- Cardoso, G. M., R. Bastos-Pereira, L. A. Souza and R. L. Ferreira (2021),
Chaimowiczia: a new Iuiuniscinae genus from Brazil (Oniscidea, Synocheta, Styloniscidae) with the description of two new troglobitic species, Subterr Biol **39** 45.
- Chifman, J. and L. Kubatko (2014), *Quartet inference from SNP data under the coalescent model*,
 Bioinformatics **30** 3317.
- Dimitrov, D., J. J. Astrin and B. A. Huber (2013), *Pholcid spider molecular systematics revisited, with new insights into the biogeography and the evolution of the group*, Cladistics **29** 132.
- Duchene, D. A., C. Mather Niklas and Van Der Wal and S. Y. W. Ho (2022),
Excluding Loci With Substitution Saturation Improves Inferences From Phylogenomic Data,
 Syst Biol **71** 676.
- Eberle, J., D. Dimitrov, A. Valdez-Mondragon and B. A. Huber (2018),
Microhabitat change drives diversification in pholcid spiders, BMC Evol Biol **18** 141.
- Esposito, L. A., H. Y. Yamaguti, C. A. Souza, R. Pinto-Da-Rocha and L. Prendini (2017),
Systematic revision of the neotropical club-tailed scorpions, Physoctonus, Rhopalurus, and Troglorhopalurus, revalidation of Heteroctenus, and descriptions of two new genera and three new species (Buthidae: Rhopalurusinae), Bull Am Mus Nat Hist **2017** 1.
- Faircloth, B. C. (2016), *PHYLUCE is a software package for the analysis of conserved genomic loci*,
 Bioinformatics **32** 786.
- Ferreira, R. L., M. Berbert-Born and M. Souza-Silva (2023), *The Água Clara cave system in northeastern Brazil: the richest hotspot of subterranean biodiversity in South America*,
 Diversity **15** 761.
- Ferreira, R. L., M. F. V. Souza, E. O. Machado and A. D. Brescovit (2011),
Description of a new Eukoenia (Palpigradi: Eukoeneniidae) and Metagonia (Araneae: Pholcidae) from Brazilian caves, with notes on their ecological interactions, J Arachnol **39** 409.
- Ginestet, C. (2011), *ggplot2: Elegant Graphics for Data Analysis*, J R Stat Soc Ser A Stat Soc **174** 245.
- Gnaspini, P., S. A. Vanin and N. M. Godoy (1998),
A new genus of troglobitic carabid beetle from Brazil (Coleoptera, Carabidae, Zuphiini),
 Papéis Avulsos de Zoologia **40** 297.
- Huber, B. A. (2000), *New World pholcid spiders (Araneae: Pholcidae): a revision at generic level*,
 Bull Am Mus Nat Hist **2000** 1.
- Huber, B. A. (2003), *Southern African pholcid spiders: revision and cladistic analysis of Quamtana gen. nov. and Sperophora Hentz (Araneae: Pholcidae), with notes on male-female covariation*,
 Zool J Linn Soc **139** 477.

-
- Huber, B. A. (2011), *Phylogeny and classification of Pholcidae (Araneae): an update*, J Arachnol **39** 211.
- Huber, B. A. (2012), *Revision and cladistic analysis of the Afrotropical endemic genus Smeringopus Simon, 1890 (Araneae: Pholcidae)*, Zootaxa **3461** 1.
- Huber, B. A. (2013), *Revision and cladistic analysis of the Guineo-Congolian spider genus Smeringopina Kraus (Araneae, Pholcidae)*, Zootaxa **3713** 1.
- Huber, B. A. (2018), *The South American spider genera Mesabolivar and Carapoia (Araneae, Pholcidae): new species and a framework for redrawing generic limits*, Zootaxa **4395** 1.
- Huber, B. A. (2021), *Beyond size: sexual dimorphisms in pholcid spiders*, Arachnology **18** 656.
- Huber, B. A. (2022), *Revisions of Holocnemus and Crossopriza: the spotted-leg clade of Smeringopinae (Araneae, Pholcidae)*, Eur J Taxon **795** 1.
- Huber, B. A. and L. S. Carvalho (2019), *Filling the gaps: descriptions of unnamed species included in the latest molecular phylogeny of Pholcidae (Araneae)*, Zootaxa **4546** 1.
- Huber, B. A., J. Eberle and D. Dimitrov (2018), *The phylogeny of pholcid spiders: a critical evaluation of relationships suggested by molecular data (Araneae, Pholcidae)*, Zookeys 51.
- Huber, B. A. and G. Meng (2023), *On the mysterious Seychellois endemic spider genus Cenemus (Araneae, Pholcidae)*, Arthropod Syst Phylogeny **81** 179, doi: [10.3897/asp.81.e86793](https://doi.org/10.3897/asp.81.e86793).
- Huber, B. A. and G. Meng (2024), *Old World Micropholcus spiders, with first records of acrocerid parasitoids in Pholcidae (Araneae)*, ZooKeys **1213** 95, doi: [10.3897/zookeys.1213.133178](https://doi.org/10.3897/zookeys.1213.133178).
- Huber, B. A. and G. Meng (2025), *Like grains of sand: Ninetis spiders on the Arabian Peninsula (Araneae, Pholcidae)*, Zootaxa **5563** 290, doi: [10.11646/zootaxa.5563.1.19](https://doi.org/10.11646/zootaxa.5563.1.19).
- Huber, B. A., G. Meng, A. E. Acurio et al. (2022), *Metagonia spiders of Galápagos: blind cave-dwellers and their epigean relatives (Araneae, Pholcidae)*, Invertebr Syst **36** 647, doi: [10.1071/IS21082](https://doi.org/10.1071/IS21082).
- Huber, B. A., G. Meng, H. L. Clark and G. Cazanove (2023), *First blind daddy long-legs spiders from Australia and Réunion (Araneae, Pholcidae)*, Subterr Biol **46** 1, doi: [10.3897/subtbiol.46.105798](https://doi.org/10.3897/subtbiol.46.105798).
- Huber, B. A., G. Meng, T. M. Dederichs et al. (2024), *Castaways: the Leeward Antilles endemic spider genus Papiamenta (Araneae: Pholcidae)*, Invertebr Syst **38** IS23052, doi: [10.1071/IS23052](https://doi.org/10.1071/IS23052).
- Huber, B. A., G. Meng, N. Dupérré, J. Astrin and M. Herrera (2023), *Andean giants: Priscula spiders from Ecuador, with notes on species groups and egg-sac troglomorphism (Araneae: Pholcidae)*, Eur J Taxon **909** 1, doi: [10.5852/ejt.2023.909.2351](https://doi.org/10.5852/ejt.2023.909.2351).
- Huber, B. A., G. Meng, N. Dupérré, M. Herrera et al. (2023), *Humpback spiders from Ecuador: relationships, prosoma ‘inflation’ and genital asymmetry (Araneae: Pholcidae: Mecolaesthus)*, Invertebr Syst **37** 117, doi: [10.1071/IS22052](https://doi.org/10.1071/IS22052).
- Huber, B. A., G. Meng, J. C. garcía and L. S. Carvalho (2024), *Thriving in dry conditions: on the Neotropical spider genus Galapa (Araneae: Pholcidae)*, Zootaxa **5419** 301, doi: [10.11646/zootaxa.5419.3.1](https://doi.org/10.11646/zootaxa.5419.3.1).
- Huber, B. A., G. Meng, J. Král, I. M. Ávila Herrera, M. A. Izquierdo and L. S. Carvalho (2023), *High and dry: integrative taxonomy of the Andean spider genus Nerudia (Araneae: Pholcidae)*, Zool J Linn Soc **198** 534, doi: [10.1093/zoolinnean/zlac100](https://doi.org/10.1093/zoolinnean/zlac100).

- Huber, B. A., **G. Meng**, J. Král, I. M. Ávila Herrera and M. A. Izquierdo (2023), *Revision of the South American Ninetinae genus Guaranita (Araneae, Pholcidae)*, Eur J Taxon **900** 32, DOI: [10.5852/ejt.2023.900.2301](https://doi.org/10.5852/ejt.2023.900.2301).
- Huber, B. A., **G. Meng**, J. Král, I. M. Á. Herrera and L. S. Carvalho (2024), *Diamonds in the rough: Ibotyporanga (Araneae, Pholcidae) spiders in semi-arid Neotropical environments*, Eur J Taxon **963** 1, DOI: [10.5852/ejt.2024.963.2687](https://doi.org/10.5852/ejt.2024.963.2687).
- Huber, B. A., **G. Meng** and A. Valdez-Mondragón (2024), *Notes on Chisosa (Araneae, Pholcidae), with the description of a new species from Mexico.*, Zootaxa **5419** 217, DOI: [10.11646/zootaxa.5419.2.3](https://doi.org/10.11646/zootaxa.5419.2.3).
- Huber, B. A., **G. Meng**, A. Valdez-Mondragón et al. (2023), *Short-legged daddy-long-leg spiders in North America: the genera Pholcophora and Tolteca (Araneae, Pholcidae)*, Eur J Taxon **880** 1, DOI: [10.5852/ejt.2023.880.2173](https://doi.org/10.5852/ejt.2023.880.2173).
- Huber, B. A. and F. Schaller (2005), *High species diversity, male-female coevolution, and metaphyly in Southeast Asian pholcid spiders: the case of Belisana Thorell 1898 (Araneae, Pholcidae)*, vol. 155, Schweizerbart 1.
- Huber, B. A. and O. Villarreal (2020), *On Venezuelan pholcid spiders (Araneae, Pholcidae)*, Eur J Taxon **718** 1.
- Jiang, T., P. Kearney and M. Li (2001), *A polynomial time approximation scheme for inferring evolutionary trees from quartet topologies and its application*, SIAM Journal on Computing **30** 1942.
- Kalyanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermini (2017), *ModelFinder: fast model selection for accurate phylogenetic estimates*, Nat Methods **14** 587.
- Katoh, K. and D. M. Standley (2013), *MAFFT multiple sequence alignment software version 7: improvements in performance and usability*, Mol Biol Evol **30** 772.
- Klimova†, A., R. Rodríguez-Estrella, **G. Meng†** et al. (2023), *Metabarcoding reveals seasonal and spatial patterns of arthropod community assemblages in two contrasting habitats: Desert and oasis of the Baja California Peninsula, Mexico*, Diversity and Distributions **29** 438, DOI: [10.1111/ddi.13672](https://doi.org/10.1111/ddi.13672).
- Kozlov, A. M., D. Darriba, T. Flouri, B. Morel and A. Stamatakis (2019), *RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference*, Bioinformatics **35** 4453.
- Lam, H. M., O. Ratmann and M. F. Boni (2018), *Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm*, Mol Biol Evol **35** 247.
- Lartillot, N., N. Rodrigue, D. Stubbs and J. Richer (2013), *PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment*, Syst Biol **62** 611.
- Liu, S., D. Lang, **G. Meng** et al. (2022), *Tracing the origin of honey products based on metagenomics and machine learning*, Food chemistry **371** 131066, DOI: [10.1016/j.foodchem.2021.131066](https://doi.org/10.1016/j.foodchem.2021.131066).
- Liu, S., C. Zhou, **G. Meng** et al. (2022), *Evolution and diversification of Mountain voles (Rodentia: Cricetidae)*, Commun Biol **5** 1417, DOI: [10.1038/s42003-022-04371-z](https://doi.org/10.1038/s42003-022-04371-z).
- Machado, E. O., R. L. Ferreira and A. D. Brescovit (2011), *A new troglomorphic Metagonia Simon 1893 (Araneae, Pholcidae) from Brazil*, Zootaxa **3135** 59.
- Magalhaes, I. L., P. H. Martins, B. T. Faleiro et al. (2024), *Complete phylogeny of Micrathena spiders suggests multiple dispersal events among Neotropical*

-
- rainforests, islands and landmasses, and indicates that Andean orogeny promotes speciation, *Cladistics* **40** 552.
- Meng, G., L. Podsiadlowski, D. Dimitrov and B. A. Huber (2025), *The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders*, submitted.
- Meng, G., L. Podsiadlowski and B. A. Huber (2025a), *Exploring Phylogenetic Discordance: The Influence of Introgression and Insights into the Evolution of the Pholcidae (Araneae) Mitochondrial Genomes*, in preparation.
- Meng, G., L. Podsiadlowski and B. A. Huber (2025b), *Low coverage genome sequencing reveals a new subfamily of daddy long-legs spiders from Brazilian Caatinga (Araneae, Pholcidae)*, in preparation.
- Meng, G., L. Podsiadlowski, B. A. Huber and D. Dimitrov (2024), “The Complex Interplay Between Evolutionary Flexibility and Diversification in a Family of Spiders”, *The 116th Annual Meeting of the German Zoological Society*, Stuttgart, Germany.
- Meng, G., C. Zhou, S. Liu and X. Zhou (2022), “Phylogenomics based on transcriptomics and short-read sequences”, *Zoological Systematics and Evolution Protocol eBook*, ed. by G. Qiao, M. Zhu, W. Wang et al., 1st ed., Bio-protocol LLC 793, DOI: [10.21769/BioProtoc.1010649](https://doi.org/10.21769/BioProtoc.1010649).
- Minh, B. Q., H. A. Schmidt, O. Chernomor et al. (2020), *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era*, *Mol Biol Evol* **37** 1530.
- Monte, B. G. O. and M. E. Bichuette (2020), *Distinção Taxonômica da fauna subterrânea do Parque Nacional Cavernas do Peruaçu, estado de Minas Gerais, leste do Brasil*, *Biota Neotropica* **20** e20190810.
- Moro, M. F., V. O. Amorim, L. P. de Queiroz et al. (2024), *Biogeographical districts of the Caatinga dominion: a proposal based on geomorphology and endemism*, *Bot. Rev.* **90** 376.
- Naser-Khdour, S., B. Q. Minh, W. Zhang, E. A. Stone and R. Lanfear (2019), *The prevalence and impact of model violations in phylogenetic analysis*, *Genome Biol Evol* **11** 3341.
- Perez-Gonzalez, A., F. S. Ceccarelli, B. G. Monte et al. (2017), *Light from dark: A relictual troglobite reveals a broader ancestral distribution for kimulid harvestmen (Opiliones: Laniatores: Kimulidae) in South America*, *PloS One* **12** e0187919.
- Polhemus, D. A. and R. L. Ferreira (2018), *Two unusual new genera of cavernicolous Hydrometridae (Insecta: Heteroptera) from eastern Brazil*, *Tijdschrift voor Entomologie* **161** 25.
- Prijbelski, A., D. Antipov, D. Meleshko, A. Lapidus and A. Korobeynikov (2020), *Using SPAdes De Novo Assembler*, *Curr Protoc Bioinformatics* **70** e102.
- R Core Team, R. (2022), *R: A language and environment for statistical computing*.
- Sayyari, E., J. B. Whitfield and S. Mirarab (2018), *DiscoVista: Interpretable visualizations of gene tree discordance*, *Mol Phylogenet Evol* **122** 110.
- Seemann, T. (2017), *Shovill: faster SPAdes assembly of Illumina reads*, Tilgjengelig fra Tilgjengelig fra.
- Simon, E. (1893), *Histoire Naturelle des Araignées*, 2nd, vol. 1, 2, Paris: Roret 256.
- Strimmer, K. and A. von Haeseler (1997), *Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment*, *Proc Natl Acad Sci U S A* **94** 6815.

- Swofford, D. L. (2002), *PAUP: phylogenetic analysis using parsimony (and other methods)*, 4.0 beta, <http://paup.csit.fsu.edu/>.
- Teixeira Jr, M., R. C. Amaro, R. S. Recoder, M. Sena and M. T. Rodrigues (2012), *A relict new species of Oreobates (Anura, Strabomantidae) from the Seasonally Dry Tropical Forests of Minas Gerais, Brazil, and its implication to the biogeography of the genus and that of South American Dry Forests*, Zootaxa **3158** 37.
- Vachaspati, P. and T. Warnow (2018), *SVDquest: Improving SVDquartets species tree estimation using exact optimization within a constrained search space*, Mol Phylogenet Evol **124** 122.
- Zhang, C. and S. Mirarab (2022), *Weighting by Gene Tree Uncertainty Improves Accuracy of Quartet-based Species Trees*, Mol Biol Evol **39** msac215.
- Zhang, C., M. Rabiee, E. Sayyari and S. Mirarab (2018), *ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees*, BMC Bioinformatics **19** 153.
- Zhou†, L., **G. Meng†**, L. Zhu, L. Ma and K. Chen (2024), *Insect Antimicrobial Peptides as Guardians of Immunity and Beyond: A Review*, Int J Mol Sci **25** 3835, doi: [10.3390/ijms25073835](https://doi.org/10.3390/ijms25073835).

Exploring Phylogenetic Discordance: The Influence of Introgression and Insights into the Evolution of the Pholcidae (Araneae) Mitochondrial Genomes

Guanliang Meng^{1,*}, Lars Podsiadlowski¹ and Bernhard A. Huber¹

¹ Zoological Research Museum Alexander Koenig, LIB, Bonn, Germany

* Corresponding author: G.Meng@leibniz-lib.de

6.1 Summary

This chapter is in preparation to be submitted as:

G. Meng, L. Podsiadlowski and B. A. Huber (2025a), *Exploring Phylogenetic Discordance: The Influence of Introgression and Insights into the Evolution of the Pholcidae (Araneae) Mitochondrial Genomes*, in preparation

Summary

The exchange of genetic material between lineages through hybridization and backcrossing is referred to as introgression (Aguillon et al. 2022). Introgressed alleles can be adaptive in specific environments (termed “adaptive introgression”), contributing to adaptation (Ge et al. 2023; Racimo et al. 2015). The formation of new genetic combinations between introgressed and existing alleles may give rise to “speciation genes,” leading to species divergence (Du et al. 2024; Rosser et al. 2024). In phylogenomics, introgression is a significant factor contributing to incongruences among gene trees, resulting in species networks rather than a traditional species tree (Rancilhac et al. 2024). Introgression has been documented in various organisms, including plants (H. Wang et al. 2024), fungi (Kinneberg et al. 2023), and vertebrates (Ge et al. 2023; Rancilhac et al. 2024). However, the occurrence of introgression in spiders (Araneae) has not yet been thoroughly assessed. While there are more than 130 spider families and more than 50,000 described species (WSC 2025), previous studies on introgression in spiders have primarily focused on a limited number of species and populations within a few spider families, mostly relying on a few loci (Garb et al. 2009; Johannessen et al. 2001; Lattimore et al. 2011). While mitochondrial genomes have been extensively studied in groups such as vertebrates and insects, only a few hundred spider mitochondrial genomes are publicly

available on NCBI. For example, there are currently only 545 spider mitogenomes (including four from the Pholcidae family) that exceed 10 Kbp available on NCBI, compared to 29,630 insect mitogenomes as of March 16, 2025. Mitochondrial genomes have important evolutionary, taxonomic, and ecological implications (Biot-Pelletier et al. 2023; Lattimore et al. 2011; Tang et al. 2015). The scarcity of mitochondrial genome data in spiders has hindered our ability to understand the architecture and evolution of their mitochondrial genomes, as well as their interactions with nuclear genomes.

In this study, we utilized ultraconserved elements (UCEs) from 90 newly sequenced genomic data obtained through the low-coverage whole genome sequencing (lcWGS) method, together with data from our previous studies (Meng, Podsiadlowski, Dimitrov et al. 2025; Meng, Podsiadlowski and Huber 2025b), resulting in a total of 193 analyzed pholcid species. We investigated the prevalence of introgression and its impact on phylogenetic inference in Pholcidae while also examining the evolution of mitochondrial genomes within the family. Our results show that while our new UCE phylogeny is identical to those of two recent studies (Meng, Podsiadlowski, Dimitrov et al. 2025; Meng, Podsiadlowski and Huber 2025b), the mitochondrial data suggest different topologies. Moreover, we found evidence of multiple inter-subfamily and inter-genus introgressions within Pholcidae. We present 193 mitochondrial genomes, 98 of which are complete or nearly complete. However, we encountered challenges in annotating some protein-coding genes (PCGs), including ATP8, ND4, ND4L, and ND5, as well as tRNA genes in pholcids. Our findings indicate that the gene order of mitochondrial PCGs is highly conserved within Pholcidae, while these PCGs exhibit strong codon usage bias (CUB) and significant purifying selection. This CUB appears to be driven mainly by selection pressure rather than mutation bias. We argue that introgression and the rapid evolution of mitogenomes contribute to the discordance observed between phylogenies constructed from different data.

This study underscores the importance of considering biological processes, including introgression, when resolving the phylogenetic relationships among organisms. It also calls for improvements in the mitochondrial genome annotation of spiders.

Personal contribution

B.H. and L.P. initiated the project. B.H. collected, taxonomically identified, and selected samples. L.P. coordinated the molecular methods. G.M. designed this study, performed molecular methods, conducted analyses, and drafted the manuscript. All the authors contributed to the manuscript revision. We thank the HPC unit of Leibniz Institute for the Analysis of Biodiversity Change (LIB) and the molecular lab, especially Anja Bodenheim and Sandra Kukowka for hands-on introduction to the laboratory experiments.

6.2 Introduction

A phylogeny aids in interpreting the evolution of genes, morphological traits, and the origins of organisms, among other aspects. Therefore, understanding the mechanisms that underlie the heterogeneity in gene trees and the discordance between gene trees and species trees is critical in phylogenomics. Biologically, different genes often have distinct evolutionary histories, which can arise from factors such as gene duplication and loss, recombination, incomplete lineage sorting (ILS), or gene flow due to horizontal gene transfer and hybridization (Maddison 1997; Morales-Briones et al. 2021; Szöllősi et al. 2015). Additionally, issues like saturation, low phylogenetic signal, hidden paralogy, and evolutionary models that fail to account for heterogeneity in evolutionary rates and base composition across loci or sites, and branch lengths can lead to long branch attraction (LBA) and incorrect gene trees (Duchene et al. 2022; Hernandez

et al. 2021). Consequently, various biological processes and technical artifacts pose significant challenges to reconstructing a reliable tree of life.

During speciation, different lineages (“parental lineages”) may form hybrid individuals. Genetic exchange between these parental lineages can occur when the hybrids backcross with the parental lineages, a phenomenon known as introgression (Aguillon et al. 2022). Introgression, a form of gene flow, among other mechanisms, leads to the formation of a phylogenetic network instead of a phylogenetic tree, and it contributes to the incongruence between the gene tree and the species tree (Rancilhac et al. 2024). Recent studies have revealed the presence of introgression in various organisms, including plants (H. Wang et al. 2024), fungi (Kinneberg et al. 2023), and vertebrates (Ge et al. 2023; Rancilhac et al. 2024). Introgression plays a crucial role in speciation, radiation, and adaptation (Ge et al. 2023; Hedrick 2013; Ma et al. 2019). For instance, it generates novel genetic combinations that enhance reproductive barriers between parental lineages (Du et al. 2024; Rosser et al. 2024) or introduces adaptive alleles to the parental lineages (known as “adaptive introgression”) (Ge et al. 2023; Racimo et al. 2015).

With more than 50,000 described species (WSC 2025), spiders (Araneae) are one of the most species-rich groups of arthropods, exhibiting a wide range of lifestyles and inhabiting diverse ecosystems worldwide (Agnarsson et al. 2013; Foelix 2010). However, the prevalence of introgression in spiders has been evaluated only in a limited number of species across several families, based on restricted data at a small scale (e.g., population; Garb et al. 2009; Johannesen et al. 2001; Lattimore et al. 2011). Furthermore, few studies have focused on the evolutionary significance of introgression in the adaptation and diversification of spiders (Ortiz et al. 2024; Oxford 2019). For example, Salgado-Roa et al. (2018) reported that introgression and Andean uplift contributed to the diversification of *Gasteracantha cancriformis* (Araneidae). Ivanov et al. (2018) proposed that the most likely evolutionary explanation for mitonuclear discordance in the *Alopecosa aculeata* and *Pardosa pullata* groups of wolf spider species (Lycosidae) is mitochondrial introgression followed by fixation. Recently, Responde et al. (2025) suggested that adaptive introgression and balancing selection played a role in the speciation of the *Argyrodes lanyuensis* species complex (Theridiidae), while mating behavior, rather than habitat overlap and timing of reproductive maturity impacts the direction and geographic extent of introgression in New Zealand fishing spiders (Dolomedidae: *Dolomedes*) (Connolly et al. 2025).

The daddy-long-leg spiders, Pholcidae, with more than 2,000 described species and approximately 6,000 estimated species, represent one of the most species-rich families (Huber and Chao 2019). They are classified into seven subfamilies, including Arteminae, Cai***inae, Modisiminae, Ninetinae, Pholcinae, Phy***inae and Smeringopinae. They have a worldwide distribution but are mostly found in tropical regions, occupying diverse microhabitats (Eberle et al. 2018). The phylogeny of the family has received considerable attention (e.g., Dimitrov et al. 2013; Eberle et al. 2018; Meng, Podsiadlowski, Dimitrov et al. 2025) and it has been used extensively to reconstruct the evolution of various character systems, such as sperm ultrastructure (Dederichs et al. 2022) and karyology (Ávila Herrera et al. 2021). However, it is unclear if introgression is present in pholcids and how it might affect pholcid evolution as well as the construction of pholcid phylogeny.

With the decreasing cost of sequencing over the past two decades, mitochondrial genomes (mitogenomes) have become invaluable for studying phylogenetic relationships across diverse lineages, as well as for investigating introgression, selection, adaptation, co-evolution between mitogenome and nuclear genome, and biodiversity monitoring (Biot-Pelletier et al. 2023; Lattimore et al. 2011; Tang et al. 2015). For instance, the evolutionary rates of mitogenomes vary significantly based on factors such as parasitic lifestyle and locomotory capacity in Bilateria, with endoparasites exhibiting the highest rates of evolution and micropredatory or free-living groups showing the lowest (Jakovlić et al. 2023). Despite the ecolo-

gical and evolutionary significance of spiders, their mitogenomes remain seriously underrepresented in public databases compared to those of other arthropods, such as insects. For instance, only 545 spider mitogenomes (including four from Pholcidae) longer than 10 Kbp are publicly available on NCBI, while there are 29,630 insect mitogenomes (access date: March 16, 2025). Therefore, we know little about the architecture and evolutionary dynamics of spider mitogenomes. Only recently have several researchers shed light on the phylogenetic relationships and divergence times among different spider families based on mitogenomes (M. Li et al. 2022; R. Zhang et al. 2025).

In this study, we applied the low-coverage whole genome sequencing (lcWGS) method to 90 pholcid species and integrated this new data with the genomic data from Meng, Podsiadlowski, Dimitrov et al. (2025) and Meng, Podsiadlowski and Huber (2025b) (see Chapters 3 and 5). Our aims are to: (1) explore the presence, extent and evolutionary implications of introgression in pholcids; (2) assemble and annotate 193 Pholcidae mitogenomes; and (3) characterize codon usage bias, genetic divergence, selection pressure, and gene order rearrangement in these mitogenomes.

6.3 Materials and methods

6.3.1 Taxonomic sampling

A total of 90 pholcid samples were selected (Table D.1). Furthermore, we integrated the UCE data from all samples in Meng, Podsiadlowski, Dimitrov et al. (2025) and Meng, Podsiadlowski and Huber (2025b), resulting in a total of 193 terminal taxa representing all currently recognized subfamilies (8 Arteminae, 3 Cai***inae, 79 Modisiminae, 61 Ninetinae, 17 Pholcinae, 13 Phy***inae, and 12 Smeringopinae). Following Meng, Podsiadlowski, Dimitrov et al. (2025), four species from Scytodidae, Sicariidae, Orsolobidae, and Hypochilidae were used as outgroups in the UCE phylogeny. For the construction of the mitochondrial phylogeny, we also downloaded the mitogenomes of ten further spider species (3 Modisiminae, 2 Pholcinae, and 5 outgroups representing Atypidae, Dipluridae, Salticidae, Sicariidae, and Sparassidae) (Table D.1).

6.3.2 Molecular methods

lcWGS sequencing. Detailed methods for DNA isolation, library construction, sequencing, and contig assembly can be found in Meng, Podsiadlowski and Huber (2025b). Potential cross-contaminants were identified and filtered out using an “all-blast-all” method described in Meng, Podsiadlowski, Dimitrov et al. (2025). The ultra-conserved elements (UCEs) were identified using Phyluce (v1.7.1; Faircloth 2016) and subsequently aligned with MAFFT (v7.475; Katoh et al. 2013) (–adjustdirection –maxiterate 1000). The alignments were then trimmed with TrimAl (v1.4.rev15; Capella-Gutierrez et al. 2009) (–automated 1). Symetry test was performed with IQ-TREE 2 (–symtest-only; v2.2.0; Minh, Schmidt et al. 2020). Based on alignment length (50 bp vs 100 bp) and taxon coverage (40%, 50%, and 60%), we created three matrices (1-3) (Figure D.1a). Intra-locus recombination anlaysis was finished with 3SEQ (v1.8.0; Boni et al. 2007; Lam et al. 2018). By removing the sequences containing intra-locus recombinants and the introgressed loci from matrix 1 (see the section **Introgression Detection** below), we further created matrix 4 (Figure D.1a). As in Meng, Podsiadlowski, Dimitrov et al. (2025) and Meng, Podsiadlowski and Huber (2025b), we evaluated systematic bias in the dataset by examining basic characteristics of the alignments, conducting likelihood mapping (Strimmer et al. 1997) and testing for symmetry (Naser-Khdour et al. 2019), respectively.

Phylogenetic analysis with UCEs. To achieve a robust phylogenetic hypothesis, we conducted phylogenetic inference using concatenation, summary and site-based methods (a total of 18 analyses), respectively (Figure D.1b). We firstly conducted maximum likelihood (ML)-based species tree search with IQ-TREE and RAxML-NG (v1.2.2; Kozlov et al. 2019), respectively. Each gene was treated as a separate partition. For IQ-TREE, the best-fitting model was determined using ModelFinder, and the partitions were automatically merged to reduce the potential over-parameterization (-m MFP+MERGE) (Kalyaanamoorthy et al. 2017). 2000 replicates were used for the ultrafast bootstrap (UFBoot) (Minh, M. A. T. Nguyen et al. 2013) and SH-aLRT test (Guindon et al. 2010) (-sampling GENESITE), respectively. The -bnni option was applied to avoid overestimating branch supports with UFBoot. To prevent being trapped in local optima, a smaller perturbation strength (-pers 0.2), a larger number of stop iterations (-nstop 500), and 10 independent runs were employed (Minh, Schmidt et al. 2020). In the RAxML-NG analysis, the model GTR+R4+I+FO was employed. Proportional branch lengths were assigned to each partition (-brlen scaled). A total of 100 non-parametric bootstraps were conducted. To mitigate the possibility of being trapped in local optima, 25 random starting trees and 25 parsimony starting trees were used. Additionally, to minimize the potential impact of compositional heterogeneity on species tree inference, Bayesian inference (BI) was performed using PhyloBayes-MPI (v1.8; Lartillot et al. 2013) with the CAT-GTR+G4 model and matrix 2 and 3, respectively (Figure D.1b). For summary methods, the gene trees were first inferred using IQ-TREE with the aforementioned parameters, and then summarized with unweighted and weighted ASTRAL-IV (C. Zhang and Mirarab 2022), and Asteroid (v1.0; Morel et al. 2023), respectively, which are all based on the multi-species coalescent (MSC) model and are statistically consistent when incomplete lineage sorting (ILS) is the only reason responsible for the discordance between species tree and gene tree (Morel et al. 2023). To ensure optimality, we adopted 16 rounds of search and 16 rounds of subsampling and exploration (-R) for ASTRAL-IV. While weighted ASTRAL-IV is more robust to gene trees with missing species than unweighted ASTRAL-IV (C. Zhang and Mirarab 2022), Asteroid is well-established for handling gene trees with a high level of missing data (Morel et al. 2023). However, such summary methods are known to be sensitive to gene tree estimation error (Mirarab et al. 2016). Therefore, we also conducted the SVDquartets analysis (Chifman et al. 2014) which infers species trees directly from the site patterns among species based on MSC. Finally, the gene concordance factor (gCF) and the site concordance factor (sCF) (-scf 2000) were calculated with IQ-TREE (Minh, Hahn et al. 2020).

Introgression Detection. The ABBA-BABA test, also known as *D*-statistics (Durand et al. 2011; Green et al. 2010), along with f_4 -ratio statistics (Patterson et al. 2012), were calculated using the Dtrios subcommand of Dsuite (v0.5 r57; Malinsky, Matschiner et al. 2021) for all possible combinations of subfamilies and genera (Figure D.1c). In a phylogenetic tree (((P1,P2), P3), O) with four taxa: P1, P2, P3, and O (outgroup), the *D*-statistic tests whether the number of shared derived alleles (denoted as “B”) between P2 and P3 is significantly different from that between P1 and P3 (or vice versa). This comparison is made using ancestral alleles (“A”) defined by the outgroup. More specifically, it evaluates if the numbers of site patterns ABBA and BABA differ from each other. If there is only ILS and no introgression, the two numbers should be equal, and the *D* statistic should be zero (Patterson et al. 2012). The f_4 -ratio estimates the proportion of introgressed loci in an admixed population (Patterson et al. 2012). Additionally, *f*-branch statistics (Malinsky, Svardal et al. 2018) were computed using the Fbranch subcommand in Dsuite. It evaluates the comparative excess of allele sharing between the descendants (taxon “B”) of branch *b* and taxon C, relative to the allele sharing observed between the descendants (taxon “A”) of branch *a* and taxon C (see Malinsky, Svardal et al. 2018). This comparison helps identify specific internal branches where introgression has occurred (Malinsky, Matschiner et al. 2021). For trios exhibiting significantly elevated

D values, we further identified the introgressed loci by estimating f_{dM} statistics within sliding windows along the concatenated MSAs, using the *Dinvestigate* subcommand in *Dsuite*. A window size of 50 base pairs (bp) and a step size of 25 bp were employed. The f_{dM} statistic ranges from -1 to 1, with a value of zero indicating the absence of introgression (Malinsky, Challis et al. 2015). Introgression between P2 and P3 leads to positive f_{dM} , while introgression between P1 and P3 result in negative f_{dM} . We identified sliding windows corresponding to the top 5% of the f_{dM} distribution. We considered loci within these windows as introgressed and then examined their biological functions. Specifically, we aligned the sequences of these loci against the genome of *Stegodyphus dumicola* (Eresidae) using the *blastn* command from BLAST+ (v2.9.0; Camacho et al. 2009) with an e-value threshold of $1e-5$. For each sequence, we retained its best hit if it had an alignment length of at least 100 bp. For each locus, we retained only the blast match with the longest alignment length and the highest similarity. The genes located within this alignment range were extracted and submitted to the web server *g:Profiler* (version: e112_eg59_p19_25aa4782; <https://biit.cs.ut.ee/gprofiler/gost>; Kolberg et al. 2023) to evaluate functional enrichment. Furthermore, for these introgressed and non-introgressed loci, we also conducted likelihood mapping to investigate their phylogenetic information distribution.

Mitogenome assembly and Phylogenetic inference. Mitogenome assembly was performed using *MitoZ* (v3.6; Meng, Y. Li et al. 2019) and *NOVOPlasty* (v4.3.1; Dierckxsens et al. 2017), while gene annotation was conducted with *MitoZ* and *MITOS2* (Donath et al. 2019). However, the annotation of some protein-coding genes (PCGs)—specifically ATP8, ND4, ND4L, and ND5—in certain samples was unsuccessful, likely due to significant divergence from the reference gene set. To address this, we created a custom database for *MitoZ* by incorporating the related PCGs from Araneae available in the NCBI GenBank database, along with the Pholcidae PCGs identified in this study. In a few cases, particularly with the ATP8 gene, we used the *nhmmer* command (Eddy 2011; <http://hmmer.org/>) and the gene HMM profile provided by *MitoZ* to estimate the gene’s rough range. We then manually curated the annotation by checking the alignments against the corresponding genes from other pholcid samples. Mitochondrial genome visualization was conducted with *OGDRAW* (v1.3.1; Greiner et al. 2019). Next, the gene fragments from Meng, Podsiadlowski, Dimitrov et al. (2025) were combined with 10 mitogenomes from NCBI and our newly assembled mitogenomes. We then performed multiple-sequence alignment (MSA) for the 13 protein-coding genes (PCGs) and the 12S and 16S rDNA genes, respectively. To avoid non-biological frame shifts in the PCG MSAs, we first translated the coding sequences (CDSs) of the 13 PCGs into protein sequences using the invertebrate mitochondrial genetic code. Subsequently, protein MSAs were constructed using the L-INS-I algorithm of MAFFT (v7.487). CDS MSAs were then constructed based on the protein MSAs with PAL2NAL (v14.1; Suyama et al. 2006). The alignments of the 12S and 16S rDNA genes were constructed based on their secondary structures with the *mafft-xinsi* algorithm of MAFFT (v7.487) and *MXSCARNA* (Tabei et al. 2008). We checked all alignments with *AliView* (v1.28; Larsson 2014) to remove outliers. Finally, we trimmed the MSAs with *Gblocks* (v0.91b; Castresana 2000), *TrimAl* (v1.5.rev0) (-automated1), and *ClipKIT* (v2.3.0; Steenwyk et al. 2020). We tested the following trimming strategies of *ClipKIT* (-m kpic, kpi, kpic, kpic-gappy, kpic-smart-gap, kpi-gappy, kpi-smart-gap, and smart-gap). Similar to the previous phylogenetic inference using UCE data, for the 13 PCGs and two rDNA genes, we employed *IQ-TREE* and *RAXML-NG* to perform partitioned concatenation analysis. Meanwhile, we utilized unweighted and weighted *ASTRAL-IV* to perform summary method-based inference.

Characterization of the mitogenomes. For each gene, Kimura’s 2-parameter (K2P) distances (Kimura 1980) were calculated within and between subfamilies using the R package *ape* (v5.8.1; Paradis 2019). The GC content of the entire gene and each codon position (GC1, GC2, GC3) was computed with *CAT*

(v1.3; Z. Zhang, J. Li, Cui et al. 2012). We also calculated GC12 which is the average of GC1 and GC2. To quantify codon usage bias (CUB) in the PCGs, RSCU (relative synonymous codon usage) and ENC (effective number of codons) were evaluated for each codon (excluding stop codons) using CodonW (v1.4.4; <https://codonw.sourceforge.net/index.html>, last accessed March 1, 2025). An RSCU value of 1 indicates equal usage of all synonymous codons encoding the same amino acid, while values greater than 1 suggest positive CUB, and values less than 1 suggest negative CUB. Invertebrate mitogenomes have only two stop codons. Their ENC values range from 20 to 62, where 20 suggests strong CUB with each amino acid using only one of its synonymous codons, and 62 indicates equal usage of all synonymous codons (Wright 1990). To ascertain the relative contribution of mutation bias and natural selection on codon usage, we constructed the neutrality plot (GC12 versus GC3) and the Enc plot (ENC versus GC3), respectively. In the neutrality plot, a regression line with a slope near zero indicates the absence of directional mutation pressure and the presence of strong selective constraints on maintaining high GC content. Conversely, a slope close to one suggests that neutral mutation is the primary factor responsible for codon selection (Sueoka 1988; Y. Zhang et al. 2022). In the Enc plot, the curve $ENC = 2 + s + (29 / (s^2 + (1 - s)^2))$ holds when there is no selection acting on the codon usage of the gene, where s represents the value of GC3 (Wright 1990). When GC-bias mutation pressure or selection is present, the data points will deviate from the reference curve (Wright 1990). For each PCG of each species, the ratio (Ka/Ks) of nonsynonymous (Ka) and synonymous (Ks) substitution rates was calculated using KaKs_Calculator (v2.0; D. Wang et al. 2010) respectively. The model averaging method (-m MA) was employed. A ratio greater than one indicates positive selection, a ratio equal to one suggests neutral mutation, and a ratio less than one suggests negative (purifying) selection (Z. Zhang, J. Li, Zhao et al. 2006). *Loxosceles reclusa* (Sicariidae) was utilized as the reference species. Finally, we compared the order of PCGs for the samples with complete or near complete mitogenomes.

6.4 Results

UCE data. Based on the newly sequenced 90 samples and 107 samples from previous studies and NCBI (including four outgroups), we created matrices 1-4 with varying taxon coverage (40%, 50%, and 60%) and minimum locus length (100 bp versus 50 bp). These matrices contain 56K, 31K, 16K, and 11K bp sites, respectively (Figure D.1a). The likelihood mapping plot indicates that approximately 32% of data points are located at the three corners of the triangle, suggesting a high level of information suitable for phylogenetic analysis (Figure D.2). Most alignments are between 100 and 300 bp in length (Figure D.3). Matrix 1 (containing 238 loci) has a relatively high level of missing data (Figure D.4), while the extent of data missing is less severe in matrix 2 (containing 133 loci; Figure D.5). However, samples from lcWGS (the “G” codes) exhibit a higher level of missing data compared to samples from the TCS method (the “E” codes; Figure D.5). A total of 47 loci (19.7%) involving 189 species contain recombinants whose length exceeds 100 bp for matrix 1 (Figure D.6), while it is 28 loci (21.1%) involving 181 species for matrix 2 (Figure D.6). A total of 42 (17.6%) and 27 (20.3%) loci violates the assumptions of stationarity and homogeneity (SymPval < 0.01) for matrix 1 and 2, respectively (Table D.2 and D.3).

Pholcidae UCE phylogeny. The monophyly of all subfamilies is largely supported by 18 phylogenetic analyses based on different matrices (1-4) and tree construction methods (ML, coalescence, BI) (Figure 6.1, 6.2 and 6.3, Figure D.7). Both SH-aLRT and UFBoot support values (%) are ≥ 99.5 for Smeringopinae, Pholcinae, Arteminae, Ninetinae, respectively. The SH-aLRT/UFBoot values for Modisiminae and Phy***inae are 90.2/91 and 85.8/80, respectively. However, SVDquartets analyses based on matrices 1-3

(Figure D.8) and unweighted and weighted ASTRAL-IV analyses based on matrix 2 (Figure D.9) placed a subgroup of Modisiminae as the sister group of Phy***inae, while the analysis of matrix 4 + IQ-TREE placed *Anopsicus banksi* in Modisiminae (Figure D.10) rather than in Phy***inae as did all other analyses. The subfamily-level phylogenetic relationships are largely supported by most analyses, which are identical to those of Meng, Podsiadlowski, Dimitrov et al. 2025. Specifically, the sister relationships of Arteminae + Ninetinae, Phy***inae + Modisiminae, and Pholcinae + Smeringopinae are supported by all 18 analyses. Inconsistency was only found in the position of Cai***inae: while 14 out of 17 analyses placed Cai***inae as sister of all pholcids except Smeringopinae and Pholcinae, the analysis of matrix 1 + Asteroid placed Cai***inae together with Arteminae + Ninetinae (Figure D.11), and analyses of matrix 2 + unweighted ASTRAL-IV (Figure D.9) and Asteroid (not shown) placed Cai***inae together with Pholcinae + Smeringopinae.

Pholcidae MT phylogeny. A total of 193 mitogenomes of Pholcidae were assembled, with 98 of them being complete or nearly complete. The phylogenies based on mitogenomes are less consistent and often contradict the UCE-based phylogeny described above (Figure D.12 and D.13). Only the monophyly of the subfamily Ninetinae is well supported. The subfamily-level phylogenetic relationships are largely similar to those of the UCE-based phylogeny, but the other subfamilies are often non-monophyletic, and Cai***inae is sometimes sister to all other pholcids except for the genus *Priscula*. Notably, the genus *Priscula* was resolved as sister group to all other pholcids (Figure D.12).

Introgression analysis. D , f_4 -ratio and f_{dM} statistics all revealed the presence of introgression in all pholcid subfamilies except Ninetinae (Figure 6.4a, D.14). However, the f -branch statistic suggests that introgression also involves Ninetinae but not Modisiminae (Figure 6.4b). Our analyses revealed the following significant D and f_4 -ratio statistics (p-values < 0.05(*), <0.01(**) and <0.001(***)): 0.08/0.29 between Arteminae and Pholcinae (*); 0.12/0.28 between Arteminae and Smeringopinae (**); 0.05/0.07 between Arteminae and Cai***inae (*); 0.08/0.40 between Arteminae and Phy***inae (**); 0.04/0.12 between Arteminae and Modisiminae (*); 0.03/0.13 between Phy***inae and Modisiminae (**). Along the loci, many regions exhibit f_{dM} values close to -1 or 1, indicating introgressed regions (Figure D.14). In the likelihood mapping analysis (Figure D.15a), introgressed loci had a higher proportion of sites (5.4%) supporting Pholcinae + Arteminae compared to non-introgressed loci (1.9%). Meanwhile, 71.1% of the quartets in the introgressed loci and 84.2% of the quartets in the non-introgressed loci supported Pholcinae + Modisiminae (Figure D.15a), respectively. Similar results were observed for Smeringopinae + Arteminae (Figure D.15b) and Cai***inae + Arteminae (Figure D.15c), respectively. Additionally, our results showed the following f -branch statistics (Figure 6.4b): 0.33 between Pholcinae and Phy***inae; 0.15 between Pholcinae and Arteminae + Ninetinae; 0.13 between Smeringopinae and Arteminae + Ninetinae; 0.32 between Smeringopinae and Phy***inae; 0.28 between Smeringopinae and Arteminae; 0.40 (or 0.28) between Phy***inae and Arteminae; and 0.38 (or 0) between Ninetinae and Phy***inae. The non-symmetry is a characteristic of the f -branch statistic (Malinsky, Matschiner et al. 2021). Since the f -branch statistic is not designed to detect introgression between sister groups, we were unable to calculate the f -branch statistic between Modisiminae and Phy***inae, which had an introgression event as indicated by the D and f_4 -ratio statistics (Figure 6.4a). These values suggest moderate levels of gene flow (introgression) between different subfamilies of Pholcidae. We identified 155 introgressed loci. Of these, 22 (14.2%) are exonic, while 102 (65.8%) are located in non-coding regions. In total, 89 genes are affected. Functional enrichment analysis revealed that the introgressed genes are associated with specific molecular functions (3 genes) and biological processes (22 genes) (Figure 6.4c, Table D.4). Three genes are involved in DNA-binding transcription factor activity and transcription regulator activity, while 22 genes are related to transcription regulation, gene expression, and the biosynthesis of various molecules.

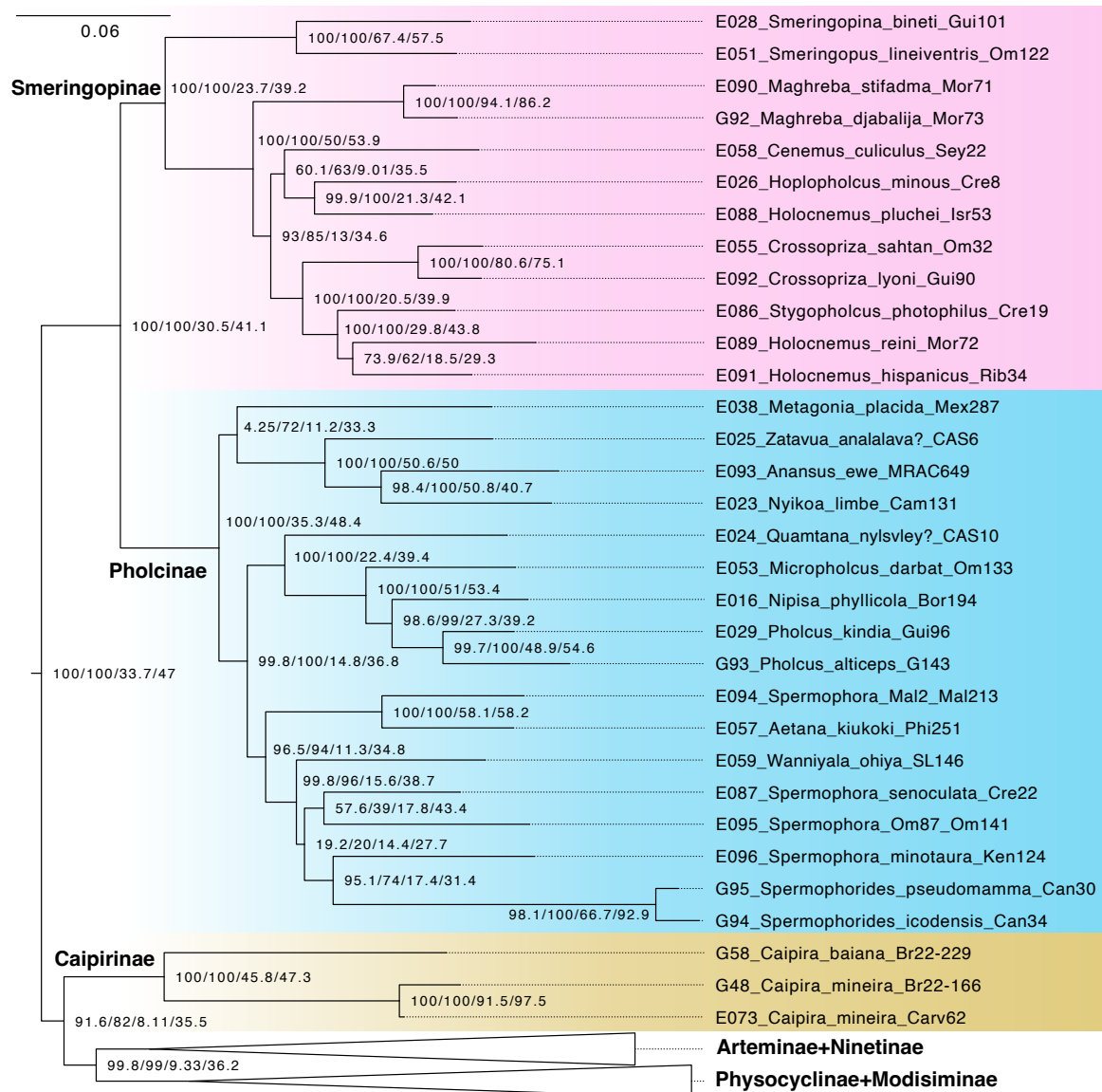
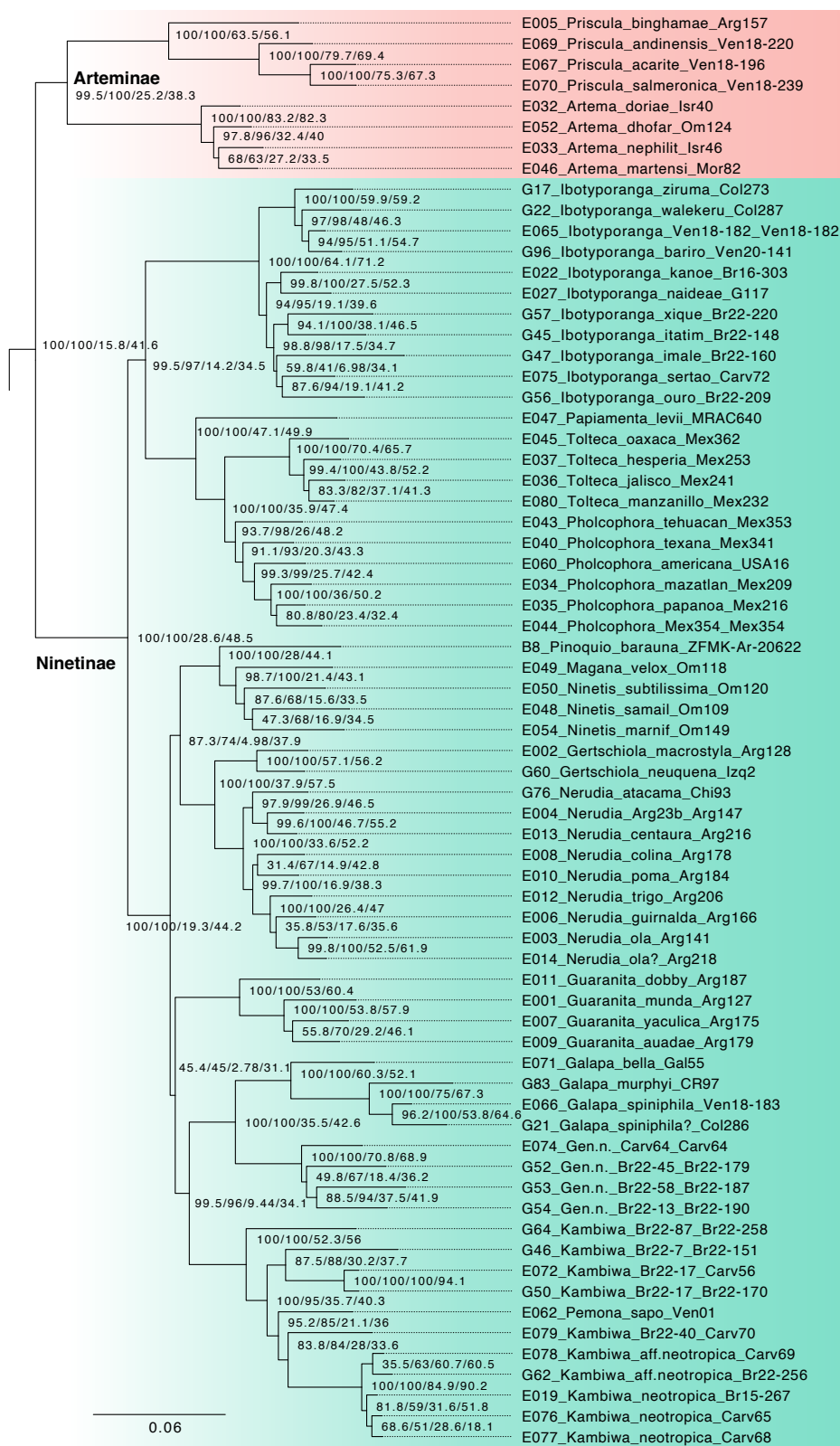


Figure 6.1. Maximum likelihood phylogeny of Pholcidae based on matrix 1 (part 1) . Partitioned concatenation analysis was performed with IQ-TREE and the best-fitting model was determined by ModelFinder (-m MFP+MERGE). The support values shown on the branches are SH-aLRT support (%) / ultrafast bootstrap support (UFBoot) (%) / gene concordance factor (gCF) / site concordance factor (sCF), respectively.



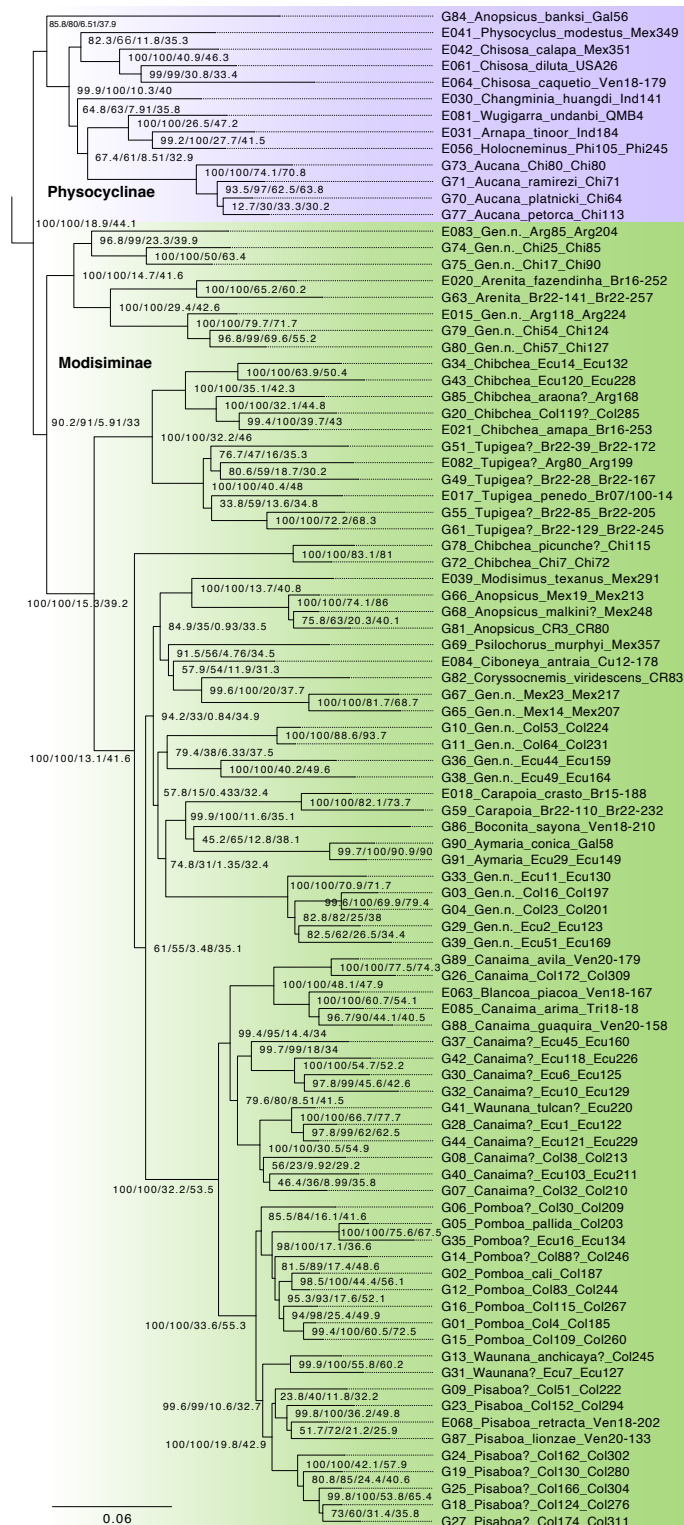


Figure 6.3. Maximum likelihood phylogeny of Pholcidae based on matrix 1 (part 3). This figure is a continuation of Figure 6.1

At the genus level, both D and f_4 -ratio statistics indicated widespread intra-subfamily introgression (except for *Cai***inae* and *Arteminae*) (Figure D.16 and D.17). Additionally, *Ninetinae* and *Modisiminae* exhibited more introgression events between different genera than other subfamilies (Figure D.16 and D.17).

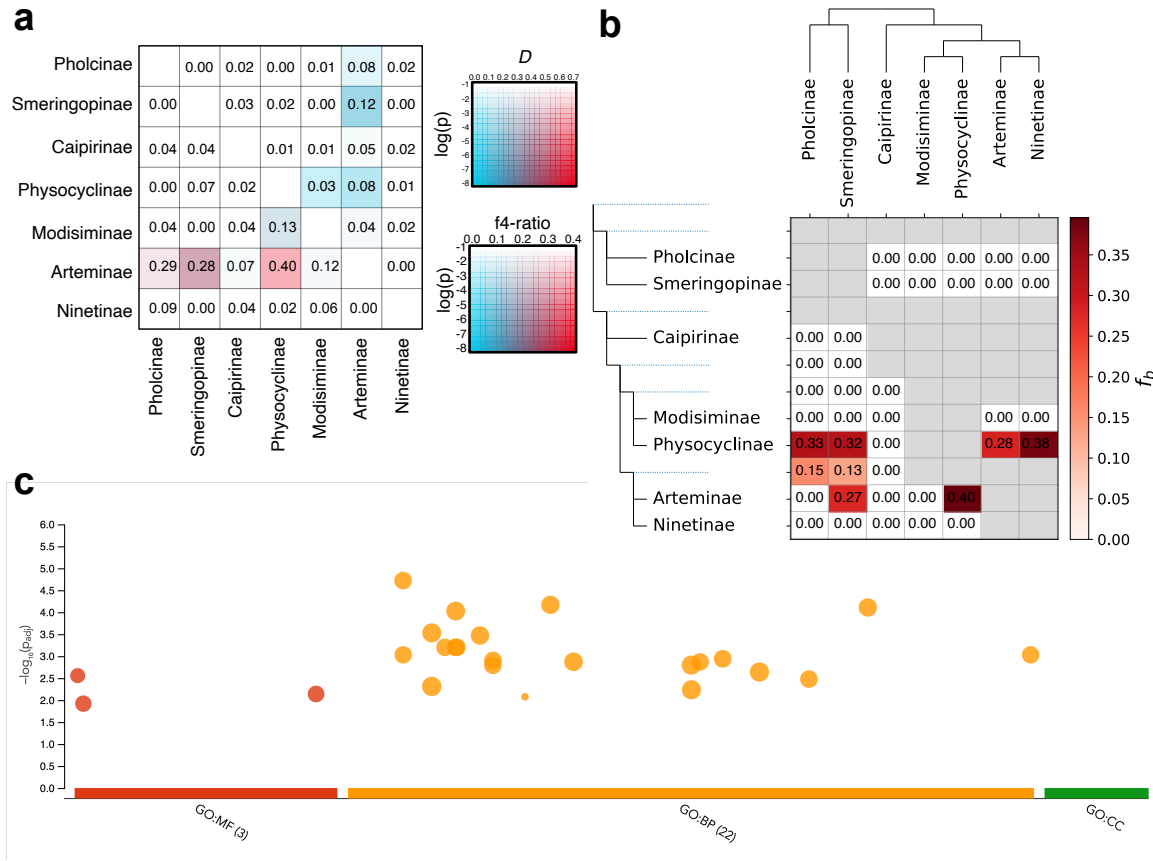


Figure 6.4. Introgression between Pholcidae subfamilies. (a) The upper triangle represents the D statistic, and the lower triangle shows the f_4 -ratio statistic. (b) Distribution of the f_4 -ratio statistic. (c) Significant loci identified in the functional enrichment analysis. The y-axis notes the logarithmically transformed adjusted p-values., while the x-axis indicates the number of significant loci in each GO class. There are 3 significant loci in the molecular function (MF) class, 22 in the biological process (BP) class, and none in the cellular component (CC) class.

Mitochondrial genetic distance. We discovered that the sizes of pholcid mitogenomes are approximately 14.5 Kbp, and the order of protein-coding genes (PCGs) and ribosomal DNA genes (rDNA), which total 15 genes, is highly conserved across species (Figure D.18). However, both MITOS2 and MitoZ were only able to annotate about half of the tRNA genes or fewer, so we excluded these genes from our analysis. Our analysis revealed that PCGs and rDNA genes exhibit significant variability within each subfamily (Figure 6.5; Table D.5). In pholcids, ATP8, 12S, ND3, 16S, ND2, and ND6 are the fastest-evolving genes, with median intra-subfamily K2P genetic distances ranging from 0.49 to 0.73 on average. In contrast, COX1, CYTB, COX2, and COX3 are the slowest-evolving genes in pholcids, with median intra-subfamily K2P genetic distances between 0.22 and 0.34. Genes with intermediate evolutionary rates include ND1, ATP6, ND4L, ND4, and ND5, with median intra-subfamily K2P genetic distances ranging

from 0.40 to 0.46. Compared to PCGs, the two rDNA genes exhibit the highest intra-subfamily K2P genetic distances, exceeding 2.3. Interestingly, in 10 out of 15 genes, Ninetinae is the subfamily with the greatest maximum intra-subfamily genetic distance (Figure 6.5; Table D.5). These genes include 12S, 16S, ATP6, COX1, CYTB, ND3, ND4, ND4L, ND5, and ND6. In contrast, likely due to its small sample size (two species and three specimens), Cai***inae has the smallest intra-subfamily genetic distance for all 13 PCGs. Among the 15 genes, COX1 exhibits the smallest inter-subfamily genetic distance, with median values ranging from 0.22 to 0.26 (Figure D.19). But ATP8 shows the highest values, ranging from 0.71 to 0.95 (Figure D.19). Surprisingly, the inter-subfamily genetic distances of ATP8 within Pholcidae can exceed those observed between Pholcidae and Sicariidae. For instance, the median K2P distance between Cai***inae and Sicariidae is 0.60, while the K2P distance between Cai***inae and Phy***inae is 0.81 (Figure D.19).

Selection pressure on mitochondrial PCGs. The Ka/Ks values for most species are less than 1, indicating strong negative (purifying) selection on these genes (Figure 6.6). Among them, COX1, COX2, COX3, and CYTB experience the strongest negative selection pressure, with Ka/Ks values mostly around 0.06, and as low as 0.015–0.020 for COX1. ATP6 and ND1 are under moderate negative selection, with Ka/Ks values around 0.10. In contrast, ATP8, ND2, ND3, ND4, ND4L, ND5, and ND6 exhibit Ka/Ks values ranging from 0.15 to 0.6. Notably, in Ninetinae, ATP8 shows evidence of positive selection, as two species (*Galapa spiniphila*; *Ibotyporanga ziruma*) have Ka/Ks values of 1.1 and 1.2, respectively (Figure 6.6).

GC content distribution of mitochondrial PCGs. The protein coding genes show variable GC contents spanning from 12% to 49% (N=2088, average 31.9%, standard deviation 6.7%). 94.7% percent of the genes have GC content between 20% to 45% (Figure 6.7 and D.20). Notably, Ninetinae has higher mean GC content (36.2% to 40.7%) than the remained pholcids for all 13 PCGs and greatest mean GC content (38%) among the pholcid subfamilies. Modisiminae is the opposite with the lowest mean GC content (27.6%) among the pholcid subfamilies. Except for ATP8 and ND5, Modisiminae also shows the lowest mean GC content (22.6% to 32.9%) among the pholcid subfamilies for PCGs. Smeringopinae shows the lowest GC content for ATP8 (19.9%) and Cai***inae shows the lowest mean GC content for ND5 (25.2%) (Figure 6.7).

Codon usage bias in mitochondrial PCGs. Figure 6.8 shows the RSCU distribution across PCGs and subfamilies. While some codons are frequently used by different genes and subfamilies (e.g., TTT, TTA, TCT), others encoding the same amino acids are rarely used (e.g., TTC; CTT, CTG, CTC; TCG, TCC) (Figure 6.8). Among the top five codons with the highest RSCU values in a gene and subfamily, COX1 and CYTB are the most “conserved,” each including seven codons, three of which are shared by all 7 pholcid subfamilies (Figure D.21). In contrast, the ND3 gene includes 20 codons, suggesting that different subfamilies prefer distinct codons for encoding the same amino acids (or different amino acid composition) (Figure D.21).

In the Nc plot (Figure 6.9), most genes across all subfamilies fall below the null (no selection) curve, with GC3s values below 0.5. This suggests that these genes have either undergone G+C-biased mutation pressure or that natural selection favors codons ending in C and/or G.

In the neutrality plot (GC12 vs GC3) (Figure 6.10), the coefficients range from 0.1467 (Smeringopinae), to 0.1585 (Ninetinae), 0.1816 (Phy***inae), 0.2241 (Pholcinae), 0.2994 (Cai***inae), 0.3741 (Modisiminae), and 0.3846 (Arteminae). The p-values for most subfamilies are $<< 0.001$, while they are 0.0026 and 0.0047 for Smeringopinae and Cai***inae, respectively. The R^2 for the regression line for most subfamilies are $0.13 \leq R^2 \leq 0.28$, while it is 0.07 for Smeringopinae. The smallest coefficients of Smeringopinae and Ninetinae indicate they are under low levels of directional mutation pressure and un-

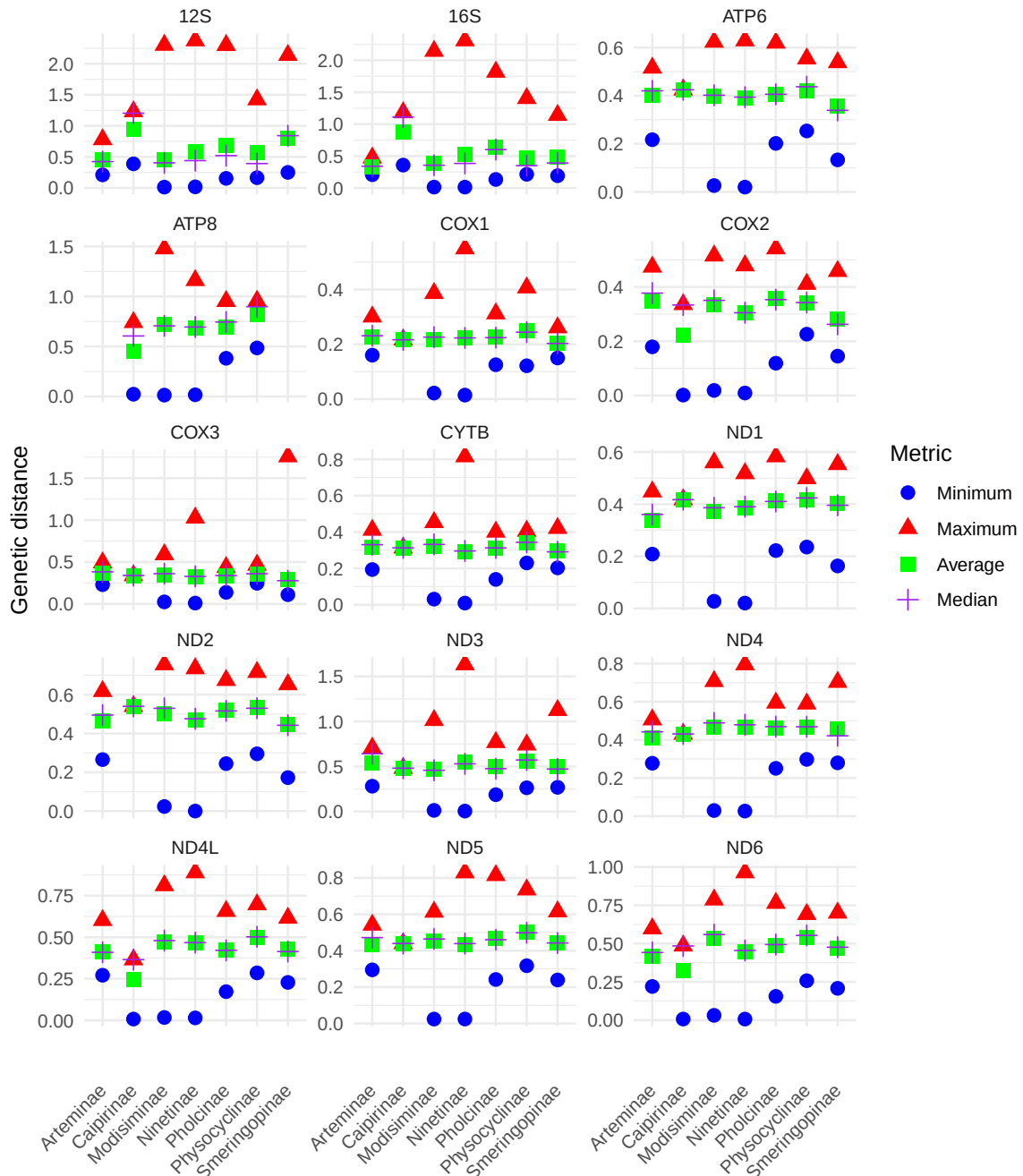


Figure 6.5. Intra-subfamily genetic distance between mitochondrial genes of Pholcidae. The analysis was conducted using the Kimura's 2-parameter (K2P) model, with a pairwise deletion strategy to eliminate gaps during calculations. The `dist.dna()` function from the `ape` package was employed for the analysis.

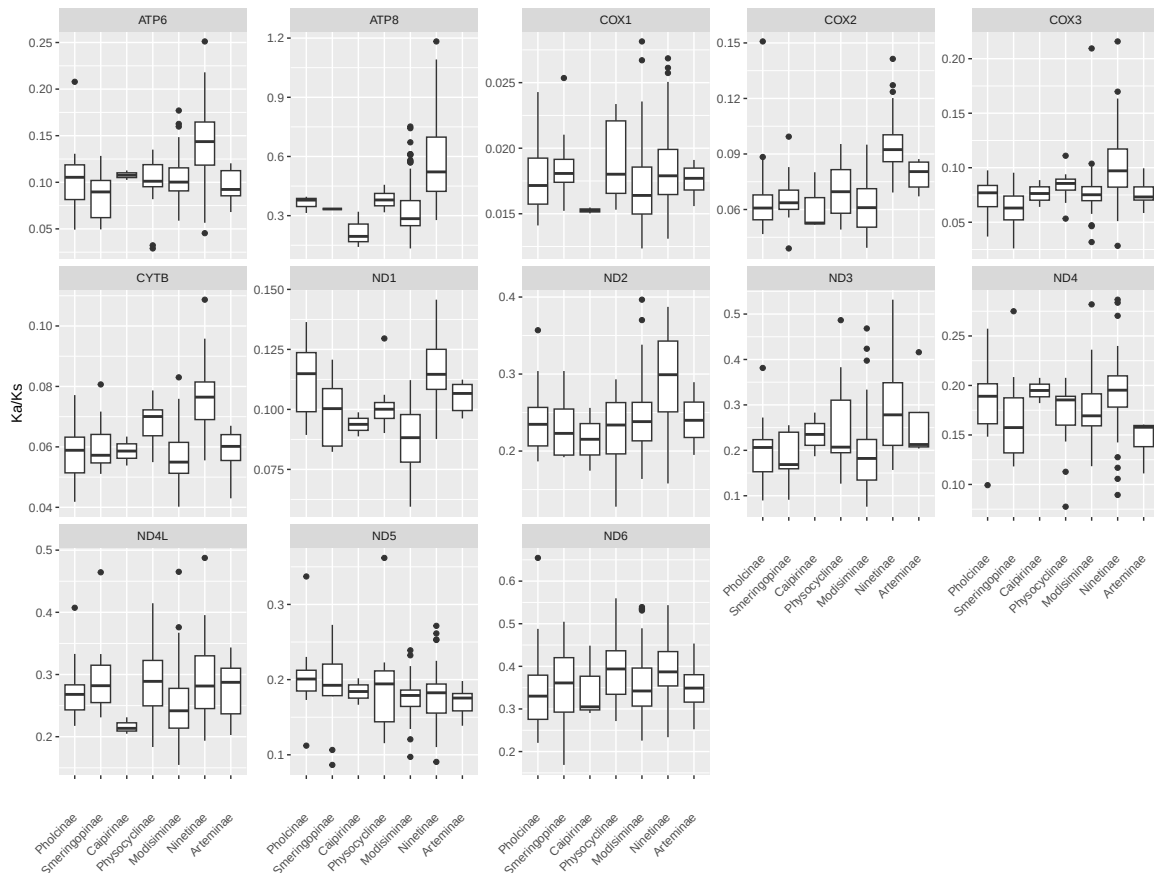


Figure 6.6. The Ka/Ks distribution of Pholcidae mitogenomes. The KaKs_Calculator and the model averaging method (-m MA) were employed for the calculation.

der stronger selective constraints while Modisiminae and Arteminae show the opposite. The coefficient for all pholcids is 0.35 ($R^2 = 0.44$, p -value < 0.001) (Figure 6.10), suggesting the evolution of codons in the mitochondrial PCGs of pholcids are under modest selective constraints.

6.5 Discussion

6.5.1 Reconciling phylogeny with introgression: challenges and considerations

With over 2,000 described species distributed among seven subfamilies, Pholcidae represents an excellent group for studying phylogeny and evolution. Their wide range of morphological characteristics (e.g., body size and leg length), alongside their diverse behaviors and global distribution across various microhabitats, creates a rich context for examining evolutionary relationships. Notably, Pholcidae exhibit their highest diversity in tropical regions, where multiple species can coexist in close vicinity (Castanheira et al. 2016; Huber and Rheims 2011). This spatial overlap increases the likelihood of hybridization and introgression, complicating the understanding of phylogenetic relationships. Indeed, the phylogenetic relationships among different subfamilies of Pholcidae spiders have long been debated, with various studies suggesting distinct topologies. Two recent studies addressed this issue using large-scale genomic

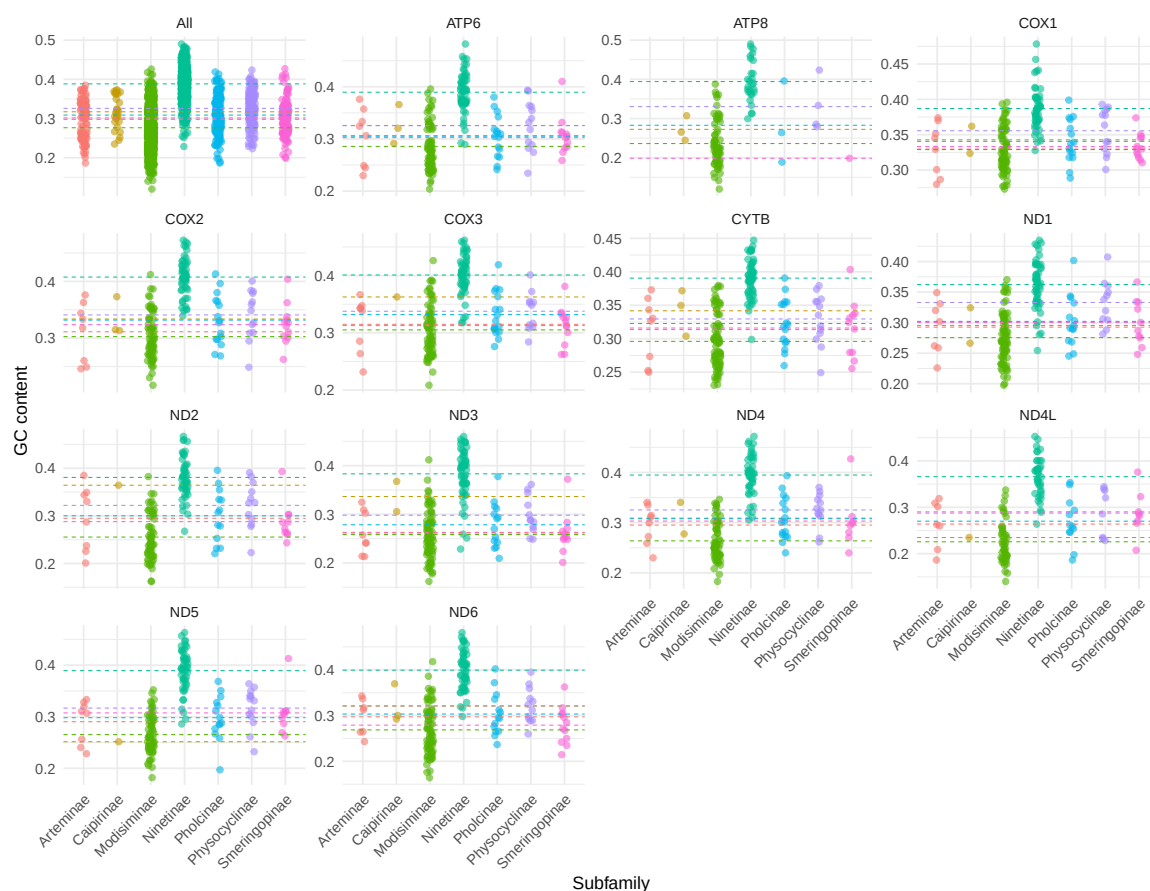


Figure 6.7. The GC content distribution of Pholcidae mitochondrial PCGs. The “All” category shows the overall GC content of all 13 PCGs. The CAT program was employed for the analysis. The horizontal dashed lines denote mean values.

sampling of ultra-conserved elements (UCEs) (Meng, Podsiadlowski, Dimitrov et al. 2025; Meng, Podsiadlowski and Huber 2025b). However, it remains unclear why different lines of evidence—morphology, genetic markers, and UCEs—often contradict each other. For instance, a study based on six genetic markers placed Ninetinae as sister to all other pholcids (Dimitrov et al. 2013). This finding is also supported by the presence of plesiomorphic synspermia in most Ninetinae, whereas all other studied pholcids have derived cleistospermia (Dederichs et al. 2022). However, in Eberle et al. (2018), which also employed six genetic markers, Ninetinae was placed as sister to the genus *Artema*, and both together were sister to the remaining pholcids. Similar inconsistencies have been reported with respect to the positions of the isolated Andean genus *Priscula* (reviewed in Huber, Eberle et al. 2018).

Phylogenetic inference can be influenced by various factors, including the number of loci and sampled taxa, the source of loci (e.g., mitochondrial or nuclear, coding or non-coding regions), incomplete lineage sorting, gene duplication and loss, as well as technical artifacts such as model assumption violations and low phylogenetic signal. In recent years, introgression has increasingly been recognized as a significant factor contributing to incongruent phylogenetic inferences, such as the discordance observed



Figure 6.8. The RSCU distribution of Pholcidae mitochondrial PCGs . The CodonW program was used for the calculation. The labels of the x-axis denotes the abbreviation of amino acids and codons, respectively.

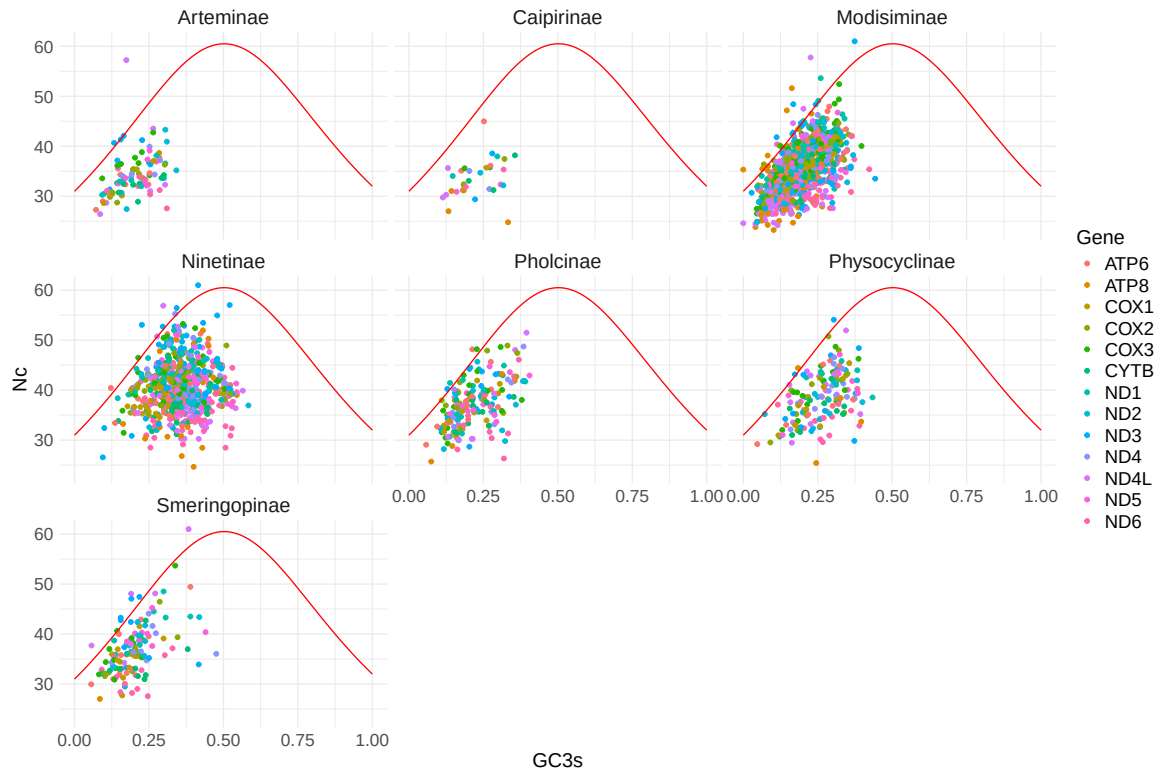


Figure 6.9. The Enc plot for Pholcidae mitochondrial PCGs . The y-axis represents the effective number of codons, while the x-axis shows the G+C content of the third codon positions. Calculations were performed using the CodonW program.

between gene trees and species trees. In this study, we explored the potential contribution of introgression to phylogenetic inference by examining 238 UCE loci of 193 pholcid species and their mitogenomes. We discovered that mitochondrial data suggest topologies similar to those of previous studies using (mitochondrial) genetic markers (Figure D.12 and D.13), while our recovered UCE phylogeny aligns well with those of Meng, Podsiadlowski, Dimitrov et al. 2025; Meng, Podsiadlowski and Huber 2025b at the subfamily level (Figure 6.1, 6.2 and 6.3, D.7). Unexpectedly, we found evidence of intra-locus recombination and multiple inter-subfamily (Figure 6.4) and prevalent intra-subfamily (Figure D.16 and D.17) introgressions. At the subfamily level, all three statistics, the D , f_4 -ratio, f_{dM} confirmed the presence of introgressions between Arteminae and Pholcinae/Smeringopinae/Cai***inae/Phy***inae/Modisiminae, and between Phy***inae and Modisiminae (Figure 6.4a and D.14). This is further corroborated by the higher percentage of quartets supporting their unification in the introgressed loci compared to the non-introgressed loci (Figure D.15). The expected minimum value of scf is 33.3%. While the low value of gcf may be attributed to ILS, the low values of both gcf and scf in the UCE phylogeny (e.g. Modisiminae, and the common ancestor of all subfamilies except Pholcinae and Smeringopinae) suggest the existence of other processes, such as introgression (Figure 6.1, 6.2 and 6.3). For example, the unification of Ninetinae with Pholcinae reported by Bruvo-Madaric et al. (2005) could be explained by the presence of a weak introgression signal between *Ninetis* and *Pholcus*. In a similar way, the fact that Ninetinae shows

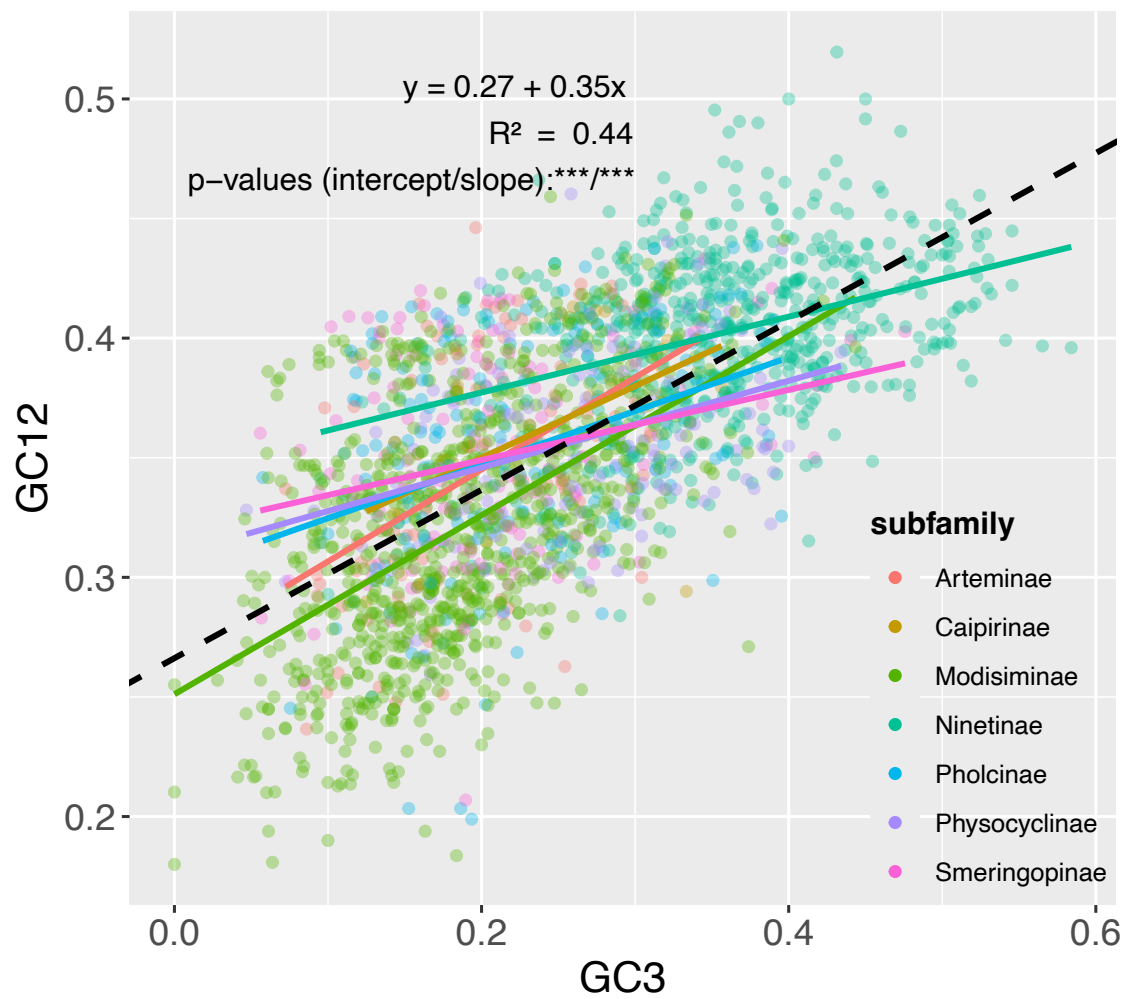


Figure 6.10. The neutrality plot for Pholcidae mitochondrial PCGs. The notation GC12 represents the GC content at the first and second codon positions, while GC3 indicates the GC content at the third codon position. The formula denotes the regression line (black dashed line) for all pholcid species. *** indicates a significance level of $p - \text{value} < 0.001$.

different degrees of introgression to certain genera of other subfamilies may explain why Ninetinae was recovered as a sister group to the remaining pholcids in Dimitrov et al. (2013) (Figure D.16 and D.17). Finally, the genus *Priscula* shows weak to strong introgression signals to Smeringopinae (e.g., *Hoplopholcus*), Pholcinae (e.g., *Spermophorides*), Cai***inae (GENUS-1), Ninetinae (e.g., *Ibotyporanga* and *Kambiwa*), Phy***inae (e.g., *Chisosa*, *Aucana*) and Modisiminae (e.g., *Arenita*, *Aymaria* and *Blancoa*) (Figure D.16 and D.17), which may explain its unstable placements in previous studies (Astrin et al. 2007; Eberle et al. 2018; Huber, Eberle et al. 2018) and our mitochondrial phylogeny (Figure D.12 and D.13). However, this example also highlights the difficulty of explaining some of our results: *Priscula* is an Andean genus, while *Hoplopholcus* is restricted to Turkey and neighboring regions. Numerical values that are here used to suggest introgression may partly result from other causes or be artifacts. For instance, the precision of our inference regarding introgression may be compromised by the relatively small number of variants (SNPs) included in our dataset (18,668 bp). This is particularly true for the f -branch statistic, whose accuracy significantly diminishes when the number of variants in the dataset is limited to less than 10,000 (Malinsky, Matschiner et al. 2021). Alternatively, horizontal gene transfer, such as that mediated by transposons from the spiders' prey (insects), may contribute to the observed introgression signatures across different subfamilies. Additionally, it is possible that UCEs are not the most suitable dataset for detecting introgression due to their high conservation and potential relaxed conservation in certain lineages, which could result in plesiomorphic traits in more distantly related subfamilies. On the other hand, ancient introgression signatures can be lost or fixed in some of the descendent lineages due to selection (Suvorov et al. 2022), genetic drift and new mutations over time. Indeed, ancient introgressions between higher taxonomic ranks (e.g., suborders, superfamilies) have been observed in Odonata (Insecta; Suvorov et al. 2022) and Gesneriaceae (Angiosperms; L. Yang et al. 2023).

While there are a few studies attempting to detect gene flow among spiders, to our knowledge, almost all of them focus on specific populations of one or a few species (Garb et al. 2009; Ivanov et al. 2018; Johannesen et al. 2001; Lattimore et al. 2011; Salgado-Roa et al. 2018). On the other hand, recent studies aiming to resolve the phylogenetic relationships at higher ranks of spiders (e.g., family) have not tested the impact of introgression on phylogenetic inference (e.g., Fernández et al. 2018; Ramírez et al. 2021). Gene flow can lead to gene tree-species tree incongruence and result in phylogenetic networks rather than phylogenetic trees. One reason for the lack of testing may be unawareness, while another reason may be methodological problems. For instance, the D and f_4 -ratio statistics assume (1) there are sufficient variants among species inherited from a common ancestor and the existence of ILS; (2) a negligible number of recurrent mutations and back mutations; and (3) similar substitution rates across different species (Patterson et al. 2012; Pease et al. 2015). For species with deep divergence times, the latter two assumptions might be violated. Another approach is to infer phylogenetic networks directly from genomic data or gene trees. This is often implemented using the multispecies network coalescent (MSNC) model within either a maximum likelihood framework or a Bayesian framework. Examples include PhyloNet (Than et al. 2008), the SpeciesNetwork package (C. Zhang, Ogilvie et al. 2018) of BEAST2, and BPP (Flouri et al. 2020; Z. Yang 2015). These methods allow for the modeling of incomplete lineage sorting (ILS) and reticulation events simultaneously. However, they are not scalable to large datasets with many loci and species. The presence of introgression in the Pholcidae family suggested by our study indicates that it is crucial to consider introgression when conducting phylogenetic analyses, particularly for spiders. Indeed, several researchers have recently emphasized the importance of accounting for gene flow in cases of phylogenetic incongruence (Bjornson et al. 2024; Komarova et al. 2022).

Additionally, UCEs have long been considered to originate from non-coding regions of genomes. In vertebrates, UCEs primarily originate from non-coding regions, where they regulate gene expression (e.g.,

through enhancers) (Baira et al. 2008; Mclean et al. 2019). Our UCEs were captured using the spider-specific probes developed by Kulkarni et al. (2020). The study by Hedin et al. (2019) revealed that most captured arachnid UCEs are exonic. However, our analysis showed that 65.8% (102 out of 155) of the introgressed loci are located in non-coding regions, which may partly explain some observed differences between the UCE-based phylogeny (Kulkarni et al. 2020) and the transcriptome-based phylogeny (Fernández et al. 2018); that is, different datasets may lead to different phylogenetic relationships to certain degrees. However, it's possible that our captured UCEs are biased to non-coding regions because the probes developed by Kulkarni et al. (2020) target approximately 2,000 loci, while we had only 238 loci in matrix 1. Non-coding regions contain cis-regulatory elements (e.g., promoters, enhancers, and silencers) and trans-regulatory elements (Hill et al. 2021), which play an important role in adaptation and speciation (Louder et al. 2024; Mack et al. 2017; Romero et al. 2012). For instance, the UCEs of *Drosophila* (Insecta) have been found to be crucial hubs for gene regulatory functions, where intronic and intergenic UCEs are primarily associated with transcriptional and epigenetic regulation, while exonic UCEs are more likely to contain protein-coding sequences, TF-binding sites, splice sites, and RNA editing sites (Warnefors et al. 2016). Our identified introgressed loci are functionally enriched in transcription, gene expression and various biosynthetic processes (Figure 6.4c, Table D.4), which may have contributed to the diversification of pholcids and their adaptations to various environments.

6.5.2 The need to improve mitochondrial gene annotation in spiders

Arachnida represent an important component of arthropods. However, compared to insects, they are less studied. For instance, there are only 2,220 Arachnida mitogenomes longer than 10 Kbp on NCBI, and only 545 (24.5%) of them belong to spiders (Araneae). In contrast, there are 29,630 insect mitogenomes longer than 10 Kbp available on NCBI (access date: March 16, 2025). The limited number of mitogenomes available has restricted our ability to study their genomic diversity, the evolution of gene arrangements, and their functional significance. Furthermore, it hampers the use of mitogenomes to study species phylogeny, gene flow, and interactions with nuclear genes.

In this study, we assembled 193 Pholcidae mitogenomes, with 98 of them being complete or nearly complete. This represents an important contribution to related research. However, we found that current mainstream methods (e.g., MITOS2, MitoZ) fail to annotate many of the mitochondrial genes of Pholcidae, including tRNA genes as well as some protein-coding genes (PCGs) such as ATP8, ND4, ND4L, and ND5. This incapability arises from the fact that most programs adopt a homology-search method, in which the genes of closely related species are used as references to delimit the gene boundaries of new species. For instance, BLAST or HMMER (Eddy 2011; <http://hmmmer.org/>) is used to identify the rough regions of a gene, and then the exact gene boundaries are fine-tuned manually or by specific algorithms. Compared to BLAST, HMMER is more sensitive due to its use of profile hidden Markov models (pHMMs) and is thus more widely used for the annotation of mitochondrial PCGs. However, this method fails if the reference species and target species are too distant, or if the genes are too divergent due to a high evolutionary rate. Therefore, for the missing PCGs of pholcid species, we manually annotated these genes in some species with the help of HMMER, and then used these annotated genes as references to annotate the remaining species. However, the annotation of tRNA genes poses more difficulty due to their high divergence at the nucleotide level, as seen in M. Li et al. (2022). To assist with the annotation, secondary structure information of tRNA genes is used and modelled by covariance models (CMs). Related tools include tRNAscan-SE 2.0 (Chan et al. 2021) and MiTFi (Jühling et al. 2012). However, tRNAscan-SE 2.0 was only trained on mammal and vertebrate mitochondrial tRNAs (mt-tRNAs) (Chan et al. 2021),

while the CMs of MiTFi were constructed based on a very limited number of species and lack updates. MITOS2, which adopts MiTFi internally, has also not expanded its CMs due to a lack of data. Thus, there is a need to incorporate these newly sequenced spider mitogenomes, curate the boundaries of PCGs and tRNA genes, and develop more sensitive methods for mitochondrial gene annotation. Since gene rearrangements frequently occur around tRNA genes in arachnids (as evidenced by Ban et al. 2022; Tyagi et al. 2020), a comprehensive annotation of tRNA genes will undoubtedly offer valuable insights into the evolution of spider mitogenomes.

6.5.3 Rapid evolution of mitogenomes in Pholcidae

While we observed a highly conserved PCG order across subfamilies of Pholcidae (Figure D.18), we found significant genetic variation within and between subfamilies. For instance, the median intra-subfamily K2P genetic distances for the ATP8, 12S, ND3, 16S, ND2, and ND6 genes range from 0.49 to 0.73 (Figure 6.5; Table D.5). The second-fastest evolving genes are ND1, ATP6, ND4L, ND4, and ND5, with their intra-subfamily K2P genetic distances ranging from 0.40 to 0.46. In contrast, COX1, CYTB, COX2, and COX3 exhibit the lowest evolutionary rates, with intra-subfamily K2P genetic distances ranging from 0.22 to 0.34 (Figure 6.5; Table D.5). Importantly, inter-subfamily genetic distances can be equal to or even smaller than intra-subfamily distances; for example, while COX1 is the slowest evolving gene, its maximum intra-subfamily genetic distance can range from 0.3 to over 0.5 (Figure 6.5), whereas the median inter-subfamily genetic distance ranges from 0.22 to 0.28 (Figure D.19). This suggests that it is challenging to assign unknown samples to a subfamily (or even genus, see below text) based solely on mitochondrial genes. Compared to vertebrates and insects, which typically display maximum intra-specific genetic distances around 0.02-0.03 (e.g., Hebert et al. 2004; Huemer et al. 2014; H. Zhang et al. 2022), the values for Pholcidae are generally around 0.07 to 0.1 (e.g., Huber, Meng, Dupérré et al. 2023; Huber, Meng, García et al. 2024), but can occasionally reach up to 0.15-0.17 (Huber and Meng 2025; Huber, Meng, Král, Ávila Herrera et al. 2023; Huber, Meng and Valdez-Mondragn 2024). More surprisingly, intra-genus genetic distances can exceed 0.27 (Huber, Meng, Král, Herrera et al. 2024) (Huber, Meng, Clark et al. 2023), while relatively small intra-genus distances have also been noted, such as 0.043 to 0.061 among species of *Mecolaesthus* (Huber, Meng, Clark et al. 2023). The high substitution rates have led to overlappings between intra- and interspecific genetic distances for some pholcids like *Belisana* (Huber, Meng, Král, Herrera et al. 2024). Rapid evolution can lead to saturation of substitutions and long branch attraction (LBA) (Ontano et al. 2021); therefore, it is highly likely that the high substitution rate is responsible for the non-monophyly of all subfamilies except for Ninetinae in our mitochondrial genome phylogenies (Figure D.12 and D.13).

It has been proposed that the elevated mutation rate in animal mitogenomes results from decreased efficiency of DNA repair mechanisms, a more mutagenic local environment probably caused by oxidative radicals, and an increased number of replications occurring during each cell division (Lynch et al. 2007). Compared to other subfamilies, Ninetinae is unique for its smaller body size and its preference for ground microhabitats in dry and often hot regions. Smaller-bodied spiders typically undergo fewer molts (Head 1995), resulting in shorter generations and higher rates of cell division. This, combined with adaptations to extreme environments, may explain why Ninetinae often exhibits the highest maximum intra-subfamily genetic variation (Figure 6.5). Indeed, Ninetinae mitogenomes also show the highest GC content among pholcid subfamilies (Figure 6.7). Chemically, G:C pairs have three hydrogen bonds while A:T pairs have two (Yakovchuk et al. 2006). Therefore, G:C pairs have higher thermal stability (Yakovchuk et al. 2006), which are important to maintain structural stability of DNA and RNA transcripts (Biro 2008; Šmarda

and Bureš 2012), especially for Ninetinae which live in dry regions. Thus, GC-rich DNA may provide a benefit in terms of cell survival during freezing and desiccation events (Šmarda et al. 2014). On the other hand, G:C pairs show higher mutability due to frequent cytosine methylation (Coulondre et al. 1978; Ossowski et al. 2010; Pfeifer 2006), which explains the high intra-subfamily genetic variation in Ninetinae. Additionally, genera such as *Chisosa* (Huber, Meng and Valdez-Mondragón 2024), *Nerudia* (Huber, Meng, Král, Ávila Herrera et al. 2023) and *Ninetis* (Huber and Meng 2025) exhibit significant intra-genus genetic variation, while certain species like *Ibotyporanga* (Huber, Meng, Král, Herrera et al. 2024) and *Belisana* (Huber, Meng, Clark et al. 2023) display substantial intra-specific genetic variation. All of these genera and species are characterized by relatively small body sizes compared to their counterparts. However, this observation may also be caused data bias.

6.5.4 Selection pressure drives codon usage bias in Pholcidae

Our analysis clearly revealed that all Pholcidae subfamilies and mitochondrial PCGs exhibit codon usage bias (see Figures 6.8 and D.21). For example, while almost all PCGs show a preference for the codon TTA to encode the amino acid leucine (L), four genes (ND1, ND4, ND4L, and ND5) also demonstrate a similar degree of preference for the synonymous codon CTA. Additionally, different subfamilies exhibit varying degrees of preference for synonymous codons that encode the same amino acid, as exemplified by the codon TCA in ND4, ND4L, and ND5, which encodes serine (S) (Figure 6.8). G+C-biased mutation and/or natural selection are responsible for codon bias, as indicated by the findings that the ENC values of most protein-coding genes (PCGs) fall below the reference (neutral) curve (Figure 6.9). Our neutrality analysis, which yielded a coefficient of 0.35 ($R^2 = 0.44$) (Figure 6.10), demonstrated that selection pressure, rather than directional mutation pressure, is predominantly responsible for the codon usage bias in the mitogenomes of Pholcidae. At the subfamily level, Smeringopinae and Ninetinae experience the strongest selection pressures, with coefficients of 0.1467 and 0.1585, respectively. In contrast, Modisiminae and Arteminae exhibit the highest levels of directional mutation pressure and the lowest levels of natural selection pressure for codon usage among pholcids, with coefficients of 0.3741 and 0.3846, respectively.

6.6 Conclusions

Introgression occurs in Pholcidae spiders and likely contributes to the incongruences observed in phylogenetic inferences in previous studies. Therefore, introgression analysis should be a routine component of phylogenomic research. Additionally, greater efforts are needed to improve the mitochondrial annotation of under-studied organisms including spiders. The rapid evolution of mitogenomes can lead to increased intra-specific genetic distances among small-bodied Pholcidae spiders, as well as mito-nuclear phylogenetic incongruence. Finally, selection pressure is the primary factor driving codon usage bias in the mitogenomes of Pholcidae.

6.7 Supplementary data

Supplementary figures and tables are available at Appendix D.

References

- Agnarsson, I., J. A. Coddington, M. Kuntner and D. Penny (2013), *Spider research in the 21st century: trends and perspectives*, Systematics: progress in the study of spider diversity and evolution 320.
- Aguillon, S. M., T. O. Dodge, G. A. Preising and M. Schumer (2022), *Introgression*, Curr Biol **32** R865.
- Astrin, J. J., B. Misof and B. A. Huber (2007), *The pitfalls of exaggeration: molecular and morphological evidence suggests Kaliana is a synonym of Mesabolivar (Araneae: Pholcidae)*, Zootaxa **1646** 17.
- Ávila Herrera, I. M., J. Král, M. Pastuchová et al. (2021), *Evolutionary pattern of karyotypes and meiosis in pholcid spiders (Araneae: Pholcidae): implications for reconstructing chromosome evolution of araneomorph spiders*, BMC Ecol Evol **21** 75.
- Baira, E., J. Greshock, G. Coukos and L. Zhang (2008), *Ultraconserved elements: genomics, function and disease*, RNA biology **5** 132.
- Ban, X.-C., Z.-K. Shao, L.-J. Wu, J.-T. Sun and X.-F. Xue (2022), *Highly diversified mitochondrial genomes provide new evidence for interordinal relationships in the Arachnida*, Cladistics **38** 452.
- Biot-Pelletier, D., S. Bettinazzi, I. Gagnon-Arsenault et al. (2023), *Evolutionary trajectories are contingent on mitonuclear interactions*, Mol Biol Evol **40** msad061.
- Bjornson, S., H. Verbruggen, N. S. Upham and J. L. Steenwyk (2024), *Reticulate evolution: Detection and utility in the phylogenomics era*, Mol Phylogenet Evol 108197.
- Boni, M. F., D. Posada and M. W. Feldman (2007), *An exact nonparametric method for inferring mosaic structure in sequence triplets*, Genetics **176** 1035.
- Bruvo-Madaric, B., B. A. Huber, A. Steinacher and G. Pass (2005), *Phylogeny of pholcid spiders (Araneae: Pholcidae): combined analysis using morphology and molecules*, Mol Phylogenet Evol **37** 661.
- Camacho, C., G. Coulouris, V. Avagyan et al. (2009), *BLAST+: architecture and applications*, BMC Bioinformatics **10** 421.
- Capella-Gutierrez, S., J. M. Silla-Martinez and T. Gabaldon (2009), *trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses*, Bioinformatics **25** 1972.
- Castanheira, P., A. Pérez-González and R. L. C. Baptista (2016), *Spider diversity (Arachnida: Araneae) in Atlantic Forest areas at Pedra Branca State Park, Rio de Janeiro, Brazil*, Biodiversity Data Journal e7055.
- Castresana, J. (2000), *Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis*, Mol Biol Evol **17** 540.
- Chan, P. P., B. Y. Lin, A. J. Mak and T. M. Lowe (2021), *tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes*, Nucleic Acids Res **49** 9077.
- Chifman, J. and L. Kubatko (2014), *Quartet inference from SNP data under the coalescent model*, Bioinformatics **30** 3317.

- Connolly, S. J., K. M. Curtis, C. J. Vink and C. J. Painting (2025), *Mating behaviour influences the direction and geographic extent of introgression in New Zealand fishing spiders (Dolomedes)*, J Evol Biol **38** 291.
- Coulondre, C., J. H. Miller, P. J. Farabaugh and W. Gilbert (1978), *Molecular basis of base substitution hotspots in Escherichia coli*, Nature **274** 775.
- Dederichs, T. M., B. A. Huber and P. Michalik (2022), *Evolutionary morphology of sperm in pholcid spiders (Pholcidae, Synspermiata)*, BMC Zoology **7** 52.
- Dierckxsens, N., P. Mardulyn and G. Smits (2017), *NOVOPlasty: de novo assembly of organelle genomes from whole genome data*, Nucleic Acids Res **45** e18.
- Dimitrov, D., J. J. Astrin and B. A. Huber (2013), *Pholcid spider molecular systematics revisited, with new insights into the biogeography and the evolution of the group*, Cladistics **29** 132.
- Donath, A., F. Jühling, M. Al-Arab et al. (2019), *Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes*, Nucleic Acids Res **47** 10543.
- Du, K., J. M. B. Ricci, Y. Lu et al. (2024), *Phylogenomic analyses of all species of swordtail fishes (genus Xiphophorus) show that hybridization preceded speciation*, Nat Commun **15** 6609.
- Duchene, D. A., C. Mather Niklas and Van Der Wal and S. Y. W. Ho (2022), *Excluding Loci With Substitution Saturation Improves Inferences From Phylogenomic Data*, Syst Biol **71** 676.
- Durand, E. Y., N. Patterson, D. Reich and M. Slatkin (2011), *Testing for ancient admixture between closely related populations*, Mol Biol Evol **28** 2239.
- Eberle, J., D. Dimitrov, A. Valdez-Mondragon and B. A. Huber (2018), *Microhabitat change drives diversification in pholcid spiders*, BMC Evol Biol **18** 141.
- Eddy, S. R. (2011), *Accelerated profile HMM searches*, PLoS Comput Biol **7** e1002195.
- Faircloth, B. C. (2016), *PHYLUCE is a software package for the analysis of conserved genomic loci*, Bioinformatics **32** 786.
- Fernández, R., R. J. Kallal, D. Dimitrov et al. (2018), *Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life*, Curr Biol **28** 1489.
- Flouri, T., X. Jiao, B. Rannala and Z. Yang (2020), *A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis*, Mol Biol Evol **37** 1211.
- Foelix, R. F. (2010), *Biology of spiders*, 3rd, Oxford university press.
- Garb, J. E. and R. G. Gillespie (2009), *Diversity despite dispersal: colonization history and phylogeography of Hawaiian crab spiders inferred from multilocus genetic data*, Mol Ecol **18** 1746.
- Ge, D., Z. Wen, A. Feijó et al. (2023), *Genomic consequences of and demographic response to pervasive hybridization over time in climate-sensitive pikas*, Mol Biol Evol **40** msac274.
- Green, R. E., J. Krause, A. W. Briggs et al. (2010), *A draft sequence of the Neandertal genome*, science **328** 710.
- Greiner, S., P. Lehwark and R. Bock (2019), *OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes*, Nucleic Acids Res **47** W59.
- Guindon, S., J.-F. Dufayard, V. Lefort et al. (2010), *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0*, Syst Biol **59** 307.

-
- Head, G. (1995), *Selection on Fecundity and Variation in the Degree of Sexual Size Dimorphism Among Spider Species (Class Araneae)*, *Evolution* **49** 776.
- Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak and C. M. Francis (2004), *Identification of birds through DNA barcodes*, *PLoS Biol* **2** e312.
- Hedin, M., S. Derkarabetian, A. Alfaro, M. J. Ramírez and J. E. Bond (2019), *Phylogenomic analysis and revised classification of atypoid mygalomorph spiders (Araneae, Mygalomorphae), with notes on arachnid ultraconserved element loci*, *PeerJ* **7** e6864.
- Hedrick, P. W. (2013), *Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation*, *Mol Ecol* **22** 4606.
- Hernandez, A. M. and J. F. Ryan (2021), *Six-state amino acid recoding is not an effective strategy to offset compositional heterogeneity and saturation in phylogenetic analyses*, *Syst Biol* **70** 1200.
- Hill, M. S., P. Vande Zande and P. J. Wittkopp (2021), *Molecular and evolutionary processes generating variation in gene expression*, *Nat Rev Genet* **22** 203.
- Huber, B. A. and A. Chao (2019), *Inferring global species richness from megatranssect data and undetected species estimates*, *Contributions to Zoology* **88** 42.
- Huber, B. A., J. Eberle and D. Dimitrov (2018), *The phylogeny of pholcid spiders: a critical evaluation of relationships suggested by molecular data (Araneae, Pholcidae)*, *Zookeys* 51.
- Huber, B. A. and G. Meng (2025), *Like grains of sand: Ninetis spiders on the Arabian Peninsula (Araneae, Pholcidae)*, *Zootaxa* **5563** 290, doi: [10.11646/zootaxa.5563.1.19](https://doi.org/10.11646/zootaxa.5563.1.19).
- Huber, B. A., G. Meng, H. L. Clark and G. Cazanove (2023), *First blind daddy long-legs spiders from Australia and Réunion (Araneae, Pholcidae)*, *Subterr Biol* **46** 1, doi: [10.3897/subtbiol.46.105798](https://doi.org/10.3897/subtbiol.46.105798).
- Huber, B. A., G. Meng, N. Dupérré, J. Astrin and M. Herrera (2023), *Andean giants: Priscula spiders from Ecuador, with notes on species groups and egg-sac troglomorphy (Araneae: Pholcidae)*, *Eur J Taxon* **909** 1, doi: [10.5852/ejt.2023.909.2351](https://doi.org/10.5852/ejt.2023.909.2351).
- Huber, B. A., G. Meng, J. C. garcía and L. S. Carvalho (2024), *Thriving in dry conditions: on the Neotropical spider genus Galapa (Araneae: Pholcidae)*, *Zootaxa* **5419** 301, doi: [10.11646/zootaxa.5419.3.1](https://doi.org/10.11646/zootaxa.5419.3.1).
- Huber, B. A., G. Meng, J. Král, I. M. Ávila Herrera et al. (2023), *High and dry: integrative taxonomy of the Andean spider genus Nerudia (Araneae: Pholcidae)*, *Zool J Linn Soc* **198** 534, doi: [10.1093/zoolinnean/zlac100](https://doi.org/10.1093/zoolinnean/zlac100).
- Huber, B. A., G. Meng, J. Král, I. M. Á. Herrera and L. S. Carvalho (2024), *Diamonds in the rough: Ibotyporanga (Araneae, Pholcidae) spiders in semi-arid Neotropical environments*, *Eur J Taxon* **963** 1, doi: [10.5852/ejt.2024.963.2687](https://doi.org/10.5852/ejt.2024.963.2687).
- Huber, B. A., G. Meng and A. Valdez-Mondragón (2024), *Notes on Chisosa (Araneae, Pholcidae), with the description of a new species from Mexico.*, *Zootaxa* **5419** 217, doi: [10.11646/zootaxa.5419.2.3](https://doi.org/10.11646/zootaxa.5419.2.3).
- Huber, B. A. and C. A. Rheims (2011), *Diversity and endemism of pholcid spiders in Brazil's Atlantic Forest, with descriptions of four new species of the Atlantic Forest endemic genus Tupigea (Araneae: Pholcidae)*, *J Nat Hist* **45** 275.

- Huemer, P., M. Mutanen, K. M. Sefc and P. D. N. Hebert (2014), *Testing DNA barcode performance in 1000 species of European Lepidoptera: large geographic distances have small genetic impacts*, PLoS One **9** e115774.
- Ivanov, V., K. M. Lee and M. Mutanen (2018), *Mitonuclear discordance in wolf spiders: Genomic evidence for species integrity and introgression*, Mol Ecol **27** 1681.
- Jakovlić, I., H. Zou, T. Ye et al. (2023), *Mitogenomic evolutionary rates in bilateria are influenced by parasitic lifestyle and locomotory capacity*, Nat Commun **14** 6307.
- Johannesen, J. and M. Veith (2001), *Population history of Eresus cinnaberinus (Araneae: Eresidae) colour variants at a putative species transition*, Heredity **87** 114.
- Jühling, F., J. Pütz, M. Bernt et al. (2012), *Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements*, Nucleic Acids Res **40** 2833.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermiin (2017), *ModelFinder: fast model selection for accurate phylogenetic estimates*, Nat Methods **14** 587.
- Katoh, K. and D. M. Standley (2013), *MAFFT multiple sequence alignment software version 7: improvements in performance and usability*, Mol Biol Evol **30** 772.
- Kimura, M. (1980), *A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences*, Journal of molecular evolution **16** 111.
- Kinneberg, V. B., D. S. Lü, D. Peris, M. Ravinet and I. Skrede (2023), *Introgression between highly divergent fungal sister species*, J Evol Biol **36** 1133.
- Kolberg, L., U. Raudvere, I. Kuzmin et al. (2023), *g: Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update)*, Nucleic Acids Res **51** W207.
- Komarova, V. and L. Lavrenchenko (2022), *Approaches to the detection of hybridization events and genetic introgression upon phylogenetic incongruence*, Biology Bulletin Reviews **12** 240.
- Kozlov, A. M., D. Darriba, T. Flouri, B. Morel and A. Stamatakis (2019), *RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference*, Bioinformatics **35** 4453.
- Kulkarni, S., H. Wood, M. Lloyd and G. Hormiga (2020), *Spider-specific probe set for ultraconserved elements offers new perspectives on the evolutionary history of spiders (Arachnida, Araneae)*, Mol Ecol Resour **20** 185.
- Lam, H. M., O. Ratmann and M. F. Boni (2018), *Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm*, Mol Biol Evol **35** 247.
- Larsson, A. (2014), *AliView: a fast and lightweight alignment viewer and editor for large datasets*, Bioinformatics **30** 3276.
- Lartillot, N., N. Rodrigue, D. Stubbs and J. Richer (2013), *PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment*, Syst Biol **62** 611.
- Lattimore, V. L., C. J. Vink, A. M. Paterson and R. H. Cruickshank (2011), *Unidirectional introgression within the genus Dolomedes (Araneae: Pisauridae) in southern New Zealand*, Invertebr Syst **25** 70.
- Li, M., W.-T. Chen, Q.-L. Zhang et al. (2022), *Mitochondrial phylogenomics provides insights into the phylogeny and evolution of spiders (Arthropoda: Araneae)*, Zoological Research **43** 566.
- Louder, M. I., H. Justen, A. A. Kimmitt et al. (2024), *Gene regulation and speciation in a migratory divide between songbirds*, Nat Commun **15** 98.

-
- Lynch, M. and B. Walsh (2007), *The origins of genome architecture*, vol. 98, Sinauer associates Sunderland, MA.
- Ma, Y., J. Wang, Q. Hu et al. (2019), *Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex*, *Commun Biol* **2** 213.
- Mack, K. L. and M. W. Nachman (2017), *Gene regulation and speciation*, *Trends in Genetics* **33** 68.
- Maddison, W. P. (1997), *Gene trees in species trees*, *Syst Biol* **46** 523.
- Malinsky, M., R. J. Challis, A. M. Tyers et al. (2015), *Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake*, *Science* **350** 1493.
- Malinsky, M., M. Matschiner and H. Svardal (2021), *Dsuite-Fast D-statistics and related admixture evidence from VCF files*, *Mol Ecol Resour* **21** 584.
- Malinsky, M., H. Svardal, A. M. Tyers et al. (2018), *Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow*, *Nat Ecol Evol* **2** 1940.
- McLean, B. S., K. C. Bell, J. M. Allen, K. M. Helgen and J. A. Cook (2019), *Impacts of inference method and data set filtering on phylogenomic resolution in a rapid radiation of ground squirrels (Xerinae: Marmotini)*, *Syst Biol* **68** 298.
- Meng, G., Y. Li, C. Yang and S. Liu (2019), *MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization*, *Nucleic Acids Res* **47** e63.
- Meng, G.**, L. Podsiadlowski, D. Dimitrov and B. A. Huber (2025), *The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders*, submitted.
- Meng, G.**, L. Podsiadlowski and B. A. Huber (2025a), *Exploring Phylogenetic Discordance: The Influence of Introgression and Insights into the Evolution of the Pholcidae (Araneae) Mitochondrial Genomes*, in preparation.
- Meng, G.**, L. Podsiadlowski and B. A. Huber (2025b), *Low coverage genome sequencing reveals a new subfamily of daddy long-legs spiders from Brazilian Caatinga (Araneae, Pholcidae)*, in preparation.
- Minh, B. Q., M. W. Hahn and R. Lanfear (2020), *New methods to calculate concordance factors for phylogenomic datasets*, *Mol Biol Evol* **37** 2727.
- Minh, B. Q., M. A. T. Nguyen and A. von Haeseler (2013), *Ultrafast approximation for phylogenetic bootstrap*, *Mol Biol Evol* **30** 1188.
- Minh, B. Q., H. A. Schmidt, O. Chernomor et al. (2020), *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era*, *Mol Biol Evol* **37** 1530.
- Mirarab, S., M. S. Bayzid and T. Warnow (2016), *Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting*, *Syst Biol* **65** 366.
- Morales-Briones, D. F., G. Kadereit, D. T. Tefarikis et al. (2021), *Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae sl*, *Syst Biol* **70** 219.
- Morel, B., T. A. Williams and A. Stamatakis (2023), *Asteroid: a new algorithm to infer species trees from gene trees under high proportions of missing data*, *Bioinformatics* **39** btac832.
- Naser-Khdour, S., B. Q. Minh, W. Zhang, E. A. Stone and R. Lanfear (2019), *The prevalence and impact of model violations in phylogenetic analysis*, *Genome Biol Evol* **11** 3341.

- Ontano, A. Z., G. Gainett, S. Aharon et al. (2021), *Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions*, *Mol Biol Evol* **38** 2446.
- Ortiz, D., S. Pekár and M. Dianat (2024), *A consequential one-night stand: Episodic historical hybridization leads to mitochondrial takeover in sympatric desert ant-eating spiders*, *Mol Phylogenet Evol* **199** 108167.
- Ossowski, S., K. Schneeberger, J. I. Lucas-Lledó et al. (2010), *The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana*, *science* **327** 92.
- Oxford, G. S. (2019), *Non-additive effects on the morphology of hybrids between two species of large house spiders, Eratigena saeva and E. duellica (Araneae: Agelenidae)*, *Arachnology* **18** 223.
- Paradis Emmanuel and Schliep, K. (2019), *ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R*, *Bioinformatics* **35** 526.
- Patterson, N., P. Moorjani, Y. Luo et al. (2012), *Ancient admixture in human history*, *Genetics* **192** 1065.
- Pease, J. B. and M. W. Hahn (2015), *Detection and polarization of introgression in a five-taxon phylogeny*, *Syst Biol* **64** 651.
- Pfeifer, G. P. (2006), *Mutagenesis at methylated CpG sequences*, *DNA methylation: basic mechanisms* **301** 259.
- Racimo, F., S. Sankararaman, R. Nielsen and E. Huerta-Sánchez (2015), *Evidence for archaic adaptive introgression in humans*, *Nat Rev Genet* **16** 359.
- Ramírez, M. J., I. L. Magalhaes, S. Derkarabetian et al. (2021), *Sequence capture phylogenomics of true spiders reveals convergent evolution of respiratory systems*, *Syst Biol* **70** 14.
- Rancilhac, L., E. D. Enbody, R. Harris et al. (2024), *Introgression underlies phylogenetic uncertainty but not parallel plumage evolution in a recent songbird radiation*, *Syst Biol* **73** 12.
- Responte, M. A., C.-Y. Wu, N. U. Elias et al. (2025), *Recent Range Expansion and Genomic Admixture in a Kleptoparasitic Spider, Argyrodes lanyuensis: A Case of Adaptive Introgression on Small, Isolated Islands of the Taiwan–Philippine Transition Zone?*, *Mol Ecol* **34** e17630.
- Romero, I. G., I. Ruvinsky and Y. Gilad (2012), *Comparative studies of gene expression and the evolution of gene regulation*, *Nat Rev Genet* **13** 505.
- Rosser, N., F. Seixas, L. M. Queste et al. (2024), *Hybrid speciation driven by multilocus introgression of ecological traits*, *Nature* **628** 811.
- Salgado-Roa, F. C., C. Pardo-Díaz, E. Lasso et al. (2018), *Gene flow and Andean uplift shape the diversification of Gasteracantha cancriformis (Araneae: Araneidae) in Northern South America*, *Ecol Evol* **8** 7131.
- Šmarda, P., P. Bureš, L. Horová et al. (2014), *Ecological and evolutionary significance of genomic GC content diversity in monocots*, *Proceedings of the National Academy of Sciences* **111** E4096.
- Steenwyk, J. L., T. J. Buida III, Y. Li, X.-X. Shen and A. Rokas (2020), *ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference*, *PLoS Biol* **18** e3001007.

-
- Strimmer, K. and A. von Haeseler (1997),
Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment,
Proc Natl Acad Sci U S A **94** 6815.
- Sueoka, N. (1988), *Directional mutation pressure and neutral molecular evolution.*,
Proceedings of the National Academy of Sciences **85** 2653.
- Suvorov, A., C. Scornavacca, M. S. Fujimoto et al. (2022),
Deep ancestral introgression shapes evolutionary history of dragonflies and damselflies,
Syst Biol **71** 526.
- Suyama, M., D. Torrents and P. Bork (2006), *PAL2NAL: robust conversion of protein sequence
alignments into the corresponding codon alignments*, Nucleic Acids Res **34** W609.
- Szöllősi, G. J., E. Tannier, V. Daubin and B. Boussau (2015),
The inference of gene trees with species trees, Syst Biol **64** e42.
- Tabei, Y., H. Kiryu, T. Kin and K. Asai (2008),
A fast structural multiple alignment method for long RNA sequences, BMC Bioinformatics **9** 1.
- Tang, M., C. J. Hardman, Y. Ji et al. (2015),
High-throughput monitoring of wild bee diversity and abundance via mitogenomics,
Methods Ecol Evol **6** 1034.
- Than, C., D. Ruths and L. Nakhleh (2008), *PhyloNet: a software package for analyzing and
reconstructing reticulate evolutionary relationships*, BMC Bioinformatics **9** 1.
- Tyagi, K., V. Kumar, N. Poddar et al. (2020), *The gene arrangement and phylogeny using
mitochondrial genomes in spiders (Arachnida: Araneae)*,
International journal of biological macromolecules **146** 488.
- Wang, D., Y. Zhang, Z. Zhang, J. Zhu and J. Yu (2010), *KaKs_Calculator 2.0: a toolkit incorporating
gamma-series methods and sliding window strategies*,
Genomics, proteomics & bioinformatics **8** 77.
- Wang, H., W. Zhang, Y. Yu et al. (2024), *Biased gene introgression and adaptation in the face of
chloroplast capture in Aquilegia amurensis*, Syst Biol **73** 886.
- Warnefors, M., B. Hartmann, S. Thomsen and C. R. Alonso (2016),
Combinatorial gene regulatory functions underlie ultraconserved elements in Drosophila,
Mol Biol Evol **33** 2294.
- Wright, F. (1990), *The 'effective number of codons' used in a gene*, Gene **87** 23.
- WSC (2025), *World Spider Catalog*, DOI: [10.24436/2](https://doi.org/10.24436/2), (visited on 20/01/2025).
- Yakovchuk, P., E. Protozanova and M. D. Frank-Kamenetskii (2006),
Base-stacking and base-pairing contributions into thermal stability of the DNA double helix,
Nucleic Acids Res **34** 564.
- Yang, L., A. J. Harris, F. Wen et al. (2023), *Phylogenomic analyses reveal an allopolyploid origin of
core Didymocarpaceae (Gesneriaceae) followed by rapid radiation*, Syst Biol **72** 1064.
- Yang, Z. (2015), *The BPP program for species tree estimation and species delimitation*,
Curr Zool **61** 854.
- Zhang, C. and S. Mirarab (2022),
Weighting by Gene Tree Uncertainty Improves Accuracy of Quartet-based Species Trees,
Mol Biol Evol **39** msac215.
- Zhang, C., H. A. Ogilvie, A. J. Drummond and T. Stadler (2018),
Bayesian inference of species networks from multilocus sequence data, Mol Biol Evol **35** 504.

-
- Zhang, H. and W. Bu (2022), *Exploring large-scale patterns of genetic variation in the COI gene among Insecta: Implications for DNA barcoding and threshold-based species delimitation studies*, *Insects* **13** 425.
- Zhang, R., N. Xiang, X. Gao et al. (2025), *Molecular Phylogenetic Relationships Based on Mitogenomes of Spider: Insights Into Evolution and Adaptation to Extreme Environments*, *Ecol Evol* **15** e70774.
- Zhang, Y., Z. Shen, X. Meng et al. (2022), *Codon usage patterns across seven Rosales species*, *BMC Plant Biology* **22** 65.
- Zhang, Z., J. Li, P. Cui et al. (2012), *Codon Deviation Coefficient: a novel measure for estimating codon usage bias and its statistical significance*, *BMC Bioinformatics* **13** 1.
- Zhang, Z., J. Li, X.-Q. Zhao et al. (2006), *KaKs_Calculator: calculating Ka and Ks through model selection and model averaging*, *Genomics, proteomics & bioinformatics* **4** 259.

Discussion and Perspectives

7.1 Review of research goals and questions

The first essential goal of this thesis was to fill the most significant knowledge gaps in the phylogeny of the spider family Pholcidae by determining the phylogenetic placements of the subfamily Ninetinae and the genera *Artema* and *Priscula* within this family.

Another important objective was to model the evolutionary history of body size and leg length, quantifying their evolutionary flexibility, infer the dynamics of speciation and extinction within Pholcidae, and test the correlations between these factors.

A further aim was to conduct extensive taxon sampling in order to explore the phylogenetic placement of two enigmatic new species from the semi-arid Brazilian Caatinga, and of the Chilean genus *Aucana*.

A fourth goal was to substantially improve our knowledge about spider mitogenomes and to study the evolution of mitogenomes in Pholcidae.

A final goal was to evaluate the prevalence of introgression and its influences on the phylogenetic inference of Pholcidae.

By addressing these questions, we aimed to enhance our understanding of Pholcidae phylogeny and evolution, the factors contributing to the uneven species diversity within Pholcidae subfamilies, the factors responsible for the discordance in Pholcidae phylogenies constructed from different datasets, and the evolutionary dynamics of the mitochondrial genomes of Pholcidae.

7.2 Summary of main findings

Our study resulted in several important findings:

Pholcidae Phylogeny. Our analysis confirmed the previously reported sister-group relationships between Pholcinae and Smeringopinae, as well as between “other Arteminae” and Modisiminae. We also validated the monophyly of the subfamily Ninetinae through comprehensive sampling of Ninetinae species while rejecting the monophyly of the conventional concept of Arteminae (“other Arteminae” + *Artema*). Our results showed that the genera *Artema* and *Priscula* are sister taxa, and together are sister of Ninetinae. This clade is sister to (“other Arteminae”, Modisiminae). Later, we redelimited the subfamily Arteminae to consist solely of *Artema* and *Priscula*, and renamed “other Arteminae” as Phy***inae. Moreover, we established a new subfamily, Cai***inae composed of a new genus and two new species, GENUS-1

sp1 and GENUS-1 sp2. Therefore, we increased the number of Pholcidae subfamilies from 5 to 7, and these subfamilies have the following relationships: ((Pholcinae, Smeringopinae), (Cai***inae, ((Arteminae, Ninetinae), (Phy***inae, Modisiminae)))). Finally, we constructed the largest known Pholcidae phylogeny, encompassing 956 pholcid species based on ultra-conserved elements (UCEs) and six genetic markers, which provides the most comprehensive insights into the generic relationships within this family.

Trait evolutionary flexibility and Pholcidae diversification. We found that evolutionary flexibility positively correlates with rates of speciation, trait evolution, and species richness in Pholcidae. Regarding the allometric scaling relationships between leg length and body size, we observed a near-isometric pattern in species-rich groups and positive allometry in species-poor groups. The former indicates stabilizing selection, whereas the latter suggests directional selection. Rapid and slow microhabitat shifts were observed in species-rich and species-poor groups, respectively, implying that environmental heterogeneity and frequent microhabitat transitions contribute to the origin and maintenance of evolutionary flexibility in pholcids. Conversely, low evolutionary flexibility—observed particularly in groups that tend towards dwarfism or gigantism—could lead to an evolutionary dead end. Finally, our results support the idea that pholcids do not in general evolve according to Cope’s rule.

The evolution of Pholcidae mitogenomes. A total of 193 mitochondrial genomes were assembled and annotated in the course of this thesis, and 98 of them are either complete or almost complete. During the gene annotation process, we found that the existing pipelines perform poorly on the annotation of ATP8, ND4, ND4L, and ND5, as well as tRNA genes. We also found that pholcid species have a conserved gene order for mitochondrial protein-coding-genes (PCGs). Strong codon usage bias was observed in these PCGs, which was likely shaped by selection pressure instead of mutation bias.

Introgression in Pholcidae. We identified several introgressions among subfamilies and genera within Pholcidae, particularly within Arteminae and Modisiminae. The introgressed loci, encompassing 89 genes, are predominantly located in non-coding regions (102 out of 155 loci, or 65.8%). Among these genes, three are functionally enriched in specific molecular functions, while 22 genes are enriched in distinct biological processes. Specifically, the three functionally enriched genes are associated with DNA-binding transcription factor activity and transcription regulator activity, whereas the 22 process-enriched genes are involved in transcription regulation, gene expression, and the biosynthesis of various molecules.

Pholcidae Biodiversity. During this thesis, I was part of a team that described more than 70 new species of Pholcidae spiders representing 16 genera. In the 14 resulting publications, I contributed molecular data and analyses in an integrative taxonomic framework (see the section **List of Publications and Contributions**).

7.3 Evaluation of methodology

This thesis presents the largest-scale taxon and gene sampling of any spiders family to date by integrating genetic markers and UCEs from targeted capture sequencing (TCS) and low-coverage whole genome sequencing (lcWGS). Compared to traditional sequencing of several genetic markers (e.g., Dimitrov, Astrin et al. 2013; Dimitrov, Benavides et al. 2017; Eberle et al. 2018; Wheeler et al. 2017), the TCS and lcWGS methods offer low-cost and efficient ways to obtain a large number of loci across a wide range of taxa for phylogenetic analysis, which is critical to solve difficult phylogenetic problems (Zhang, Lin et al. 2025). These methods have been applied in numerous studies, which demonstrate the significant implications of UCE data in various taxa, including plants (Phang et al. 2023), spiders (Kulkarni, Kallal et al. 2021; Ramírez et al. 2021; Xu et al. 2021), insects (Faircloth, Branstetter et al. 2015; Gustafson et al. 2020),

and vertebrates (Andersen et al. 2019; S. Liu et al. 2022). UCEs typically consist of a conserved region in the center and variable flanking regions (Carter et al. 2023). They are widely present in taxa that diverged long ago and are mostly single-copy elements spread throughout the genome (thus, the analysis should not be biased towards certain loci), without involving issues related to paralogs or retrotransposons (which violate the model assumptions of phylogenetic inference) (Faircloth, McCormack et al. 2012; Harvey et al. 2016). The assembly of UCEs does not rely on reference genomes (Faircloth 2016). The conserved regions facilitate multiple-sequence alignment (MSA) of divergent species, while the higher level of genetic variability in the flanking regions of UCEs provides additional information for phylogenetic inference (Carter et al. 2023). These characteristics make UCEs suitable loci for elucidating dubious subfamily relationships within Pholcidae, as the age of their common ancestor dates back approximately 117 million years (Ma).

It has been suggested that TCS data, including UCEs, may suffer from greater irreproducibility in gene tree inference compared to data from transcriptomic sequencing and whole genome sequencing (Shen et al. 2020). This is because TCS methods often generate short alignments that contain low phylogenetic information, which can lead to unreliable species trees when using the gene-tree-then-species-tree (i.e., summary) method for phylogenetic inference (Shen et al. 2020). On the other hand, higher recombination rates in GC-rich regions of UCE data may result in less accurate phylogenies (Bossert et al. 2017). We addressed these issues by creating data matrices with varying MSA lengths, filtering loci with detected intra-locus recombination, and using a site-based coalescence method for phylogenetic inference (thus avoiding gene tree errors). Our results remained consistent regardless of the phylogenetic analysis methods employed (maximum likelihood and Bayesian-based; concatenation and coalescence; models considering rate and composition heterogeneity) and the data matrices varying in gene completeness and taxon coverage, suggesting that our findings are robust.

7.4 Theoretical and practical significance

Our review of the methods for phylogenomic analysis (Chapter 2) provides both theoretical and practical guidance for researchers who are not familiar with this field, thereby facilitating the application of phylogenomic methods in evolutionary biology and taxonomy.

Our results not only provide new insights into the subfamily phylogenetic relationships of Pholcidae (Chapters 3, 5 and 6) but also offer testable hypotheses regarding the generic relationships within each subfamily (Chapter 4). This is important because generic relationships within Pholcidae have only been inferred from a limited number of genetic markers and/or morphological characters, which possess low resolution potential (Astrin et al. 2007; Bruvo-Madaric et al. 2005; Dimitrov, Astrin et al. 2013; Eberle et al. 2018). By introducing two new subfamilies, Cai****inae and Phy****nae, this thesis represents a significant update and improvement of Pholcidae systematics.

The newly established subfamily phylogenetic relationships and the largest Pholcidae phylogeny have important implications for understanding trait evolution, biogeography, and other related fields. For instance, our results refute the idea that small body size was the ancestral state of pholcids and reject Cope's rule for all extant pholcids as a whole. Instead, our analysis demonstrates that pholcid ancestors generally evolved to various body sizes (both larger and smaller) within each subfamily. Pholcidae is a subgroup of Synspermiata, a group of spiders with a specific sperm transfer form (synspermia) (Kulkarni, H. M. Wood et al. 2023). The presence of synspermia was thus thought to be a plesiomorphic state of pholcids (Huber, Meng et al. 2023). However, our results demonstrate that cleistospemia (a derived sperm transfer

form) may have evolved in pholcid ancestors, with a reversal to synspermia in Ninetinae. Alternatively, Pholcidae experienced at least four times of convergent evolution from synspermia toward cleistospermia. Similarly, the genus *Artema* and “other Arteminae” (i.e., *Phy***nae*) share a unique morphology of the male procursus (intromittent organ) present in both *Artema* and “other Arteminae” that is related to asymmetric insertion of the male genitalia. Our results suggest that this character may have evolved independently in each of the two groups or been lost repeatedly in the other groups.

Our study further emphasizes the strong correlation between trait evolution and species diversification, as well as the diversity heterogeneity in the tree of life (Chapter 4). However, unlike most studies that focus either on identifying key innovations or on the correlation between absolute trait values and speciation or extinction rates, our study stresses the significance of trait evolution rates, particularly the evolutionary flexibility of traits. For the first time, we observed near-isometry between body size and leg length in species-rich groups and positive allometry in species-poor regions, suggesting that different selection pressures are exerted on these groups. The rapid microhabitat shifts in species-rich groups, along with environmental heterogeneity, may be responsible for the differences in evolutionary flexibility among various groups.

Our study also highlights the importance of assessing the influence of biological (e.g., recombination and incomplete lineage sorting) and technical factors on the construction of phylogenetic trees, particularly in the case of a high degree of incongruence between gene trees. Some of these genes may have different evolutionary histories compared to the species tree. Furthermore, our results suggest the widespread occurrence of introgression in Pholcidae, which could also be true for other spider families.

7.5 Study Limitations

Despite the efforts we made in the course of this thesis, several limitations must be acknowledged. For instance, the number of UCE loci (approximately 130) used for phylogenetic analysis (Chapter 3) was somewhat limited compared to other studies that employed the same spider-specific probe set (Kulkarni, H. Wood et al. 2020), which may be attributed to an insufficient amount of sequencing data, low efficiency during probe-DNA hybridization, or high PCR duplication rates. The lcWGS method can sample more loci when sequencing coverage is high, allowing for a broader range of genomic regions to be sampled instead of focusing only on conserved regions such as UCEs.

Another potential limitation is the relatively low number of taxa sampled compared to the high species diversity within the Pholcidae family. It has been suggested that more comprehensive taxon sampling may yield alternative phylogenetic results. For instance, by combining denser taxon sampling with transcriptome and genome data, Ballesteros et al. (2022) propose that Arachnida is paraphyletic, with horseshoe crabs (*Xiphosura*) nested within Arachnida. Since the year 2000, more than 1,400 new species have been formally described in Pholcidae, highlighting the high abundance of pholcids and emphasizing the importance of extensive sampling. Thus, while our study has the currently highest percentage of sequenced taxa of any major spider family, future studies should aim at a further substantial increase of taxon sampling.

We are also aware of the large confidence intervals of the estimated species divergence times for Pholcidae. This is largely due to the fact that very few Pholcidae fossils have been documented; these fossils are typically found in relatively young amber and are not evenly distributed across subfamilies, with 20% belonging to Ninetinae (Dunlop et al. 2025; Huber, Meng et al. 2023).

Incomplete annotation of some PCGs and many tRNA genes, along with gene rearrangements often occurring around tRNA genes in arachnids (Ban et al. 2022; Tyagi et al. 2020), has limited our ability to

fully study the evolution of mitochondrial rearrangements within Pholcidae.

The accuracy of introgression detection is compromised when the dataset is small, which is especially the case for the f -branch statistic (Malinsky et al. 2021). With only 18,668 bp SNPs in our dataset, it is possible that our introgression analysis was affected by insufficient data.

7.6 Future research directions

Species description and delimitation are essential for many studies at the species level. With only one-third of pholcid diversity documented (Huber and Chao 2019) and approximately 60 new species being described each year (Figure 1.3), it may take another 60 years to fully describe all pholcids. However, it is uncertain how much time remains for us to discover and protect these species, given that climate change, pollution, and habitat destruction are becoming increasingly prevalent (Bernatchez et al. 2024; Cowie et al. 2022; Wagner et al. 2021). To accelerate the species discovery process, more resources should be allocated to taxonomy, more taxonomists should be trained, and more advanced methods should be developed. The concept of integrative taxonomy was introduced two decades ago (Dayrat 2005), and several tools (ABGD, ASAP, PTP, BPP) have been devised to delimit species based on single-locus or multi-locus data (Puillandre, Brouillet et al. 2021; Puillandre, Lambert et al. 2012; Z. Yang 2015; Zhang, Kapli et al. 2013). Additionally, methods based on machine learning and molecular data have also been developed (Derkarabetian et al. 2019; Pei et al. 2018; Smith et al. 2020). However, only recently have a few integrative frameworks been proposed to combine evidence from various sources (molecular, morphological, ecological, etc.) to assist in species delimitation and identification (e.g., Oberprieler 2023; B. Yang et al. 2022). It is also critical to acknowledge which clades have significant knowledge gaps regarding new species and to identify species hotspots for these clades. In Pholcidae, the species biodiversity of the newly recognized subfamily Cai***inae remains unclear, as it currently consists of only two species from the same genus, both originating from a poorly sampled area in semiarid Brazilian environments. The subfamily Ninetinae equally suffers from sampling bias due to their extremely small body sizes, fast-running and cryptic habitats in dry regions (Huber, Meng et al. 2023). However, also humid tropical regions should be further targeted for sampling; for instance, while around 1,000 new spider species were nominated in the past few years (Figure 1.2), about one-third of these were from China, many of which were from tropical areas (Li 2020). Finally, museum specimens should also be digitized and examined, considering that many specimens in museums have yet to be identified.

While there are 7,786 arthropod genomes available on NCBI (<https://www.ncbi.nlm.nih.gov/datasets/genome/>; accessed on February 17, 2025), 6,939 of these are insects. Among the spider (Araneae) genomes, which total only 87, the majority (47) belong to the Araneoidea superfamily; notably, no genomes from the Pholcidae family have been released. The lack of reference genomes hinders our ability to conduct studies in comparative genomics, functional genomics, and evolutionary developmental biology, limiting our understanding of the genetic basis of various morphological traits. Future directions for Pholcidae research may include using Genome-Wide Association Studies (GWAS) to investigate the loci responsible for specific traits, employing pangenomes to study the extent of genomic variation within species, and applying comparative genomics to elucidate genome evolution and its contribution to pholcid diversity. Moreover, using RNA-seq, epigenetics, single-cell transcriptomics, and 3D genomes will enable us to understand how pholcids adapt to various environments. On the other hand, studying the evolution of gene expression and regulatory networks will provide a better understanding of the evolution of novel traits.

While microhabitat, body size, and leg length have been found to be related to pholcid diversification, further studies are needed to explore the influence of other factors, such as climate and geographic factors, as well as sexual dimorphism, on pholcid diversification. Given that microbes play an important role in the evolution of many species (e.g., Kirsch et al. 2025), it would be particularly interesting to study the evolutionary dynamics of host-microbe interactions in Pholcidae.

Unlike insects or vertebrates, spider mitochondrial genomes are understudied, and only several hundred of them are publicly available on NCBI. The annotation of mitochondrial genes (tRNA and PCGs) heavily relies on the available homologous gene sets of closely related species. This poses challenges to the gene annotation of underexplored groups, such as spiders. Thus, we advocate a focus on improving the database and algorithms for the annotation of spider mitochondrial genes.

References

- Andersen, M. J., J. M. McCullough, N. R. Friedman et al. (2019), *Ultraconserved elements resolve genus-level relationships in a major Australasian bird radiation (Aves: Meliphagidae)*, *Emu-Austral Ornithology* **119** 218.
- Astrin, J. J., B. Misof and B. A. Huber (2007), *The pitfalls of exaggeration: molecular and morphological evidence suggests Kaliana is a synonym of Mesabolivar (Araneae: Pholcidae)*, *Zootaxa* **1646** 17.
- Ballesteros, J. A., C. E. Santibáñez-López, C. M. Baker et al. (2022), *Comprehensive Species Sampling and Sophisticated Algorithmic Approaches Refute the Monophyly of Arachnida*, *Mol Biol Evol* **39** msac021.
- Ban, X.-C., Z.-K. Shao, L.-J. Wu, J.-T. Sun and X.-F. Xue (2022), *Highly diversified mitochondrial genomes provide new evidence for interordinal relationships in the Arachnida*, *Cladistics* **38** 452.
- Bernatchez, L., A.-L. Ferchaud, C. S. Berger, C. J. Venney and A. Xuereb (2024), *Genomics for monitoring and understanding species responses to global climate change*, *Nat Rev Genet* **25** 165.
- Bossert, S., E. A. Murray, B. B. Blaimer and B. N. Danforth (2017), *The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data*, *Mol Phylogenet Evol* **111** 149.
- Bruvo-Madarc, B., B. A. Huber, A. Steinacher and G. Pass (2005), *Phylogeny of pholcid spiders (Araneae: Pholcidae): combined analysis using morphology and molecules*, *Mol Phylogenet Evol* **37** 661.
- Carter, J. K., R. T. Kimball, E. R. Funk et al. (2023), *Estimating phylogenies from genomes: A beginners review of commonly used genomic data in vertebrate phylogenomics*, *Journal of Heredity* **114** 1.
- Cowie, R. H., P. Bouchet and B. Fontaine (2022), *The Sixth Mass Extinction: fact, fiction or speculation?*, *Biological Reviews* **97** 640.
- Dayrat, B. (2005), *Towards integrative taxonomy*, *Biological journal of the Linnean society* **85** 407.
- Derkarabetian, S., S. Castillo, P. K. Koo, S. Ovchinnikov and M. Hedin (2019), *A demonstration of unsupervised machine learning in species delimitation*, *Mol Phylogenet Evol* **139** 106562.
- Dimitrov, D., J. J. Astrin and B. A. Huber (2013), *Pholcid spider molecular systematics revisited, with new insights into the biogeography and the evolution of the group*, *Cladistics* **29** 132.

- Dimitrov, D., L. R. Benavides, M. A. Arnedo et al. (2017), *Rounding up the usual suspects: a standard target-gene approach for resolving the interfamilial phylogenetic relationships of ecribellate orb-weaving spiders with a new family-rank classification (Araneae, Araneoidea)*, *Cladistics* **33** 221.
- Dunlop, J. A., D. Penney and D. Jekel (2025), *A summary list of fossil spiders and their relatives*, (visited on 20/01/2025).
- Eberle, J., D. Dimitrov, A. Valdez-Mondragon and B. A. Huber (2018), *Microhabitat change drives diversification in pholcid spiders*, *BMC Evol Biol* **18** 141.
- Faircloth, B. C. (2016), *PHYLUCE is a software package for the analysis of conserved genomic loci*, *Bioinformatics* **32** 786.
- Faircloth, B. C., M. G. Branstetter, N. D. White and S. G. Brady (2015), *Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera*, *Mol Ecol Resour* **15** 489.
- Faircloth, B. C., J. E. McCormack, N. G. Crawford et al. (2012), *Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales*, *Syst Biol* **61** 717.
- Gustafson, G. T., S. M. Baca, A. M. Alexander and A. E. Z. Short (2020), *Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance*, *Syst Entomol* **45** 552.
- Harvey, M. G., B. T. Smith, T. C. Glenn, B. C. Faircloth and R. T. Brumfield (2016), *Sequence capture versus restriction site associated DNA sequencing for shallow systematics*, *Syst Biol* **65** 910.
- Huber, B. A. and A. Chao (2019), *Inferring global species richness from megatranssect data and undetected species estimates*, *Contributions to Zoology* **88** 42.
- Huber, B. A., G. Meng, J. Král et al. (2023), *High and dry: integrative taxonomy of the Andean spider genus Nerudia (Araneae: Pholcidae)*, *Zool J Linn Soc* **198** 534, doi: [10.1093/zoolinnean/zlac100](https://doi.org/10.1093/zoolinnean/zlac100).
- Kirsch, R., Y. Okamura, M. García-Lozano et al. (2025), *Symbiosis and horizontal gene transfer promote herbivory in the megadiverse leaf beetles*, *Curr Biol* **35** 640.
- Kulkarni, S., R. J. Kallal, H. Wood et al. (2021), *Interrogating Genomic-Scale Data to Resolve Recalcitrant Nodes in the Spider Tree of Life*, *Mol Biol Evol* **38** 891.
- Kulkarni, S., H. Wood, M. Lloyd and G. Hormiga (2020), *Spider-specific probe set for ultraconserved elements offers new perspectives on the evolutionary history of spiders (Arachnida, Araneae)*, *Mol Ecol Resour* **20** 185.
- Kulkarni, S., H. M. Wood and G. Hormiga (2023), *Advances in the reconstruction of the Spider Tree of Life: a roadmap for spider systematics and comparative studies*, *Cladistics* **39** 479.
- Li, S. (2020), *Spider taxonomy for an advanced China*, *Zoological Systematics* **45** 73.
- Liu, S., C. Zhou, G. Meng et al. (2022), *Evolution and diversification of Mountain voles (Rodentia: Cricetidae)*, *Commun Biol* **5** 1417, doi: [10.1038/s42003-022-04371-z](https://doi.org/10.1038/s42003-022-04371-z).
- Malinsky, M., M. Matschiner and H. Svardal (2021), *Dsuite-Fast D-statistics and related admixture evidence from VCF files*, *Mol Ecol Resour* **21** 584.

-
- Oberprieler, C. (2023), *The Wettstein tesseract: a tool for conceptualising species-rank decisions and illustrating speciation trajectories*, *Taxon* **72** 1.
- Pei, J., C. Chu, X. Li, B. Lu and Y. Wu (2018), *CLADES: A classification-based machine learning method for species delimitation from population genetic data*, *Mol Ecol Resour* **18** 1144.
- Phang, A., F. F. Pezzini, D. F. Burslem et al. (2023), *Target capture sequencing for phylogenomic and population studies in the Southeast Asian genus Palaquium (Sapotaceae)*, *Botanical Journal of the Linnean Society* **203** 134.
- Puillandre, N., S. Brouillet and G. Achaz (2021), *ASAP: assemble species by automatic partitioning*, *Mol Ecol Resour* **21** 609.
- Puillandre, N., A. Lambert, S. Brouillet and G. Achaz (2012), *ABGD, Automatic Barcode Gap Discovery for primary species delimitation*, *Mol Ecol* **21** 1864.
- Ramírez, M. J., I. L. Magalhaes, S. Derkarabetian et al. (2021), *Sequence capture phylogenomics of true spiders reveals convergent evolution of respiratory systems*, *Syst Biol* **70** 14.
- Shen, X.-X., Y. Li, C. T. Hittinger, X.-x. Chen and A. Rokas (2020), *An investigation of irreproducibility in maximum likelihood phylogenetic inference*, *Nat Commun* **11** 6096.
- Smith, M. L. and B. C. Carstens (2020), *Process-based species delimitation leads to identification of more biologically relevant species*, *Evolution* **74** 216.
- Tyagi, K., V. Kumar, N. Poddar et al. (2020), *The gene arrangement and phylogeny using mitochondrial genomes in spiders (Arachnida: Araneae)*, *International journal of biological macromolecules* **146** 488.
- Wagner, D. L., E. M. Grames, M. L. Forister, M. R. Berenbaum and D. Stopak (2021), *Insect decline in the Anthropocene: Death by a thousand cuts*, *Proceedings of the National Academy of Sciences* **118** e2023989118.
- Wheeler, W. C., J. A. Coddington, L. M. Crowley et al. (2017), *The spider tree of life: phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling*, *Cladistics* **33** 574.
- Xu, X., Y.-C. Su, S. Y. Ho et al. (2021), *Phylogenomic analysis of ultraconserved elements resolves the evolutionary and biogeographic history of segmented trapdoor spiders*, *Syst Biol* **70** 1110.
- Yang, B., Z. Zhang, C.-Q. Yang et al. (2022), *Identification of species by combining molecular and morphological data using convolutional neural networks*, *Syst Biol* **71** 690.
- Yang, Z. (2015), *The BPP program for species tree estimation and species delimitation*, *Curr Zool* **61** 854.
- Zhang, J., P. Kapli, P. Pavlidis and A. Stamatakis (2013), *A general species delimitation method with applications to phylogenetic placements*, *Bioinformatics* **29** 2869.
- Zhang, J., L. Lin, Y. Mu, A. Brelsford and J. Purcell (2025), *A comparison of phylogenomic inference pipelines for low-coverage whole-genome sequencing in Formica ants*, *Syst Entomol*.

Acknowledgements

I started my PhD journey during the pandemic COVID-19. More than four years have flown by, and now I have finally completed my doctoral thesis.

It was more than 20 years ago that I began primary school in a remote mountain village in Wuzhou City, China, where transportation was extremely limited, with just two tractors and two motorcycles in the village. At that time, the villagers still lived in houses with tile roofs that were several decades old, some of which were built with adobe bricks. No one could afford a telephone, and the most expensive item in a household was a black-and-white television. The only refrigerator in the village belonged to the sole small shop. To reach the market, villagers had to travel by boat, which ran only every few days. The nearest market was a half-hour boat ride away, while more distant ones could take several hours. For most people, the farthest they ever traveled was to one of these markets.

The village primary school was a fragile adobe-brick building with a tile roof, serving four grades of students, two of which shared a single classroom. There was even a year when the school had to move to an abandoned building in a forest farm deep in the mountains, far from the village. We had to walk an hour every day to reach the school, having only our self-packed congee and pickled vegetables for lunch. With only two teachers, the school could only offer Chinese and Math. As a child, I was introverted, and primary school served as my main window to the outside world.

My parents worked from dawn to dusk in the fields and mountains. After school, although I longed to watch cartoons and play games like other children, I had to start fires, cook, feed the pigs and chickens, look after my younger sister, or go to the fields to collect mulberry leaves to feed the silkworms. I was fortunate to have a few old comic books and storybooks, passed down by my father and uncles. They became a precious window for me to glimpse the outside world and a source of spiritual nourishment.

In grades five and six, I attended the central primary school in the nearest township. Each morning, after quickly eating leftovers from the previous day, I would walk 40 minutes to school. During windy or rainy seasons, or when floods hit, the roads turned into thick mud, making the journey more difficult and often leaving the clothes and shoes covered in dirt.

Later, I attended middle school in the county and high school in the city. Due to the long distances and poor transportation, I had to board at school, returning home only once or twice per semester. Eventually, I went to university in Wuhan City, becoming one of the first few students from our village to attend college. After graduation, I worked in Shenzhen and Beijing for several years before moving to Germany to pursue my PhD.

If we start from the point when I began attending school, it took me more than four years to leave the small village where I was born and raised, six years to reach the county, nine years to reach Wuzhou City, twelve years to travel over 1,000 kilometers to Wuhan City, and 22 years to arrive in Germany, over 13,000 kilometers from my home village. Along the way, I was fortunate to meet many kind-hearted people—

teachers, classmates and security guards at school, colleagues and friends, as well as both acquaintances and strangers—who offered me invaluable help and support. My family’s selfless sacrifices, combined with government education aid policies, allowed me to complete high school and university without much financial burden.

Over the past 20 years, our village has undergone remarkable transformations. The once dirt roads have been replaced by smooth concrete ones, now lit by streetlights. Most families have built modern flat-roofed houses, own color TVs and refrigerators, and many have cars and motorcycles. Nearly everyone now has a mobile phone.

Twenty-five years ago, downstream of our village, a hydraulic project was constructed, protecting the people on both sides of the Xijiang River from the devastation of annual floods. Twenty years ago, the village’s small shop installed its first landline phone. Fifteen years ago, Wuzhou City opened its first train station, ending its trainless history. I remember taking the slowest train for more than 16 hours to travel over 1,000 kilometers to Wuhan City for university. Ten years ago, Wuzhou City welcomed its high-speed rail station, sparing me the grueling eight-hour bus journey from Shenzhen, 500 kilometers away, to my hometown. Five years ago, an airport was built, making it possible to reach Beijing, 2,200 kilometers away, in just three hours. Both the high-speed rail station and the airport are within 20 kilometers of our village, finally fulfilling the childhood rhyme we often sang: “The big airplane lands for me, taking me to Beijing.” Meanwhile, the village has been connected to the internet, linking it to the world and integrating it into our increasingly interconnected “global village.”

I have always grown up at a slower pace than others. I was not gifted and often envied my classmates who consistently earned awards during childhood. Growing up in a poor family, I struggled with feelings of inferiority, shyness, and timidity. At school, I rarely mustered the courage to express myself, let alone pursue the desires hidden deep within my heart. I frequently admired classmates who came from well-off families and excelled in studies, envying their confidence, multitalent and the vibrant world they had access to from an early age.

But I was also fortunate. At my village’s small primary school, there were only 13 students in my grade, many of whom dropped out early for various reasons. I became one of the few who completed their middle school education. When I transitioned from primary school to middle school, I was fortunate to be one of the seven students from the entire school to be admitted to the county’s key middle school, which was an unprecedented event in our township. On the day the high school entrance exam grades were released, a relative rode a motorcycle at 120 km/h to take me to the registration point of the city’s key high school, making me the last one to complete the registration. In my third year of university, a teacher encouraged me to apply for an interview for the BGI Innovation Class, which led me to join BGI, the company that once represented China in the Human Genome Project. There, I not only received invaluable guidance from my several mentors, Professor Xin Zhou, Professor Shanlin Liu and Dr. Min Tang, but also met a group of enthusiastic peers. They were the first to show me the joy of combining academia with practice and helped me understand the importance of teamwork. After arriving in Germany, my supervisors, fellow doctoral students at the LIB, and landlady/landlord provided tremendous support, helping me gradually adapt to the new environment. Here, I also made some close friends, whose companionship has given me warmth and support.

With the above experiences in mind, I am deeply grateful to the Deutsche Forschungsgemeinschaft (DFG) for funding my PhD position, to my supervisors, Dr. Bernhard A. Huber, Dr. Lars Podsiadlowski, and Prof. Dr. Dimitar Dimitrov, for their invaluable guidance, constructive feedback, suggestions, and encouragement throughout this research project. Their expertise and unwavering support have been instrumental in shaping this thesis. I would also like to thank the Leibniz Institute for the Analysis of

Biodiversity Change (LIB) and the University of Bonn for providing the essential resources and facilities that made this research possible. Additionally, I appreciate the technical support and contributions of Sandra Kukowka, Anja Bodenheimer, Jonas Eberle, and Hans-Joachim Krammer to various aspects of this study. Finally, I also sincerely thank the members of my doctoral committee: Prof. Dr. Bernhard Misof, Dr. Lars Podsiadlowski, Prof. Dr. Nicolas Gompel and Prof. Dr. Thomas Döring. I deeply appreciate them for taking the time to review my thesis and serving as examiners during the defense amidst their demanding schedules.

Special thanks go to my family and friends, as well as my former colleagues and mentors, Prof. Dr. Xin Zhou, Prof. Dr. Shanlin Liu, Dr. Min Tang, Dr. Chengran Zhou, and Dr. Chentao Yang, for their understanding, encouragement, and unwavering support throughout my academic journey.

Finally, I'm grateful for my perseverance and my enduring curiosity about the unknown.

中文致谢

我在疫情期间开始了我的读博生涯。四年多的时光匆匆而过，如今，我终于完成了我的博士毕业论文。

时光回溯至二十多年前，在梧州市下属的一个偏远的山村，我开始上小学。那时候村民居住的还是几十年老的瓦片房，其中一些还是用生砖砌成的。交通极为闭塞，村里仅有的机动车只有两台拖拉机和两辆摩托车。没有人有电话，稍微昂贵一点的设备就是黑白电视机了。除了村里唯一的小卖部，没有家庭拥有冰箱。村民们赶集的时候只能坐船，最近的小集市需要坐半个小时的船，远点的集市则需要坐上几个小时。船每几天才运行一趟。大部分人一辈子去过的最远的地方也就是这几个集市了。

村里的小学是一栋危危欲坠的生砖瓦片房，里面有四个年级的学生，每两个年级共享一个教室。甚至有一年，学校不得不搬到距离村庄很远的大山里的林场的废弃房子。学生们每天步行一个小时才能到达学校，吃着自带的白粥和咸菜作为午餐。全校只有两位老师，仅能教授语文和数学两门课程。我从小内向，小学就是我与外界接触的地方。父母每天早出晚归在田野、山间劳作，每天放学回家后，我需要生火做饭、喂猪、喂鸡，带妹妹，或者去地里摘桑叶养蚕，尽管我也很希望能像其他孩子一样看动画片、戏耍。很庆幸的是，家里还有一些破旧的、父亲和叔叔们保留下来的连环画和故事书，它们是我了解这个世界的重要途径和精神粮食。

五六年级的时候，我去了乡里的中心小学上学。每天早上随便吃点前一天的剩饭剩菜之后就得出门，需要走路四十分钟才能到达学校。每年刮风下雨、发洪水的时候，路是最难走的，泥泞不堪，常常会弄脏衣服和鞋子。

后来，我去了县城读初中、市里读高中。由于路途太遥远和交通不便，平时只能住校，每个学期回家一两次。再后来，我去武汉上了大学，成为了我们村最早考上大学的几个学生之一。毕业后，我先后在深圳、北京工作了一些年。最后，我来到了德国读博士。

如果从上小学开始算，我花了4年多的时间走出了生我养我的小村庄，6年的时间去到了县城，9年的时间去到了梧州市，12年的时间去到了距离家乡1000多公里以外的武汉，22年的时间来到了距离家乡13万多公里的德国。这一路上，我遇到了很多友好的人，包括学校里的老师和同学、门卫大叔、同事和朋友，以及很多认识的和不认识的人，他们给予了我很多的帮助和支持。家人的无私付出、国家的教育资助政策，使得我能够顺利读完大学。

过去的20多年间，我们的村庄也发生了翻天覆地的变化。村里的泥土路变成了水泥路，安装了路灯，大部分家庭新建了现代化的平顶房，拥有了彩色电视机和冰箱，很多家庭开上了小车和摩托车，几乎人人都有手机。25年前，村里的小卖部开始有了第一部有线电话；20年前，村庄下游修建了长洲水利枢纽，西江两岸的人民再也不用担心洪水的肆虐；15年前，梧州市结束了没有火车的历史，我坐着绿皮火车，花16个小时去1000多公里之外的武汉上大学；10年前，梧州市开通了高铁站，结束了我需要坐8个小时大巴车从500公里远的深圳回家的历史；5年前，又开通了飞机场，从此3个小时可以抵达2200公里之外的北京。高铁

站和飞机场距离我们村均不到 20 公里，我也终于实现了儿童时代的顺口谣：“大飞机停一停，带我去北京”。与此同时，村里接通了互联网络，成为了“地球村”的一隅。

我成长的步伐总是比别人慢一些。我并不聪慧，小时候常常羡慕那些能够拿到奖状的同学。因为家境不好，我内心充满了自卑、害羞和胆怯。在学校里，我很少有勇气表达自己，更不敢去追求那些深藏心底的渴望。我常常羡慕那些家庭条件优越、学习成绩出色的同学，羡慕他们的自信、多才多艺和从小便能接触到的丰富世界。

但我又是幸运的。在村里的小学班上，只有 13 名同学，其中许多人因各种原因早早辍学，而我成为了极少数完成初中学业的人之一。小学升初中时，我幸运地成为全校仅有的 7 名考上县重点初中的学生之一。这在我们当地是前所未有的事情。中考成绩公布的那天，一位亲戚骑着摩托车以 120 公里的时速，及时将我送到了市重点高中的报名点，让我成为最后一个完成报名的学生。在那里，我遇到了关爱学生的老师和真挚的同学。大三时，一位老师鼓励我报名参加华大基因创新班的面试，这次机遇让我得以走进曾代表中国参与人类基因组计划的华大基因。在那里，我不仅得到了几位恩师——周欣老师、刘山林老师和唐敏老师——的悉心指导，还结识了一群充满朝气的小伙伴。他们让我第一次体会到学术与实践结合的乐趣，也让我明白了团队合作的重要性。来到德国后，我的导师、同学和房东给予了我极大的帮助，让我逐渐适应了全新的环境。在这里，我还结识了一些挚友，他们的陪伴让我感受到温暖与支持。

考虑到上述经历，我尤其感谢德国科学基金会（DFG）资助了我的博士研究岗位，感谢我的导师 Bernhard A. Huber 博士、Lars Podsiadlowski 博士和 Dimitar Dimitrov 教授，在整个研究项目中提供了宝贵的指导、建设性的反馈、建议和鼓励。他们的专业知识和坚定支持在这篇论文写作过程中起到了关键作用。我还要感谢莱布尼茨生物多样性变化研究所（LIB）和波恩大学提供了支持本研究所所需的资源和设施。此外，我也非常感激 Sandra Kukowka、Anja Bodenheimer、Jonas Eberle 和 Hans-Joachim Krammer 在本研究的各个方面提供的技术支持和贡献。最后，我也非常感谢我的博士答辩委员会成员：Bernhard Misof 教授，Lars Podsiadlowski 博士，Nicolas Gompel 教授和 Thomas Döring 教授。感谢他们在百忙之中审核了我的论文和作为评委出席答辩。

特别感谢我的家人和朋友，以及我之前的同事和导师，周欣教授、刘山林教授、唐敏博士、周程冉博士和杨琛涛博士，他们在我的学术旅程中给予了我无私的指导、充分的理解、鼓励和始终如一的支持。

最后，感谢自己的持之以恒和一直以来对未知世界的好奇心。

Supplementary information for Chapter 3

Supplementary text

A total of 96 specimens (Table A.1) covering all the seven subfamily-level groups (i.e. Arteminae, Ninetinae, Modisiminae, Smeringopinae and Pholcinae, *Artema*, *Priscula*) of Pholcidae were collected with local permits.

DNA extraction. Tissues were stored in non-denatured pure ethanol (~99%) at -20° C before DNA extraction. All extracted genomic DNA is deposited at the LIB Biobank, Museum Koenig, Bonn. DNA of the 95 specimens was isolated with Dneasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instruction. DNA concentrations were evaluated with Quantus™ Fluorometer (Cat # E6150, Promega Corporation, USA), and fragment size ranges were determined with Fragment Analyzer (Advanced Analytical Technologies, Inc) and HS Genomic DNA Kit (Agilent Technologies, USA), respectively, to ensure that extracted DNA meets the requirements of subsequent library constructions.

Library preparation, enrichment and sequencing. For each specimen, a total of 100 ng DNA was used for library preparation. If the specimen had a total mass of DNA less than 100 ng, we would use all the DNA instead.

DNA standardization. Different specimens may have different DNA concentrations. To ensure all specimens have 100 ng of DNA and 14 µl in volume, we may evaporate the excessive water using a Savant SpeedVac® SPD111V (ThermoFisher Scientific, Waltham, MA; USA) under 40 °C, and/or add nuclease-free water to dilute the DNA.

DNA fragmentation. Next, Bioruptor® Pico sonication device (Diagenode S.A.) was used to fragment DNA following the manufacturer's instruction. Depending on the original DNA fragment size ranges, 12, 24 (12+12), 36 (12+12+12) or 45 (15+15+15) cycles were applied, with one cycle consisting of 30s sonication on and 30s sonication off. To achieve optimum DNA fragment size ranges of all DNA, after every 12 (or 15) cycles, the DNA was mixed and briefly spinned using a centrifuge (VMR international Galaxy Mini Centrifuge C1413V). Subsequently, Fragment Analyzer (Advanced Analytical Technologies, Inc) and HS NGS Fragment (1-6000bp) Kit (Cat # 5191-6578, Lot # 0006481688, Agilent Technologies, USA) were used to determine fragment size ranges according to the manufacturer's protocol. We repeated the fragmentation process for each specimen if necessary, until >90% of the DNA had fragment size in the range of 1 to 500 bp in length.

FFPE DNA repair. For two pools (pool A and pool B) of samples, due to their low DNA quality, we applied FFPE DNA repair procedure using the NEBNext® FFPE DNA Repair Mix (Cat # M6630L,

New England Biolabs, Inc) following the manufacturer's protocol, to fix the potential DNA damages (e.g. caused by sample storage conditions) and thus improve subsequent library quality and quality of sequencing data. The repaired DNA were then purified with AMPure® XP Beads (Beckman Coulter, Inc. #A63881) and 80% ethanol following the manufacturer's instruction, the volume ratio of beads to DNA sample was 3:1.

End repair, dA-Tailing and ligation of adaptors. The sheared DNA were then subject to end-repair, dA-tailing and then the ligation of molecular-barcoded adaptors following manufacturer's protocols, using SureSelect XT HS2 Library Preparation Kit (Cat # 5500-0147, Lot # 0006583499, Agilent Technologies, USA). Next, the DNA were purified with AMPure® XP Beads (Beckman Coulter, Inc. #A63881) and 80% ethanol, the volume ratio of beads to DNA sample was 1:0.8. Next, DNA concentrations of all specimens were evaluated with Quantus™ Fluorometer (Cat # E6150, Promega Corporation, USA), respectively. After that, the adaptor-ligated libraries were amplified using the SureSelect XT HS2 Library Preparation Kit (Cat # 5500-0147, Lot # 0006583499, Agilent Technologies, USA) and SureSelect XT HS2 Index Primer Pairs 97-192 (Cat # 5191-5689, Lot # 0006613734), 9 or 16 PCR cycles were applied based on experience. Lastly, the indexed PCR products were purified with AMPure® XP Beads (Beckman Coulter, Inc. #A63881) and 80% ethanol, the volume ratio of beads to DNA sample was 1:0.8, followed by DNA concentration evaluation with Quantus™ Fluorometer (Cat # E6150, Promega Corporation, USA). Pooling of indexed PCR products, hybridization and capture. The indexed PCR products were pooled together with equal mass (93.75 ng) for every 8 specimens, resulting in 750 ng of total gDNA in each pool. Subsequently, the indexed gDNA library pools were hybridized to our designed target-specific probes using the SureSelect XT HS2 Target Enrichment Kit (Cat # 5191-6690, Lot # 0006599172, Agilent Technologies, USA) following the manufacturer's instruction. The hybridized DNA were then captured using streptavidin-coated beads from the SureSelect XT HS2 Target Enrichment Kit (Cat # 5191-6690, Lot # 0006599172, Agilent Technologies, USA).

Post-capture sample processing. The SureSelect-enriched DNA libraries were then PCR amplified using the SureSelect XT HS2 Target Enrichment Kit (Cat # 5191-6690, Lot # 0006599172, Agilent Technologies, USA) and a PCR cycle number of 20. Subsequently, the PCR products were purified with AMPure® XP Beads (Beckman Coulter, Inc. #A63881) and 80% ethanol following the manufacturer's instruction, the volume ratio of beads to DNA sample was 1:0.8. Lastly, DNA concentration evaluations were finished with Quantus™ Fluorometer (Cat # E6150, Promega Corporation, USA) and fragment size distributions were analyzed with Fragment Analyzer (Advanced Analytical Technologies, Inc) and HS NGS Fragment (1-6000bp) Kit (Cat # 5191-6578, Lot # 0006481688, Agilent Technologies, USA).

Sequencing. Sequencing of captured DNA libraries were performed by MacroGen Europe (Amsterdam, NL) using pair-end (PE) 150 bp strategy. Each specimen was sequenced with ca. 1.5 Gb of data.

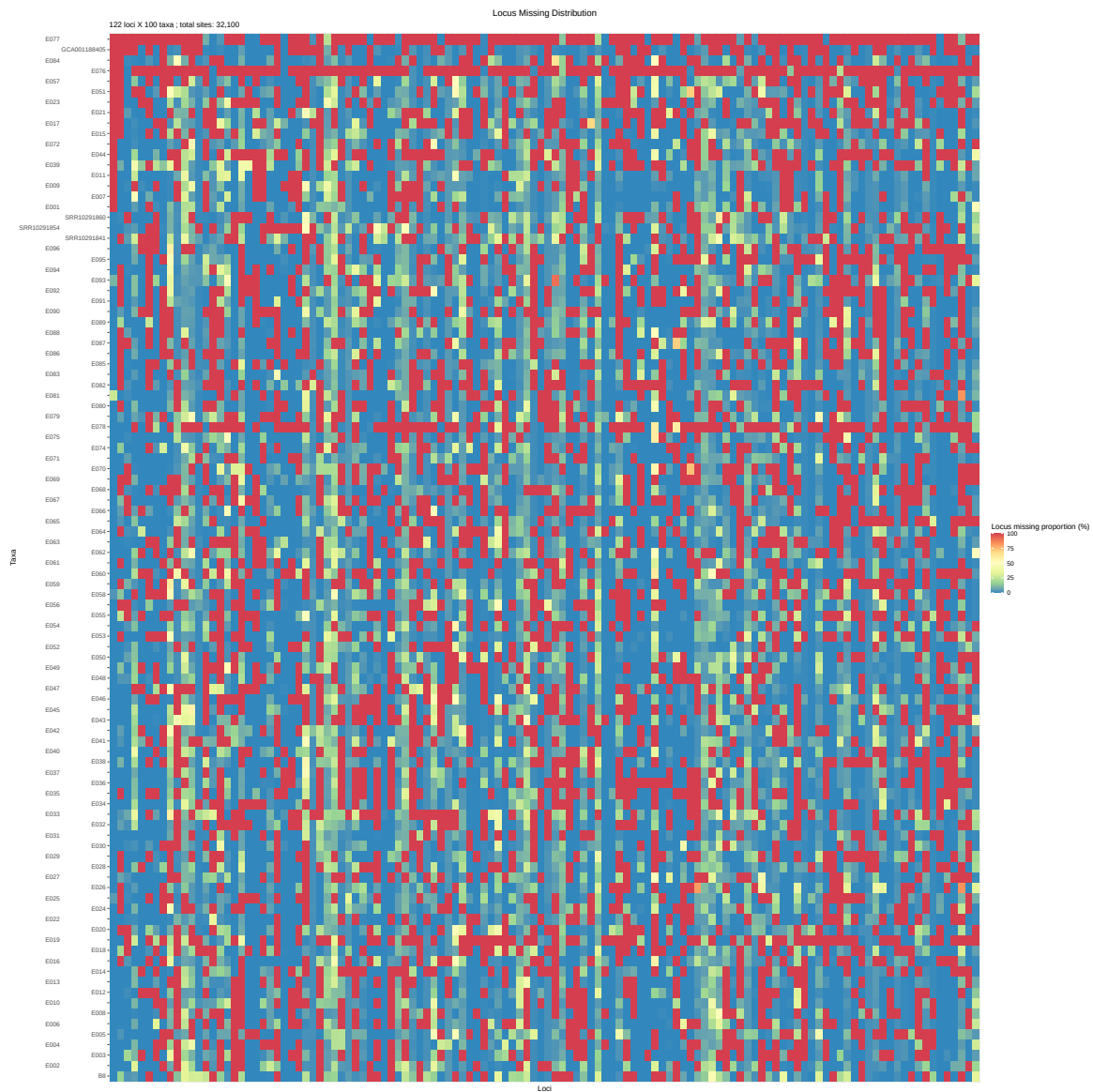


Figure A.1. Missing data distribution of matrix 3. Blue and red color indicate low and high level of missing data, respectively.

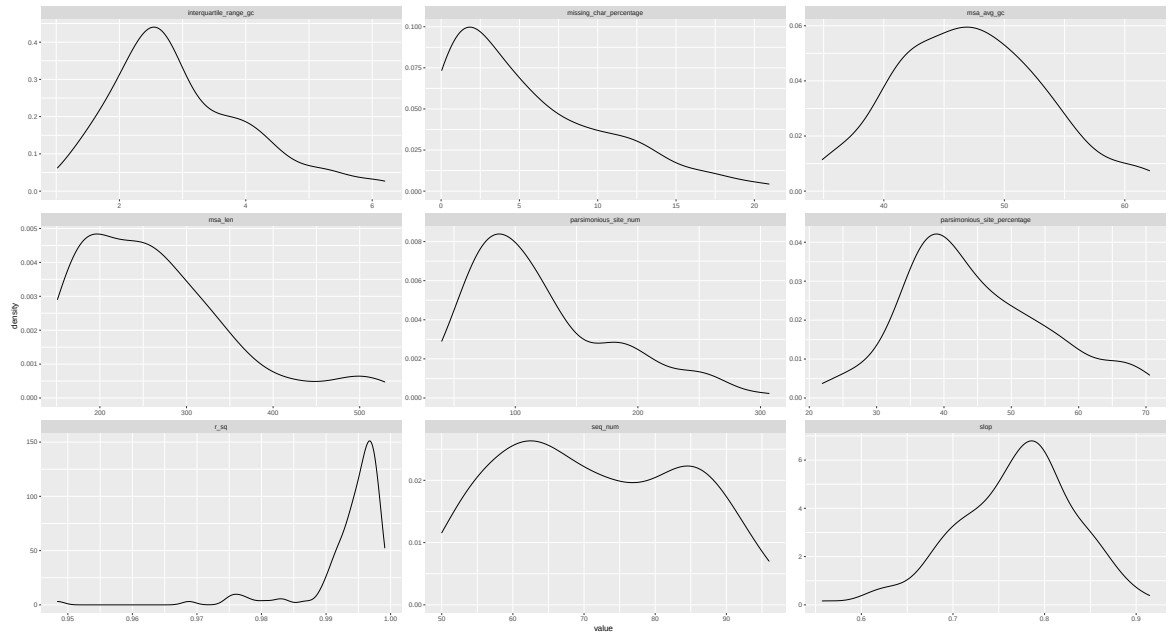


Figure A.2. Density distribution of statistics of matrix 3. For each locus, we calculated its interquartile range of GC content, the number and proportion of missing data (i.e., gaps and Ns), average GC content, alignment length (in bp), the number and proportion of parsimonious sites, the coefficient of linear regression fitting between the p-distance and the corrected genetic distance (with the Kimura-2-Parameter model), the taxon number in the alignment, and fitted slop between the p-distance and the corrected genetic distance.

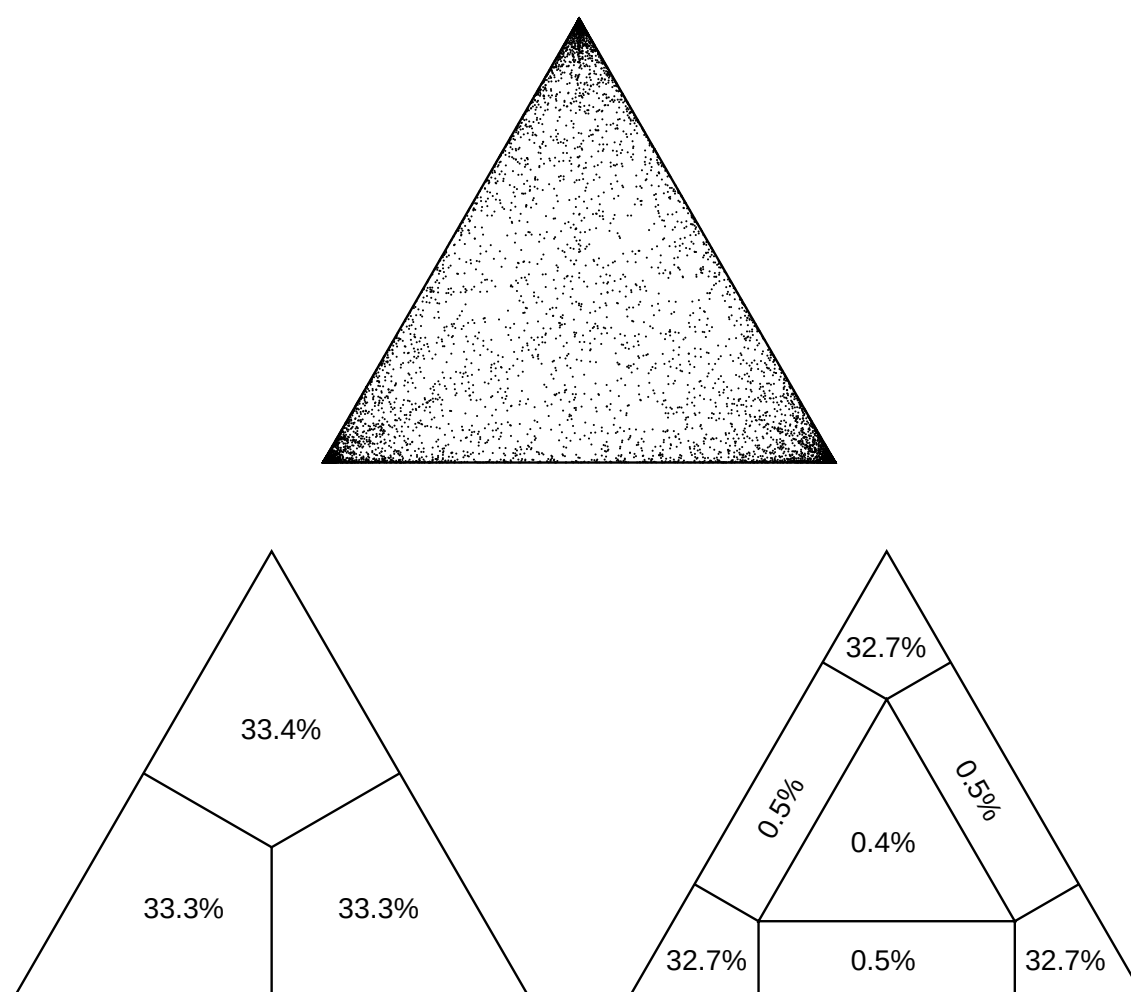


Figure A.3. Likelihood mapping of matrix 3. The likelihood mapping method (Strimmer and von Haeseler 1997) in IQ-TREE (v2.2.0) (Minh et al. 2020) was used and 100,000 random quartets of taxa were drawn from the concatenated MSA of all loci and visualized in a single triangular graph. The more quartets located at the three corners of the triangular graph, the more suitable the data is to construct a well-resolved phylogeny (Strimmer and von Haeseler 1997).

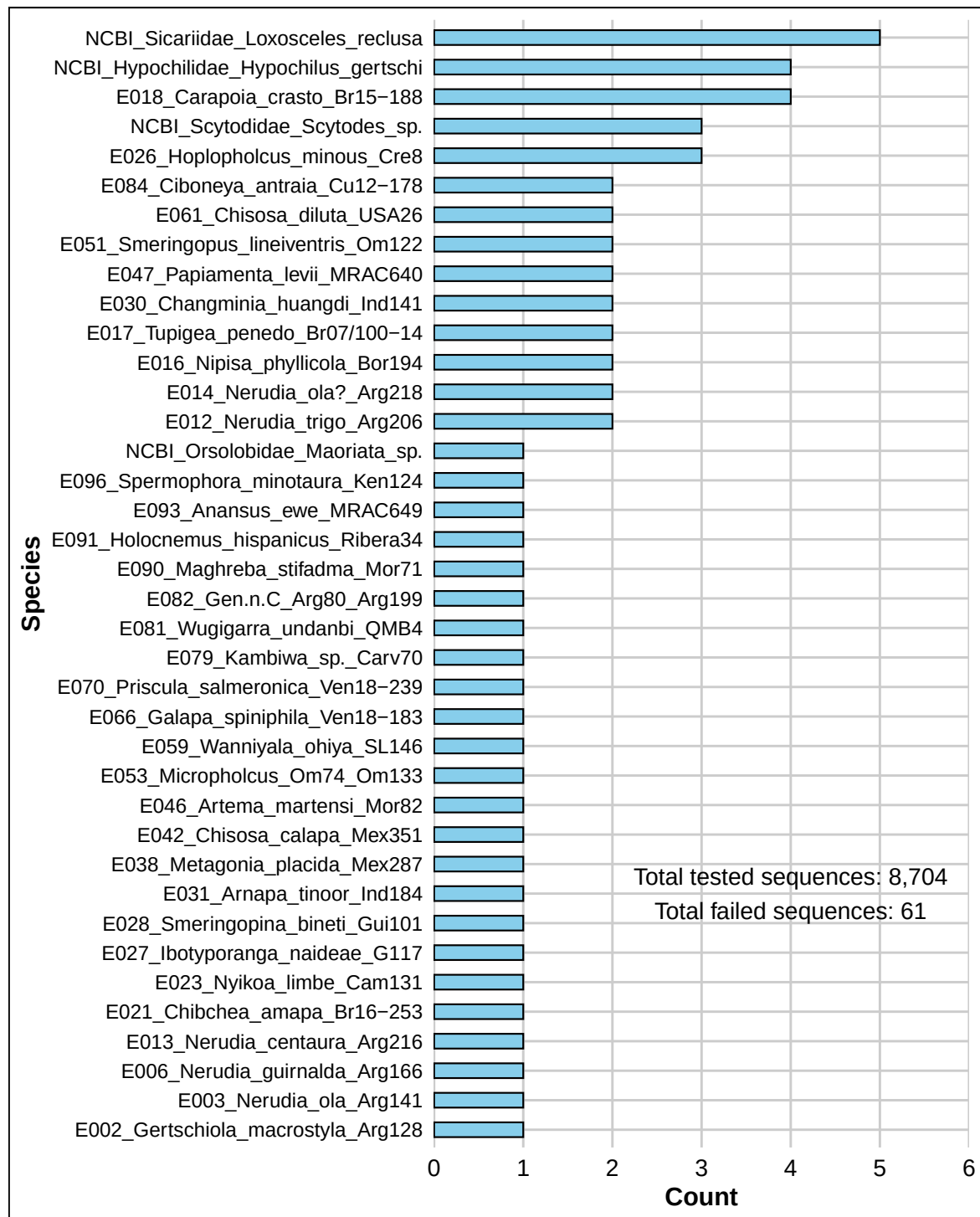


Figure A.4. Compositional homogeneity test of matrix 3. IQ-TREE (v2.2.0) (Minh et al. 2020) was used to perform a composition chi-square test for every sequence in matrix 3. A sequence is labeled as “failed” if its character composition differs significantly from the average composition of the alignment. This plot shows only the number of failed sequences in all 8,704 tests.

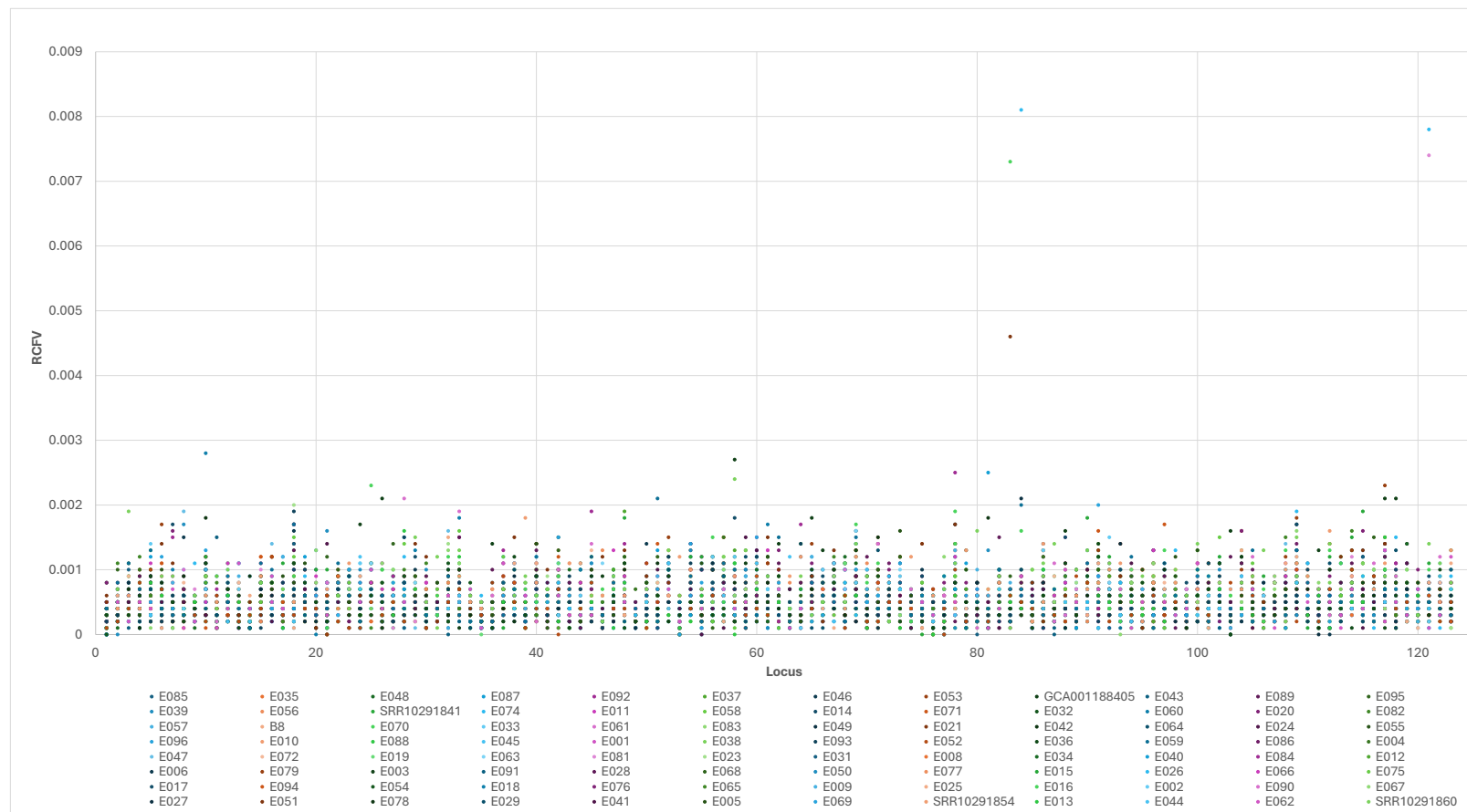


Figure A.5. Gene-wise RCFV distribution based on BaCoCa analysis

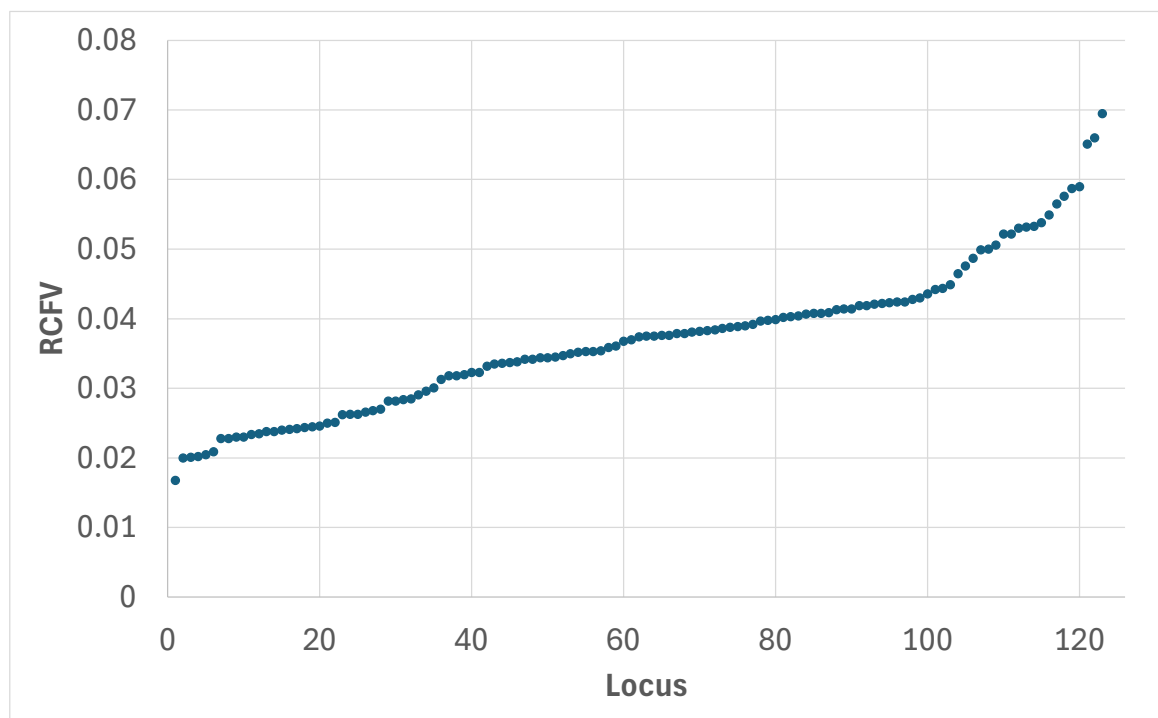


Figure A.6. Taxon-wise RCFV distribution based on BaCoCa analysis

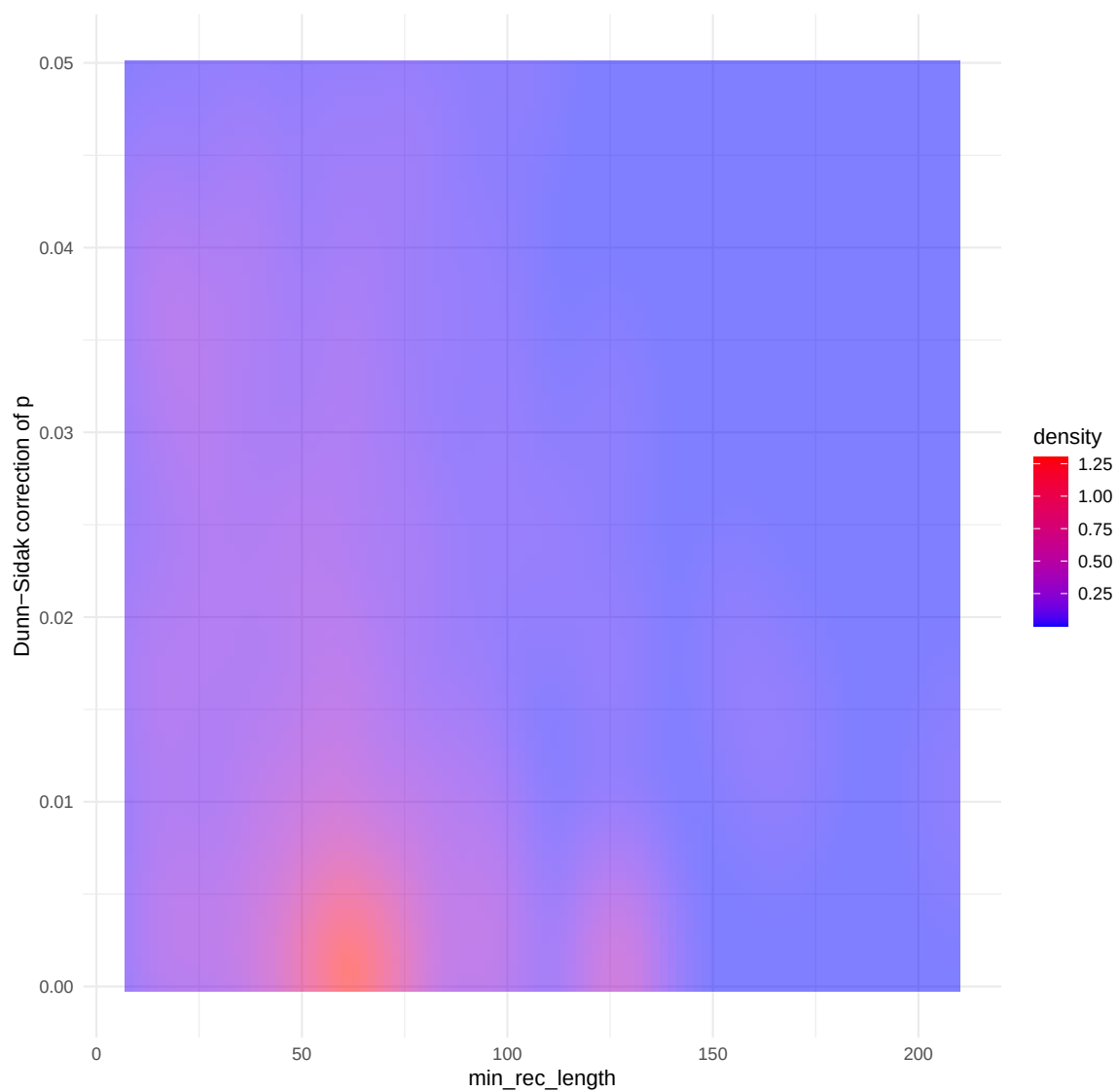


Figure A.7. Intra-locus aecombination analysis based on data matrix 3. min_rec_length: the minimum length of the recombinant segments; i.e. the minimum length of the shorter of the two recombinant segments

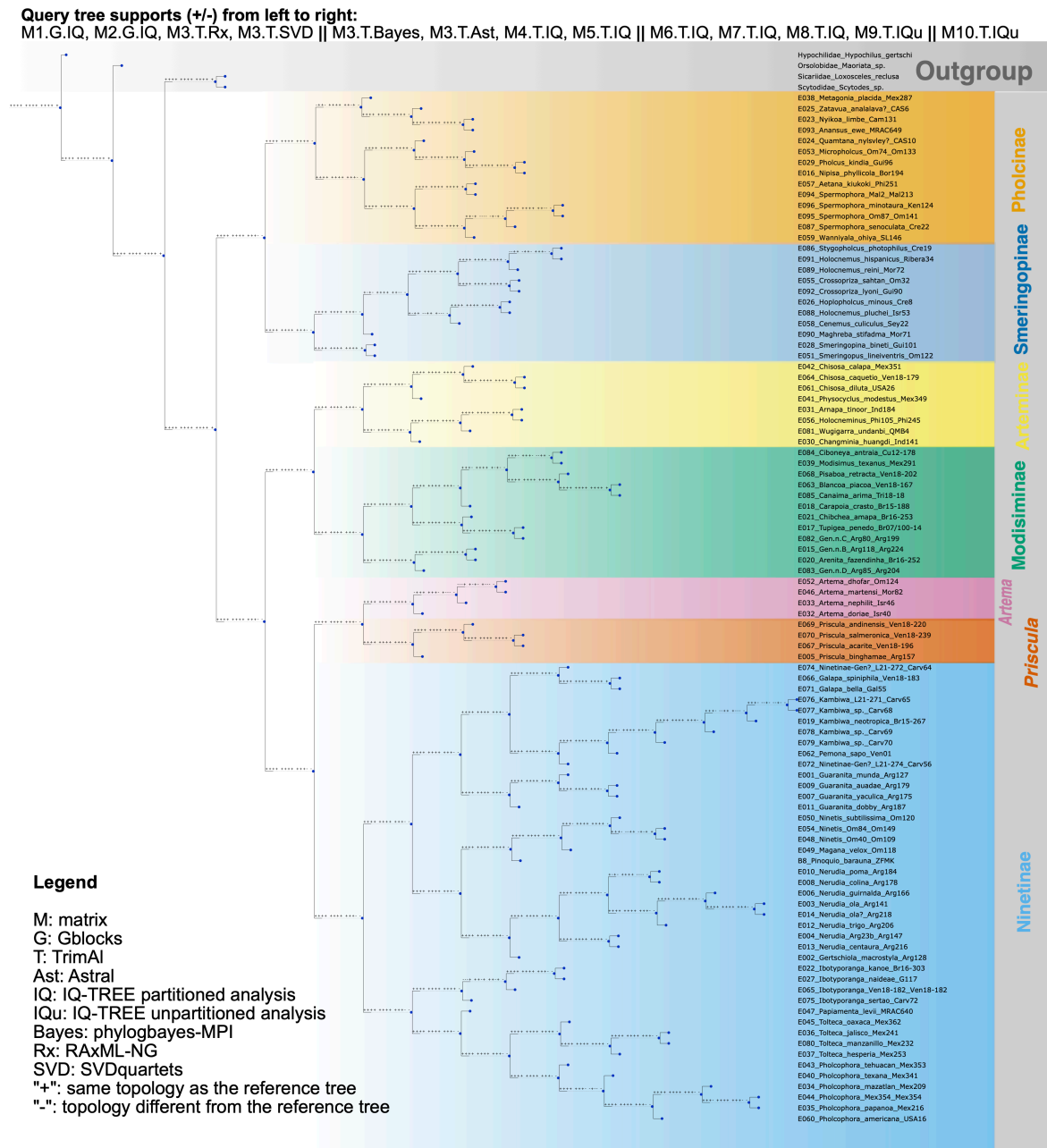


Figure A.8. Congruence and incongruence between phylogenetic trees. This figure summarizes the topological similarities and differences between trees constructed from matrices 1 to 10, using the alignment trimming methods TrimAl and Gblocks, along with various tree construction methods. The comparison is made between each of the 'query' trees and the reference tree (i.e., Figure 1). A '+' indicates that the query tree shares the same topology as the reference tree, while a '-' indicates a difference in topology.

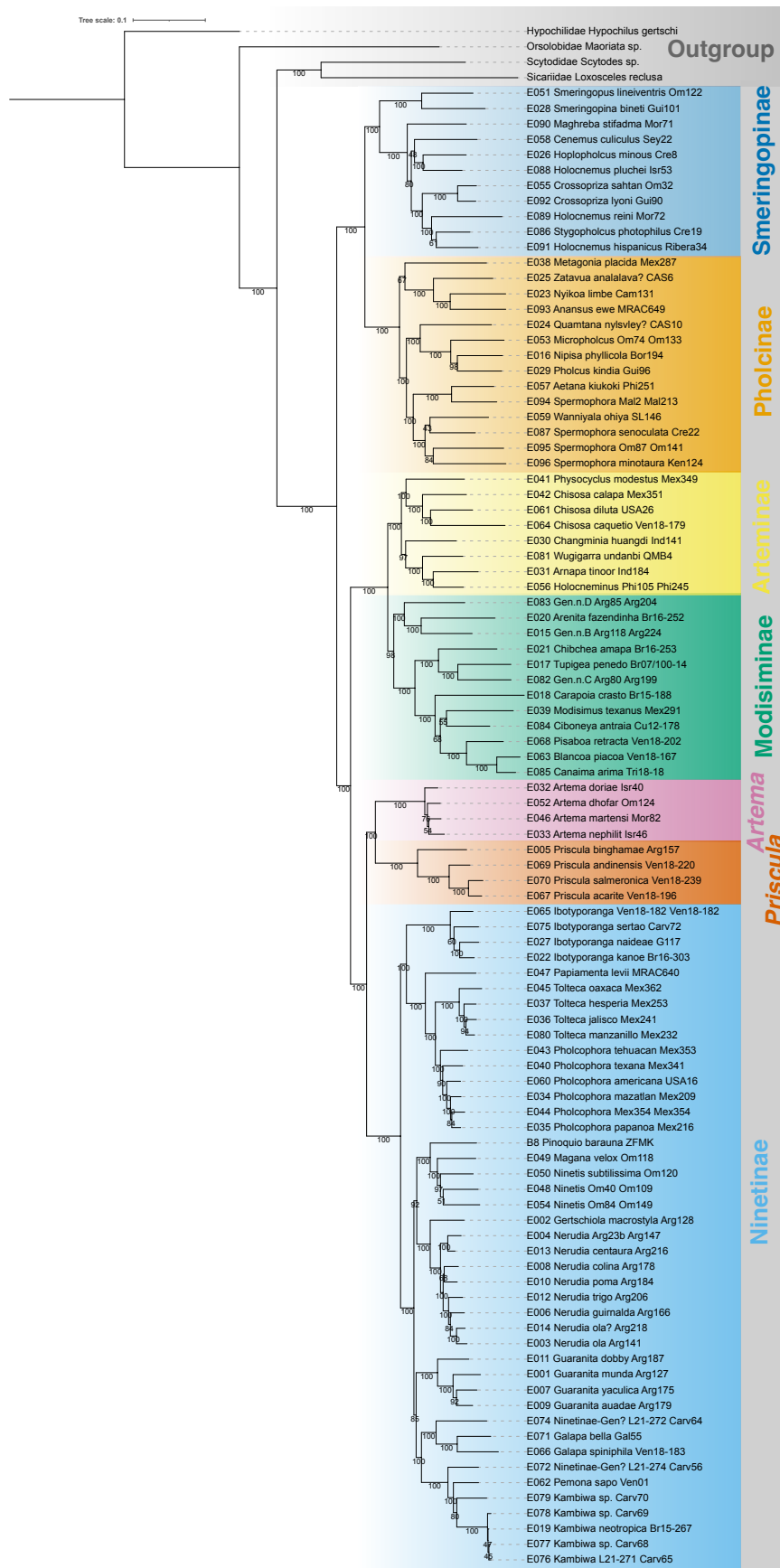


Figure A.9. Matrix-3 maximum likelihood tree from RAxML-NG partitioned analysis. Numbers on the branches are bootstrap supports (%). The GTR+R4+I+F0 model, and 25 parsimony starting trees + 25 random starting trees were used, and the number of bootstrap replicates was automatically determined (--bs-trees autoMRE) (Pattengale et al. 2010).

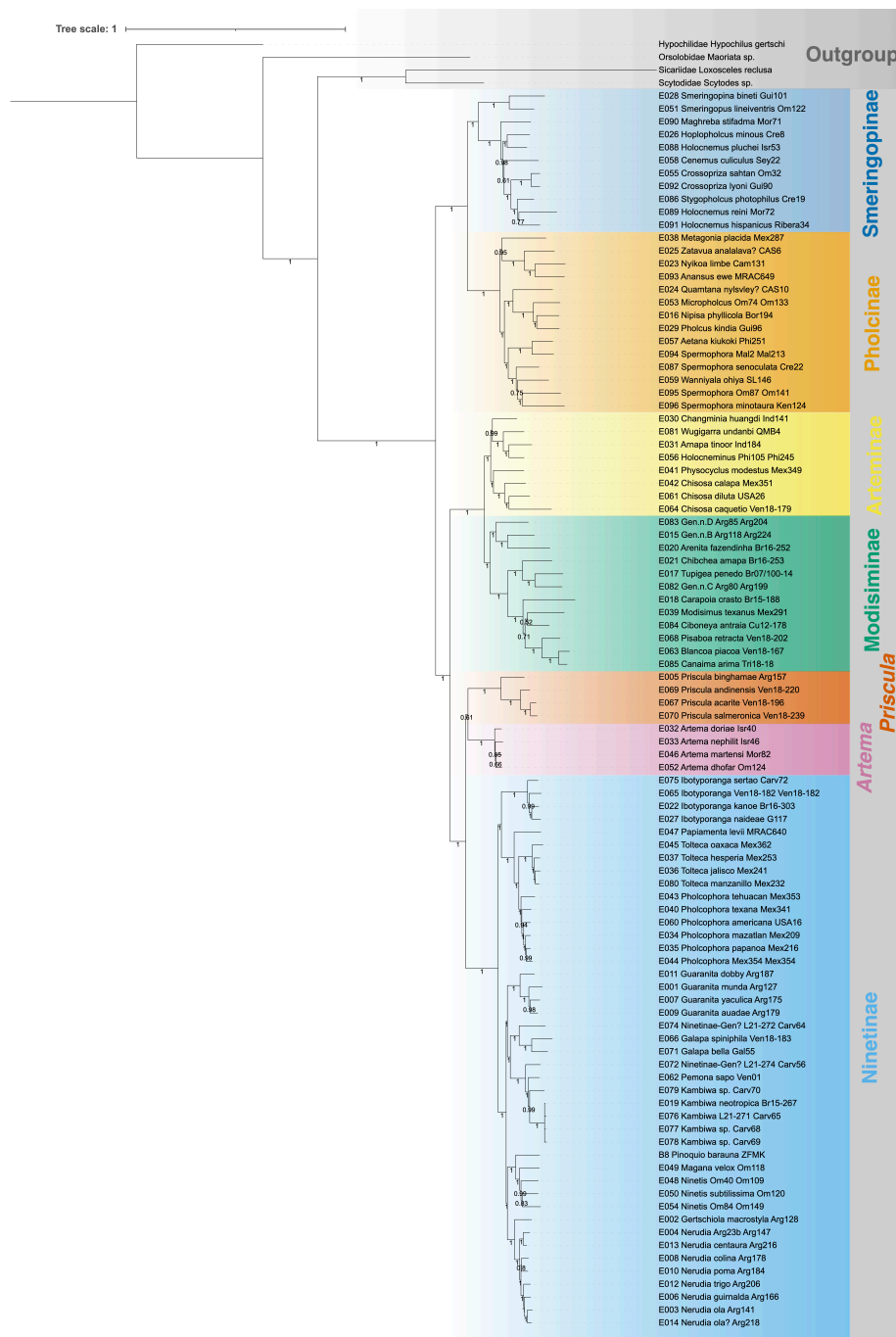


Figure A.10. Matrix-3 Bayesian tree from phylogbayes-MPI analysis. Numbers on the branches are posterior probabilities. The CAT-GTR+G4 model (Lartillot and Philippe 2004) was used, which can model both rate-heterogeneity across sites and pattern-heterogeneity across sites (i.e., across-site compositional heterogeneity), and is less sensitive to systematic errors than classical empirical models (Lartillot et al. 2013). Two independent PhyloBayes-MPI runs were performed, and convergences of the continuous parameters of the model as well as the tree spaces were evaluated using the `tracecomp` and `bpcomp` commands in PhyloBayes-MPI. One-third of samplings of each run was used as burn-in, and a majority consensus tree was generated by combining the sampling trees of the two runs using the `bpcomp` command.

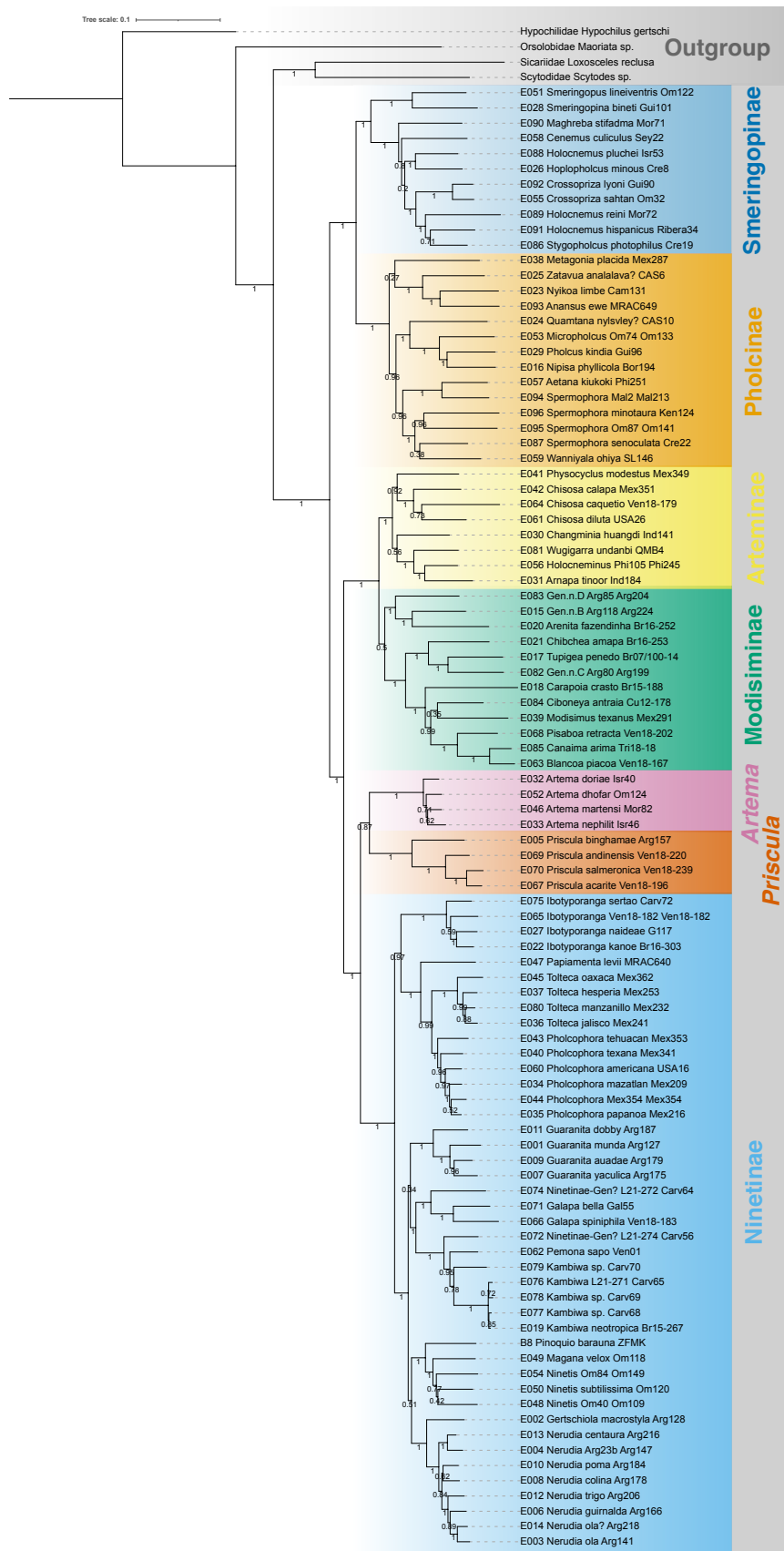


Figure A.11. Matrix-3 coalescence tree from ASTRAL analysis. Numbers on the branches are posterior probabilities. The individual gene trees were firstly estimated by IQ-TREE (v2.2.0) (Minh et al. 2020) and then summarized by ASTRAL (v5.7.8) (Zhang et al. 2018) with default parameters. To reduce noise, internal nodes of gene trees with low branch supports (≤ 30) were contracted (Zhang, Sayyari and Mirarab 2017).

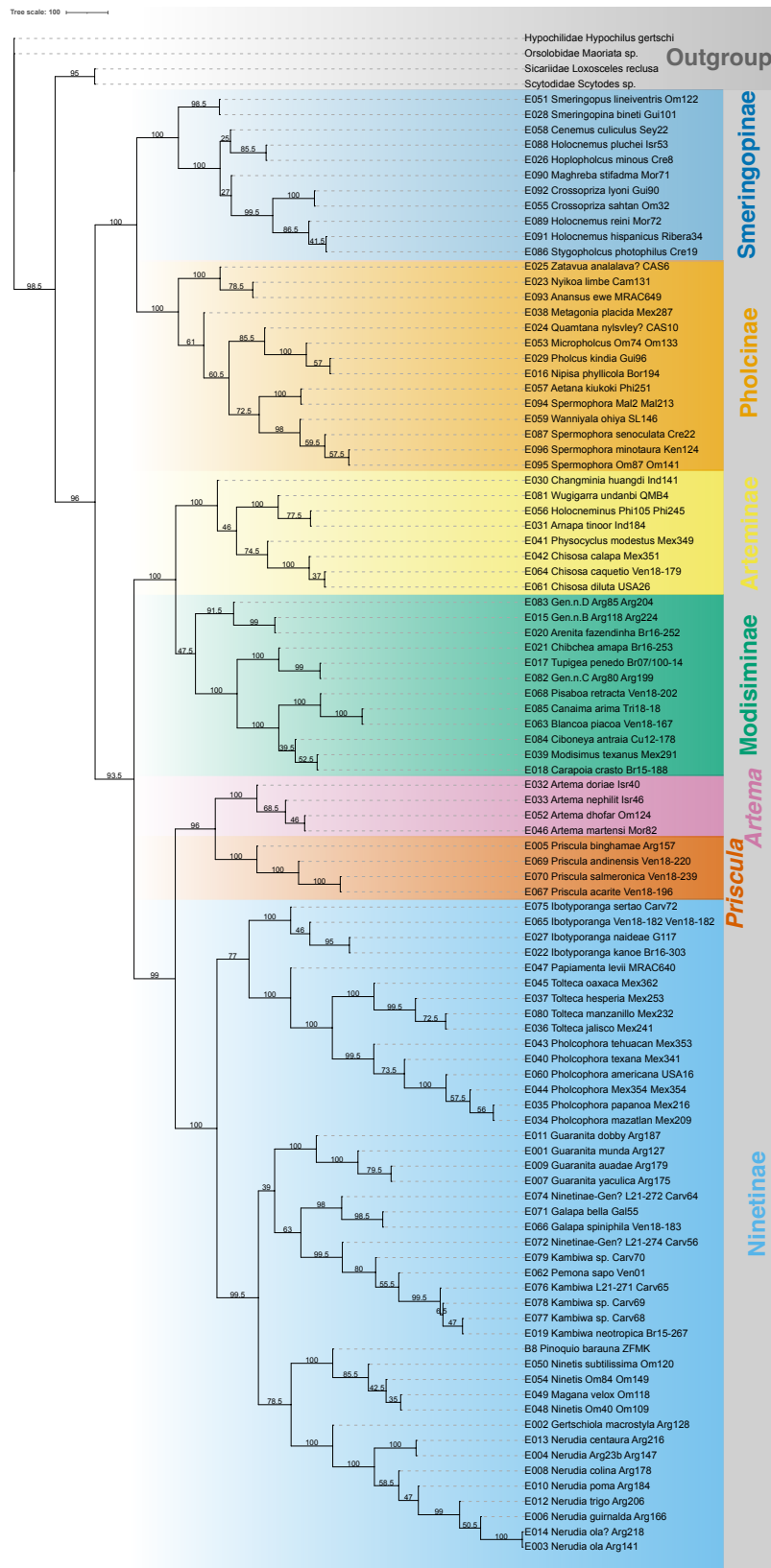


Figure A.12. Matrix-3 coalescence tree from SVDquartets analysis. Numbers on the branches are bootstrap supports. We evaluated all quartets (`evalQuartets=all`) under the multispecies coalescent model (`tree-model=mscoalescent`) to allow for gene tree discordance due to incomplete lineage sorting, the “Erik+2” normalization was used (`plus2=yes`) to consider the heterogeneity among sites and lineages (Fernandez-Sanchez and Casanellas 2016). The QFM algorithm (Reaz, Bayzid and Rahman 2014) was used to search for an optimal tree under the maximum quartet compatibility algorithm (`treeInf=QFM`), and 200 multilocus bootstrap replicates were sampled (`bootstrap=multilocus nreps=200`).

Table A.1. Samples for target capture sequencing and outgroups

Sample ID	Family	Subfamily	Bioproject	Biosample	SRA	Volume (Gbp)
E001_Guaranita_munda_Arg127	Pholcidae	Ninetinae	PRJNA851037	SAMN29208292	SRR19792659	1.57
E002_Gertschiola_macrostyla_Arg128	Pholcidae	Ninetinae	PRJNA851037	SAMN29208293	SRR19792658	1.30
E003_Nerudia_ola_Arg141	Pholcidae	Ninetinae	PRJNA851037	SAMN29208294	SRR19792647	1.33
E004_Nerudia_Arg23b_Arg147	Pholcidae	Ninetinae	PRJNA851037	SAMN29208295	SRR19792636	1.21
E005_Priscula_binghamae_Arg157	Pholcidae	Modisiminae	PRJNA851037	SAMN29208296	SRR19792625	1.48
E006_Nerudia_guirnalda_Arg166	Pholcidae	Ninetinae	PRJNA851037	SAMN29208297	SRR19792614	1.45
E007_Guaranita_yaculica_Arg175	Pholcidae	Ninetinae	PRJNA851037	SAMN29208298	SRR19792603	1.45
E008_Nerudia_colina_Arg178	Pholcidae	Ninetinae	PRJNA851037	SAMN29208299	SRR19792592	1.48
E009_Guaranita_auadae_Arg179	Pholcidae	Ninetinae	PRJNA851037	SAMN29208300	SRR19792581	1.60
E010_Nerudia_poma_Arg184	Pholcidae	Ninetinae	PRJNA851037	SAMN29208301	SRR19792570	1.60
E011_Guaranita_dobby_Arg187	Pholcidae	Ninetinae	PRJNA851037	SAMN29208302	SRR19792657	1.12
E012_Nerudia_trigo_Arg206	Pholcidae	Ninetinae	PRJNA851037	SAMN29208303	SRR19792656	1.81
E013_Nerudia_centaura_Arg216	Pholcidae	Ninetinae	PRJNA851037	SAMN29208304	SRR19792655	1.57
E014_Nerudia_ola?_Arg218	Pholcidae	Ninetinae	PRJNA851037	SAMN29208305	SRR19792654	1.57
E015_Gen.n.B_Arg118_Arg224	Pholcidae	Modisiminae	PRJNA851037	SAMN29208306	SRR19792653	1.48
E016_Nipisa_phyllicola_Bor194	Pholcidae	Pholcinae	PRJNA851037	SAMN29208307	SRR19792652	1.90
E017_Tupigea_penedo_Br07/100-14	Pholcidae	Modisiminae	PRJNA851037	SAMN29208308	SRR19792651	1.42
E018_Carapoia_craeto_Br15-188	Pholcidae	Modisiminae	PRJNA851037	SAMN29208309	SRR19792650	1.72
E019_Kambiwa_neotropica_Br15-267	Pholcidae	Ninetinae	PRJNA851037	SAMN29208310	SRR19792649	1.36
E020_Arenita_fazendinha_Br16-252	Pholcidae	Modisiminae	PRJNA851037	SAMN29208311	SRR19792648	1.39
E021_Chibchea_amapa_Br16-253	Pholcidae	Modisiminae	PRJNA851037	SAMN29208312	SRR19792646	1.63
E022_Ibotyporanga_kanoe_Br16-303	Pholcidae	Ninetinae	PRJNA851037	SAMN29208313	SRR19792645	1.30
E023_Nyikoa_limbe_Cam131	Pholcidae	Pholcinae	PRJNA851037	SAMN29208314	SRR19792644	1.09
E024_Quamtana_nylsvley?_CAS10	Pholcidae	Pholcinae	PRJNA851037	SAMN29208315	SRR19792643	1.33
E025_Zatavua_analalava?_CAS6	Pholcidae	Pholcinae	PRJNA851037	SAMN29208316	SRR19792642	0.97
E026_Hoplopholcus_minous_Cre8	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208317	SRR19792641	3.74
E027_Ibotyporanga_naideae_G117	Pholcidae	Ninetinae	PRJNA851037	SAMN29208318	SRR19792640	1.06
E028_Smeringopina_bineti_Gui101	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208319	SRR19792639	1.24
E029_Pholcus_kindia_Gui96	Pholcidae	Pholcinae	PRJNA851037	SAMN29208320	SRR19792638	2.30
E030_Changminia_huangdi_Ind141	Pholcidae	Arteminae	PRJNA851037	SAMN29208321	SRR19792637	2.45
E031_Arnapa_tinoor_Ind184	Pholcidae	Arteminae	PRJNA851037	SAMN29208322	SRR19792635	1.30
E032_Artema_doriae_Isr40	Pholcidae	Arteminae	PRJNA851037	SAMN29208323	SRR19792634	2.51
E033_Artema_nephilit_Isr46	Pholcidae	Arteminae	PRJNA851037	SAMN29208324	SRR19792633	2.87
E034_Pholcophora_mazatlan_Mex209	Pholcidae	Ninetinae	PRJNA851037	SAMN29208325	SRR19792632	1.18
E035_Pholcophora_papanao_Mex216	Pholcidae	Ninetinae	PRJNA851037	SAMN29208326	SRR19792631	1.21
E036_Tolteca_jalisco_Mex241	Pholcidae	Ninetinae	PRJNA851037	SAMN29208327	SRR19792630	1.03
E037_Tolteca_hesperia_Mex253	Pholcidae	Ninetinae	PRJNA851037	SAMN29208328	SRR19792629	0.97
E038_Metagonia_placida_Mex287	Pholcidae	Pholcinae	PRJNA851037	SAMN29208329	SRR19792628	1.36
E039_Modisimus_texanus_Mex291	Pholcidae	Modisiminae	PRJNA851037	SAMN29208330	SRR19792627	1.48

Continued on next page

Sample ID	Family	Subfamily	Bioproject	Biosample	SRA	Volume (Gbp)
E040_Pholcophora_texana_Mex341	Pholcidae	Ninetinae	PRJNA851037	SAMN29208331	SRR19792626	1.45
E041_Physocylus_modestus_Mex349	Pholcidae	Arteminae	PRJNA851037	SAMN29208332	SRR19792624	1.90
E042_Chisosa_calapa_Mex351	Pholcidae	Arteminae	PRJNA851037	SAMN29208333	SRR19792623	1.66
E043_Pholcophora_tehuacan_Mex353	Pholcidae	Ninetinae	PRJNA851037	SAMN29208334	SRR19792622	1.51
E044_Pholcophora_Mex354_Mex354	Pholcidae	Ninetinae	PRJNA851037	SAMN29208335	SRR19792621	1.15
E045_Tolteca_oaxaca_Mex362	Pholcidae	Ninetinae	PRJNA851037	SAMN29208336	SRR19792620	1.36
E046_Artema_martensi_Mor82	Pholcidae	Arteminae	PRJNA851037	SAMN29208337	SRR19792619	3.02
E047_Papiamenta_levii_MRAC640	Pholcidae	Ninetinae	PRJNA851037	SAMN29208338	SRR19792618	1.51
E048_Ninetis_Om40_Om109	Pholcidae	Ninetinae	PRJNA851037	SAMN29208339	SRR19792617	1.21
E049_Magana_velox_Om118	Pholcidae	Ninetinae	PRJNA851037	SAMN29208340	SRR19792616	1.21
E050_Ninetis_subtilissima_Om120	Pholcidae	Ninetinae	PRJNA851037	SAMN29208341	SRR19792615	1.21
E051_Smeringopus_lineiventris_Om122	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208342	SRR19792613	1.54
E052_Artema_dhofar_Om124	Pholcidae	Arteminae	PRJNA851037	SAMN29208343	SRR19792612	2.69
E053_Micropholcus_Om74_Om133	Pholcidae	Pholcinae	PRJNA851037	SAMN29208344	SRR19792611	1.81
E054_Ninetis_Om84_Om149	Pholcidae	Ninetinae	PRJNA851037	SAMN29208345	SRR19792610	1.27
E055_Crossopriza_sahtan_Om32	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208346	SRR19792609	1.54
E056_Holocnemius_Phi105_Phi245	Pholcidae	Arteminae	PRJNA851037	SAMN29208347	SRR19792608	1.66
E057_Aetana_kiukoki_Phi251	Pholcidae	Pholcinae	PRJNA851037	SAMN29208348	SRR19792607	1.63
E058_Cenemus_culiculus_Sey22	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208349	SRR19792606	1.72
E059_Wanniyala_ohiya_SL146	Pholcidae	Pholcinae	PRJNA851037	SAMN29208350	SRR19792605	1.57
E060_Pholcophora_americanana_USA16	Pholcidae	Ninetinae	PRJNA851037	SAMN29208351	SRR19792604	1.45
E061_Chisosa_diluta_USA26	Pholcidae	Arteminae	PRJNA851037	SAMN29208352	SRR19792602	1.15
E062_Pemona_sapo_Ven01	Pholcidae	Ninetinae	PRJNA851037	SAMN29208353	SRR19792601	1.12
E063_Blancoa_piacoa_Ven18-167	Pholcidae	Modisiminae?	PRJNA851037	SAMN29208354	SRR19792600	1.27
E064_Chisosa_caquetio_Ven18-179	Pholcidae	Arteminae	PRJNA851037	SAMN29208355	SRR19792599	1.90
E065_Ibotyporanga_Ven18-182_Ven18-182	Pholcidae	Ninetinae	PRJNA851037	SAMN29208356	SRR19792598	1.18
E066_Galapa_spiniphila_Ven18-183	Pholcidae	Ninetinae	PRJNA851037	SAMN29208357	SRR19792597	1.42
E067_Priscula_acarite_Ven18-196	Pholcidae	Modisiminae	PRJNA851037	SAMN29208358	SRR19792596	1.39
E068_Pisaboa_retracta_Ven18-202	Pholcidae	Modisiminae	PRJNA851037	SAMN29208359	SRR19792595	1.42
E069_Priscula_andinensis_Ven18-220	Pholcidae	Modisiminae	PRJNA851037	SAMN29208360	SRR19792594	1.57
E070_Priscula_salmeronica_Ven18-239	Pholcidae	Modisiminae	PRJNA851037	SAMN29208361	SRR19792593	1.66
E071_Galapa_bella_Gal55	Pholcidae	Ninetinae	PRJNA851037	SAMN29208362	SRR19792591	1.45
E072_Ninetinae-Gen?_L21-274_Carv56	Pholcidae	Ninetinae	PRJNA851037	SAMN29208363	SRR19792590	1.72
E074_Ninetinae-Gen?_L21-272_Carv64	Pholcidae	Ninetinae	PRJNA851037	SAMN29208365	SRR19792588	1.09
E075_Ibotyporanga_sertao_Carv72	Pholcidae	Ninetinae	PRJNA851037	SAMN29208366	SRR19792587	1.57
E076_Kambiwa_L21-271_Carv65	Pholcidae	Ninetinae	PRJNA851037	SAMN29208367	SRR19792586	1.09
E077_Kambiwa_sp._Carv68	Pholcidae	Ninetinae	PRJNA851037	SAMN29208368	SRR19792585	1.21
E078_Kambiwa_sp._Carv69	Pholcidae	Ninetinae	PRJNA851037	SAMN29208369	SRR19792584	1.15
E079_Kambiwa_sp._Carv70	Pholcidae	Ninetinae	PRJNA851037	SAMN29208370	SRR19792583	1.42
E080_Tolteca_manzanillo_Mex232	Pholcidae	Ninetinae	PRJNA851037	SAMN29208371	SRR19792582	1.75
E081_Wugigarra_undambi_QMB4	Pholcidae	Arteminae	PRJNA851037	SAMN29208372	SRR19792580	1.06
E082_Gen.n.C_Arg80_Arg199	Pholcidae	Modisiminae	PRJNA851037	SAMN29208373	SRR19792579	1.45

Continued on next page

Sample ID	Family	Subfamily	Bioproject	Biosample	SRA	Volume (Gbp)
E083_Gen.n.D_Arg85_Arg204	Pholcidae	Modisiminae	PRJNA851037	SAMN29208374	SRR19792578	1.39
E084_Ciboneya_antraia_Cu12-178	Pholcidae	Modisiminae	PRJNA851037	SAMN29208375	SRR19792577	1.51
E085_Canaïma_arima_Tri18-18	Pholcidae	Modisiminae?	PRJNA851037	SAMN29208376	SRR19792576	1.72
E086_Stygopholcus_photophilus_Cre19	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208377	SRR19792575	1.66
E087_Spermophora_senoculata_Cre22	Pholcidae	Pholcinae	PRJNA851037	SAMN29208378	SRR19792574	1.48
E088_Holocnemus_pluchei_Isr53	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208379	SRR19792573	1.15
E089_Holocnemus_reini_Mor72	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208380	SRR19792572	1.54
E090_Maghreba_stifadma_Mor71	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208381	SRR19792571	1.24
E091_Holocnemus_hispanicus_Ribera34	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208382	SRR19792569	1.24
E092_Crossopriza_lyoni_Gui90	Pholcidae	Smeringopinae	PRJNA851037	SAMN29208383	SRR19792568	1.27
E093_Anansus_ewe_MRAC649	Pholcidae	Pholcinae	PRJNA851037	SAMN29208384	SRR19792567	1.42
E094_Spermophora_Mal2_Mal213	Pholcidae	Pholcinae	PRJNA851037	SAMN29208385	SRR19792566	1.87
E095_Spermophora_Om87_Om141	Pholcidae	Pholcinae	PRJNA851037	SAMN29208386	SRR19792565	1.48
E096_Spermophora_minotaura_Ken124	Pholcidae	Pholcinae	PRJNA851037	SAMN29208387	SRR19792564	1.42
B8_Pinoquio_baraua_ZFMK-Ar-20622	Pholcidae	Ninetinae	PRJNA851037	SAMN29208388	SRR19792563	0.88
NCBI_Scytodidae_Scytodes_sp.	Scytodidae	Kulkarni et al. 2020	PRJNA575576	SAMN13006101	SRR10291841	-
NCBI_Sicariidae_Loxosceles_reclusa	Sicariidae	i5k Consortium 2013	PRJNA167404	SAMN02641903	GCA_001188405.1	-
NCBI_Orsolobidae_Maoriata_sp.	Orsolobidae	Kulkarni et al. 2020	PRJNA575576	SAMN13006089	SRR10291854	-
NCBI_Hypochilidae_Hypochilus_gertschi	Hypochilidae	Kulkarni et al. 2020	PRJNA575576	SAMN13006084	SRR10291860	-

Table A.2. Fossil records used for divergence times estimation. A-H correspond to the remarks on Figure 3.2. The Age (Mya) has been updated according to the latest International chronostratigraphic chart <http://www.stratigraphy.org/>

ID	Superfamily	Family	Subfamily	Species	Fossil type	Stratum
A	-	Plectreuridae	-	<i>Eoplectreurys gertschi</i> Selden & Huang	-	Middle Jurassic
B	-	Plectreuridae	-	<i>Montsecarachne amicum</i> Selden	-	earliest Barremian age (Cretaceous)
C	Scytodoidea	Ochyroceratidae	Psilodercinae	<i>Priscacleclercera paucispinae</i> Wunderlich	Cretaceous Burmese amber	Cenomanian age (Cretaceous)
D	Scytodoidea	Scytodidae	-	<i>Scytodes weitschati</i> Wunderlich	Eocene Baltic amber	Lutetian age (early Middle Eocene, Cenozoic)
E	Pholcoidea	Pholcidae	Pholcinae	<i>Metagonia esquincacanoi</i> García-Villafuerte, 2019	Chiapas amber	Aquitanian age (Early Miocene, Cenozoic)
F	Pholcoidea	Pholcidae	Pholcinae	<i>Quamtana huberi</i> Penney, 2007a	Le Quesnoy amber	Ypresian age (Lowermost Eocene, Cenozoic)
G	Pholcoidea	Pholcidae	Modisiminae	<i>Modisimus chiapanecus</i> sp. nov.	Miocene Aquitanian amber	Aquitanian age (Early Miocene, Cenozoic)
H	Pholcoidea	Pholcidae	Ninetinae	<i>Pholcophora brevipes</i> Wunderlich, 1988	Dominican amber	Burdigalian age (Cenozoic)

ID	Dating method	Age (Mya)	Justifications of the assignment	Ref.
A	Radiometric	164	<i>Eoplectreurys</i> is a stem Synspermiata. It is here used as a minimum age to the Synspermiata + Hypochilidae node ^l .	a
B	-	121.4 - 129.4	<i>Montsecarachne</i> displays two synapomorphies of the Scytodoidea + Lost Tracheae Clade lineage: basally fused chelicerae and convergent endites (both lacking in the Dysderoidea). Thus, <i>Montsecarachne</i> is here considered a crown Synspermiata and used to set a minimum age to the Synspermiata node ^l .	b
C	Radiometric	98.79 ± 0.62	This is taken as evidence that <i>P. paucispinae</i> is a stem Ochyroceratidae, and this fossil is used here as a minimum age for the Scytodoidea + Lost Tracheae Clade lineage ^l .	c
D	Radiometric	43-47.8	it is a stem Scytodes. Here it is used as a minimum bound to the age of Scytodoidea ^l	d
E	Radiometric	22.8	It is used here as the minimum bound to the age of <i>Metagonia</i> + <i>Zatavua</i> + <i>Anansus</i> + <i>Nyikoa</i> lineage.	e, f
F	-	53	It is used here as the minimum bound to the age of <i>Quamtana</i> + <i>Micropholcus</i> + <i>Nipisa</i> + <i>Pholcus</i> lineage.	g
G	-	23.03 - 20.44	It is used here as the minimum bound to the age of <i>Modisimus</i> + Modisiminae lineage.	h
H	-	20.44 - 13.82	BAH said that this and the other two (not shown here) Dominican amber <i>Pholcophora</i> are probably closer to <i>Papiamenta</i> than to true <i>Pholcophora</i> , as we point out in our <i>Pholcophora-Tolteca</i> manuscript; thus it is used here as the minimum bound to the age of <i>Papiamenta</i> + <i>Pholcophora</i> + <i>Tolteca</i> + <i>Ibotyporanga</i> lineage.	i

Reference:

- a. Ke-Qin, G. A. O., and Ren Dong. "Radiometric dating of ignimbrite from Inner Mongolia provides no indication of a post-Middle Jurassic age for the Daohugou Beds." *Acta Geologica Sinica-English Edition* 80.1 (2006): 42-45.
- b. Selden, Paul A. "A new spider (Araneae: Haplogynae: Plectreuridae) from the Cretaceous Fossil-Lagerstätte of El Montsec, Spain." *The Journal of Arachnology* 42.1 (2014): 16-23.
- c. Shi, Guanghai, et al. "Age constraint on Burmese amber based on U-Pb dating of zircons." *Cretaceous research* 37 (2012): 155-163. d. Ritzkowski, S. "K-ar-alterbestimmungen der bernstein-führenden sedimente des samlandes (paläogen, bezirk kaliningrad)." *Metalla (Sonderheft)* 66 (1997): 19-23.
- e. de Lourdes Serrano-Sánchez, María, et al. "The aquatic and semiaquatic biota in Miocene amber from the Campo La Granja mine (Chiapas, Mexico): paleoenvironmental implications." *Journal of South American Earth Sciences* 62 (2015): 243-256.
- f. García-Villafuerte, Miguel Ángel. "Metagonia esquincacanoi sp. nov.: el primer registro fósil de la familia Pholcidae (Arachnida: Araneae) en el ámbar de Chiapas, México." *Revista mexicana de biodiversidad* 90 (2019).
- g. Penney, David. "The oldest fossil pholcid and selenopid spiders (Araneae) in lowermost eocene amber from the Paris Basin, France." *The Journal of Arachnology* 34.3 (2006): 592-598.
- h. García-Villafuerte, Miguel Ángel, and Alejandro Valdez-Mondragón. "The oldest fossils of the spider subfamily Modisiminae from the Americas: Description of a new species of the genus Modisimus Simon (Araneae: Pholcidae) from the amber of Mexico with a checklist of the extant Mexican species." *Journal of South American Earth Sciences* 103 (2020): 102702.
- i. Wunderlich, Jörg. "Die fossilen Spinnen im dominikanischen Bernstein." *Beit. Araneol.* 2 (1988): 1-378.
- j. Magalhaes, Ivan LF, et al. "The fossil record of spiders revisited: implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic." *Biological Reviews* 95.1 (2020): 184-217.

Table A.3. Saturation test by PhyloMad based on matrix 3

Locus	N_ sites	N_ taxa	t_ Entropy	t_pre- dicted _thre- shold _En- tropy	t_pre- dicted _TPR_ Entropy	t_pre- dicted _FPR_ Entropy	Risk_ Entropy	t_ Entro- pyvar	t_pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR_ Entro- pyvar	t_ pre- dicted _FPR_ Entro- pyvar	Risk_ Entro- pyvar
uce-3288	147	52	30.75	10.57	0.65	0.15	low.risk	10.54	13.56	0.92	0.16	high.risk
uce-8071	176	85	34.83	11.14	0.66	0.16	low.risk	18.47	18.27	0.92	0.13	low.risk
uce-110	162	84	49.82	10.54	0.65	0.15	low.risk	28.14	18.19	0.92	0.13	low.risk
uce-788	144	53	48.64	10.39	0.65	0.15	low.risk	11.25	13.96	0.91	0.16	high.risk
uce-9701	177	63	43.25	11.64	0.67	0.17	low.risk	20.72	15.98	0.9	0.14	low.risk
uce-105	178	75	37.96	11.42	0.67	0.17	low.risk	17.54	17.54	0.91	0.13	high.risk
uce-1959	191	61	40.7	12.28	0.68	0.18	low.risk	12.08	15.47	0.91	0.15	high.risk
uce-1457	166	58	37.62	11.28	0.66	0.17	low.risk	12.63	15.03	0.91	0.15	high.risk
uce-7414	156	85	40.76	10.25	0.65	0.14	low.risk	20.53	18.14	0.92	0.13	low.risk
uce-5380	188	51	37.57	12.42	0.68	0.19	low.risk	13.81	13.71	0.91	0.16	low.risk
uce-9134	165	82	29.4	10.71	0.66	0.15	low.risk	14.43	17.94	0.92	0.13	high.risk
uce-4318	148	60	28.64	10.4	0.65	0.15	low.risk	12.47	14.75	0.92	0.16	high.risk
uce-8327	180	52	39.09	12.05	0.67	0.18	low.risk	18.1	14.41	0.9	0.15	low.risk
uce-6304	159	71	69.88	10.66	0.66	0.15	low.risk	36.5	16.95	0.91	0.14	low.risk
uce-4021	174	56	29.95	11.69	0.67	0.17	low.risk	16.09	14.83	0.9	0.15	low.risk
uce-3313	239	72	62.08	13.9	0.7	0.21	low.risk	21.36	17.36	0.9	0.13	low.risk
uce-7690	154	69	29.42	10.48	0.65	0.15	low.risk	20.73	16.83	0.91	0.14	low.risk
uce-1715	126	54	32.54	9.47	0.64	0.13	low.risk	14.93	13.85	0.92	0.16	low.risk
uce-5664	212	93	69.05	12.48	0.68	0.18	low.risk	13.3	19.02	0.92	0.13	high.risk
uce-9809	130	85	62.32	8.99	0.63	0.12	low.risk	21.47	18.37	0.91	0.13	low.risk
uce-4753	344	79	48.39	17.21	0.74	0.27	low.risk	17.51	20.01	0.88	0.11	high.risk
uce-8467	157	82	46.25	10.35	0.65	0.14	low.risk	23	17.61	0.92	0.14	low.risk
uce-8990	187	75	37.24	11.8	0.67	0.17	low.risk	20.38	17.59	0.91	0.13	low.risk
uce-6115	207	69	40.85	12.74	0.68	0.19	low.risk	14.94	16.83	0.91	0.14	high.risk
uce-129	159	55	44.77	11.04	0.66	0.16	low.risk	22.16	14.11	0.91	0.16	low.risk
uce-4900	173	60	45.83	11.54	0.67	0.17	low.risk	17.95	15.33	0.91	0.15	low.risk
uce-7458	172	74	31.93	11.18	0.66	0.16	low.risk	18.26	17.33	0.91	0.13	low.risk
uce-2181	241	92	50.38	13.59	0.7	0.2	low.risk	20.16	19.61	0.91	0.12	low.risk
uce-2287	189	82	46.14	11.75	0.67	0.17	low.risk	18.85	18.37	0.91	0.13	low.risk
uce-6485	117	55	40.98	8.97	0.63	0.12	low.risk	18.71	13.68	0.92	0.17	low.risk
uce-5126	147	67	30.82	10.19	0.65	0.14	low.risk	26.36	16.34	0.91	0.14	low.risk
uce-258	191	70	73.21	12.08	0.68	0.18	low.risk	25.03	17.38	0.9	0.13	low.risk

Continued on next page

Locus	N_ sites	N_ taxa	t_ Entropy	t_pre- dicted _thre- shold _En- tropy	t_pre- dicted _TPR _Entropy	t_pre- dicted _FPR _Entropy	Risk_ Entropy	t_ Entro- pyvar	t_pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR _Entro- pyvar	t_ pre- dicted _FPR _Entro- pyvar	Risk_ Entro- pyvar
uce-1018	209	63	49.49	12.96	0.69	0.2	low.risk	26.61	16.22	0.9	0.14	low.risk
uce-6284	136	63	29.7	9.75	0.64	0.14	low.risk	11.77	15.07	0.92	0.16	high.risk
uce-387	174	81	39.09	11.13	0.66	0.16	low.risk	20.44	18.47	0.9	0.12	low.risk
uce-3800	188	70	70.04	11.95	0.67	0.18	low.risk	15.43	16.6	0.91	0.14	high.risk
uce-2645	133	73	30.84	9.38	0.64	0.13	low.risk	14.74	16.41	0.92	0.15	high.risk
uce-5617	161	76	45.74	10.65	0.66	0.15	low.risk	21.32	17.44	0.91	0.13	low.risk
uce-508	152	66	34.18	10.45	0.65	0.15	low.risk	12.11	15.67	0.92	0.15	high.risk
uce-5934	228	70	41.75	13.53	0.69	0.21	low.risk	23.52	17.91	0.89	0.12	low.risk
uce-1582	155	80	41.57	10.3	0.65	0.14	low.risk	11.37	16.83	0.93	0.15	high.risk
uce-4864	112	78	42.61	8.17	0.62	0.1	low.risk	20.2	16.74	0.93	0.15	low.risk
uce-2701	181	63	37.9	11.81	0.67	0.18	low.risk	15.67	15.52	0.91	0.15	low.risk
uce-3001	162	70	35.81	10.82	0.66	0.16	low.risk	22.5	16.62	0.91	0.14	low.risk
uce-5379	232	84	49.66	13.4	0.69	0.2	low.risk	15.88	18.27	0.91	0.13	high.risk
uce-9340	169	63	36.39	11.29	0.66	0.17	low.risk	14.57	15.54	0.91	0.15	high.risk
uce-8073	196	87	51.58	11.95	0.67	0.17	low.risk	18.58	18.28	0.92	0.13	low.risk
uce-1956	154	71	40.79	10.43	0.65	0.15	low.risk	13.15	15.81	0.93	0.16	high.risk
uce-106	112	56	45.24	8.67	0.63	0.11	low.risk	13.53	14.13	0.92	0.16	high.risk
uce-7805	164	75	32.83	10.81	0.66	0.15	low.risk	13.77	16.87	0.92	0.14	high.risk
uce-8902	206	61	32.79	12.89	0.69	0.2	low.risk	15.82	15.65	0.9	0.14	low.risk
uce-149	165	56	42.27	11.29	0.66	0.17	low.risk	17.43	14.73	0.91	0.15	low.risk
uce-5793	186	88	47.64	11.52	0.67	0.17	low.risk	15.2	19.04	0.91	0.12	high.risk
uce-1373	282	90	40.21	15.05	0.71	0.23	low.risk	16.31	20.13	0.89	0.11	high.risk
uce-2226	152	96	35.01	9.87	0.65	0.13	low.risk	18.03	19.02	0.92	0.13	high.risk
uce-8168	169	62	50.34	11.32	0.66	0.17	low.risk	15.88	15.92	0.9	0.14	high.risk
uce-9459	155	62	36.01	10.68	0.66	0.15	low.risk	20.52	15.75	0.91	0.14	low.risk
uce-4286	187	85	59.14	11.61	0.67	0.17	low.risk	14.8	17.49	0.93	0.14	high.risk
uce-115	299	89	76.62	15.63	0.72	0.24	low.risk	32.52	19.94	0.9	0.11	low.risk
uce-2948	286	95	43.79	15.1	0.71	0.23	low.risk	15.1	20.51	0.9	0.11	high.risk
uce-10293	255	87	37.29	14.18	0.7	0.21	low.risk	18.23	19.64	0.9	0.11	high.risk
uce-6482	209	68	44.35	12.85	0.69	0.19	low.risk	21.41	16.47	0.91	0.14	low.risk
uce-6443	130	74	44.04	9.2	0.64	0.12	low.risk	8.71	15.95	0.94	0.16	high.risk
uce-6634	171	81	53.18	11	0.66	0.16	low.risk	33.09	18.23	0.91	0.13	low.risk
uce-10337	169	62	44.19	11.32	0.66	0.17	low.risk	19.01	15.25	0.92	0.15	low.risk
uce-4628	155	59	31.07	10.76	0.66	0.16	low.risk	12.33	15.01	0.91	0.15	high.risk

Continued on next page

Locus	N_ sites	N_ taxa	t_ Entropy	t_pre- dicted _thre- shold _En- tropy	t_pre- dicted _TPR_ Entropy	t_pre- dicted _FPR_ Entropy	Risk_ Entropy	t_ Entro- pyvar	t_pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR_ Entro- pyvar	t_ pre- dicted _FPR_ Entro- pyvar	Risk_ Entro- pyvar
uce-694	169	69	31.7	11.16	0.66	0.16	low.risk	14.13	16.57	0.91	0.14	high.risk
uce-1245	159	63	46.29	10.84	0.66	0.16	low.risk	18.37	15.64	0.91	0.15	low.risk
uce-2998	126	86	33.72	8.77	0.63	0.11	low.risk	12.12	18.18	0.92	0.13	high.risk
uce-8404	188	53	33.98	12.36	0.68	0.19	low.risk	13.1	14.02	0.91	0.16	high.risk
uce-834	136	76	42.8	9.46	0.64	0.13	low.risk	22.5	17.05	0.92	0.14	low.risk
uce-4863	103	88	35.44	7.47	0.61	0.09	low.risk	16.77	17.42	0.94	0.15	high.risk
uce-6385	226	83	47.49	13.2	0.69	0.2	low.risk	27.06	19.14	0.9	0.12	low.risk
uce-7758	199	76	40.17	12.28	0.68	0.18	low.risk	25.34	18.24	0.9	0.12	low.risk
uce-1736	303	89	60.51	15.76	0.72	0.24	low.risk	20.71	20.64	0.89	0.11	low.risk
uce-5731	204	90	37.23	12.22	0.68	0.18	low.risk	18.96	19.21	0.91	0.12	high.risk
uce-161	171	65	34.22	11.33	0.66	0.17	low.risk	14.47	16.19	0.91	0.14	high.risk
uce-9683	123	59	41.36	9.18	0.63	0.12	low.risk	15.19	14.66	0.92	0.16	low.risk
uce-10316	172	57	40.19	11.57	0.67	0.17	low.risk	15.75	14.61	0.91	0.15	low.risk
uce-2980	142	51	33.16	10.35	0.65	0.15	low.risk	12.1	14.01	0.9	0.15	high.risk
uce-2170	188	64	43.97	12.09	0.68	0.18	low.risk	18.72	16.36	0.9	0.14	low.risk
uce-1246	238	64	60.08	14.04	0.7	0.22	low.risk	31.6	16.83	0.89	0.13	low.risk
uce-3802	234	84	42.29	13.48	0.69	0.2	low.risk	19.62	18.59	0.91	0.13	low.risk
uce-231	231	89	72.67	13.28	0.69	0.2	low.risk	35.56	19.24	0.91	0.12	low.risk
uce-2296	190	59	60.33	12.29	0.68	0.19	low.risk	22.23	15.71	0.9	0.14	low.risk
uce-2199	145	50	31.67	10.53	0.65	0.15	low.risk	8.62	13.39	0.91	0.16	high.risk
uce-7775	126	52	41	9.52	0.64	0.13	low.risk	16.47	13.51	0.92	0.16	low.risk
uce-3342	221	80	44.72	13.06	0.69	0.2	low.risk	9.56	18.14	0.91	0.13	high.risk
uce-6262	195	94	41.16	11.79	0.67	0.17	low.risk	15.45	19.79	0.91	0.12	high.risk
uce-5394	202	66	35.63	12.61	0.68	0.19	low.risk	15.61	16.56	0.9	0.14	high.risk
uce-9305	175	68	35.66	11.44	0.67	0.17	low.risk	18.85	16.94	0.9	0.13	low.risk
uce-6448	160	64	43.65	10.87	0.66	0.16	low.risk	18.43	15.65	0.91	0.15	low.risk
uce-2764	187	66	62.41	12	0.67	0.18	low.risk	26.37	17.24	0.89	0.13	low.risk
uce-850	158	72	38.86	10.6	0.65	0.15	low.risk	14.97	16.56	0.92	0.14	high.risk
uce-10308	161	82	39.01	10.53	0.65	0.15	low.risk	13.39	17.61	0.92	0.14	high.risk
uce-583	200	50	36.32	12.94	0.69	0.2	low.risk	15.22	13.81	0.9	0.15	low.risk
uce-813	179	58	40.47	11.85	0.67	0.18	low.risk	15.32	14.92	0.91	0.15	low.risk
uce-9018	162	54	31.25	11.21	0.66	0.17	low.risk	13.21	14.53	0.9	0.15	high.risk
uce-1656	223	90	53.55	12.96	0.69	0.19	low.risk	20.85	18.39	0.92	0.13	low.risk
uce-111	170	74	70	11.09	0.66	0.16	low.risk	19.87	17.07	0.91	0.14	low.risk

Continued on next page

Locus	N_ sites	N_ taxa	t_ Entropy	t_pre- dicted _thre- shold _En- tropy	t_pre- dicted _TPR_ Entropy	t_pre- dicted _FPR_ Entropy	Risk_ Entropy	t_ Entro- pyvar	t_pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR_ Entro- pyvar	t_ pre- dicted _FPR_ Entro- pyvar	Risk_ Entro- pyvar
uce-838	200	66	44.07	12.53	0.68	0.19	low.risk	18.72	16.58	0.9	0.14	low.risk
uce-5132	207	58	44.74	13	0.69	0.2	low.risk	19.66	15.49	0.9	0.14	low.risk
uce-1017	240	83	57.53	13.72	0.7	0.21	low.risk	24.57	18.42	0.91	0.13	low.risk
uce-8731	205	51	53.7	13.11	0.69	0.2	low.risk	16.69	14.34	0.9	0.15	low.risk
uce-582	175	56	44.54	11.73	0.67	0.18	low.risk	19.01	14.77	0.91	0.15	low.risk
uce-10324	165	76	36.99	10.83	0.66	0.15	low.risk	17.71	17.42	0.91	0.13	low.risk
uce-3162	190	62	62.73	12.22	0.68	0.18	low.risk	26.04	16.18	0.9	0.14	low.risk
uce-6579	206	94	79.22	12.23	0.68	0.18	low.risk	28.27	19.39	0.91	0.12	low.risk
uce-4577	153	87	37.67	10.07	0.65	0.14	low.risk	18.64	18.24	0.92	0.13	low.risk
uce-10306	173	92	39.31	10.89	0.66	0.15	low.risk	13.99	18.89	0.92	0.13	high.risk
uce-8854	128	78	54.23	9.02	0.63	0.12	low.risk	18.49	17.05	0.92	0.14	low.risk
uce-7286	187	84	42	11.63	0.67	0.17	low.risk	21.34	18.44	0.91	0.13	low.risk
uce-3423	138	62	41.01	9.87	0.64	0.14	low.risk	17.3	15.1	0.92	0.15	low.risk
uce-6777	280	86	49.57	15.06	0.71	0.23	low.risk	20.97	19.03	0.9	0.12	low.risk
uce-8739	296	87	54.15	15.57	0.72	0.24	low.risk	20.63	19.18	0.9	0.12	low.risk
uce-4543	274	87	70.13	14.84	0.71	0.23	low.risk	37.01	20.01	0.89	0.11	low.risk
uce-5646	180	63	33.83	11.77	0.67	0.17	low.risk	17.5	15.96	0.9	0.14	low.risk
uce-7838	178	64	32.69	11.66	0.67	0.17	low.risk	19.01	16.02	0.91	0.14	low.risk
uce-1647	219	69	35.76	13.21	0.69	0.2	low.risk	13.83	16.99	0.9	0.13	high.risk
uce-7352	333	56	44.48	17.4	0.74	0.27	low.risk	25.69	16.89	0.87	0.13	low.risk
uce-8877	170	57	40.32	11.49	0.67	0.17	low.risk	20.68	14.7	0.91	0.15	low.risk
uce-789	271	76	69.23	14.94	0.71	0.23	low.risk	22.94	18.51	0.89	0.12	low.risk

Table A.4. Tests of symmetry of matrix 3. IQ-TREE (v2.2.0) (Minh et al. 2020) was used to test the assumptions of stationarity (constant nucleotide/amino-acid frequencies over time) and homogeneity (constant substitution rates over time). The two assumptions are rejected when p-values (SymPval) < 0.01 are found. **Locus:** Partition name; **SymSig:** Number of significant sequence pairs by test of symmetry; **SymNon:** Number of non-significant sequence pairs by test of symmetry; **SymPval:** P-value for maximum test of symmetry; **MarSig:** Number of significant sequence pairs by test of marginal symmetry; **MarNon:** Number of non-significant sequence pairs by test of marginal symmetry; **MarPval:** P-value for maximum test of marginal symmetry; **IntSig:** Number of significant sequence pairs by test of internal symmetry; **IntNon:** Number of non-significant sequence pairs by test of internal symmetry; **IntPval:** P-value for maximum test of internal symmetry

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-4753	1,204	1,877	0.0000	1,393	1,688	3.3E-16	116	2,965	0.3470
uce-1373	1,551	2,454	0.0000	1,784	2,221	2.9E-11	119	3,886	0.5725
uce-105	553	2,222	0.0000	827	1,945	3.9E-07	89	2,683	0.8292
uce-850	572	1,984	0.0005	823	1,733	2.1E-04	73	2,482	0.1895
uce-6385	202	3,201	0.0010	259	3,142	2.9E-03	177	3,195	0.0355
uce-2170	451	1,565	0.0025	522	1,494	3.4E-04	137	1,879	0.6277
uce-4900	336	1,434	0.0033	392	1,371	5.9E-04	104	1,641	0.5283
uce-4863	131	3,687	0.0040	339	3,390	2.5E-03	83	3,517	0.1893
uce-8739	199	3,542	0.0057	320	3,404	8.9E-03	99	3,596	0.0848
uce-129	71	1,413	0.0063	130	1,311	5.9E-02	22	1,275	0.0146
uce-1736	484	3,432	0.0067	678	3,212	7.6E-04	136	3,724	0.8105
uce-387	294	2,946	0.0071	319	2,916	4.4E-02	178	3,049	0.0227
uce-10306	390	3,796	0.0083	534	3,651	4.6E-03	205	3,979	0.2379
uce-4864	110	2,893	0.0115	193	2,799	3.8E-03	93	2,876	0.3819
uce-7352	367	1,173	0.0126	529	1,005	3.8E-02	48	1,486	0.0503
uce-5379	449	3,037	0.0158	682	2,799	2.9E-03	110	3,369	0.6442
uce-4286	51	3,502	0.0206	67	3,234	5.4E-02	129	2,774	0.0631
uce-3802	426	3,060	0.0233	675	2,811	1.1E-01	129	3,354	0.0346
uce-8168	16	1,874	0.0235	49	1,784	1.3E-02	44	1,678	0.2876
uce-3342	354	2,806	0.0243	484	2,676	2.7E-03	128	3,031	0.9435
uce-8990	328	2,447	0.0247	499	2,275	5.1E-03	102	2,665	0.6391
uce-1246	239	1,777	0.0287	365	1,648	1.4E-02	73	1,938	0.3204
uce-110	81	3,405	0.0385	72	3,406	3.4E-01	151	3,312	0.0193
uce-6634	373	2,867	0.0409	615	2,625	9.9E-02	73	3,164	0.0762
uce-9701	125	1,828	0.0438	169	1,777	1.9E-02	74	1,823	0.3918
uce-2287	173	3,148	0.0482	224	3,091	7.9E-03	117	3,183	0.8421
uce-4021	128	1,412	0.0516	216	1,320	2.3E-02	37	1,490	0.3907
uce-7690	133	2,213	0.0536	141	2,202	1.1E-02	145	2,195	0.7310
uce-2701	242	1,711	0.0571	334	1,597	3.7E-01	53	1,818	0.0281
uce-111	112	2,587	0.0584	130	2,549	8.9E-02	105	2,495	0.1297
uce-6304	13	2,447	0.0614	63	2,040	3.2E-02	15	1,767	0.3620
uce-5380	108	1,167	0.0680	131	1,143	1.8E-01	55	1,215	0.0768
uce-1018	175	1,778	0.0687	244	1,706	1.5E-02	89	1,854	0.7422
uce-5664	66	4,211	0.0716	216	4,026	3.8E-01	79	4,088	0.0368
uce-6448	231	1,785	0.0724	325	1,682	1.5E-02	95	1,883	0.7657
uce-115	125	3,790	0.0828	323	3,557	3.8E-02	57	3,693	0.4279
uce-1715	159	1,272	0.0884	222	1,205	5.5E-01	52	1,356	0.0307
uce-583	116	1,109	0.0936	159	1,060	4.9E-02	53	1,150	0.3954
uce-8073	265	3,476	0.0951	485	3,244	2.2E-02	82	3,595	0.7608
uce-2998	128	3,527	0.0963	215	3,436	1.7E-01	87	3,537	0.1223
uce-7286	106	3,380	0.1048	214	3,262	2.2E-02	90	3,295	0.8211
uce-9305	200	2,077	0.1222	350	1,914	8.8E-02	58	2,190	0.3197
uce-6777	279	3,376	0.1231	430	3,225	3.1E-02	156	3,496	0.7567
uce-5132	156	1,493	0.1239	186	1,435	4.7E-02	84	1,515	0.5559
uce-4543	245	3,496	0.1315	428	3,270	4.6E-02	92	3,469	0.6059
uce-4628	170	1,541	0.1369	272	1,432	5.6E-02	68	1,627	0.6646
uce-7758	218	2,632	0.1387	231	2,613	2.1E-01	150	2,691	0.1571

Continued on next page

Appendix A Supplementary information for Chapter 3

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-4318	160	1,610	0.1501	257	1,509	8.3E-02	44	1,716	0.4274
uce-1959	333	1,497	0.1514	426	1,387	3.2E-02	50	1,737	0.8917
uce-2226	241	4,319	0.1542	337	4,217	7.3E-02	160	4,383	0.4953
uce-5394	221	1,924	0.1709	344	1,798	7.1E-02	66	2,070	0.5698
uce-2980	8	1,265	0.1716	28	1,216	1.3E-01	25	1,184	0.3338
uce-1582	90	3,067	0.1801	169	2,647	3.8E-01	58	2,490	0.0731
uce-149	55	1,485	0.1811	106	1,428	1.6E-01	45	1,487	0.2897
uce-1457	88	1,565	0.1831	199	1,441	6.9E-02	16	1,564	0.6277
uce-106	25	1,515	0.1947	42	1,455	1.5E-01	79	1,352	0.3338
uce-6482	150	2,127	0.1999	190	2,067	6.5E-02	35	2,049	0.9766
uce-7414	347	3,223	0.2054	645	2,923	1.0E-01	91	3,466	0.5257
uce-694	96	2,250	0.2255	166	2,179	3.0E-01	83	2,261	0.2078
uce-4577	175	3,566	0.2415	312	3,429	1.7E-01	97	3,639	0.3923
uce-8731	105	1,170	0.2440	143	1,132	3.6E-01	48	1,226	0.1928
uce-2764	45	2,100	0.2462	52	2,083	1.3E-01	43	2,066	0.5205
uce-8877	114	1,482	0.2549	169	1,427	1.3E-01	53	1,539	0.5643
uce-582	154	1,386	0.2677	154	1,386	6.8E-02	119	1,421	0.9233
uce-1656	251	3,754	0.2679	329	3,675	3.1E-01	168	3,830	0.2618
uce-8071	512	3,058	0.2687	844	2,704	2.4E-01	67	3,428	0.3309
uce-789	57	2,792	0.2727	173	2,637	1.0E-01	55	2,716	0.7297
uce-3162	69	1,819	0.2763	203	1,641	5.8E-01	11	1,748	0.1351
uce-8902	369	1,461	0.2865	473	1,355	6.4E-01	80	1,743	0.1261
uce-3001	42	2,372	0.3012	95	2,304	2.1E-01	62	2,324	0.4405
uce-3423	238	1,653	0.3141	354	1,531	3.4E-01	73	1,800	0.2932
uce-7775	46	1,280	0.3327	41	1,282	3.2E-01	86	1,201	0.3415
uce-6115	78	2,267	0.3390	88	2,171	4.5E-01	103	2,015	0.2470
uce-813	118	1,535	0.3439	212	1,416	1.3E-01	18	1,601	0.7951
uce-2199	22	1,203	0.3440	78	1,144	2.9E-01	26	1,189	0.3906
uce-838	84	2,060	0.3512	128	2,006	5.7E-01	93	2,018	0.1960
uce-508	155	1,989	0.3623	248	1,895	2.8E-01	53	2,087	0.4286
uce-8404	46	1,332	0.3667	96	1,273	3.5E-01	33	1,317	0.3561
uce-8854	121	2,873	0.3793	218	2,347	3.9E-01	41	2,185	0.3304
uce-10324	152	2,697	0.3805	303	2,541	1.4E-01	67	2,767	0.8221
uce-9809	43	3,527	0.3843	151	3,370	3.0E-01	80	3,270	0.4450
uce-8467	257	3,053	0.3914	447	2,725	2.1E-01	96	2,739	0.6239
uce-1245	101	1,851	0.4075	140	1,810	7.0E-01	75	1,875	0.1950
uce-5731	388	3,617	0.4079	485	3,505	1.5E-01	215	3,753	0.8268
uce-1647	233	2,113	0.4121	375	1,971	1.8E-01	74	2,269	0.7545
uce-6262	139	4,224	0.4325	269	4,021	8.5E-01	107	4,083	0.1635
uce-2645	70	2,558	0.4436	98	2,523	4.9E-01	105	2,507	0.3312
uce-1017	183	3,219	0.4572	349	3,027	3.9E-01	73	3,244	0.4358
uce-8327	110	1,216	0.4823	213	1,113	4.9E-01	30	1,291	0.3836
uce-7458	177	2,524	0.4873	295	2,402	1.8E-01	112	2,577	0.9181
uce-834	88	2,762	0.4877	91	2,741	4.2E-01	145	2,657	0.4392
uce-6284	617	1,336	0.4920	731	1,222	3.7E-01	80	1,871	0.5223
uce-161	71	2,008	0.4950	167	1,906	7.6E-01	27	2,011	0.2372
uce-9683	45	1,662	0.5433	74	1,545	nan	66	1,504	nan
uce-788	11	1,367	0.5539	40	1,330	4.0E-01	26	1,315	0.5732
uce-10293	208	3,533	0.5542	238	3,499	4.6E-01	150	3,578	0.5015
uce-3288	122	1,204	0.5737	147	1,167	3.1E-01	32	1,274	0.7545
uce-5617	83	2,763	0.5764	150	2,661	5.5E-01	116	2,647	0.4545
uce-9018	83	1,348	0.5852	113	1,317	2.7E-01	57	1,373	0.8595
uce-9134	594	2,727	0.5887	872	2,434	4.2E-01	47	3,204	1.0000
uce-258	14	2,401	0.6007	30	2,340	8.6E-01	51	2,227	0.2837
uce-5646	363	1,590	0.6125	471	1,466	5.4E-01	106	1,812	0.5077
uce-1956	51	2,434	0.6669	113	2,351	4.0E-01	74	2,286	0.8698
uce-9459	15	1,872	0.6758	37	1,817	8.5E-01	43	1,786	0.3633
uce-3800	6	2,382	0.6874	4	2,064	6.5E-01	110	1,797	0.5138
uce-5793	203	3,624	0.7178	317	3,422	7.8E-01	124	3,388	0.4542
uce-231	20	3,893	0.7302	55	3,828	8.5E-01	32	3,793	0.4242

Continued on next page

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-9340	115	1,838	0.7614	153	1,790	8.4E-01	114	1,743	0.4690
uce-5934	130	2,285	0.7760	233	2,179	7.7E-01	91	2,316	0.5469
uce-6579	41	4,324	0.7790	78	4,204	7.2E-01	155	3,933	0.5905
uce-7805	189	2,586	0.7858	344	2,426	5.0E-01	62	2,696	0.8427
uce-6443	22	2,672	0.7968	65	2,514	7.3E-01	43	2,383	0.5371
uce-6485	28	1,457	0.8013	69	1,397	8.0E-01	13	1,421	nan
uce-5126	-	2,200	0.8032	2	1,889	8.3E-01	38	1,681	0.5395
uce-10308	229	3,092	0.8086	386	2,923	5.0E-01	78	3,177	0.8937
uce-2948	1,185	3,280	0.8241	1,622	2,837	6.5E-01	60	4,375	0.7429
uce-10337	27	1,841	0.8301	63	1,711	9.1E-01	21	1,431	0.5178
uce-2296	44	1,667	0.8911	71	1,630	5.5E-01	48	1,605	0.9822
uce-2181	253	3,931	0.9138	492	3,666	9.8E-01	119	4,002	0.5933
uce-3313	103	2,443	0.9143	172	2,210	9.4E-01	33	1,949	0.6493
uce-10316	171	1,425	0.9261	162	1,416	6.8E-01	117	1,406	0.9324
uce-7838	149	1,867	0.9710	219	1,781	7.7E-01	29	1,938	0.9775

Table A.5. Topology test. IQ-TREE (v2.2.0) (Minh et al. 2020) was used to perform the analysis.

Matrix	Tree method or constraints	logL	deltaL	bp-RELL	Plus/Minus	c-ELW	Plus/Minus
matrix 1	IQ-TREE.partitioned	-693236.87	129.12	0.000633	-	0.000662	-
matrix 2	IQ-TREE.partitioned	-693199.96	92.212	0.0143	-	0.0143	-
matrix 3	SVDquartets	-693462.03	354.27	0	-	9.88E-26	-
matrix 3	(Ninetinae, Priscula), (Artema,Arteminae)	-693270.14	162.38	0	-	1.22E-21	-
matrix 3	(Ninetinae, Artema), (Priscula, Arteminae)	-693240.35	132.59	0	-	1.52E-05	-
matrix 3	monophyletic group (other Arteminae + Artema)	-693146.15	38.406	0.0108	-	0.0125	-
matrix 3	(Ninetinae, all other pholcids)	-693146.15	38.406	0.0124	-	0.0125	-
matrix 3	phylogbayes-MPI	-693188.09	80.334	0.0176	-	0.0175	-
matrix 3	ASTRAL	-693154.22	46.471	0.0684	+	0.0685	+
matrix 3	RAxML-NG.partitioned	-693124.97	17.218	0.0934	+	0.0945	+
matrix 3	IQ-TREE.partitioned	-693128.76	21.011	0.164	+	0.164	+
matrix 4	IQ-TREE.partitioned	-693107.75	0	0.51	+	0.509	+
matrix 5	IQ-TREE.partitioned	-693224.23	116.48	0.00413	-	0.0041	-
matrix 6	IQ-TREE.partitioned	-693159.08	51.325	0.0794	+	0.0793	+
matrix 7	IQ-TREE.partitioned	-693190.29	82.541	0.0453	+	0.0455	+
matrix 8	IQ-TREE.partitioned	-693409.37	301.61	0	-	2.13E-29	-
matrix 9	IQ-TREE.unpartitioned	-768899.72	75792	0	-	0	-
matrix 10	IQ-TREE.unpartitioned	-897429.21	2.04E+05	0	-	0	-

Supplementary information for Chapter [4](#)

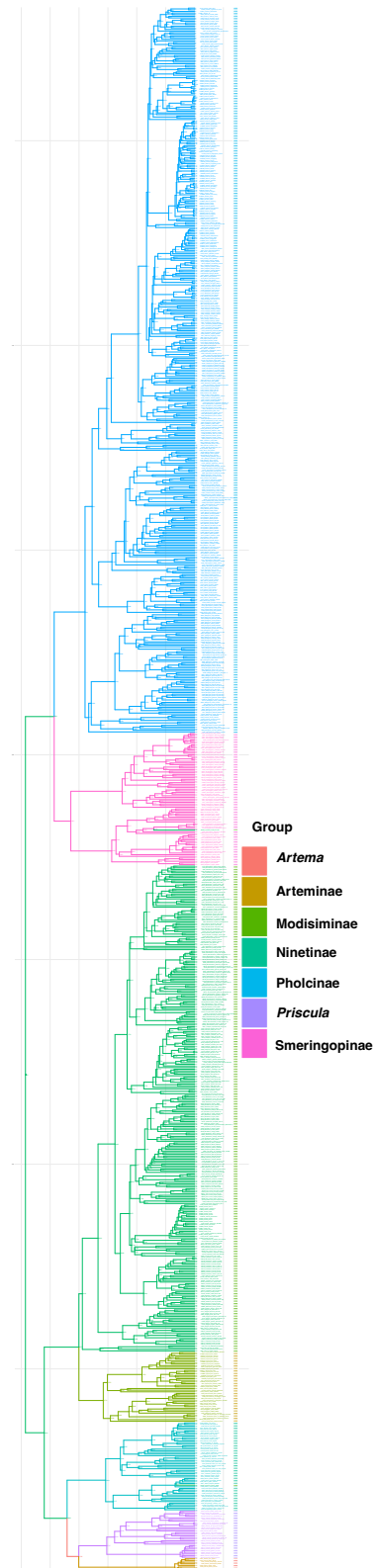


Figure B.1. ClipKIT smart-gap tree. The tree search (constrained on the dated tree Figure 3.2, Meng et al. (2025)) was performed with IQ-TREE (v2.2.0) (Minh et al. 2020) using the partitioned strategy and the same parameter settings as for Figure 3.2. Then the divergence times were calibrated using the TreePL (v 1.0) (Smith et al. 2012) and the `congruify.phylo()` R function in `geiger` (v2.0.11) (Pennell et al. 2014) based on Figure 3.2. Outgroups are not shown. Numbers on the nodes are node ids.

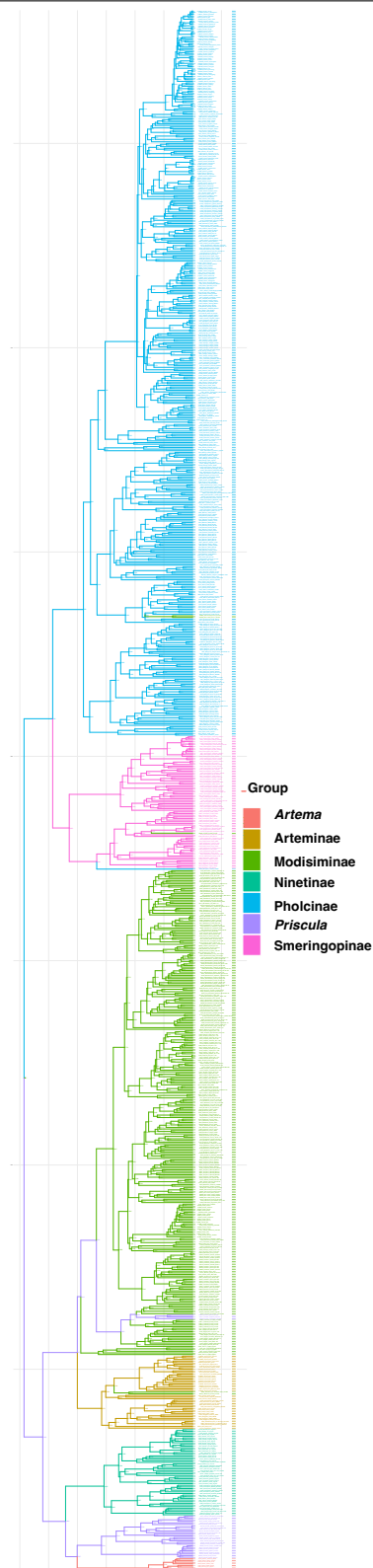


Figure B.2. ClipKIT gappy tree. The tree search (constrained on the dated tree Figure 3.2, Meng et al. (2025)) was performed with IQ-TREE (v2.2.0) (Minh et al. 2020) using the partitioned strategy and the same parameter settings as for Figure 3.2. Then the divergence times were calibrated using the TreePL (v 1.0) (Smith et al. 2012) and the `congruify.phylo()` R function in `geiger` (v2.0.11) (Pennell et al. 2014) based on Figure 3.2. Outgroups are not shown. Numbers on the nodes are node ids.

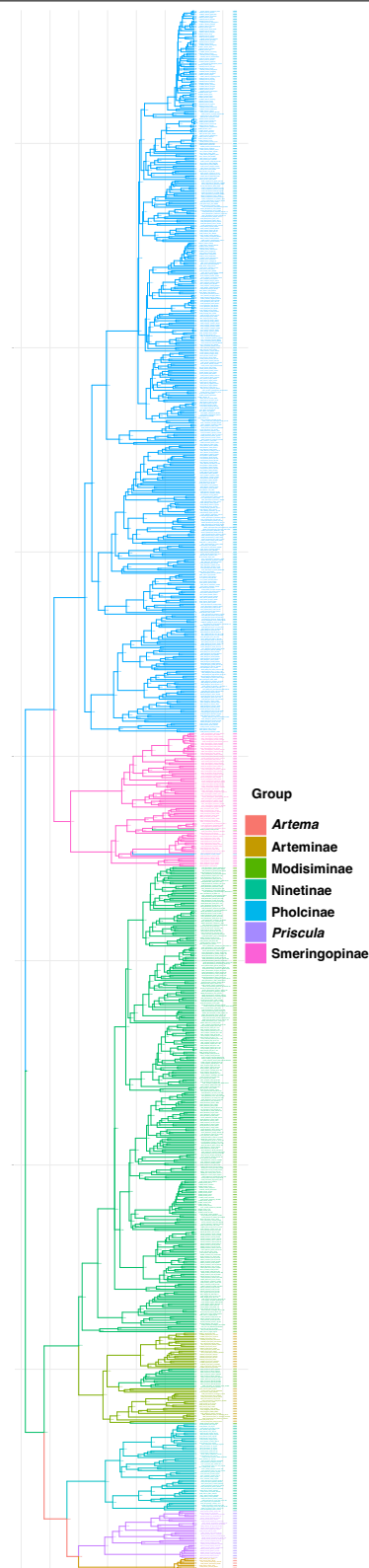


Figure B.3. ClipKIT kpic-smart-gap tree. The tree search (constrained on the dated tree Figure 3.2, Meng et al. (2025)) was performed with IQ-TREE (v2.2.0) (Minh et al. 2020) using the partitioned strategy and the same parameter settings as for Figure 3.2. Then the divergence times were calibrated using the TreePL (v 1.0) (Smith et al. 2012) and the `congruify.phylo()` R function in `geiger` (v2.0.11) (Pennell et al. 2014) based on Figure 3.2. Outgroups are not shown. Numbers on the nodes are node ids.

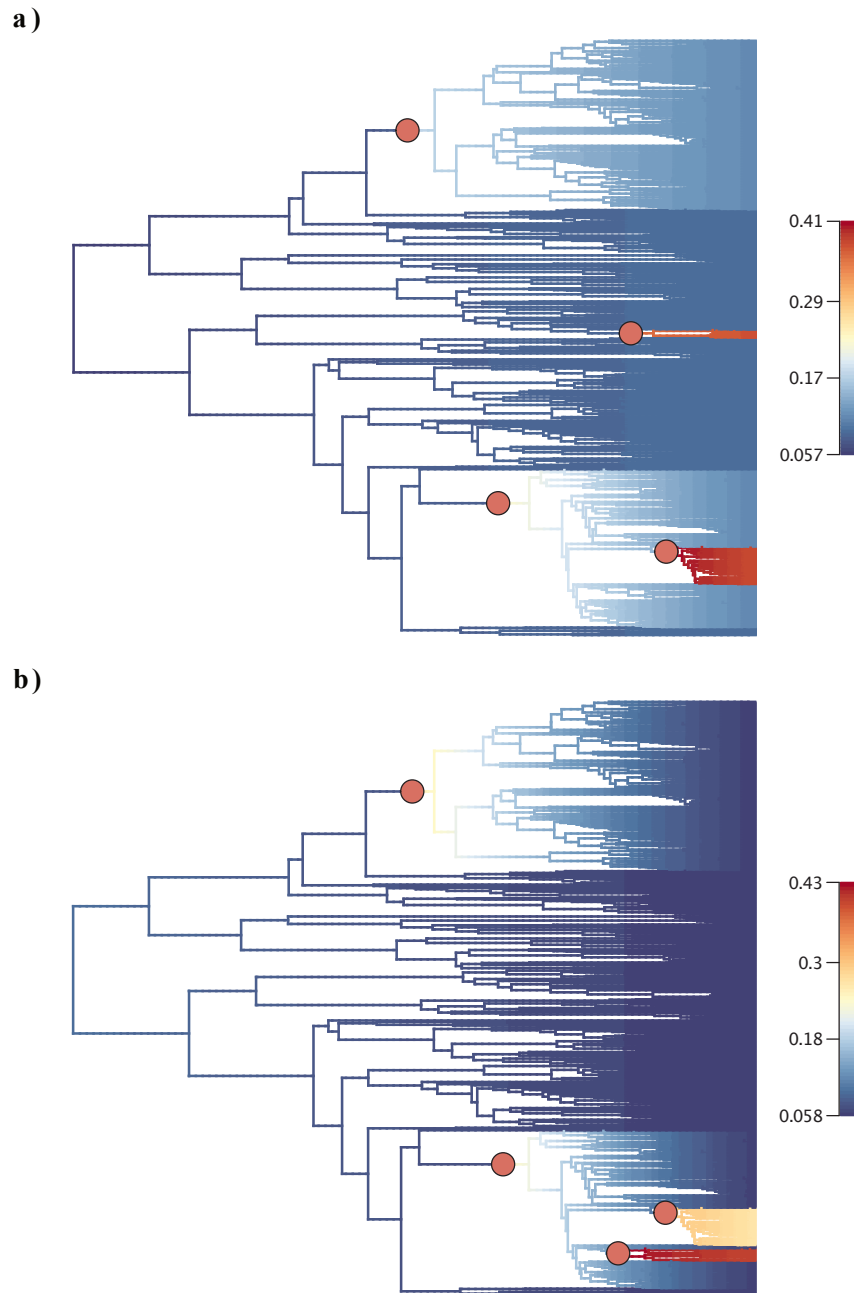
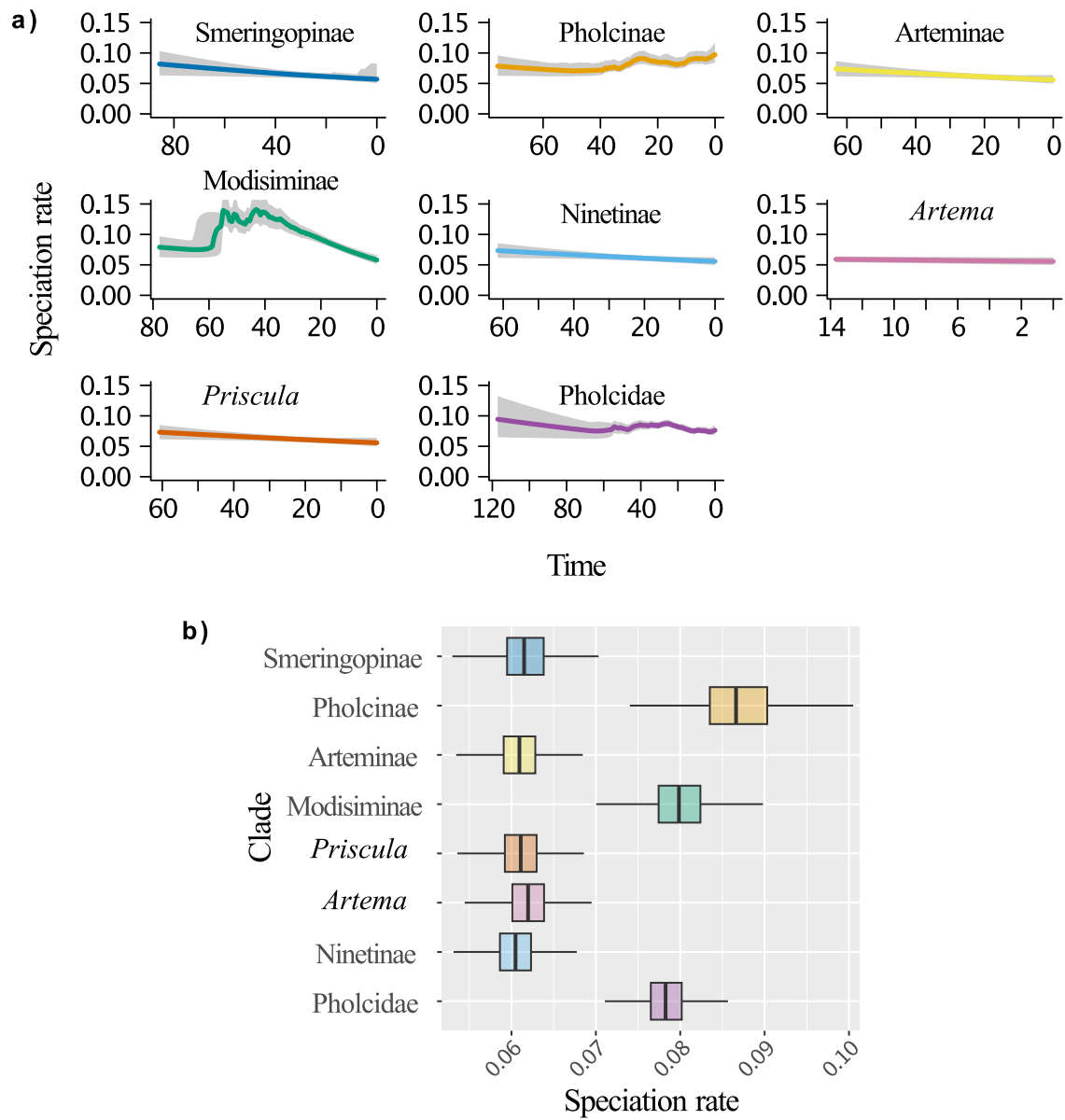


Figure B.4. BAMM single best speciation rate shift configuration. The analysis was based on the ClipKIT smart-gap tree, and genus-specific sampling fractions were applied. The prior on the number of shifts in diversification was 50, while the priors for the $\lambda_{\text{InitPrior}}$, $\lambda_{\text{ShiftPrior}}$, and $\mu_{\text{InitPrior}}$ parameters were estimated with BAMMtools (Rabosky 2014); 50 million generations for MCMC (Markov Chain Monte Carlo) simulation were applied. This tree was generated with the `getBestShiftConfiguration()` R function in BAMMtools. (a) Analysis based on estimated species richness. (b) Analysis based on currently known species richness.



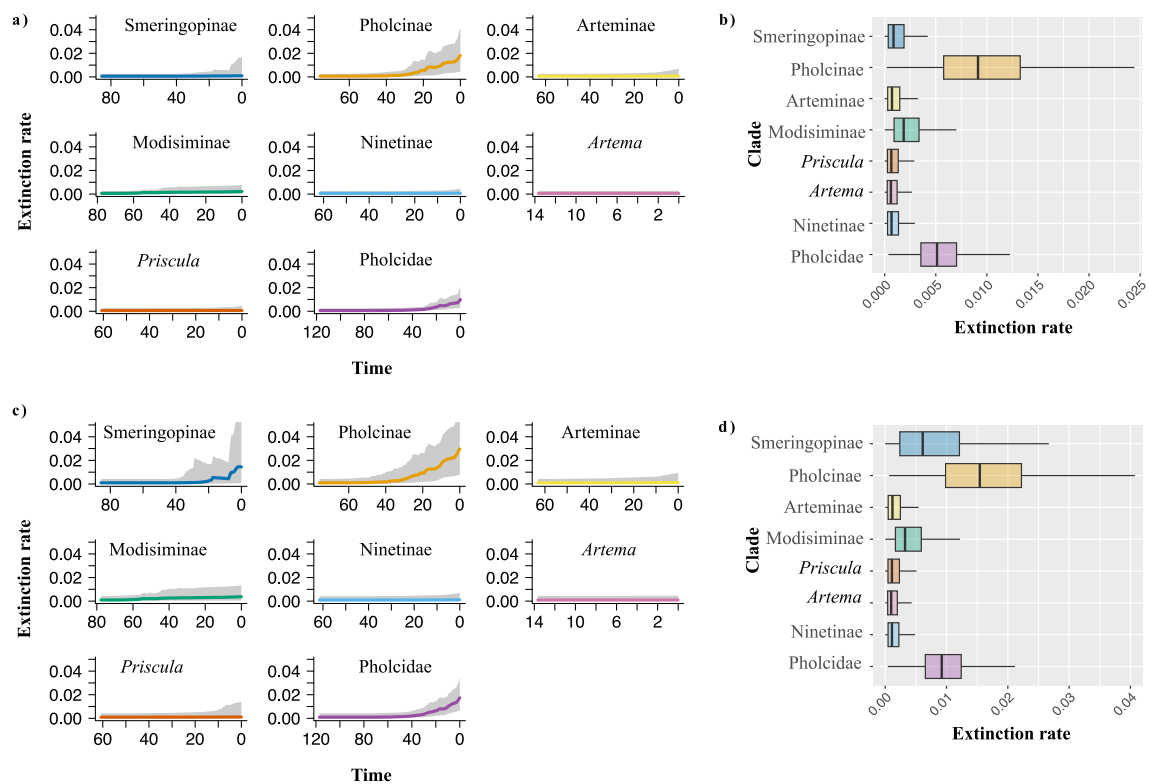


Figure B.6. Extinction rate dynamics for different groups. The extinction rates were retrieved from the BAMM analyses (based on the ClipKIT smart-gap tree), in which genus-specific sampling fractions were applied. **(a and b)** BAMM analyses based on currently known species richness. **(c and d)** BAMM analyses based on estimated species richness.

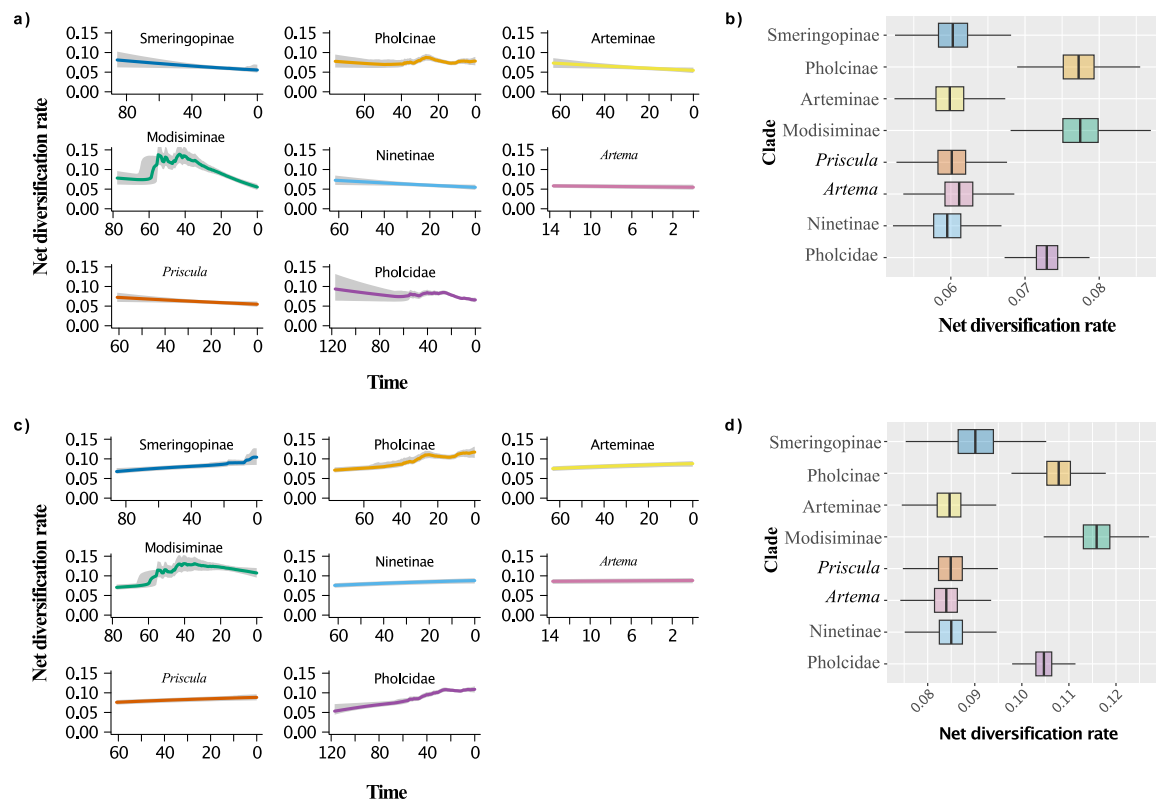


Figure B.7. Net-diversification rate dynamics for different groups. The net-diversification rates were retrieved from the BMM analyses (based on the ClipKIT smart-gap tree), in which genus-specific sampling fractions were applied. (a and b) BMM analyses based on currently known species richness. (c and d) BMM analyses based on estimated species richness.

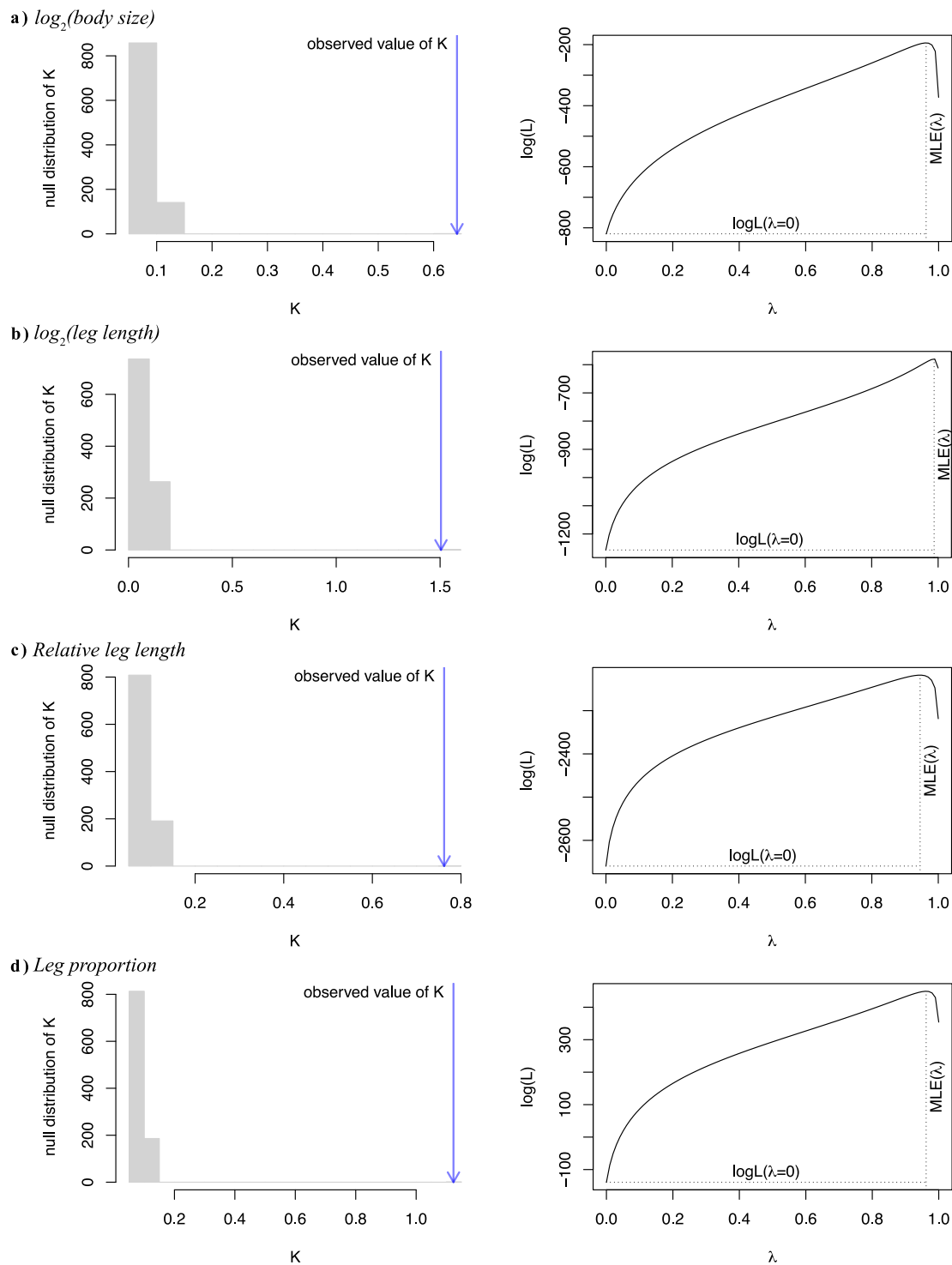


Figure B.8. Phylogenetic signal of different traits. These analyses were based on the ClipKIT smart-gap tree. Left: Blomberg's K (Blomberg et al. 2003) (p - values = 0.001); Right: Pagel's lambda (Pagel 1999) (p - values << 0.001)

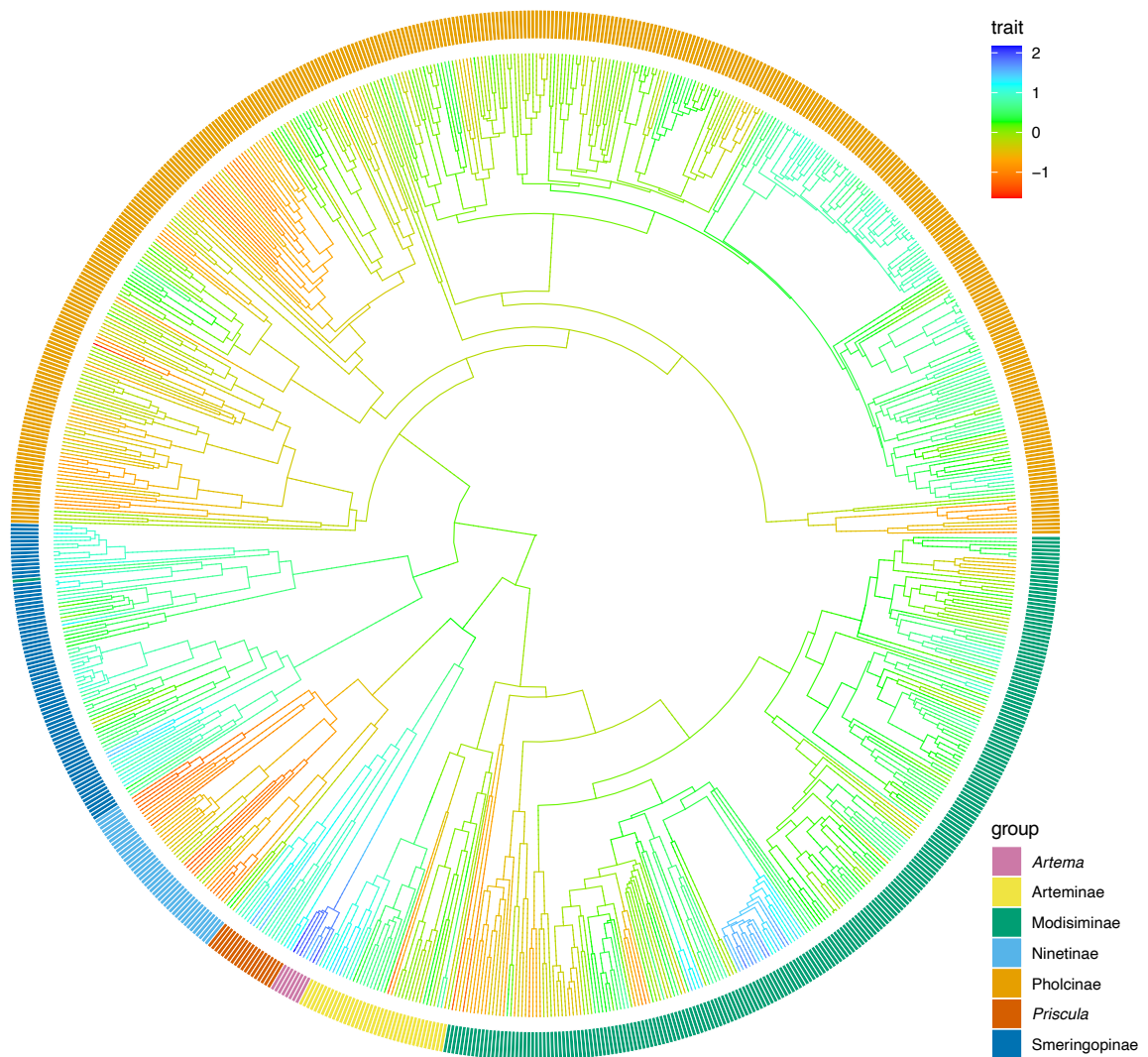


Figure B.9. Ancestral states of $\log_2(\text{body size})$. The ClipKIT smart-gap tree, Brownian motion (BM) model and the `anc.ML()` function in `phytools` (v1.9-16) (Revell 2012) were used.

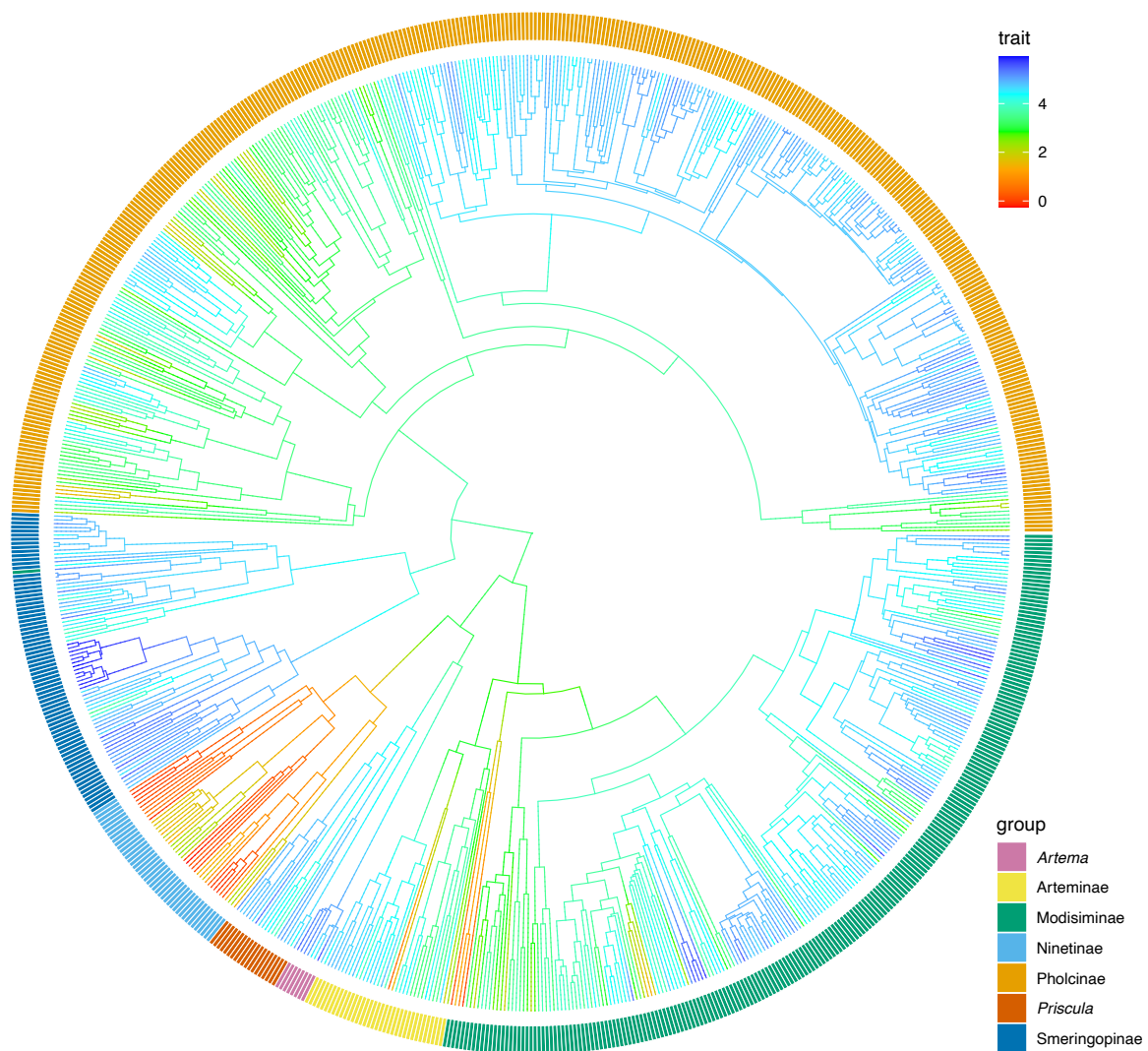


Figure B.10. Ancestral states of $\log_2(\text{leg length})$. The ClipKIT smart-gap tree, Brownian motion (BM) model and the `anc.ML()` function in `phytools` (v1.9-16) (Revell 2012) were used.

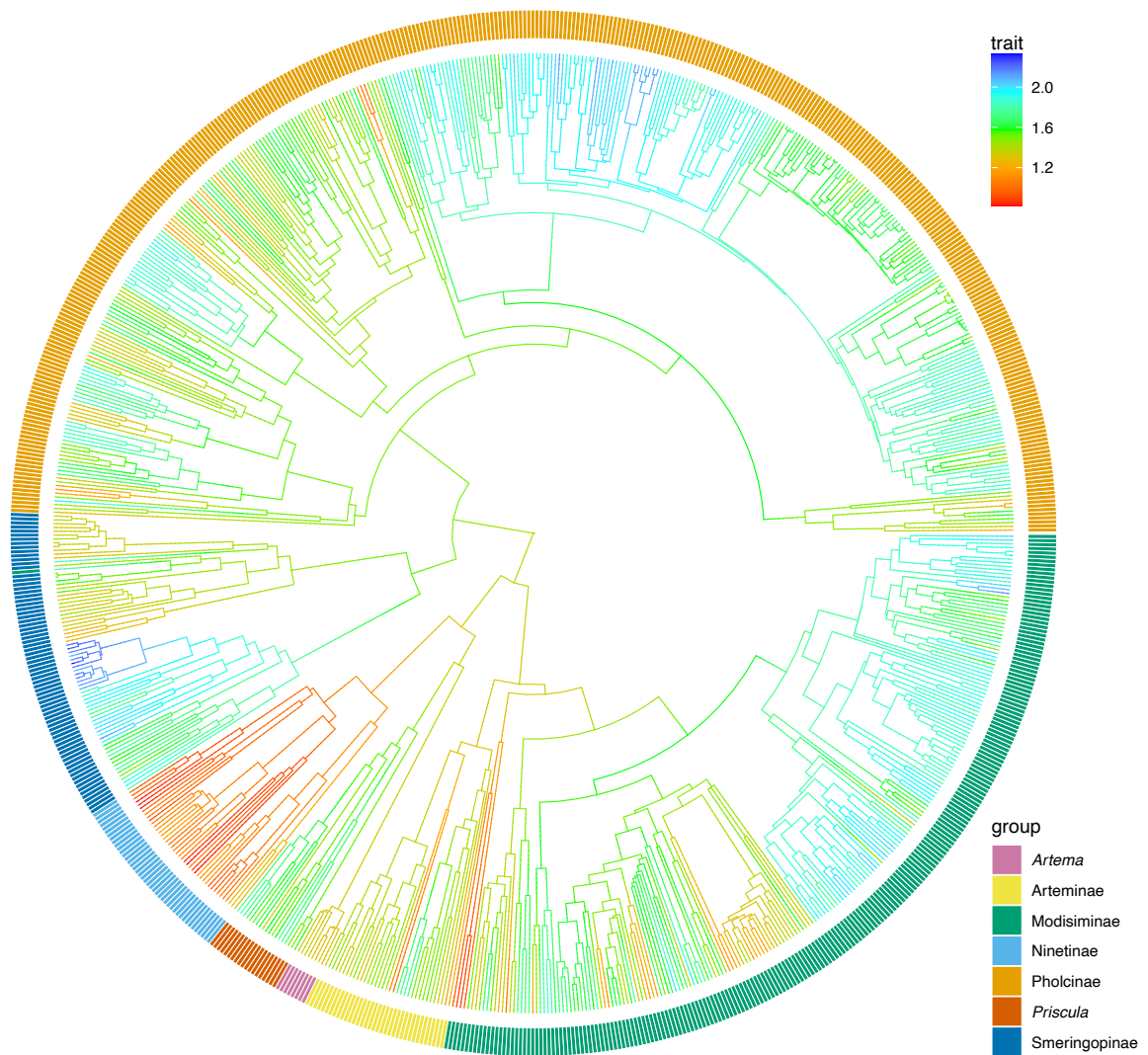


Figure B.11. Ancestral states of leg proportion. The ClipKIT smart-gap tree, Brownian motion (BM) model and the anc.ML() function in phyttools (v1.9-16) (Revell 2012) were used.

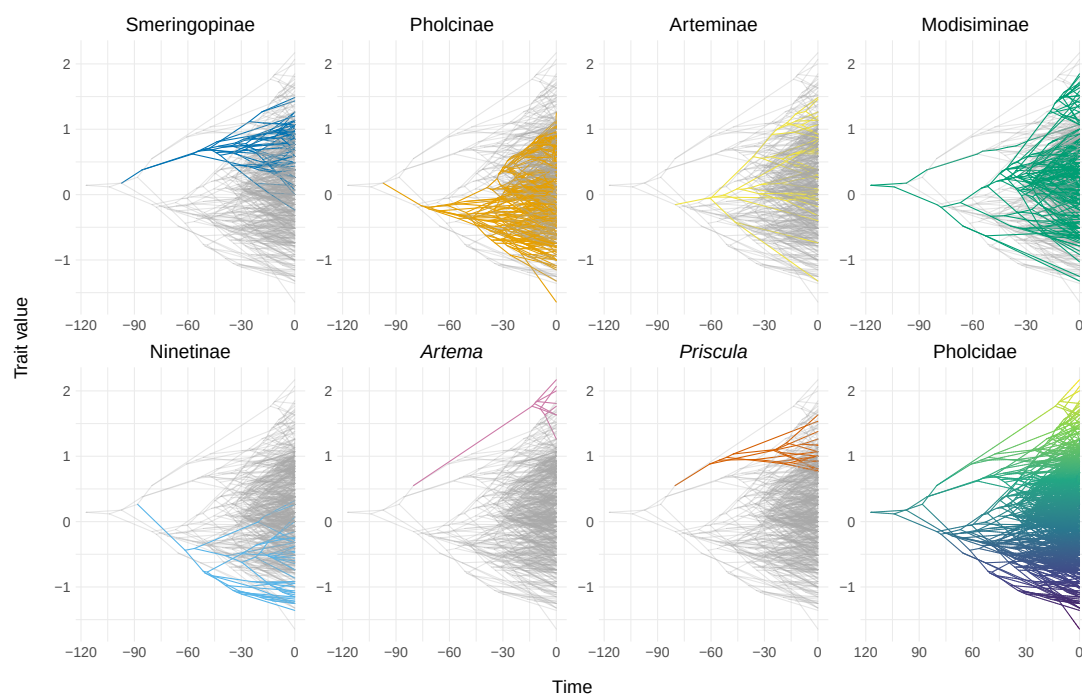


Figure B.12. Phenogram of $\log_2(\text{body size})$ for different groups. Ancestral states were constructed based on the `ClipKIT smart-gap tree`, Brownian motion (BM) model and the `anc.ML()` function in `phytools` (v1.9-16) (Revell 2012) were used; The R package `ggtree` (Yu et al. 2017) was used for visualization.

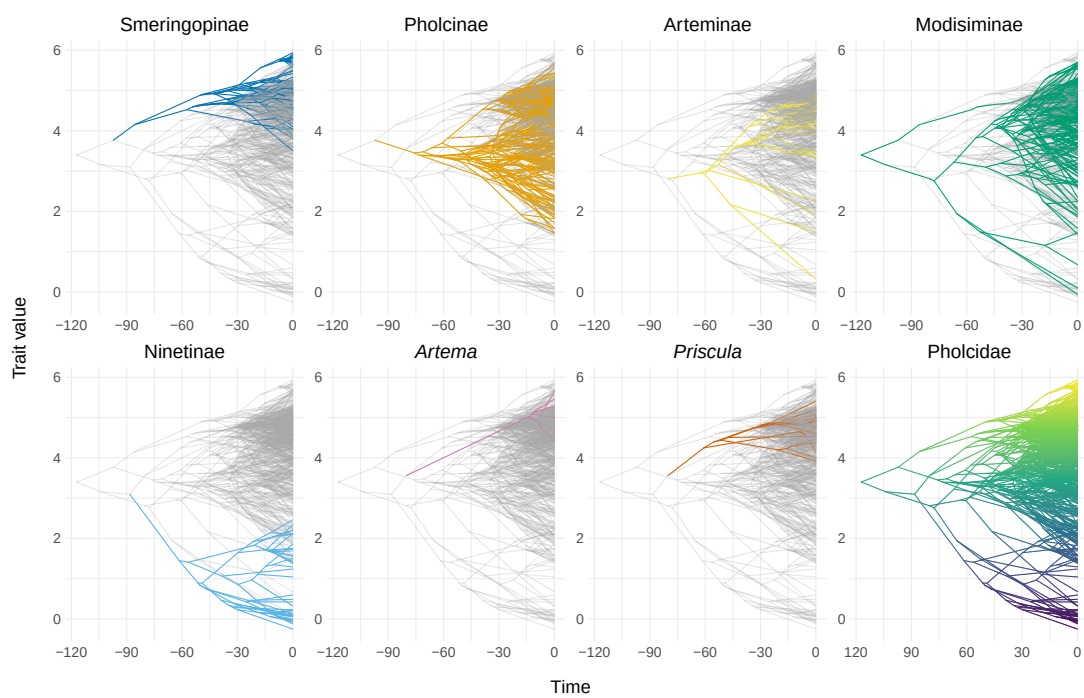


Figure B.13. Phenogram of $\log_2(\text{leg length})$ for different groups. Ancestral states were constructed based on the ClipKIT smart-gap tree, Brownian motion (BM) model and the `anc.ML()` function in `phytools` (v1.9-16) (Revell 2012) were used; The R package `ggtree` (Yu et al. 2017) was used for visualization.

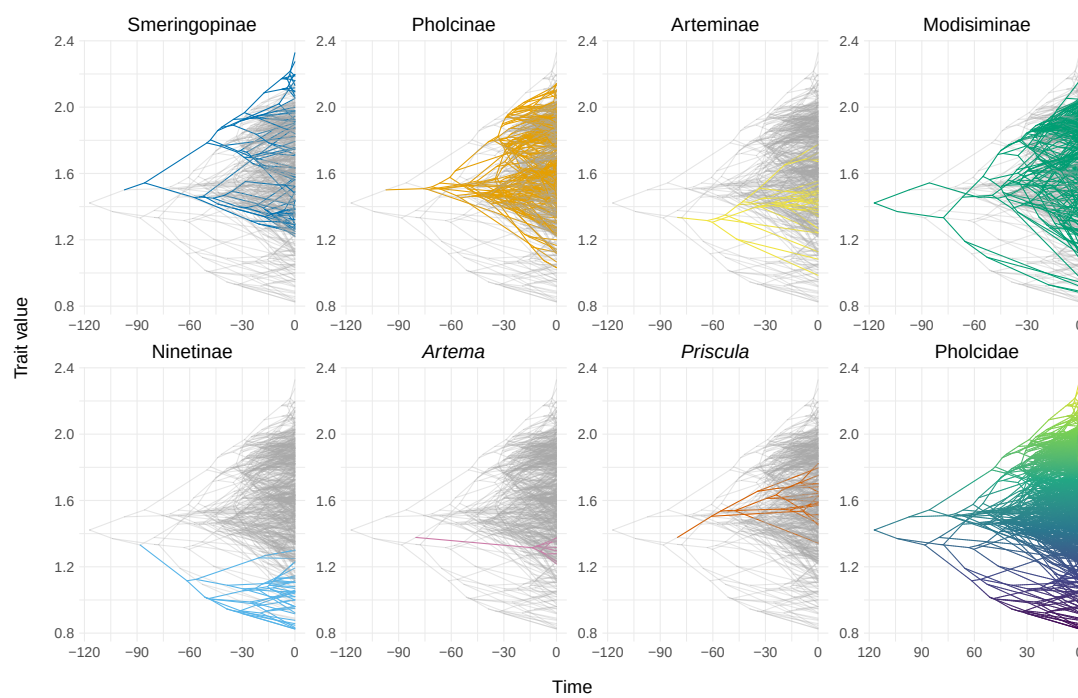


Figure B.14. Phenogram of leg proportion for different groups. Ancestral states were constructed based on the `ClipKIT smart-gap tree`, Brownian motion (BM) model and the `anc.ML()` function in `phytools` (v1.9-16) (Revell 2012) were used; The R package `ggtree` (Yu et al. 2017) was used for visualization.

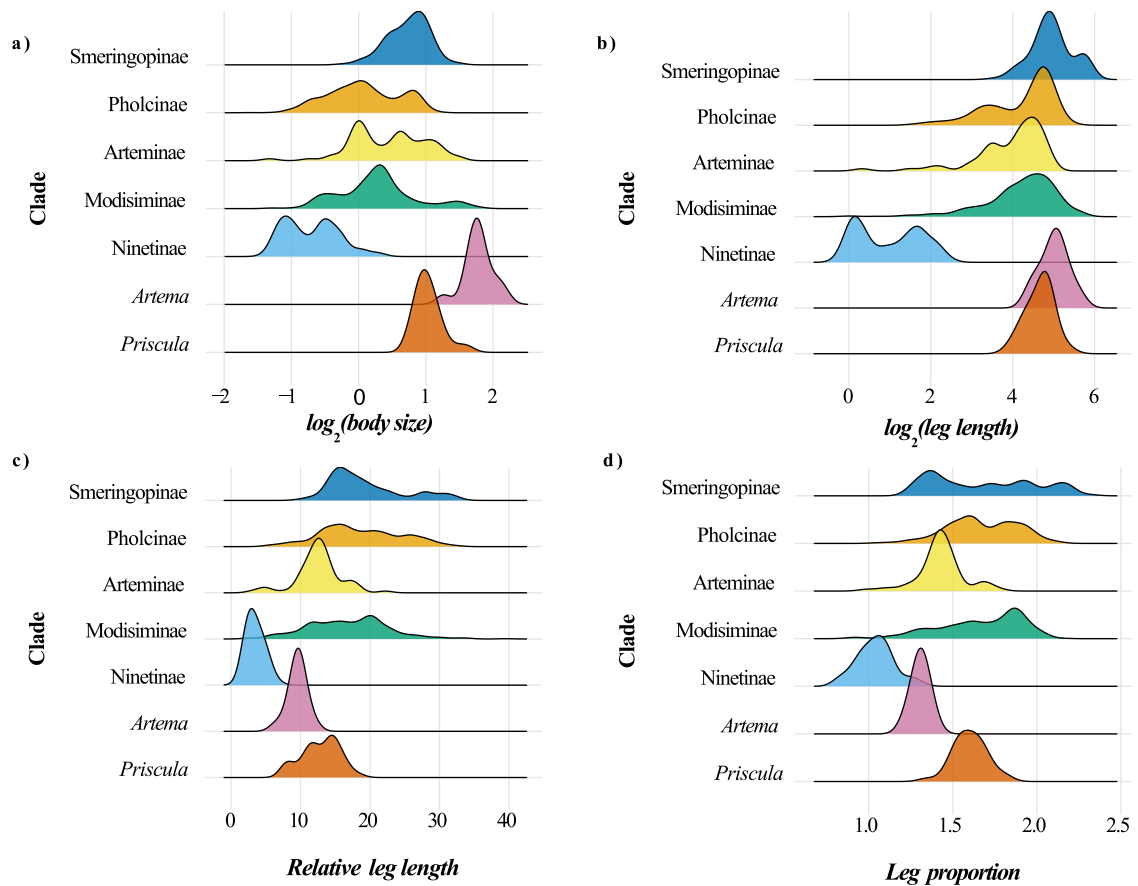


Figure B.15. Trait value density distribution for the four traits. The group-specific trait values were extracted from the BMM trait analysis, which was based on genus-specific sampling fraction and estimated species richness. The trait leg proportion (d) is a proxy of microhabitats (ground, leaf and space) according to Eberle et al. 2018.

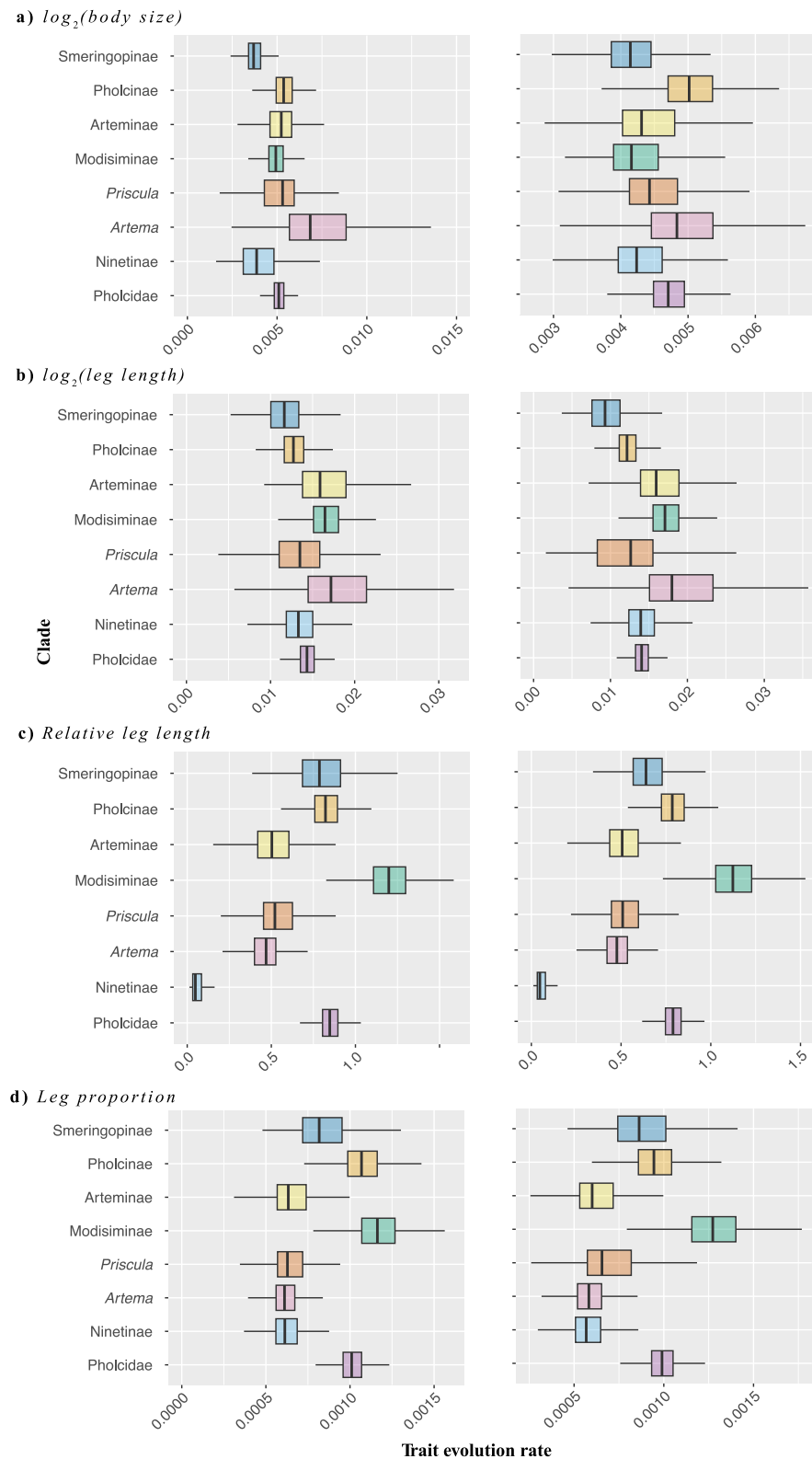


Figure B.16. Trait evolution rate results. The BAMM was used for the analysis. Left: based on all tips with traits (ca. 900 spp.). Right: based on randomly sampled 650 taxa.

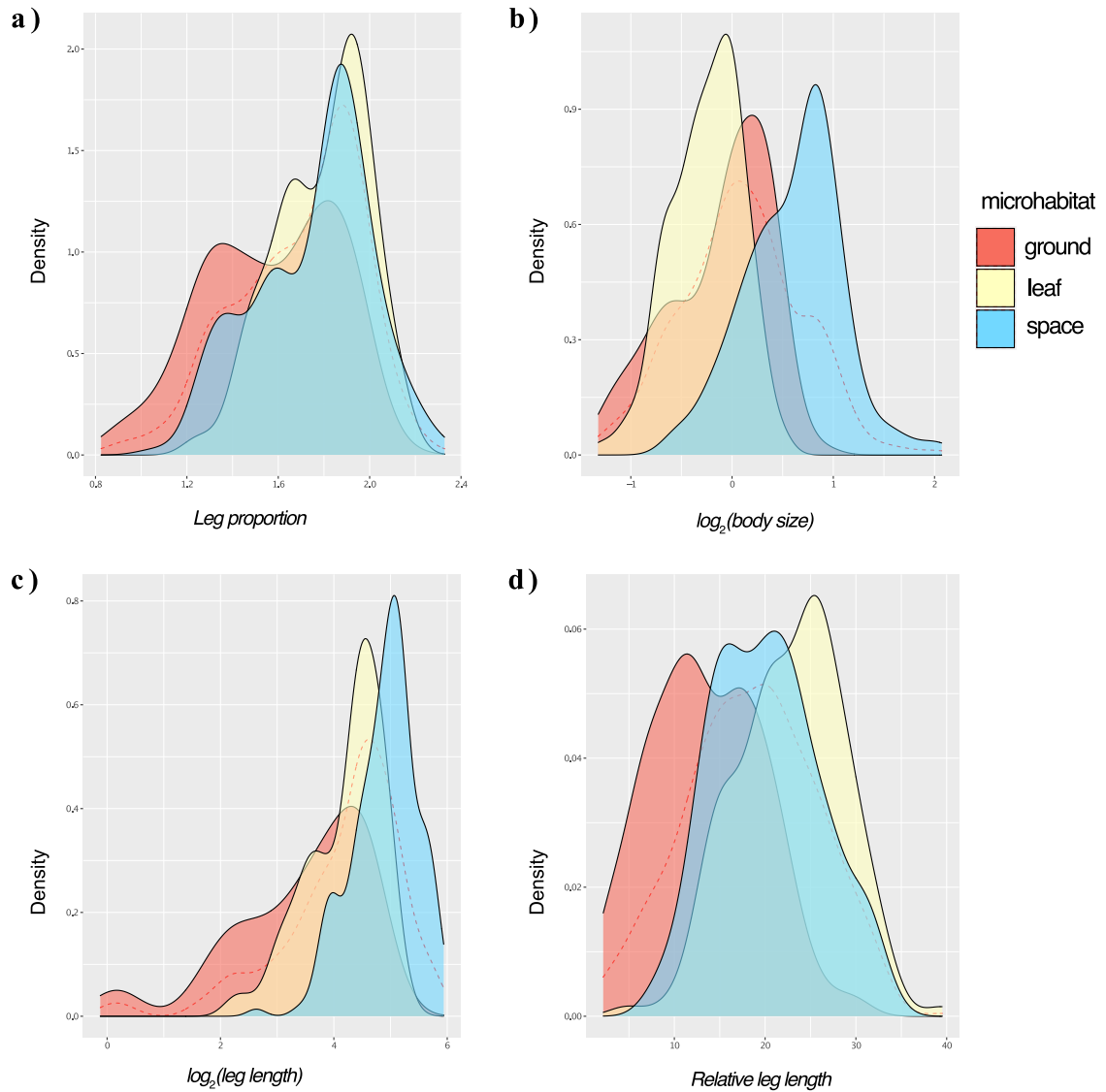


Figure B.17. Trait distribution among microhabitats. These analyses were based on the ClipKIT smart-gap tree and the species with microhabitat data (ground/leaf/space) available. **(a)** leg proportion distribution. **(b)** $\log_2(\text{body size})$ distribution. **(c)** $\log_2(\text{leg length})$ distribution. **(d)** relative leg length distribution.

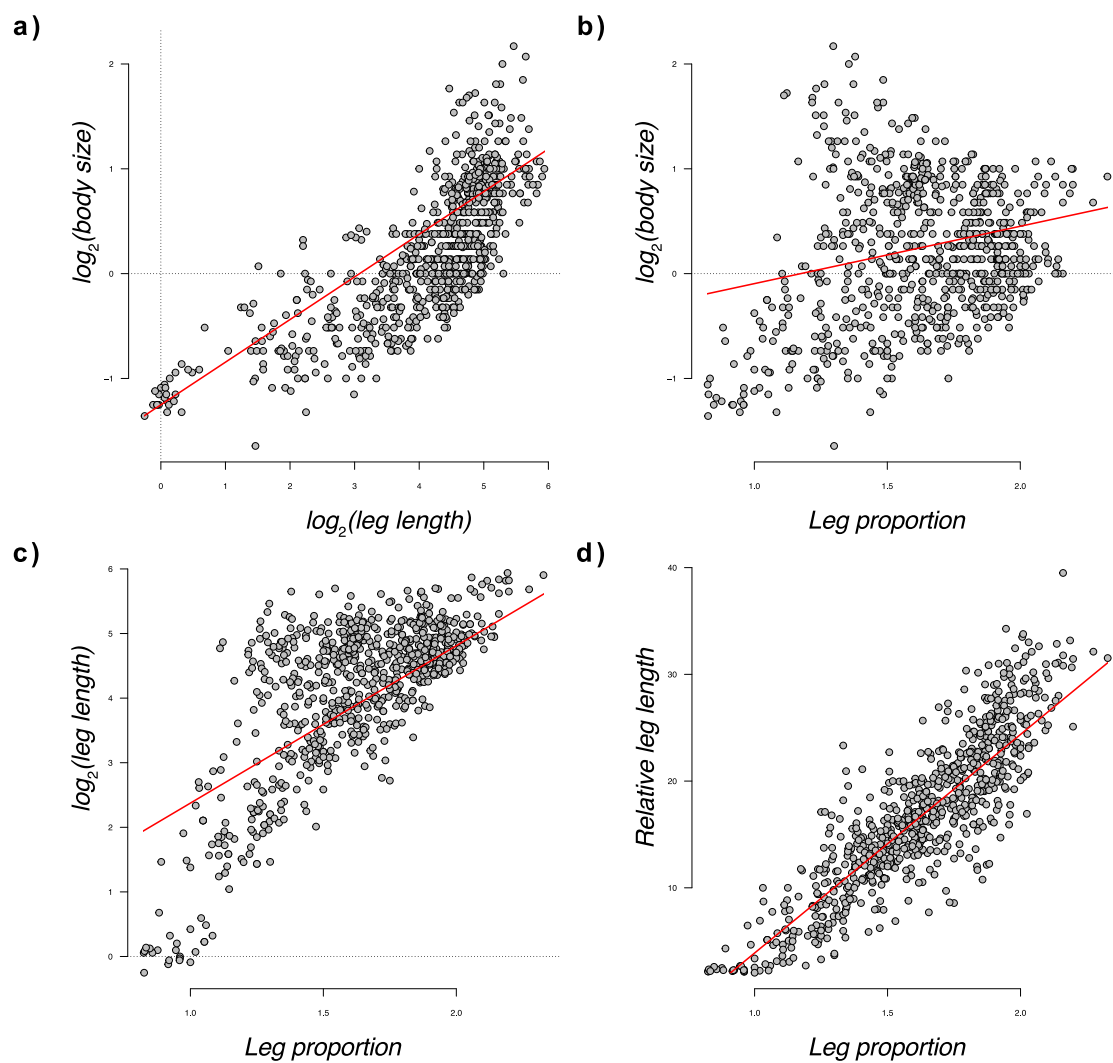
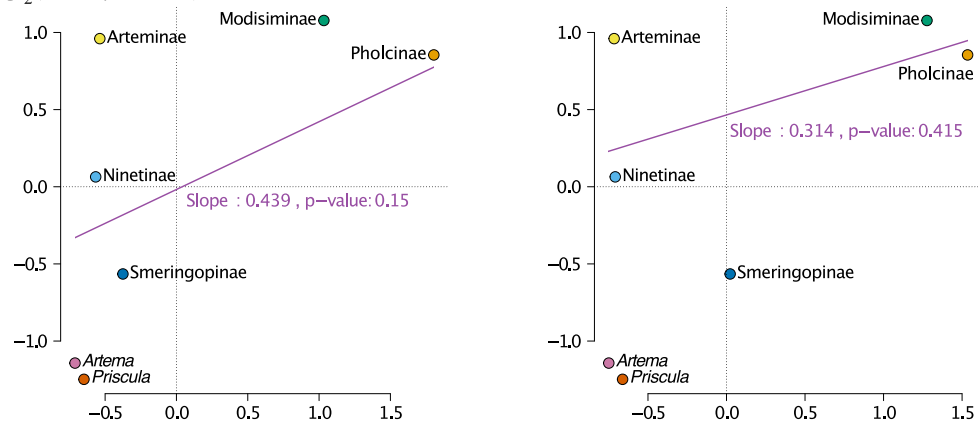
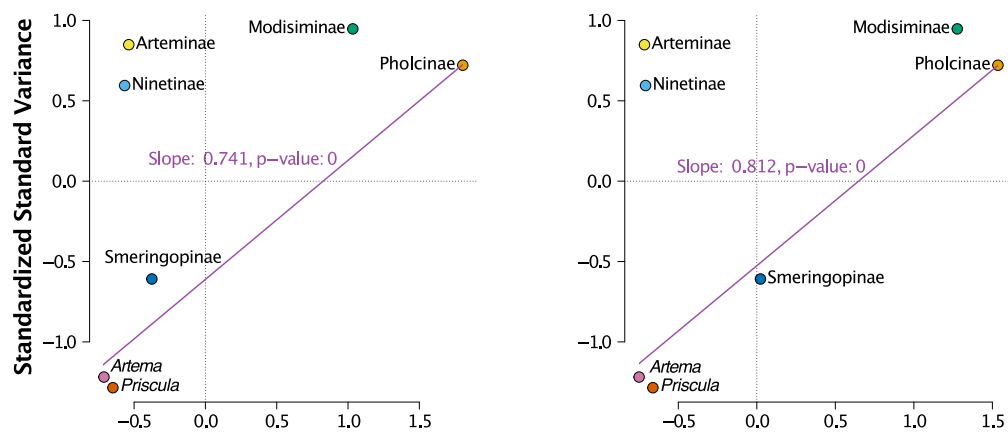


Figure B.18. Relationship between different traits. These analyses (Phylogenetic Generalized Least Squares (PGLS) regression) were based on the `ClipKIT smart-gap` tree. **(a)** $\log_2(\text{body size})$ vs. $\log_2(\text{leg length})$. **(b)** $\log_2(\text{body size})$ vs. *leg proportion*. **(c)** $\log_2(\text{leg length})$ vs. *leg proportion*. **(d)** *relative leg length* vs. *leg proportion*.

a) $\log_2(\text{body size})$



b) $\log_2(\text{leg length})$



c) *Leg proportion*

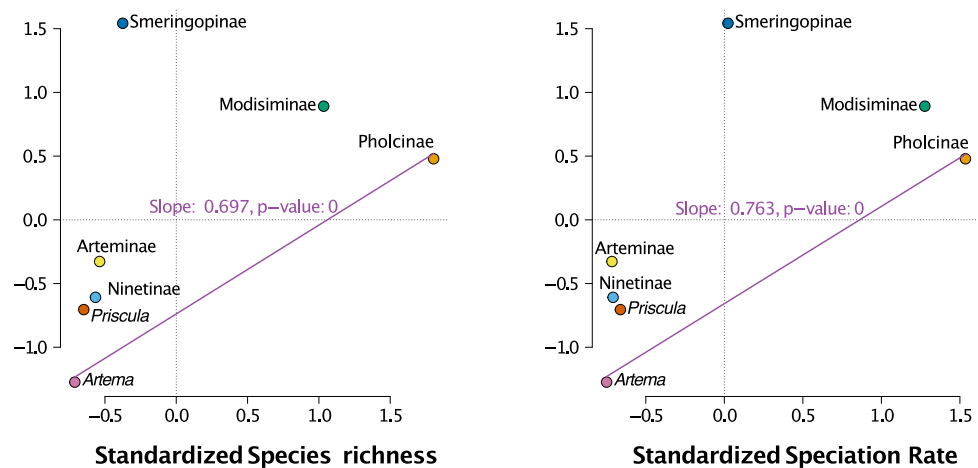


Figure B.19. Relationships between disparity, species richness and speciation rate. Left: Relationship between disparity (Standardized Standard Variance) and species richness. Right: Relationship between disparity and speciation rate. These relationships were determined by the PGLS regression based on the ClipKIT smart-gap tree.

Table B.1. Sequences used to create the six-gene matrices. Sample codes with “OG” prefix are outgroup, obtained from Eberle et al. 2018. All sequences have been uploaded to <https://datadryad.org/>

Code	CO1	H3	12S	16S	18S	28S
BB01	AY560771	None	AY560685	AY560663	None	None
BB09a	AY560778	None	AY560695	AY560669	None	AY560735
BB10	AY560780	None	AY560698	None	None	AY560738
BB11	AY560781	None	AY560699	AY560671	None	AY560739
BB13	AY560783	None	AY560702	None	None	AY560741
BB14	AY560784	None	AY560703	AY560673	None	AY560742
BB21	AY560788	None	AY560712	None	None	AY560751
BB23	AY560789	None	AY560714	None	None	AY560753
BB26	AY560793	None	AY560721	None	None	AY560759
BB27	AY560792	None	AY560720	None	None	AY560758
BB28	None	None	AY560717	None	None	AY560755
BB32	AY560772	None	AY560686	AY560664	None	AY560733
BH02	BH02	None	None	None	None	BH02
BH04	None	None	None	None	None	BH04
BH06	BH06	None	None	None	None	None
BH099	BH099	None	None	None	None	None
BH10	OP288810	None	None	None	None	OP289815
BH109	OP288812	None	None	None	None	None
BH111	BH111	None	None	None	None	BH111
BH115	None	None	None	None	None	OL872163
BH12	BH12	None	None	None	None	None
BH13	BH13	None	None	None	None	None
BH134	OQ301948	None	None	None	None	None
BH142	OP288817	None	None	None	None	OP289819
BH148	OQ301950	None	None	None	None	OQ311352
BH158	BH158	None	None	None	None	BH158_28S.1
BH159	BH159	None	None	None	None	BH159_28S.1
BH16	OP288821	None	None	None	None	OP289821
BH163	OP288822	None	None	None	None	OP289800
BH17	OP288829	None	None	None	None	OP289827
BH176	BH176	None	None	None	None	None
BH20	OP288833	None	None	None	None	OP289830
BH21	OP288834	None	None	None	None	OP289831
BH24	OP288837	None	None	None	None	None
BH25	OP288838	None	None	None	None	OP289833
BH27	OP288840	None	None	None	None	OP289835
BH29	OP288842	None	None	None	None	OP289837
BH30	OP288843	None	None	None	None	OP289838
BH31	OP288844	None	None	None	None	None
BH32	OP288845	None	None	None	None	OP289799
BH37	BH37	None	None	None	None	None
BH49	OL870339	None	None	None	None	OL872161
BH55	BH55	None	None	None	None	None
BH57	BH57	None	None	None	None	None
BH60	BH60	None	None	None	None	None
BH70	OQ301940	None	None	None	None	None
BH78	OQ301943	None	None	None	None	None
BH82	OQ301947	None	None	None	None	None
BH83	OP288847	None	None	None	None	OP289839
BH86	OP288850	None	None	None	None	OP289841
BH87	OP288851	None	None	None	None	OP289842
BH88	OP288852	None	None	None	None	OP289843
BH89	OP288853	None	None	None	None	OP289844
E001	None	E001	None	None	E001	E001
E005	E005	E005	None	E005	None	None
E015	None	E015	None	None	E015	None

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
E023	E023	E023	None	E023	None	None
E027	E027	E027	None	E027	E027	E027
E033	E033	None	E033	E033	E033	None
E035	E035	None	None	E035	None	None
E039	E039	E039	None	None	E039	E039
E041	E041	E041	E041	E041	None	E041
E042	E042	E042	E042	E042	None	None
E046	E046	E046	E046	E046	E046	E046
E051	E051	E051	E051	E051	E051	None
E052	E052	None	E052	E052	E052	None
E053	E053	E053	None	None	None	E053
E055	E055	E055	E055	E055	E055	E055
E057	E057	E057	None	None	None	None
E058	None	E058	None	None	E058	E058
E061	E061	E061	None	E061	E061	None
E065	None	E065	None	E065	None	None
E072	E072	E072	None	E072	E072	None
E074	E074	E074	None	E074	E074	None
E075	E075	E075	E075	E075	E075	None
E076	E076	E076	E076	E076	E076	None
E079	E079	E079	E079	E079	E079	None
E082	None	E082	E082	E082	None	None
E083	E083	E083	None	None	None	None
E084	E084	E084	None	E084	None	None
E088	E088	E088	E088	E088	E088	None
E089	E089	E089	None	None	E089	E089
E092	E092	E092	E092	E092	E092	None
E093	None	E093	None	None	None	None
E096	E096	E096	None	E096	E096	E096
EB002	OL870350	None	None	None	None	None
EB003	OQ301958	None	None	None	None	None
EB005	EB005	None	None	None	None	None
EB008	EB008	None	None	None	None	None
EB010	EB010	None	None	None	None	None
EB016	EB016	None	None	None	None	None
EB018	EB018	None	None	None	None	None
EB026	EB026	None	None	None	None	None
EB027	OP288861	None	None	None	None	None
EB030	EB030	None	None	None	None	None
EB032	OQ301961	None	None	None	None	None
EB036	OP288863	None	None	None	None	None
EB038	EB038	None	None	None	None	None
EB039	EB039	None	None	None	None	None
EB040	EB040	None	None	None	None	None
EB041	EB041	None	None	None	None	None
EB042	OQ301962	None	None	None	None	None
EB044	OL870351	None	None	None	None	None
EB045	EB045	None	None	None	None	None
EB047	EB047	None	None	None	None	None
EB048	EB048	None	None	None	None	None
EB050	EB050	None	None	None	None	None
EB051	EB051	None	None	None	None	None
EB054	EB054	None	None	None	None	None
EB055	EB055	None	None	None	None	None
EB056	OL870353	None	None	None	None	None
EB059	EB059	None	None	None	None	None
EB062	OL870354	None	None	None	None	None
EB065	EB065	None	None	None	None	None
EB068	EB068	None	None	None	None	None
EB069	EB069	None	None	None	None	None

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
EB071	EB071	None	None	None	None	None
EB074	EB074	None	None	None	None	None
EB075	EB075	None	None	None	None	None
EB076	EB076	None	None	None	None	None
EB078	OQ301965	None	None	None	None	None
EB079	EB079	None	None	None	None	None
EB080	EB080	None	None	None	None	None
EB082	EB082	None	None	None	None	None
EB083	OQ301966	None	None	None	None	None
EB084	OQ301967	None	None	None	None	None
EB086	EB086	None	None	None	None	None
EB087	OL870356	None	None	None	None	None
EB089	EB089	None	None	None	None	None
EB091	EB091	None	None	None	None	None
EB094	OP288864	None	None	None	None	OP289801
EB095	OQ301969	None	None	None	None	None
EB097	EB097	None	None	None	None	None
EB101	EB101	None	None	None	None	None
EB103	OL870360	None	None	None	None	None
EB109	EB109	None	None	None	None	None
EB115	OQ301973	None	None	None	None	None
EB116	EB116	None	None	None	None	None
EB117	EB117	None	None	None	None	None
EB119	OQ301974	None	None	None	None	None
EB120	EB120	None	None	None	None	None
EB122	EB122	None	None	None	None	None
EB123	EB123	None	None	None	None	None
EB125	EB125	None	None	None	None	None
EB126	EB126	None	None	None	None	None
EB128	EB128	None	None	None	None	None
EB129	EB129	None	None	None	None	None
EB131	EB131	None	None	None	None	None
EB132	EB132	None	None	None	None	None
EB134	EB134	None	None	None	None	None
EB146	OP288799	None	None	None	None	None
EB147	OP288868	None	None	None	None	OP289805
EB150	OP288869	None	None	None	None	OP289806
EB151	OP288800	None	None	None	None	OP289807
EB152	OP288801	None	None	None	None	OP289808
EB153	OP288802	None	None	None	None	OP289809
EB154	OP288803	None	None	None	None	OP289810
EB155	OP288804	None	None	None	None	OP289811
EB156	OP288805	None	None	None	None	OP289812
EB157	OP288870	None	None	None	None	OP289813
EB158	OP288806	None	None	None	None	OP289814
EB159	OP288807	None	None	None	None	None
EB160	OP288808	None	None	None	None	None
EB164	OQ301930	None	None	None	None	None
EB165	OQ301929	None	None	None	None	None
EB166	OQ301928	None	None	None	None	None
EB171	OQ301920	None	None	None	None	None
EB173	OQ301924	None	None	None	None	None
EB177	OQ301926	None	None	None	None	OQ311349
EB179	OQ301921	None	None	None	None	None
GB01	MG268802	None	None	None	None	None
GB07	MG268727	None	None	None	None	None
GB08	MG268728	None	None	None	None	None
GB09	MG268695	None	None	None	None	None
GB10	MG268694	None	None	None	None	None
GB11	MG268629	None	None	None	None	None

Continued on next page

Appendix B Supplementary information for Chapter 4

Code	CO1	H3	12S	16S	18S	28S
GB14	MG268662	None	None	None	None	None
GB15	MG268665	None	None	None	None	None
GB16	MG268663	None	None	None	None	None
GB17	MG268664	None	None	None	None	None
GB18	MG268654	None	None	None	None	None
GB19	MG268645	None	None	None	None	None
GB20	MG268652	None	None	None	None	None
GB21	MG268644	None	None	None	None	None
GB22	MG268658	None	None	None	None	None
GB24	MG268870	None	None	None	None	None
GB30	MG268716	None	None	None	None	None
IA01	MG268708	None	None	MG268024	None	MG268607
IA02	MG268671	None	None	MG268022	None	MG268608
IA04	MG268710	None	None	MG268025	None	MG268609
IA05	MG268689	None	None	MG268023	None	MG268610
Ixc29	KF150114	None	MG267480	KF178420	None	MG268421
Ixc39	KF150089	None	None	KF178395	None	None
Ixc51	KF150126	MG268979	MG267481	KF178432	MG268104	MG268422
Ixc56	KF150128	None	None	KF178434	None	None
JA10	DQ667880	None	None	DQ667772	None	None
JA100	DQ667858	None	None	DQ667753	None	None
JA102	DQ667856	None	None	DQ667750	None	DQ667840
JA107	DQ667859	None	MG267466	DQ667752	None	None
JA11	DQ667891	None	None	DQ667785	None	None
JA13	DQ667942	None	None	DQ667835	None	None
JA131	FJ799781	None	None	FJ799766	None	None
JA134	FJ799782	None	None	FJ799769	None	None
JA15	DQ667889	None	None	DQ667783	None	None
JA19	DQ667876	None	None	DQ667770	None	None
JA23	DQ667887	None	None	DQ667781	None	None
JA27	DQ667890	None	None	DQ667890	None	None
JA29	DQ667874	None	None	DQ667767	None	None
JA41	DQ667873	None	None	DQ667766	None	DQ667846
JA50	DQ667875	None	None	DQ667769	None	None
JA52	DQ667865	None	None	None	None	None
JA53	DQ667882	None	None	DQ667775	None	None
JA54	DQ667884	None	None	None	None	None
JA70	DQ667930	None	MG267482	DQ667825	MG268105	MG268423
JA71	DQ667928	None	MG267437	None	MG268106	MG268383
JA8	DQ667894	None	None	DQ667788	None	None
JA80	DQ667853	None	None	DQ667838	None	None
JA85	DQ667900	None	None	DQ667793	None	None
JA89	DQ667861	None	AY560696	DQ667756	None	AY560736
JA93	DQ667872	None	None	DQ667764	None	DQ667845
M003	OL870363	None	None	None	None	None
M004	M004	None	None	None	None	None
M008	OL870367	None	None	None	None	None
M009	OL870368	None	None	None	None	None
M011	OL870370	None	None	None	None	None
M014	OL870373	None	None	None	None	None
M016	OL870375	None	None	None	None	None
M018	OL870377	None	None	None	None	None
M019	OL870378	None	None	None	None	None
M023	OL870381	None	None	None	None	None
M024	OL870382	None	None	None	None	None
M025	OL870383	None	None	None	None	None
M026	M026	None	None	None	None	None
M029	OL870386	None	None	None	None	None
M031	OL870388	None	None	None	None	None
M032	OL870389	None	None	None	None	None

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
M036	M036	None	None	None	None	None
M037	OL870393	None	None	None	None	None
M039	OL870395	None	None	None	None	None
M042	M042	None	None	None	None	None
M043	M043	None	None	None	None	None
M044	M044	None	None	None	None	None
M046	M046	None	None	None	None	None
M047	M047	None	None	None	None	None
M048	OL870397	None	None	None	None	None
M049	OL870398	None	None	None	None	None
M051	OL870400	None	None	None	None	None
M052	OL870401	None	None	None	None	None
M053	M053	None	None	None	None	None
M054	OL870402	None	None	None	None	None
M055	OL870403	None	None	None	None	None
M059	OL870406	None	None	None	None	None
M060	M060	None	None	None	None	None
M061	M061	None	None	None	None	None
M062	OL870407	None	None	None	None	None
M063	OL870408	None	None	None	None	None
M064	M064	None	None	None	None	None
M066	OL870409	None	None	None	None	None
M067	M067	None	None	None	None	None
M068	OL870410	None	None	None	None	None
M073	ON504284	None	None	None	None	None
M075	ON504286	None	None	None	None	None
M078	ON504289	None	None	None	None	None
M079	ON504290	None	None	None	None	None
M080	ON504291	None	None	None	None	None
M081	ON504292	None	None	None	None	None
M084	ON504294	None	None	None	None	None
M085	ON504295	None	None	None	None	None
M094	OL870415	None	None	None	None	None
MRL307	OQ525972	None	None	None	None	None
Mic02	KF715606	KF715577	None	None	KF715594	KF715612
Mic16	None	KF715575	None	KF715619	KF715596	KF715614
N001	ON063450	None	None	None	None	None
N015	ON063456	None	None	None	None	None
N019	ON067461	None	None	None	None	None
N020	ON063459	None	None	None	None	None
N023	ON063461	None	None	None	None	None
N024	ON067462	None	None	None	None	None
N042	ON063472	None	None	None	None	None
N046	OQ453161	None	None	None	None	None
N059	ON970481	None	None	None	None	None
N088	PP192791	None	None	None	None	None
N091	N091	None	None	None	None	None
N092	N092	None	None	None	None	None
N095	N095	None	None	None	None	None
NA001	MG268903	None	None	None	None	MG268611
NA002	MG268902	None	None	None	None	MG268612
NF1	FJ228016	None	None	FJ227980	None	None
NF101	FJ228008	None	None	FJ227973	None	None
NF13	FJ228018	None	None	FJ227982	None	None
NF24	FJ228032	None	None	FJ227964	None	None
NF3	FJ228017	None	None	FJ227981	None	None
NF37	FJ228021	None	None	FJ227987	None	None
NF41	FJ228025	None	None	FJ227991	None	None
NF53	FJ228009	None	None	FJ227974	None	None
NF64	FJ228000	None	None	FJ227962	None	None

Continued on next page

Appendix B Supplementary information for Chapter 4

Code	CO1	H3	12S	16S	18S	28S
NF66	FJ228006	None	None	FJ227970	None	None
NF68	FJ228003	None	None	FJ227967	None	None
NF69	FJ228030	None	None	FJ227995	None	None
NF71	FJ228019	None	None	FJ227983	None	None
NF80	FJ228011	None	None	FJ227976	None	None
NF81	FJ228026	None	None	FJ227992	None	None
NF87	FJ227997	None	None	FJ227959	None	None
NF89	FJ227998	None	None	FJ227960	None	None
NF9	FJ228013	None	None	FJ227978	None	None
P0145	None	JX023610	JX023759	JX023852	JX023953	JX024066
P0146	JX023545	JX023611	None	JX023853	JX023954	JX024067
P0147	JX023546	JX023612	JX023760	JX023854	None	None
P0148	None	JX023613	JX023761	None	None	None
P0149	JX023547	JX023614	JX023762	JX023855	None	None
P0150	JX023548	JX023615	JX023763	JX023856	JX023955	JX024068
P0151	JX023549	JX023616	JX023764	JX023857	JX023956	JX024069
P0153	JX023550	JX023618	JX023766	JX023859	None	None
P0155	None	JX023620	JX023768	JX023861	JX023958	None
P0156	JX023552	None	JX023769	JX023862	JX023959	JX024072
P0157	None	None	JX023770	JX023863	JX023960	JX024073
P0160	JX023554	JX023622	JX023773	None	JX023962	JX024076
P0161	None	None	JX023774	JX023866	JX023963	None
P0163	JX023556	None	JX023776	JX023868	None	None
P0165	JX023558	JX023625	JX023778	None	JX023966	None
P0167	None	JX023627	JX023780	None	JX023968	JX024078
P0170	None	JX023630	JX023782	JX023871	JX023971	JX024080
P0171	JX023561	JX023631	JX023783	JX023872	None	None
P0172	None	JX023632	JX023784	JX023873	None	None
P0173	None	JX023633	None	JX023874	None	None
P0177	JX023563	JX023637	JX023788	JX023877	JX023974	None
P0178	None	None	None	JX023878	JX023975	JX024084
P0180	JX023565	JX023639	JX023790	JX023880	None	None
P0182	JX023566	None	JX023792	JX023882	JX023977	None
P0183	JX023567	JX023641	JX023793	JX023883	JX023978	JX024086
P0184	JX023568	None	JX023794	JX023884	JX023979	JX024087
P0187	JX023570	JX023644	JX023797	None	JX023982	JX024089
P0189	None	None	JX023798	None	JX023984	JX024091
P0190	JX023572	None	JX023799	None	JX023985	JX024092
P0191	JX023573	None	None	JX023885	JX023986	JX024093
P0193	JX023574	JX023647	JX023801	JX023887	JX023988	JX024095
P0196	JX023575	JX023649	JX023803	JX023890	JX023989	None
P0197	JX023576	None	None	None	JX023990	JX024098
P0198	JX023577	None	None	JX023891	JX023991	JX024099
P0199	JX023578	JX023650	None	JX023892	JX023992	JX024100
P0201	None	JX023652	JX023804	JX023894	JX023994	JX024102
P0202	JX023579	JX023653	JX023805	None	JX023995	None
P0204	JX023581	JX023655	None	JX023896	JX023996	JX024104
P0205	JX023582	JX023656	JX023807	JX023897	JX023997	JX024105
P0208	JX023584	JX023659	JX023808	None	JX024000	JX024107
P0209	JX023585	JX023660	JX023809	JX023900	None	JX024108
P0210	JX023586	JX023661	JX023810	JX023901	JX024001	JX024109
P0211	JX023587	JX023662	JX023811	None	None	None
P0212	JX023588	JX023663	JX023812	JX023902	None	None
P0214	JX023589	JX023665	None	JX023904	JX024002	JX024110
P0215	JX023590	JX023666	JX023814	JX023905	JX024003	JX024111
P0216	JX023591	JX023667	JX023815	JX023906	JX024004	JX024112
P0217	JX023592	JX023668	JX023816	None	JX024005	None
P0218	None	JX023669	None	JX023907	JX024006	None
P0220	None	JX023671	None	JX023908	JX024008	JX024114
P0221	JX023594	None	None	None	JX024009	JX024115

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
P0222	JX023595	None	JX023817	JX023909	JX024010	JX024116
P0223	JX023596	JX023672	JX023818	None	JX024011	None
P0224	None	JX023673	None	JX023910	JX024012	JX024117
P0225	None	JX023674	JX023819	JX023911	JX024013	None
P0226	None	None	None	None	JX024014	JX024118
P0227	None	JX023675	JX023820	JX023912	JX024015	JX024119
P0228	JX023597	JX023676	None	JX023913	JX024016	JX024120
P0230	JX023598	JX023677	None	JX023914	JX024017	JX024121
P0231	None	JX023678	JX023821	JX023915	JX024018	None
P0232	JX023599	JX023679	JX023822	JX023916	JX024019	JX024122
P0234	JX023601	JX023681	JX023823	JX023917	JX024021	None
P0235	None	None	None	JX023918	JX024022	JX024123
P0237	None	None	None	JX023919	JX024024	JX024125
P0238	None	None	None	JX023920	JX024025	None
P0241	None	JX023683	JX023825	None	JX024028	JX024127
P0244	None	None	None	JX023922	JX024030	None
P0245	None	None	JX023828	JX023923	None	None
P0247	JX023603	JX023685	JX023829	JX023924	JX024032	JX024129
P0248	None	None	JX023830	None	JX024033	JX024130
P0256	None	None	None	None	JX024037	JX024133
P0257	None	JX023688	None	JX023927	JX024038	None
P0259	None	JX023690	JX023834	None	JX024040	JX024134
P0262	None	JX023693	None	None	JX024042	None
P0264	None	JX023695	JX023836	JX023931	JX024044	JX024136
P0265	None	JX023696	None	JX023932	None	None
P0266	None	JX023697	JX023837	JX023933	JX024045	None
P0267	None	JX023698	JX023838	JX023934	JX024046	JX024138
P0277	JX023605	JX023705	None	JX023942	JX024053	JX024142
P0281	JX023606	None	JX023846	JX023945	None	None
P0282	None	JX023708	None	JX023946	JX024058	None
P0284	None	JX023710	None	None	JX024060	None
P0285	JX023607	JX023711	None	JX023948	None	None
P0287	JX023608	JX023713	JX023848	None	JX024062	JX024145
P0288	JX023609	JX023714	JX023849	JX023950	JX024063	JX024146
S002	None	MG268969	None	None	MG268035	MG268389
S009	MG268811	None	MG267447	None	MG268084	None
S010	MG268881	None	MG267455	None	MG268107	None
S012	None	MG269173	MG267441	MG267763	MG268032	None
S015	None	MG269076	MG267452	None	MG268044	MG268370
S017	None	MG269061	MG267458	MG267779	MG268039	MG268382
S019	None	MG269058	MG267454	None	MG268072	MG268369
S020	None	MG269059	MG267453	None	MG268071	None
S021	MG268822	MG268993	MG267434	MG267759	MG268041	None
S022	None	None	None	MG267781	MG268053	None
S024	None	MG268974	MG267448	MG267741	None	None
S025	None	MG269095	MG267449	MG267742	MG268070	None
S026	MG268774	MG269032	None	MG267749	MG268060	MG268357
S028	MG268775	MG269033	None	MG267748	MG268085	MG268358
S029	None	MG269030	MG267495	MG267750	MG268064	None
S031	MG268788	MG269036	MG267431	MG267744	MG268061	MG268361
S032	MG268706	None	None	MG267770	MG268065	MG268367
S034	MG268680	MG269043	MG267432	MG267752	MG268058	MG268359
S035	None	MG269041	MG267433	MG267743	MG268059	MG268360
S036	MG268758	MG269048	MG267427	MG267755	MG268054	None
S037	MG268756	MG269046	MG267428	MG267754	MG268055	MG268355
S038	MG268783	MG269038	MG267430	MG267745	MG268062	MG268353
S039	None	None	None	None	MG268036	MG268368
S040	MG268772	MG269044	MG267429	MG267746	MG268057	None
S041	MG268622	MG269119	MG267459	MG267774	MG268078	MG268387
S042	None	None	None	MG267773	MG268079	None

Continued on next page

Appendix B Supplementary information for Chapter 4

Code	CO1	H3	12S	16S	18S	28S
S043	None	MG269066	None	MG267776	MG268082	None
S044	MG268660	None	MG267460	MG267775	MG268080	None
S045	None	None	None	MG267782	MG268034	MG268366
S047	DQ667927	None	None	None	MG268027	MG268352
S049	None	None	MG267462	MG267777	MG268077	MG268386
S051	None	None	None	MG267764	MG268031	MG268351
S052	None	MG269134	MG267445	MG267768	None	None
S053	None	None	None	MG267766	MG268030	MG268350
S056	MG268880	None	MG267467	MG267765	MG268029	None
S057	None	None	MG267446	MG267758	MG268083	None
S063	None	MG269138	None	None	MG268075	None
S064	None	None	MG267444	MG267762	MG268033	MG268388
S066	MG268899	None	None	MG267760	None	MG268385
S067	None	MG269086	None	None	MG268047	MG268374
S068	None	MG269126	None	MG267757	MG268069	None
S070	MG268865	MG269174	None	MG267769	MG268259	None
S071	None	None	None	MG267761	MG268260	None
S072	MG268862	None	MG267496	MG267798	MG268108	None
S075	None	None	MG267497	MG267738	MG268052	MG268379
S076	None	None	None	None	None	MG268380
S077	MG268795	MG269084	None	MG267739	MG268050	MG268378
S078	MG268712	MG269125	None	MG267756	MG268068	None
S079	MG268797	None	None	MG267740	MG268049	MG268376
S080	None	MG269075	None	None	None	MG268371
S101	MG268907	None	None	None	None	MG268424
S102	MG268964	None	MG267483	None	None	MG268425
S103	MG268904	None	MG267484	None	MG268109	None
S104	MG268885	MG269172	MG267485	None	MG268110	None
S105	KJ650076	None	MG267486	None	MG268111	MG268426
S108	MG268938	None	MG267487	None	MG268112	MG268428
S109	MG268936	None	MG267488	None	None	None
S110	MG268935	None	MG267489	None	MG268113	MG268429
S113	MG268709	None	None	None	MG268115	MG268431
S115	MG268703	None	None	None	MG268116	None
S117	None	MG269117	MG267498	None	MG268095	MG268403
S118	MG268803	MG269079	None	None	MG268096	MG268404
S119	MG268800	MG269080	None	None	MG268097	MG268405
S126	None	None	None	None	MG268101	MG268408
S127	MG268627	MG269100	None	MG267786	MG268102	MG268409
S128	None	MG269102	None	MG267787	MG268103	None
S129	None	MG269101	MG267499	MG267788	None	None
S133	None	MG269164	MG267500	None	None	None
S136	MG268919	MG269163	MG267503	None	None	None
S172	None	None	MG267507	None	None	None
S173	MG268931	None	MG267491	MG267935	None	MG268415
S174	MG268930	None	MG267492	None	None	MG268416
S182	MG268841	MG269157	None	None	MG268117	MG268417
S183	MG268926	MG269176	None	None	None	None
S184	MG268920	MG269165	None	None	None	None
S185	MG268842	MG269158	None	None	MG268118	MG268418
S186	MG268835	None	None	None	None	None
S188	MG268863	None	MG267493	None	MG268119	MG268419
S218	MG268901	None	MG267479	None	None	MG268390
S219	None	MG269099	MG267478	None	None	MG268391
S222	MG268900	None	MG267477	None	None	MG268393
S224	MG268898	None	MG267475	None	None	MG268395
S226	MG268948	None	MG267473	None	None	MG268396
S227	MG268850	None	None	None	MG268090	MG268397
S228	None	MG269112	None	None	None	MG268398
S229	MG268918	None	None	MG267800	None	None

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
S230	MG268868	MG269096	MG267472	None	MG268091	MG268399
S231	MG268624	None	None	MG267785	None	MG268400
S232	MG268749	MG268973	MG267471	MG267784	MG268092	MG268401
S233	None	MG268971	None	None	MG268093	None
S234	MG268869	MG269132	MG267470	None	None	MG268402
S236	MG268906	MG269131	MG267469	None	None	None
S237	None	None	MG267468	None	MG268094	MG268474
S239	None	None	MG267508	None	MG268122	MG268433
S241	None	MG269104	MG267509	None	MG268123	MG268434
S242	MG268833	MG269103	MG267510	None	MG268124	MG268435
S243	None	MG269108	MG267511	MG267789	MG268125	MG268436
S244	MG268825	None	MG267512	None	MG268126	None
S246	None	None	None	MG267790	MG268127	MG268437
S247	MG268823	None	MG267513	MG267791	None	MG268438
S248	MG268731	MG269094	MG267514	MG267792	MG268128	MG268439
S249	MG268623	None	MG267515	MG267793	MG268129	None
S251	None	None	MG267516	MG267794	None	None
S252	None	None	MG267517	None	MG268130	MG268440
S254	None	MG269124	MG267519	MG267796	MG268132	MG268442
S255	None	MG269141	MG267520	None	None	MG268443
S256	MG268858	MG269180	MG267563	MG267857	MG268133	None
S257	MG268859	None	MG267521	MG267858	MG268134	None
S258	MG268879	None	MG267522	None	None	MG268444
S259	MG268839	MG269151	MG267523	MG267801	None	MG268445
S261	MG268915	None	MG267525	MG267803	None	MG268447
S262	MG268843	MG269159	MG267526	MG267804	None	MG268448
S263	MG268917	None	MG267527	MG267805	None	MG268449
S264	MG268834	None	MG267528	None	MG268135	MG268450
S265	MG268830	MG269156	MG267529	MG267806	MG268136	MG268451
S266	None	MG269171	MG267530	None	None	None
S267	MG268836	None	MG267531	MG267807	MG268137	MG268452
S268	MG268832	MG269155	MG267532	MG267808	MG268138	None
S269	MG268840	MG269153	MG267533	MG267809	MG268139	MG268453
S271	MG268916	None	MG267535	MG267810	MG268141	MG268455
S272	MG268913	None	MG267536	MG267811	None	None
S273	MG268861	MG269179	MG267537	MG267812	MG268142	None
S274	MG268912	None	MG267538	MG267813	None	MG268456
S275	MG268911	None	None	MG267814	None	None
S278	MG268939	None	None	MG267817	MG268144	MG268458
S279	None	None	None	MG267818	MG268145	MG268459
S280	MG268760	None	None	None	MG268146	MG268460
S281	MG268941	None	None	MG267819	MG268147	MG268461
S283	MG268962	None	MG267541	MG267820	MG268149	MG268463
S284	MG268966	None	None	MG267821	MG268150	MG268464
S285	MG268965	None	MG267542	MG267822	MG268151	MG268465
S287	MG268737	None	None	MG267823	MG268182	None
S288	MG268787	MG269034	MG267544	MG267824	MG268153	MG268467
S289	MG268619	MG269147	MG267545	MG267825	MG268154	MG268468
S290	MG268894	None	MG267546	MG267826	MG268155	MG268469
S291	MG268806	MG269148	MG267547	MG267827	MG268156	MG268470
S293	None	None	MG267548	MG267828	MG268157	None
S294	OP288828	MG269016	None	MG267829	MG268158	MG268471
S295	MG268632	None	MG267549	MG267830	MG268159	MG268472
S296	MG268723	MG269144	MG267550	MG267831	MG268160	None
S297	MG268724	MG269145	MG267551	MG267832	MG268161	MG268473
S298	MG268741	None	MG267552	MG267833	None	MG268475
S299	MG268925	MG269177	MG267553	None	MG268162	MG268476
S300	MG268633	MG269118	MG267554	MG267834	MG268163	None
S301	MG268661	MG269067	None	MG267835	MG268164	None
S302	MG268635	None	MG267555	MG267836	MG268165	None

Continued on next page

Appendix B Supplementary information for Chapter 4

Code	CO1	H3	12S	16S	18S	28S
S303	None	None	MG267556	MG267837	MG268166	MG268477
S304	None	MG269013	None	MG267838	None	MG268478
S305	MG268690	MG269007	None	MG267839	MG268167	None
S306	MG268942	None	None	MG267840	MG268168	MG268479
S308	MG268707	None	None	MG267842	MG268169	MG268480
S309	MG268693	None	None	MG267843	MG268186	MG268481
S310	None	MG269022	MG267564	MG267862	MG268187	MG268482
S312	MG268734	None	None	MG267845	MG268183	None
S313	None	None	None	MG267846	None	None
S315	None	None	MG267559	MG267848	MG268171	None
S316	MG268808	None	MG267560	MG267849	MG268172	None
S318	MG268730	MG269122	None	MG267851	MG268174	None
S319	MG268763	MG269133	None	MG267852	MG268175	None
S321	MG268754	MG268987	None	MG267854	MG268177	MG268484
S322	MG268651	MG269064	None	None	MG268181	None
S323	None	None	None	MG267855	MG268178	None
S324	None	None	MG267562	MG267856	MG268179	None
S330	MG268790	MG269081	MG267565	MG267863	MG268188	MG268542
S331	MG268761	MG269085	MG267626	MG267881	MG268193	MG268492
S332	MG268739	None	MG267589	MG267864	MG268244	None
S333	MG268686	MG269097	MG267627	MG267888	MG268197	None
S334	MG268636	MG269098	MG267628	MG267883	MG268198	None
S335	MG268891	MG269057	MG267634	MG267882	MG268212	MG268493
S336	MG268767	MG269105	MG267632	MG267884	MG268199	MG268494
S337	None	None	MG267588	None	MG268189	MG268487
S338	MG268764	MG269106	MG267633	MG267885	MG268200	MG268495
S339	MG268952	MG269072	MG267629	MG267886	MG268208	MG268505
S340	MG268953	None	MG267630	None	MG268209	MG268506
S341	MG268954	MG269071	MG267631	MG267887	MG268210	None
S342	MG268785	MG269040	MG267635	MG267889	MG268202	MG268496
S343	MG268755	MG269045	MG267638	MG267890	MG268203	MG268497
S344	MG268784	MG269039	MG267566	MG267865	MG268190	MG268488
S345	MG268786	MG269042	MG267636	MG267892	MG268204	MG268499
S349	MG268752	MG269140	MG267572	MG267867	MG268248	MG268491
S350	MG268643	MG269008	MG267640	MG267894	MG268207	MG268500
S351	MG268640	MG268995	MG267641	MG267895	MG268196	None
S352	MG268744	MG269025	MG267567	MG267866	MG268191	MG268489
S353	MG268687	MG269014	MG267642	MG267897	MG268211	None
S354	MG268782	MG269050	MG267639	MG267893	MG268205	MG268501
S355	MG268630	MG269146	MG267573	MG267868	MG268249	MG268535
S356	MG268625	MG268989	MG267644	MG267898	MG268195	MG268502
S357	MG268621	MG269186	None	MG267896	MG268201	MG268504
S358	MG268618	MG268978	MG267610	MG267917	MG268229	MG268508
S359	MG268818	None	MG267574	MG267869	None	MG268543
S360	MG268878	None	MG267575	MG267870	None	MG268544
S361	MG268909	None	MG267576	MG267872	MG268251	MG268536
S362	MG268958	None	MG267577	MG267871	MG268253	MG268490
S363	MG268638	None	MG267611	MG267899	MG268235	None
S364	MG268860	MG269178	MG267578	MG267876	MG268257	None
S365	MG268838	None	MG267579	MG267874	None	None
S366	MG268959	MG269139	MG267581	MG267873	MG268252	MG268537
S367	MG268949	None	MG267582	None	None	None
S368	MG268711	MG269135	MG267583	MG267877	MG268258	MG268538
S369	MG268828	None	MG267584	MG267878	MG268250	None
S370	MG268844	MG269160	MG267625	MG267916	MG268242	MG268521
S371	MG268837	None	MG267580	MG267875	None	None
S372	MG268702	MG269020	None	MG267904	MG268227	MG268509
S373	MG268704	MG269021	MG267590	MG267905	MG268226	MG268510
S374	MG268699	MG269018	None	MG267906	MG268228	None
S375	MG268743	MG269185	MG267624	MG267903	MG268241	MG268519

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
S376	MG268746	MG269056	MG267614	MG267902	MG268236	MG268507
S378	MG268961	None	MG267585	None	MG268254	MG268539
S379	MG268957	None	MG267586	None	MG268255	MG268540
S380	MG268960	None	MG267587	None	MG268256	MG268541
S381	MG268613	MG269053	MG267568	None	MG268192	None
S383	MG268951	MG269092	MG267613	MG267900	MG268237	None
S384	MG268720	MG269190	MG267591	MG267909	None	None
S385	MG268718	MG269188	MG267616	MG267910	None	None
S386	MG268765	MG269107	MG267612	MG267901	MG268240	MG268520
S387	MG268639	MG269001	MG267617	MG267911	MG268230	MG268511
S388	None	MG269002	MG267618	MG267912	MG268231	MG268512
S389	MG268674	MG269003	MG267621	MG267913	MG268232	MG268513
S390	MG268684	MG269005	MG267619	MG267914	MG268233	MG268514
S391	MG268683	MG269006	MG267620	MG267915	MG268234	MG268515
S392	MG268778	MG269027	MG267622	MG267907	MG268239	MG268516
S393	MG268777	MG269028	MG267623	MG267908	MG268238	MG268517
S394	MG268768	None	MG267592	MG267918	MG268213	MG268522
S395	MG268681	MG269011	MG267593	MG267919	None	MG268523
S396	MG268685	MG269004	MG267594	MG267922	MG268214	MG268524
S397	MG268637	MG269000	MG267595	MG267920	None	None
S398	MG268715	MG268981	MG267596	MG267924	MG268215	None
S399	MG268816	MG269113	MG267597	MG267927	MG268216	MG268531
S400	MG268826	None	MG267598	MG267926	MG268217	MG268532
S401	MG268883	None	MG267599	MG267925	MG268218	MG268526
S402	None	MG269181	MG267600	MG267928	MG268219	MG268527
S403	MG268655	MG269182	MG267601	MG267929	None	None
S404	MG268666	MG269069	MG267602	MG267930	MG268220	None
S405	MG268669	MG269070	MG267603	MG267931	None	None
S406	None	MG269060	MG267604	MG267932	MG268221	MG268528
S407	MG268871	MG269062	MG267605	MG267933	MG268222	MG268530
S408	MG268872	MG269063	MG267606	MG267934	MG268223	MG268529
S409	MG268719	MG269189	MG267607	None	MG268224	None
S410	MG268682	MG269009	MG267608	MG267921	MG268225	MG268525
S411	MG268721	MG269191	MG267609	MG267923	None	None
S413	MG268677	None	MG267645	MG267936	None	MG268545
S414	MG268805	MG268984	MG267646	MG268004	MG268261	MG268546
S416	None	None	MG267648	MG267937	MG268262	MG268547
S417	MG268804	MG269093	MG267649	None	MG268263	MG268548
S418	MG268923	None	MG267650	MG267938	MG268264	None
S419	MG268849	None	MG267651	MG267939	None	None
S420	MG268924	None	MG267652	MG267940	MG268265	None
S421	MG268854	None	MG267653	MG267941	MG268266	MG268549
S422	MG268853	None	MG267654	MG267942	MG268267	MG268550
S425	MG268769	MG269049	MG267656	MG267944	MG268269	MG268552
S426	MG268780	MG269023	MG267657	MG267945	None	MG268589
S427	MG268781	MG269024	MG267658	MG267946	MG268270	MG268553
S428	MG268770	MG269051	MG267659	MG267947	MG268333	MG268554
S429	MG268779	MG269026	MG267660	MG267948	MG268271	MG268555
S431	MG268646	MG268991	None	MG267949	MG268273	None
S433	MG268641	MG268999	MG267663	None	MG268275	None
S436	MG268813	None	MG267666	MG267952	MG268278	None
S437	MG268814	None	MG267667	MG267953	MG268279	None
S438	MG268815	None	MG267668	MG267954	MG268280	None
S439	MG268807	None	None	MG267955	MG268281	None
S440	MG268812	None	MG267669	MG267956	MG268282	None
S441	MG268809	MG269183	MG267670	MG267957	MG268283	None
S442	MG268810	None	MG267671	MG267958	MG268284	MG268558
S444	MG268697	MG268996	MG267673	MG267960	None	None
S445	MG268698	MG268997	MG267674	MG267961	MG268286	MG268560
S446	MG268692	MG268976	MG267675	MG267962	MG268287	MG268561

Continued on next page

Appendix B Supplementary information for Chapter 4

Code	CO1	H3	12S	16S	18S	28S
S447	MG268705	MG268990	None	MG267963	MG268288	MG268562
S448	MG268675	MG269015	MG267676	MG267964	MG268289	MG268597
S449	MG268688	MG269017	None	MG267965	MG268290	MG268563
S450	MG268650	None	MG267677	None	MG268291	MG268564
S451	MG268648	None	MG267678	MG267966	MG268292	MG268565
S452	MG268673	MG269012	MG267679	MG267967	MG268293	MG268591
S453	MG268649	None	None	MG267968	MG268294	MG268592
S454	MG268647	None	MG267680	MG267969	MG268295	MG268593
S455	MG268656	MG269029	None	None	MG268296	MG268594
S456	MG268771	MG269052	MG267681	MG267970	MG268297	None
S457	MG268884	None	MG267682	MG267971	MG268298	None
S458	MG268762	MG269128	None	MG267972	MG268299	MG268566
S459	MG268875	MG268983	None	MG267973	MG268300	MG268595
S460	MG268817	None	None	MG267974	MG268301	MG268567
S461	MG268893	MG269130	MG267683	MG267975	MG268302	MG268568
S462	None	None	MG267684	MG267976	MG268303	MG268569
S463	MG268824	MG268988	MG267685	MG267977	MG268304	MG268570
S465	MG268631	MG268992	MG267686	MG267979	MG268306	MG268572
S466	MG268908	None	MG267687	MG267980	MG268307	MG268573
S467	MG268940	MG269091	MG267688	MG267981	MG268308	MG268596
S468	MG268732	MG269065	MG267689	MG267982	MG268309	MG268574
S470	MG268855	None	MG267691	MG267983	MG268310	MG268575
S471	MG268827	MG269161	MG267692	MG267984	MG268311	None
S472	MG268856	None	MG267693	MG267985	None	MG268576
S473	MG268864	MG269166	MG267694	MG267986	None	None
S474	MG268722	None	MG267695	None	MG268312	None
S476	MG268867	None	MG267697	MG267988	MG268314	None
S477	None	None	None	None	MG268315	MG268578
S478	MG268628	None	MG267715	MG267989	MG268316	None
S479	MG268882	MG269121	MG267698	MG267990	MG268317	None
S480	MG268670	None	MG267699	MG268003	MG268318	None
S482	MG268950	None	MG267701	None	MG268320	MG268580
S483	MG268955	MG269073	MG267702	None	MG268321	MG268581
S484	MG268956	MG269074	MG267703	None	MG268322	MG268582
S485	MG268851	None	MG267704	MG267991	MG268323	MG268583
S486	MG268713	MG269123	MG267705	MG267992	MG268324	None
S487	MG268696	MG269010	MG267706	MG267993	MG268325	MG268584
S488	MG268873	MG268980	MG267707	MG267994	MG268326	MG268585
S489	MG268726	MG268982	MG267708	MG267995	MG268327	MG268586
S490	MG268892	MG269129	MG267709	MG267996	MG268328	MG268587
S491	None	MG269120	MG267710	MG267997	None	None
S492	MG268659	MG269068	MG267711	MG267998	None	None
S493	MG268667	None	MG267716	MG267999	MG268329	None
S494	MG268759	None	MG267712	MG268000	MG268330	None
S495	MG268751	None	MG267713	MG268001	MG268331	None
S496	MG268740	None	MG267714	MG268002	MG268332	None
S497	MG268735	None	MG267717	MG268005	MG268334	MG268598
S498	MG268676	None	None	MG268006	MG268335	None
S499	MG268821	MG269116	MG267718	MG268017	MG268336	MG268599
S500	MG268820	MG269115	MG267719	MG268007	MG268337	MG268600
S502	MG268888	None	MG267721	MG268009	S502	None
S503	MG268887	None	MG267727	MG268010	None	None
S504	MG268845	MG269168	MG267722	MG268011	MG268339	MG268602
S505	MG268829	MG269167	MG267723	MG268012	MG268340	None
S506	MG268846	MG269169	MG267724	MG268013	MG268341	MG268603
S507	MG268944	MG269089	MG267725	MG268014	MG268342	MG268604
S508	MG268943	MG269090	None	MG268015	MG268343	MG268605
SB052	MG268946	None	MG267731	None	MG268348	None
SB070	None	MG269142	MG267729	MG268019	MG268346	None
PUB11	EU215704.1	None	None	EU215654.1	None	None

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
PUB13	EU215694.1	None	None	EU215655.1	None	None
PUB156	MN434080.1	None	None	None	None	None
PUB166	ON380801.1	ON375147.1	None	None	None	None
PUB169	ON380807.1	ON375153.1	None	None	None	None
PUB170	ON380838.1	ON375184.1	None	None	None	None
PUB175	ON380837.1	ON375183.1	None	None	None	None
PUB18	EU215679.1	None	None	EU215642.1	None	None
PUB181	ON380804.1	ON375150.1	None	None	None	None
PUB183	ON380770.1	ON375116.1	None	None	None	None
PUB185	ON380808.1	ON375154.1	None	None	None	None
PUB188	ON380847.1	ON375193.1	None	None	None	None
PUB191	ON380829.1	ON375175.1	None	None	None	None
PUB195	ON380821.1	ON375167.1	None	None	None	None
PUB198	ON380824.1	ON375170.1	None	None	None	None
PUB20	EU215705.1	None	None	EU215630.1	None	None
PUB202	ON380819.1	ON375165.1	None	None	None	None
PUB204	ON380833.1	ON375179.1	None	None	None	None
PUB207	ON380815.1	ON375161.1	None	None	None	None
PUB209	ON380811.1	ON375157.1	None	None	None	None
PUB214	ON380846.1	ON375192.1	None	None	None	None
PUB216	ON380773.1	ON375119.1	None	None	None	None
PUB219	ON380776.1	ON375122.1	None	None	None	None
PUB222	ON380794.1	ON375140.1	None	None	None	None
PUB225	ON380797.1	ON375143.1	None	None	None	None
PUB229	ON380786.1	ON375132.1	None	None	None	None
PUB232	ON380780.1	ON375126.1	None	None	None	None
PUB234	ON380788.1	ON375134.1	None	None	None	None
PUB236	ON380790.1	ON375136.1	None	None	None	None
PUB240	ON380842.1	ON375188.1	None	None	None	None
PUB243	ON380782.1	ON375128.1	None	None	None	None
PUB245	ON380850.1	ON375196.1	None	None	None	None
PUB248	ON380853.1	ON375199.1	None	None	None	None
PUB254	OQ706185.1	OQ719659.1	None	None	None	None
PUB259	OQ706195.1	OQ719669.1	None	None	None	None
PUB264	OQ706172.1	OQ719646.1	None	None	None	None
PUB270	OQ706178.1	OQ719652.1	None	None	None	None
PUB271	OQ706162.1	OQ719636.1	None	None	None	None
PUB275	OQ706166.1	OQ719640.1	None	None	None	None
PUB28	EU215688.1	None	None	EU215617.1	None	None
PUB280	OQ706181.1	OQ719655.1	None	None	None	None
PUB283	OQ706189.1	OQ719663.1	None	None	None	None
PUB288	OQ706159.1	OQ719633.1	None	None	None	None
PUB291	MW721825.1	ON375203.1	None	None	None	None
PUB292	MW721826.1	ON375204.1	None	None	None	None
PUB30	EU215696.1	None	None	EU215659.1	None	None
PUB31	EU215703.1	None	None	EU215640.1	None	None
PUB32	EU215718.1	None	None	EU215627.1	None	None
PUB33	EU215692.1	None	None	EU215660.1	None	None
PUB355	KX980744.1	None	None	None	None	KX980960.1
PUB36	EU215676.1	None	None	EU215621.1	None	None
PUB360	KX980599.1	None	None	None	None	KX980815.1
PUB394	KX980588.1	None	None	None	None	KX980804.1
PUB424	KX980718.1	None	None	None	None	KX980934.1
PUB43	None	None	None	EU215631.1	None	None
PUB432	KX980648.1	None	None	None	None	KX980864.1
PUB44	EU215726.1	None	None	EU215663.1	None	None
PUB441	KX980749.1	None	None	None	None	KX980965.1
PUB443	KX980731.1	None	None	None	None	KX980947.1
PUB47	EU215731.1	None	None	EU215667.1	None	None
PUB471	KX980664.1	None	None	None	None	KX980880.1

Continued on next page

Appendix B Supplementary information for Chapter 4

Code	CO1	H3	12S	16S	18S	28S
PUB500	KX980735.1	None	None	None	None	KX980951.1
PUB503	KX980726.1	None	None	None	None	KX980942.1
PUB510	MW721782.1	None	None	None	None	None
PUB513	MW721819.1	None	None	None	None	None
PUB515	MW721802.1	None	None	None	None	None
PUB517	MW721806.1	None	None	None	None	None
PUB518	MW721784.1	None	None	None	None	None
PUB521	MW721823.1	None	None	None	None	None
PUB525	MW721794.1	None	None	None	None	None
PUB526	MW721820.1	None	None	None	None	None
PUB528	MW721812.1	None	None	None	None	None
PUB531	MW721796.1	None	None	None	None	None
PUB533	MW721816.1	None	None	None	None	None
PUB535	MW721786.1	None	None	None	None	None
PUB538	MW721804.1	None	None	None	None	None
PUB539	MW721792.1	None	None	None	None	None
PUB541	MW721789.1	None	None	None	None	None
PUB544	MW721790.1	None	None	None	None	None
PUB547	MW721808.1	None	None	None	None	None
PUB549	MW721800.1	None	None	None	None	None
PUB555	OP293160.1	None	None	None	None	OP295412.1
PUB558	OP293163.1	None	None	None	None	OP295414.1
PUB560	OP293165.1	None	None	None	None	OP295416.1
PUB572	OP293175.1	None	None	None	None	OP295423.1
PUB579	OP293180.1	None	None	None	None	OP295427.1
PUB582	OP293183.1	None	None	None	None	OP295430.1
PUB586	OP293187.1	None	None	None	None	OP295434.1
PUB590	OP293191.1	None	None	None	None	OP295438.1
PUB593	OP293194.1	None	None	None	None	OP295441.1
PUB6	EU215669.1	None	None	None	None	None
PUB600	OP293201.1	None	None	None	None	None
PUB605	OP293206.1	None	None	None	None	OP295449.1
PUB608	OP293209.1	None	None	None	None	None
PUB610	OP293211.1	None	None	None	None	None
PUB611	OP293212.1	None	None	None	None	OP295453.1
PUB615	OP293216.1	None	None	None	None	OP295457.1
PUB620	OP293221.1	None	None	None	None	OP295461.1
PUB625	OP293226.1	None	None	None	None	OP295463.1
PUB65	EU215711.1	None	None	EU215625.1	None	None
PUB67	EU215728.1	None	None	EU215664.1	None	None
PUB68	EU215700.1	None	None	EU215624.1	None	None
PUB72	EU215713.1	None	None	EU215635.1	None	None
PUB8	EU215717.1	None	None	EU215615.1	None	None
PUB80	KF150084.1	None	None	KF178390.1	None	None
PUB81	KF150127.1	None	None	KF178433.1	None	None
PUB82	KF150111.1	None	None	KF178417.1	None	None
PUB83	KF150101.1	None	None	KF178407.1	None	None
PUB84	KF150098.1	None	None	KF178404.1	None	None
PUB85	KF150108.1	None	None	KF178414.1	None	None
PUB86	KF150090.1	None	None	KF178396.1	None	None
PUB87	KF150125.1	None	None	KF178431.1	None	None
PUB88	KF150105.1	None	None	KF178411.1	None	None
PUB9	EU215715.1	None	None	EU215647.1	None	None
PUB90	KF150129.1	None	None	KF178435.1	None	None
PUB92	KF150123.1	None	None	KF178429.1	None	None
PUB94	KF150079.1	None	None	KF178385.1	None	None
PUB95	KF150131.1	None	None	KF178437.1	None	None
PUB96	KF150130.1	None	None	KF178436.1	None	None
PUB97	KF150119.1	None	None	KF178425.1	None	None
PUB98	KF150097.1	None	None	KF178403.1	None	None

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
PUB629	MW721824.1	None	None	None	None	None
CH01	DQ667855	None	AY560723	DQ667749	None	DQ667839
CH02	JX023562	None	JX023800	JX023886	JX023987	JX024094
CH03	DQ667860	None	AY560694	DQ667754	None	DQ667844
CH04	AY560786	None	AY560706	DQ667800	None	AY560749
CH05	DQ667933	MG268972	None	MG267841	None	None
CH06	DQ667939	None	AY560722	None	None	AY560761
CH07	MG268691	None	None	MG268026	None	MG268427
CH08	MG268700	MG269019	None	MG268021	None	None
CH09	JX023593	JX023670	AY560701	AY560672	JX024007	JX024113
CH10	MG268794	MG269087	MG267464	MG267735	MG268048	MG268375
CH11	MG268801	None	None	None	MG268098	MG268406
CH12	MG268757	MG269047	None	MG267753	MG268056	MG268356
CH13	MG268947	None	MG267440	None	None	None
CH14	MG268748	None	MG267456	MG267780	MG268043	None
CH15	MG268793	MG269082	None	MG267736	MG268086	MG268377
CH16	MG268796	MG269083	None	MG267737	MG268051	MG268381
CH17	MG268921	MG269162	MG267502	None	None	None
CH18	MG268776	MG269035	MG267494	MG267747	MG268087	MG268354
CH19	MG268773	MG269031	None	MG267751	MG268063	None
CH20	MG268729	JX023691	None	None	JX024041	JX024135
CH21	MG268886	JX023626	JX023779	None	JX023967	None
CH22	MG268905	KF715567	None	KF715615	KF715588	KF715608
CH23	MG268847	KF715572	None	None	KF715589	None
CH24	MG268929	None	MG267504	None	None	MG268410
CH25	MG268932	None	MG267506	None	None	MG268412
CH26	MG268933	None	MG267505	None	None	MG268411
CH27	MG268848	JX023640	JX023791	JX023881	None	None
CH28	MG268866	None	MG267474	MG267799	None	None
CH29	MG268819	JX023701	JX023840	JX023937	None	None
CH30	MG268857	None	MG267443	MG267797	MG268028	MG268414
CH31	MG268910	MG269170	MG267539	MG267815	None	MG268457
CH32	MG268831	MG269154	MG267540	MG267816	MG268143	None
CH33	MG268914	None	MG267534	None	MG268140	MG268454
CH34	MG268745	None	AY560719	None	None	AY560757
CH35	MG268634	None	None	MG267772	MG268081	None
CH36	MG268668	JX023635	JX023786	JX023876	None	None
CH37	MG268653	None	MG267461	MG267778	MG268076	None
CH38	MG268747	MG269055	MG267457	None	MG268042	None
CH39	MG268678	None	JX023842	JX023939	JX024049	None
CH40	MG268877	JX023694	JX023835	JX023930	JX024043	None
CH41	MG268679	JX023703	JX023844	None	JX024051	JX024140
CH42	MG268792	JX023642	JX023795	None	JX023980	None
CH43	MG268876	None	MG267647	None	MG268121	MG268588
CH44	MG268928	MG269175	MG267690	None	None	None
CH45	DQ667941	MG269187	MG267700	DQ667834	MG268319	MG268579
CH46	MG268791	MG269078	MG267450	None	MG268046	MG268373
CH48	MG268897	MG269137	MG267571	None	MG268246	MG268413
CH49	AY560790	None	AY560715	MG267771	MG268040	AY560754
CH50	JX023583	JX023658	None	JX023899	JX023999	JX024106
CH51	DQ667920	None	AY560708	None	None	None
CH52	JX023553	JX023621	JX023771	JX023864	JX023961	None
CH101	E050	None	AY560705	E050	E050	AY560744
CH102	S051	MG268968	MG267720	MG268008	E049	MG268601
CH103	None	E014	None	E014	E014	MG268365
CH106	EB137	MG268994	None	MG267853	MG268176	None
CH108	E022	MG268975	E022	E022	MG268285	MG268559
CH111	DQ667903	E062	None	None	E062	DQ667847
CH113	E030	MG268998	MG267570	MG267879	E030	MG268534
CH115	E026	E026	MG267558	MG267847	E026	E026

Continued on next page

Appendix B Supplementary information for Chapter 4

Code	CO1	H3	12S	16S	18S	28S
CH116	E086	MG269114	MG267561	E086	MG268173	None
CH60	MG268614	None	None	MG267861	B8	B8
CH61	E081	JX023686	E081	E081	E081	E081
CH62	E031	MG269054	MG267569	E031	MG268247	None
CH63	E056	E056	MG267463	E056	MG268067	None
CH65	E021	MG268986	E021	None	E021	None
CH66	E017	E017	E017	E017	MG268243	MG268518
CH67	E018	MG269037	MG267637	E018	MG268206	MG268498
CH68	None	E025	JX023824	JX023921	JX024027	JX024126
CH69	MG268889	MG269184	E094	E094	MG268148	MG268462
CH70	E059	None	MG267726	MG268016	E059	MG268606
CH71	E087	E087	E087	E087	E087	MG268483
CH72	E024	JX023684	JX023827	None	JX024029	JX024128
CH73	JX023580	JX023654	E029	E029	E029	E029
CH74	MG268922	E016	E016	None	E016	E016
CH75	OQ453158	E011	None	None	E011	None
CH76	OQ453157	E009	None	None	E009	None
CH77	OQ453156	E007	None	None	E007	None
CH80	ON970480	E060	E060	None	E060	E060
CH84	ON970482	E080	None	None	None	None
CH86	ON970485	E045	None	None	None	None
CH92	ON063462	E006	E006	E006	E006	None
CH93	ON063453	E003	E003	E003	E003	None
CH94	ON067457	E002	E002	E002	E002	None
CH99	ON504283	E090	None	None	E090	E090
CH120	E085	E085	None	None	E085	BH139
CH121	E063	E063	None	E063	None	BH03
CH123	BH155	None	MG267696	MG267987	MG268313	BH155
CH124	BH161	None	MG267435	None	MG268073	BH161
CH125	BH160	None	MG267436	None	MG268074	MG268363
CH126	OR830621	MG268985	MG267438	None	MG268066	MG268384
CH127	OR726373	None	None	E066	None	None
CH129	N074	E048	None	None	E048	None
CH130	N082	E054	None	E054	None	None
CH132	ON380855.1	ON375201.1	JX023831	JX023925	JX024034	None
CH133	KX980761.1	JX023712	JX023847	JX023949	JX024061	KX980977.1
CHBH144	OL870346	None	None	None	None	OL872166
CH100	E091	E091	None	E091	E091	E091
CH107	E047	E047	E047	None	E047	None
CH112	E032	None	E032	E032	E032	E032
CH114	None	JX023617	JX023765	JX023858	None	None
CH122	E068	E068	None	None	None	None
CH128	E071	E071	None	E071	E071	None
CH131	E095	E095	None	None	None	E095
CH64	S464	MG269110	S464	MG267978	MG268305	MG268571
CH78	E034	E034	E034	E034	None	None
CH79	E044	E044	E044	E044	None	None
CH81	E040	E040	E040	E040	None	E040
CH82	E043	E043	None	None	None	None
CH83	E036	E036	None	None	None	None
CH85	E037	None	None	None	E037	E037
CH87	E004	None	None	E004	E004	None
CH88	E013	E013	None	E013	E013	E013
CH89	E008	E008	None	E008	E008	E008
CH90	E010	E010	E010	E010	None	None
CH91	E012	E012	E012	None	E012	None
CH95	E067	None	E067	None	E067	E067
CH96	E070	None	E070	None	E070	None
CH97	E069	None	E069	E069	E069	None
CH98	E038	E038	E038	E038	E038	E038

Continued on next page

Code	CO1	H3	12S	16S	18S	28S
SRR10291841	None	SRR10291841	None	None	None	None
SRR10291854	None	SRR10291854	None	None	SRR10291854	None
SRR10291860	None	None	None	None	None	SRR10291860
OG1	FJ607564	FJ607602	None	FJ607454	FJ607490	FJ607529
OG2	FJ607579	FJ607617	None	FJ607466	FJ607505	FJ607540
OG3	EU139636	EU139754	None	EU139684	None	EU139822
OG4	FJ607557	FJ607596	None	None	FJ607483	FJ607522
OG6	FJ607555	FJ607594	None	FJ607447	FJ607481	FJ607520
OG7	FJ607553	FJ607592	None	FJ607445	FJ607479	FJ607518
OG8	EU003302	EU003333	None	EU003276	EU003378	None
OG9	FJ525318	FJ525337	None	FJ525351	FJ525388	FJ525370
OG10	AF244233	EU139752	None	AF244148	None	EU139820
OG11	FJ607568	FJ607606	None	FJ607457	FJ607494	FJ607533
OG12	EU003310	EU003340	EU003247	None	EU003366	EU003438
OG13	FJ949021	FJ949058	None	None	FJ948939	FJ948980
OG15	FJ607585	FJ607623	None	FJ607472	FJ607511	FJ607546
OG17	GQ279223	None	None	EU817794	None	GQ279182
OG18	EU003288	EU003320	EU003236	EU003260	EU003386	EU003433
OG19	AF244236	EU139755	None	AF244151	None	None
OG20	GQ285630	GQ285625	None	GQ285606	None	GQ285615
OG21	EU817704	None	None	EU817810	None	EU817740
OG22	FJ607563	None	None	None	FJ607489	FJ607628
OG24	AY560796	None	AY560726	AY560684	None	AY560764
OG25	AY560802	None	AY560731	None	None	AY560765
OG26	FJ607552	FJ607591	None	FJ607444	FJ607478	FJ607517
OG27	FJ607566	FJ607604	None	FJ607455	FJ607492	FJ607531
OG29	None	None	None	JN010167	JN010185	JN010190
OG30	EU003292	AY078702	EU003239	AY078664	None	EU003410
OG31	EU003308	None	EU003246	None	None	EU003429
OG32	FJ607584	FJ607622	None	FJ607471	FJ607510	FJ607545
OG33	EU003311	EU003341	EU003248	EU003251	EU003367	EU003431
OG34	FJ607554	FJ607593	None	FJ607446	FJ607480	FJ607519
OG35	AY560795	None	AY560724	AY560681	None	AY560766
OG36	AF244235	EU139753	None	None	None	EU139821
OG37	FJ607559	FJ607598	None	FJ607450	FJ607485	FJ607524

Table B.2. Pairwise comparisons of rates across clades using Kruskal-Wallis H-test. Hedges' g provides a standardized measure of the difference between two group means. A larger value indicates a more significant difference. A positive value means the indicator value of group A is higher than that of group B, while a negative value indicates the opposite. A g value of 0.2 is generally considered a small effect. A g value around 0.5 indicates a moderate effect. A g value of 0.8 or above is considered a large effect.

Rate type	Group A	Group B	Paired	Parametric	U.val	alternative	p.unc	hedges'g
Speciation Rate	Pholcinae	Smeringopinae	FALSE	FALSE	250,064,564	two-sided	0	2.8455
	Pholcinae	Modisiminae	FALSE	FALSE	169,075,595	two-sided	0	0.6461
	Pholcinae	Arteminae	FALSE	FALSE	255,984,225	two-sided	0	5.8439
	Pholcinae	Ninetinae	FALSE	FALSE	255,994,024	two-sided	0	5.8988
	Pholcinae	Priscula	FALSE	FALSE	254,592,823	two-sided	0	4.9333
	Pholcinae	Artema	FALSE	FALSE	254,648,359	two-sided	0	4.5892
	Smeringopinae	Modisiminae	FALSE	FALSE	8,046,675	two-sided	0	-2.7773
	Smeringopinae	Arteminae	FALSE	FALSE	229,294,675	two-sided	0	1.6874
	Smeringopinae	Ninetinae	FALSE	FALSE	228,973,402	two-sided	0	1.6875
	Smeringopinae	Priscula	FALSE	FALSE	222,547,687	two-sided	0	1.3932
	Smeringopinae	Artema	FALSE	FALSE	234,251,000	two-sided	0	1.4319
	Modisiminae	Arteminae	FALSE	FALSE	255,981,034	two-sided	0	7.5494
	Modisiminae	Ninetinae	FALSE	FALSE	255,986,161	two-sided	0	7.7282
	Modisiminae	Priscula	FALSE	FALSE	254,151,260	two-sided	0	5.6045
	Modisiminae	Artema	FALSE	FALSE	254,564,362	two-sided	0	4.9262
	Arteminae	Ninetinae	FALSE	FALSE	124,570,867	two-sided	3.06E-05	-0.0302
	Arteminae	Priscula	FALSE	FALSE	124,456,664	two-sided	1.65E-05	-0.1642
	Arteminae	Artema	FALSE	FALSE	147,589,963	two-sided	4.80E-124	0.0858
	Ninetinae	Priscula	FALSE	FALSE	127,790,899	two-sided	0.785305099	-0.1453
	Ninetinae	Artema	FALSE	FALSE	151,228,162	two-sided	1.27E-173	0.1047
	Priscula	Artema	FALSE	FALSE	150,434,323	two-sided	4.35E-162	0.1931
Extinction Rate	Pholcinae	Smeringopinae	FALSE	FALSE	199,543,947	two-sided	0	0.8940
	Pholcinae	Modisiminae	FALSE	FALSE	236,553,824	two-sided	0	1.7477
	Pholcinae	Arteminae	FALSE	FALSE	251,041,424	two-sided	0	2.1750
	Pholcinae	Ninetinae	FALSE	FALSE	252,822,715	two-sided	0	2.2472
	Pholcinae	Priscula	FALSE	FALSE	246,917,652	two-sided	0	1.8864
	Pholcinae	Artema	FALSE	FALSE	252,092,063	two-sided	0	1.6232
	Smeringopinae	Modisiminae	FALSE	FALSE	166,489,587	two-sided	0	0.6332
	Smeringopinae	Arteminae	FALSE	FALSE	210,176,179	two-sided	0	1.0341
	Smeringopinae	Ninetinae	FALSE	FALSE	214,560,044	two-sided	0	1.0931
	Smeringopinae	Priscula	FALSE	FALSE	208,995,395	two-sided	0	0.8522
	Smeringopinae	Artema	FALSE	FALSE	217,346,397	two-sided	0	0.7484
	Modisiminae	Arteminae	FALSE	FALSE	193,012,245	two-sided	0	0.7500
	Modisiminae	Ninetinae	FALSE	FALSE	199,373,696	two-sided	0	0.8935
	Modisiminae	Priscula	FALSE	FALSE	194,087,461	two-sided	0	0.4150
	Modisiminae	Artema	FALSE	FALSE	204,689,951	two-sided	0	0.3393
	Arteminae	Ninetinae	FALSE	FALSE	133,330,149	two-sided	1.27E-10	0.1197
	Arteminae	Priscula	FALSE	FALSE	131,553,801	two-sided	1.86E-05	-0.1140
	Arteminae	Artema	FALSE	FALSE	140,706,859	two-sided	3.12E-53	-0.0173
	Ninetinae	Priscula	FALSE	FALSE	126,577,274	two-sided	0.081664274	-0.1851
	Ninetinae	Artema	FALSE	FALSE	135,714,342	two-sided	1.20E-20	-0.0598
	Priscula	Artema	FALSE	FALSE	136,752,100	two-sided	4.01E-26	0.0504
Diversification Rate	Pholcinae	Smeringopinae	FALSE	FALSE	254,964,796	two-sided	0	3.7679
	Pholcinae	Modisiminae	FALSE	FALSE	18,806,036	two-sided	0	-2.0403
	Pholcinae	Arteminae	FALSE	FALSE	256,028,855	two-sided	0	6.1882
	Pholcinae	Ninetinae	FALSE	FALSE	256,029,629	two-sided	0	6.1569
	Pholcinae	Priscula	FALSE	FALSE	255,641,317	two-sided	0	5.6745
	Pholcinae	Artema	FALSE	FALSE	255,631,893	two-sided	0	5.6093
	Smeringopinae	Modisiminae	FALSE	FALSE	24,490	two-sided	0	-5.3342
	Smeringopinae	Arteminae	FALSE	FALSE	207,210,256	two-sided	0	1.2477
	Smeringopinae	Ninetinae	FALSE	FALSE	203,245,882	two-sided	0	1.1725

Continued on next page

Rate type	Group A	Group B	Paired	Parametric	U.val	alternative	p.unc	hedges'g
	Smeringopinae	Priscula	FALSE	FALSE	202,080,087	two-sided	0	1.1029
	Smeringopinae	Artema	FALSE	FALSE	214,584,166	two-sided	0	1.2795
	Modisiminae	Arteminae	FALSE	FALSE	256,031,991	two-sided	0	7.9424
	Modisiminae	Ninetinae	FALSE	FALSE	256,032,001	two-sided	0	7.9282
	Modisiminae	Priscula	FALSE	FALSE	255,995,493	two-sided	0	7.3623
	Modisiminae	Artema	FALSE	FALSE	255,745,525	two-sided	0	7.2222
	Arteminae	Ninetinae	FALSE	FALSE	120,586,414	two-sided	2.45E-19	-0.1056
	Arteminae	Priscula	FALSE	FALSE	122,463,940	two-sided	1.83E-11	-0.1074
	Arteminae	Artema	FALSE	FALSE	141,421,143	two-sided	3.49E-59	0.1621
	Ninetinae	Priscula	FALSE	FALSE	129,751,005	two-sided	0.03575896	-0.0101
	Ninetinae	Artema	FALSE	FALSE	148,985,061	two-sided	4.62E-142	0.2562
	Priscula	Artema	FALSE	FALSE	146,786,121	two-sided	3.18E-114	0.2494

Table B.3. Akaike weights Table for model fitting of traits. The ClipKIT smart-gap tree was used.

Traits	Models	Rank	AIC	diff	wi	AICw	Parameters
$\log_2(\text{body size})$	rate_trend	1	310	0	1.00E+00	1.00E+00	slope = -0.008984 sigsq = 0.030652
	EB	2	340	30	3.02E-07	3.02E-07	a = -0.041438 sigsq = 0.174066
	delta	3	375	64.6	0.00E+00	0.00E+00	delta = 0.091005 sigsq = 0.029054
	BM	4	395	84.8	0.00E+00	0.00E+00	sigsq = 0.003216
	OU	5	397	86.8	0.00E+00	0.00E+00	alpha = 0.000000 sigsq = 0.003216
	lambda	6	397	86.8	0.00E+00	0.00E+00	lambda = 0.962475 sigsq = 0.003342
	white	7	1646	1335.6	0.00E+00	0.00E+00	sigsq = 0.432384
$\log_2(\text{leglength})$	rate_trend	1	1063	0	1.00E+00	1.00E+00	slope = -0.008718 sigsq = 0.089557
	EB	2	1088	24.9	3.97E-06	3.97E-06	a = -0.044851 sigsq = 0.824346
	delta	3	1136	72.8	0.00E+00	0.00E+00	delta = 0.066476 sigsq = 0.127963
	BM	4	1166	102.8	0.00E+00	0.00E+00	sigsq = 0.010355
	OU	5	1168	104.8	0.00E+00	0.00E+00	alpha = 0.000000 sigsq = 0.010355
	lambda	6	1168	104.8	0.00E+00	0.00E+00	lambda = 0.999887 sigsq = 0.010356
	white	7	2519	1456.6	0.00E+00	0.00E+00	sigsq = 1.433046
Relative leg length	rate_trend	1	4437	0	1	0.5581	slope = -0.007881 sigsq = 2.351035

Continued on next page

Traits	Models	Rank	AIC	diff	wi	AICw	Parameters
	EB	2	4441	4.29	0.1171	0.0654	a = -0.010168 sigsq = 1.372987
	BM	3	4442	4.71	0.0951	0.0531	sigsq = 0.506748
	delta	4	4443	6.3	0.0429	0.024	delta = 0.747545 sigsq = 0.645813
	OU	5	4444	6.73	0.0346	0.0193	alpha = 0.000000 sigsq = 0.506748
	lambda	6	4444	6.73	0.0346	0.0193	lambda = 0.945396 sigsq = 0.536017
	white	7	5325	887.96	0	0	sigsq = 50.869049
Leg proportion	rate_trend	1	-960	0	1.00E+00	1.00E+00	slope = -0.008936 sigsq = 0.005808
	EB	2	-930	30.4	2.48E-07	2.48E-07	a = -0.038774 sigsq = 0.026541
	delta	3	-904	56.3	0.00E+00	0.00E+00	delta = 0.135155 sigsq = 0.003895
	BM	4	-892	68	0.00E+00	0.00E+00	sigsq = 0.000631
	OU	5	-890	70	0.00E+00	0.00E+00	alpha = 0.000000 sigsq = 0.000631
	lambda	6	-890	70	0.00E+00	0.00E+00	lambda = 0.997171 sigsq = 0.000632
	white	7	284	1244.2	0.00E+00	0.00E+00	sigsq = 0.083375

Table B.4. Result for the ES-sim analysis

Traits	Trees	Pearson's correlation coefficient $\rho = 0.34$	p-value
Traits	Trees	Pearson's correlation coefficient	p-value
$\log_2(\text{bodysize})$	smart-gap	0.3388434	0.0149
	gappy	0.3360932	0.0139
	kplic-smart-gap	0.3397531	0.0142
$\log_2(\text{leglength})$	smart-gap	0.3694154	0.0064
	gappy	0.3573017	0.006
	kplic-smart-gap	0.3611333	0.0053
Relative leg length	smart-gap	0.2564518	0.1044
	gappy	0.2376823	0.1084
	kplic-smart-gap	0.2389544	0.1166
Leg proportion	smart-gap	0.24451	0.1244
	gappy	0.230834	0.1251
	kplic-smart-gap	0.226326	0.1386

Table B.5. Model fitting for the QuaSSE analysis. In the analysis, speciation rate (λ) and extinction rate (μ) are functions (constant/linear/sgmoidal/hump) of trait values. (1) “minimal” means both λ and μ are constant. (2) “constant_extinction.linear” means constant μ but λ is a linear function of trait. (3) “variable_speciation_extinction.linear” means both μ but λ are linear functions of trait. (4) “drift” means the drift parameter in make.quasse() function and it allows the mean of our trait evolution process to change through time. The most fitted model (with the smallest AIC values) are in bold.

Traits	Trees	Pearson’s correlation coefficient	p-value
Traits	Trees	Pearson’s correlation coefficient	p-value
log2(body size)	smart-gap	0.3388434	0.0149
	gappy	0.3360932	0.0139
	kplic-smart-gap	0.3397531	0.0142
log2(leg length)	smart-gap	0.3694154	0.0064
	gappy	0.3573017	0.006
	kplic-smart-gap	0.3611333	0.0053
Relative leg length	smart-gap	0.2564518	0.1044
	gappy	0.2376823	0.1084
	kplic-smart-gap	0.2389544	0.1166
Leg proportion	smart-gap	0.24451	0.1244
	gappy	0.230834	0.1251
	kplic-smart-gap	0.226326	0.1386

References

- Blomberg, S. P., T. J. Garland and A. R. Ives (2003),
Testing for phylogenetic signal in comparative data: behavioral traits are more labile,
Evolution **57** 717.
- Eberle, J., D. Dimitrov, A. Valdez-Mondragon and B. A. Huber (2018),
Microhabitat change drives diversification in pholcid spiders, *BMC Evol Biol* **18** 141.
- Meng, G., L. Podsiadlowski, D. Dimitrov and B. A. Huber (2025), *The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders*, submitted.
- Minh, B. Q., H. A. Schmidt, O. Chernomor et al. (2020),
IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era,
Mol Biol Evol **37** 1530.
- Pagel, M. (1999), *Inferring the historical patterns of biological evolution*, *Nature* **401** 877.
- Pennell, M. W., J. M. Eastman, G. J. Slater et al. (2014), *geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees*, *Bioinformatics* **30** 2216.
- Rabosky, D. L. (2014), *Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees*, *PloS One* **9** e89543.
- Revell, L. J. (2012), *phytools: an R package for phylogenetic comparative biology (and other things)*,
Methods Ecol Evol **3** 217.
- Smith, S. A. and B. C. O’Meara (2012),
treePL: divergence time estimation using penalized likelihood for large phylogenies,
Bioinformatics **28** 2689.
- Yu, G., D. K. Smith, H. Zhu, Y. Guan and T. T.-Y. Lam (2017), *ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data*,
Methods Ecol Evol **8** 28.

Supplementary information for Chapter 5

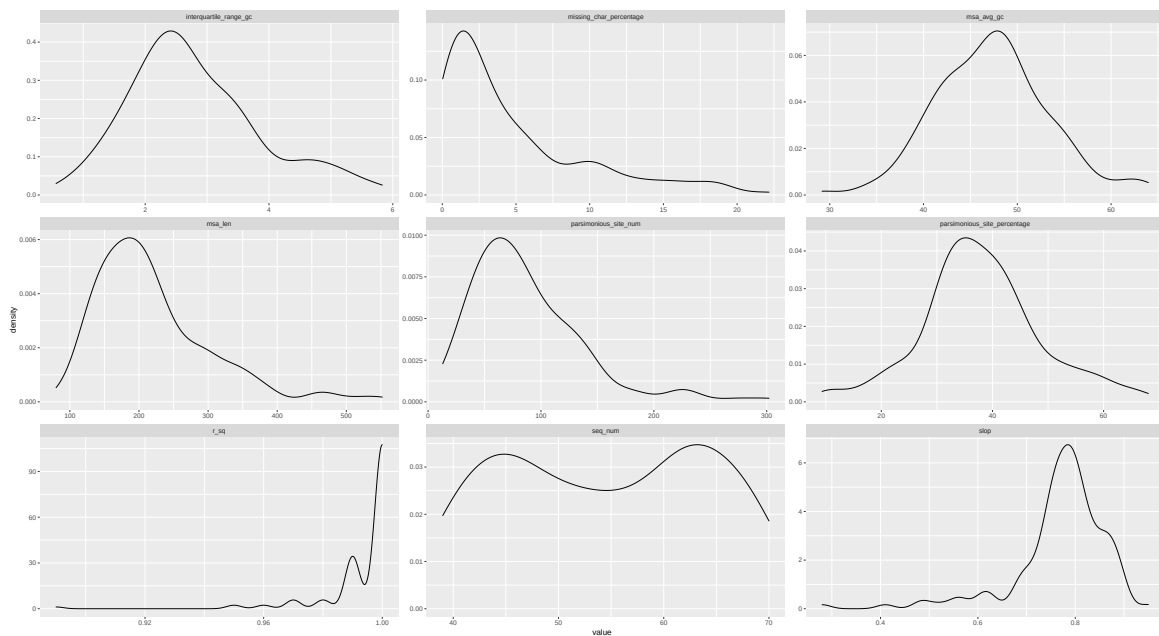


Figure C.1. Statistics of data matrix 1. Density distributions of characteristics including average GC content, GC content interquartile range, number and proportion of missing data (gaps and Ns), number and proportion of parsimony-informative sites, length of the multiple sequence alignment (MSA), and number of taxa in the MSA. Visualizations were performed with ggplot2.

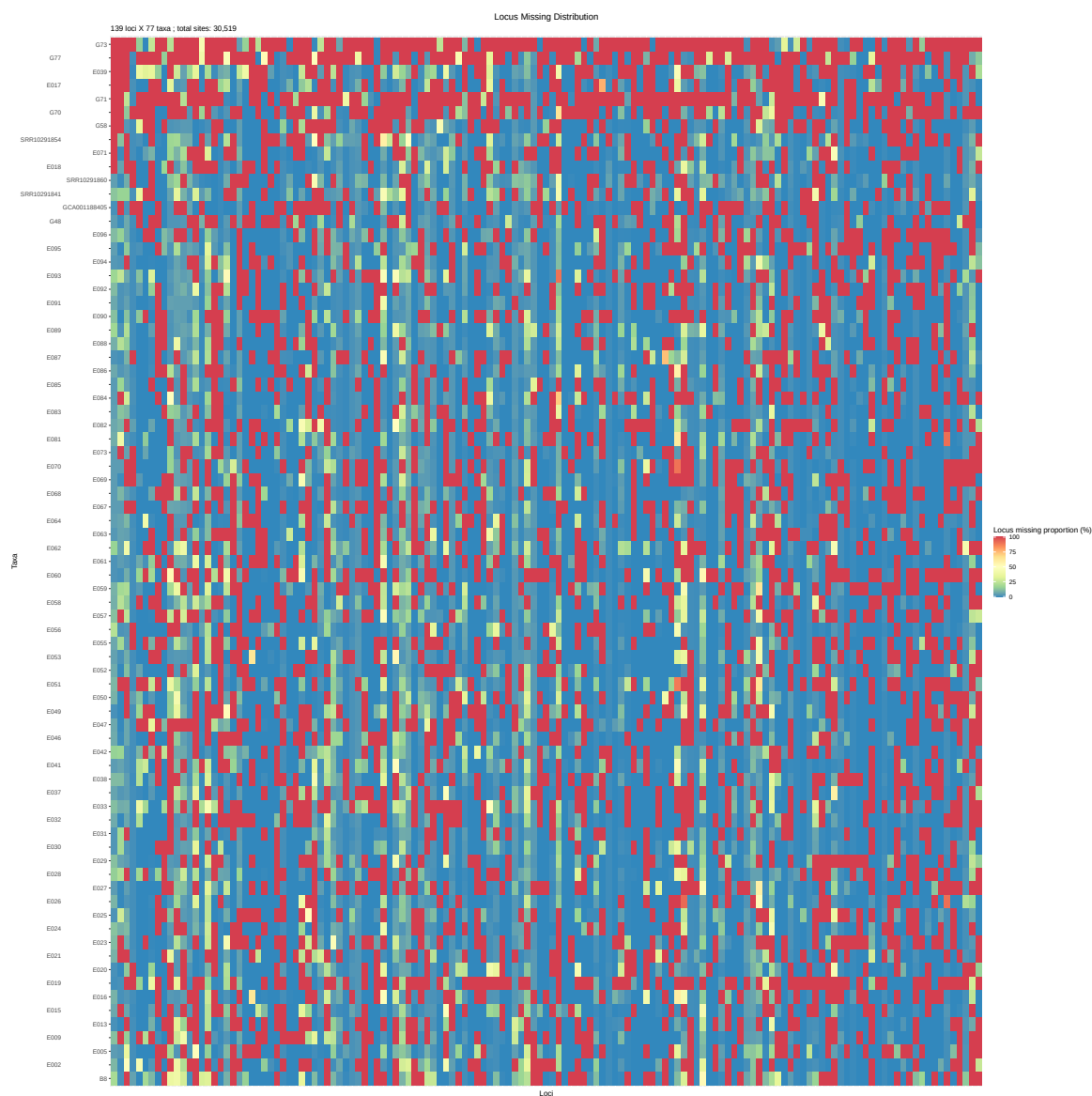


Figure C.2. Missing data distribution of matrix 1. The missing proportion of a contig of a locus was obtained by dividing the number of gaps (dashes and Ns) of the contig by the locus length. Visualization was performed with ggplot2.

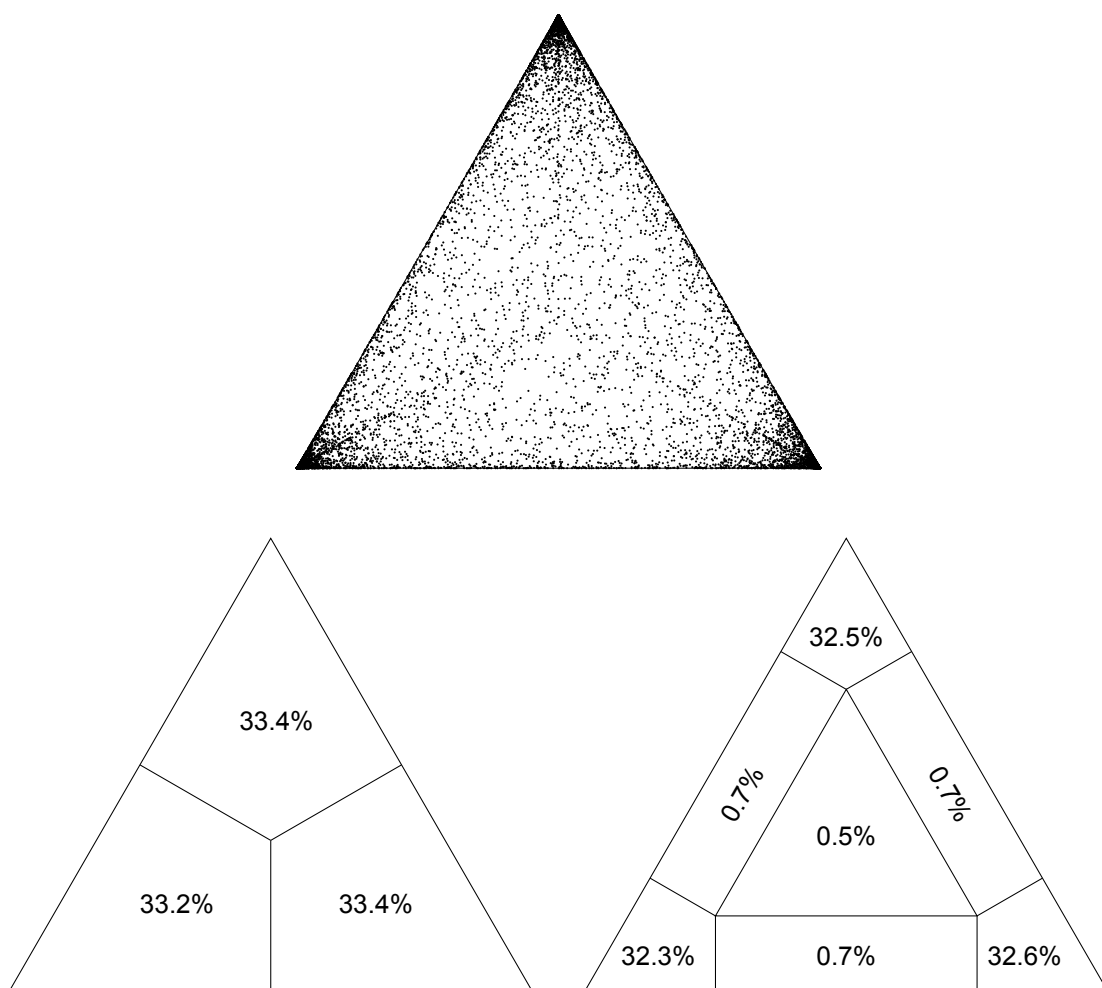


Figure C.3. Likelihood mapping analysis of matrix 1. The IQ-TREE was used for the analysis.

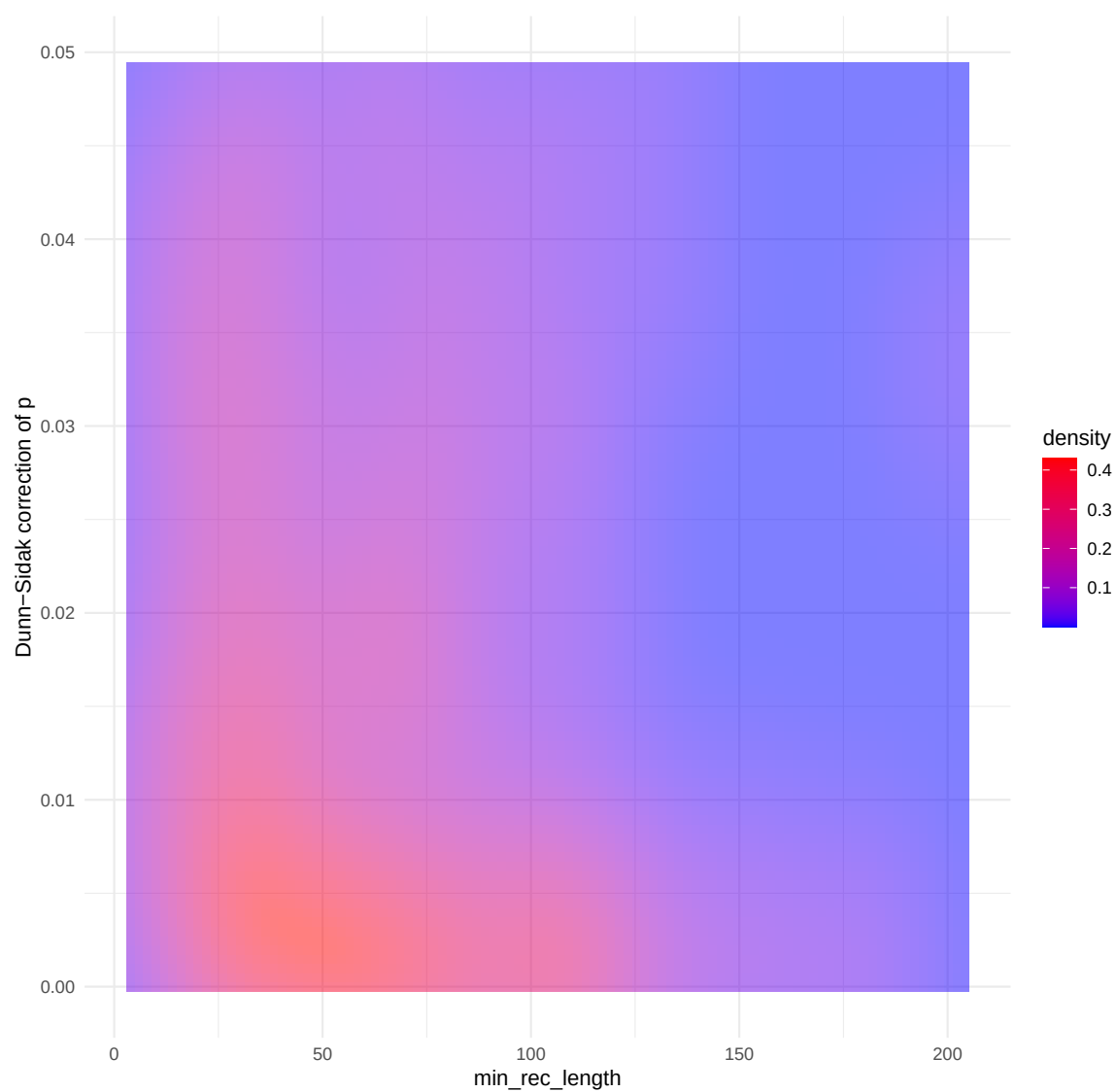


Figure C.4. Intralocus recombination analysis with 3SEQ based on matrix 1. **min_rec_length:** the minimum length of the recombinant segments; i.e. the minimum length of the shorter of the two recombinant segments;

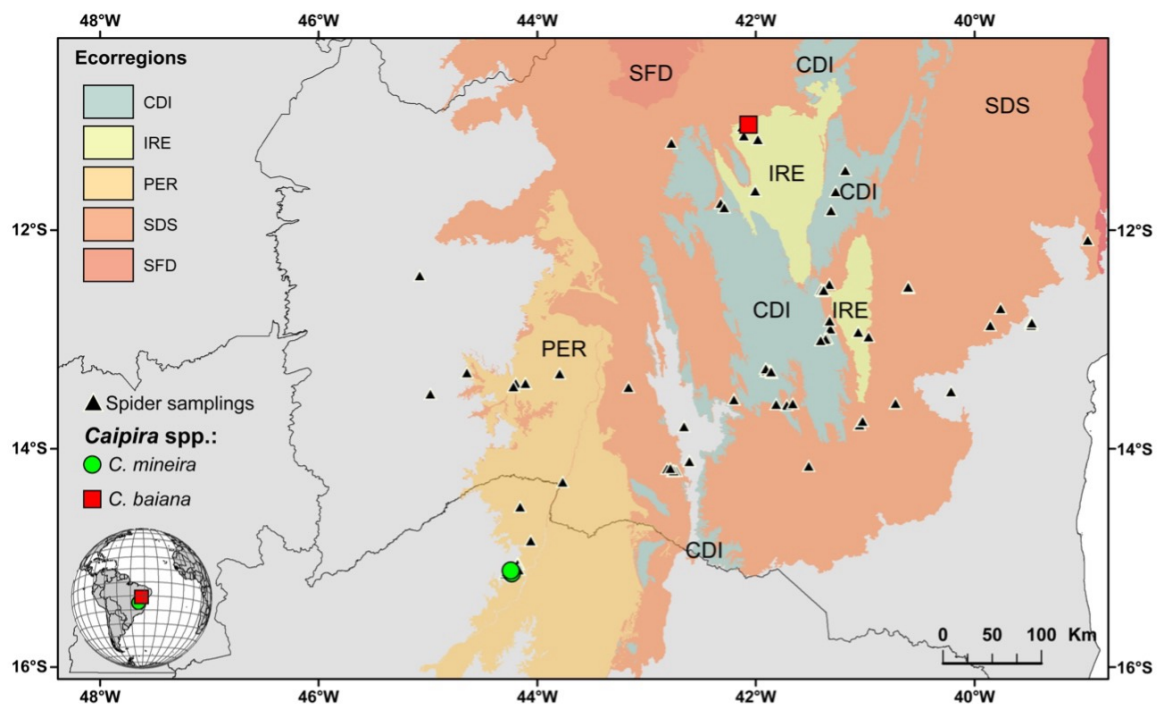


Figure C.5. Known geographic distribution of *GENUS-1* gen. nov., including *GENUS-1 sp2* sp. nov. (red squares) and *GENUS-1 sp1* sp. nov. (green circles). Colored areas represent the main districts of the Caatinga domain: CDI –Chapada Diamantina; IRE –Irecê; PER –Peruaçu; SDS –Southern Depressão Sertaneja; and SFD –São Francisco Dunes. Other biomes are not shown. Black triangles are sampled localities by BAH and LSC, in the states of Bahia and Minas Gerais.

Table C.1. Sample information for UCE-based phylogenetic analysis

Sample ID	Family	Subfamily	Source	Accession	Total bases (Gbp)
NCBI_Hypochilidae_Hypochilus_gertschi	Hypochilidae	-	Kulkarni et al. 2020	SRR10291860	-
NCBI_Sicariidae_Loxosceles_reclusa	Sicariidae	-	i5k Consortium 2013	GCA_001188405.1	-
NCBI_Orsolobidae_Maoriata_sp.	Orsolobidae	-	Kulkarni et al. 2020	SRR10291854	-
NCBI_Scytodidae_Scytodes_sp.	Scytodidae	-	Kulkarni et al. 2020	SRR10291841	-
E032_Artema_doriae_Isr40	Pholcidae	Arteminae	Meng et al. 2025	SRR19792634	2.51
E033_Artema_nephilit_Isr46	Pholcidae	Arteminae	Meng et al. 2025	SRR19792633	2.87
E046_Artema_martensi_Mor82	Pholcidae	Arteminae	Meng et al. 2025	SRR19792619	3.02
E052_Artema_dhofar_Om124	Pholcidae	Arteminae	Meng et al. 2025	SRR19792612	2.69
E005_Priscula_binghamae_Arg157	Pholcidae	Arteminae	Meng et al. 2025	SRR19792625	1.48
E067_Priscula_acarite_Ven18-196	Pholcidae	Arteminae	Meng et al. 2025	SRR19792596	1.39
E069_Priscula_andinensis_Ven18-220	Pholcidae	Arteminae	Meng et al. 2025	SRR19792594	1.57
E070_Priscula_salmeronica_Ven18-239	Pholcidae	Arteminae	Meng et al. 2025	SRR19792593	1.66
E073_GENUS-1_sp1_Carv62	Pholcidae	SUBFAM-1	this study	SRR19792589	1.39
G48_GENUS-1_sp1_Br22-166	Pholcidae	SUBFAM-1	this study		11.48
G58_GENUS-1_sp2_Br22-229	Pholcidae	SUBFAM-1	this study		11.43
E020_Arenita_fazendinha_Br16-252	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792648	1.39
E063_Blancoa_piacoa_Ven18-167	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792600	1.27
E085_Canaima_arima_Tri18-18	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792576	1.72
E018_Carapoia_crasto_Br15-188	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792650	1.72
E021_Chibchea_amapa_Br16-253	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792646	1.63
E084_Ciboneya_antraia_Cu12-178	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792577	1.51
E015_Gen.n.B_Arg118_Arg224	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792653	1.48
E082_Gen.n.C_Arg80_Arg199	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792579	1.45
E083_Gen.n.D_Arg85_Arg204	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792578	1.39
E039_Modisimus_texanus_Mex291	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792627	1.48
E068_Pisaboa_retracta_Ven18-202	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792595	1.42
E017_Tupigea_penedo_Br07/100-14	Pholcidae	Modisiminae	Meng et al. 2025	SRR19792651	1.42
E071_Galapa_bella_Gal55	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792591	1.45
E002_Gertschiola_macrostyla_Arg128	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792658	1.30
E009_Guaranita_auadae_Arg179	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792581	1.60
E027_Ibotyporanga_naideae_G117	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792640	1.06
E019_Kambiwa_neotropica_Br15-267	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792649	1.36
E049_Magana_velox_Om118	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792616	1.21
E013_Nerudia_centaura_Arg216	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792655	1.57
E050_Ninetis_subtilissima_Om120	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792615	1.21
E047_Papiamenta_levii_MRAC640	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792618	1.51
E062_Pemona_sapo_Ven01	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792601	1.12
E060_Pholcophora_americana_USA16	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792604	1.45
B8_Pinoquio_baraua_ZFMK Ar 20622	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792563	0.88

Continued on next page

Sample ID	Family	Subfamily	Source	Accession	Total bases (Gbp)
E037_Tolteca_hesperia_Mex253	Pholcidae	Ninetinae	Meng et al. 2025	SRR19792629	0.97
E057_Aetana_kiukoki_Phi251	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792607	1.63
E093_Anansus_ewe_MRAC649	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792567	1.42
E038_Metagonia_placida_Mex287	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792628	1.36
E053_Micropholcus_Om74_Om133	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792611	1.81
E016_Nipisa_phyllicola_Bor194	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792652	1.90
E023_Nyikoa_limbe_Cam131	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792644	1.09
E029_Pholcus_kindia_Gui96	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792638	2.30
E024_Quamtana_nylsvley?_CAS10	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792643	1.33
E087_Spermophora_senoculata_Cre22	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792574	1.48
E094_Spermophora_Mal2_Mal213	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792566	1.87
E095_Spermophora_Om87_Om141	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792565	1.48
E096_Spermophora_minotaura_Ken124	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792564	1.42
E059_Wanniyala_ohiya_SL146	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792605	1.57
E025_Zatavua_analalava?_CAS6	Pholcidae	Pholcinae	Meng et al. 2025	SRR19792642	0.97
E031_Arnapa_tinoor_Ind184	Pholcidae	SUBFAM-2	Meng et al. 2025	SRR19792635	1.30
G70_Aucana_platnicki_Chi64	Pholcidae	SUBFAM-2	this study		13.48
G71_Aucana_ramirezi_Chi71	Pholcidae	SUBFAM-2	this study		10.26
G77_Aucana_petorca_Chi113	Pholcidae	SUBFAM-2	this study		7.80
G73_Aucana_Chi80_Chi80	Pholcidae	SUBFAM-2	this study		11.83
E030_Changminia_huangdi_Ind141	Pholcidae	SUBFAM-2	Meng et al. 2025	SRR19792637	2.45
E042_Chisosa_calapa_Mex351	Pholcidae	SUBFAM-2	Meng et al. 2025	SRR19792623	1.66
E061_Chisosa_diluta_USA26	Pholcidae	SUBFAM-2	Meng et al. 2025	SRR19792602	1.15
E064_Chisosa_caquetio_Ven18-179	Pholcidae	SUBFAM-2	Meng et al. 2025	SRR19792599	1.90
E056_Holocneminus_Phi105_Phi245	Pholcidae	SUBFAM-2	Meng et al. 2025	SRR19792608	1.66
E041_Physocyclus_modestus_Mex349	Pholcidae	SUBFAM-2	Meng et al. 2025	SRR19792624	1.90
E081_Wugigarra_undanbi_QMB4	Pholcidae	SUBFAM-2	Meng et al. 2025	SRR19792580	1.06
E058_Cenemus_culiculus_Sey22	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792606	1.72
E055_Crossopriza_sahtan_Om32	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792609	1.54
E092_Crossopriza_lyoni_Gui90	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792568	1.27
E088_Holocnemus_pluchei_Isr53	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792573	1.15
E089_Holocnemus_reini_Mor72	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792572	1.54
E091_Holocnemus_hispanicus_Ribera34	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792569	1.24
E026_Hoplopholcus_minous_Cre8	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792641	3.74
E090_Maghreba_stifadma_Mor71	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792571	1.24
E028_Smeringopina_bineti_Gui101	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792639	1.24
E051_Smeringopus_lineiventris_Om122	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792613	1.54
E086_Stygopholcus_photophilus_Cre19	Pholcidae	Smeringopinae	Meng et al. 2025	SRR19792575	1.66

Table C.2. Data matrix (1–5) statistics

Matrix	Minimum contig length (bp)	Taxon coverage (%)	Total sites (bp)	Total loci	Number of contigs
1	50	50	30,519	139	7,573
2	150	50	27,156	113	6,221
3	150	60	20,806	85	5,038
4	150	70	14,983	62	3,909
5	150	80	9,517	40	2,632

Table C.3. Tests of symmetry of matrix 1. IQ-TREE (v2.2.0) (Minh et al. 2020) was used to test the assumptions of stationarity (constant nucleotide/amino-acid frequencies over time) and homogeneity (constant substitution rates over time). The two assumptions are rejected when p-values (SymPval) < 0.01 are found. **Locus:** Partition name; **SymSig:** Number of significant sequence pairs by test of symmetry; **SymNon:** Number of non-significant sequence pairs by test of symmetry; **SymPval:** P-value for maximum test of symmetry; **MarSig:** Number of significant sequence pairs by test of marginal symmetry; **MarNon:** Number of non-significant sequence pairs by test of marginal symmetry; **MarPval:** P-value for maximum test of marginal symmetry; **IntSig:** Number of significant sequence pairs by test of internal symmetry; **IntNon:** Number of non-significant sequence pairs by test of internal symmetry; **IntPval:** P-value for maximum test of internal symmetry

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-10145	42	2,169	0.2710	143	2,058	0.1543	35	2,160	0.5080
uce-1017	112	1,718	0.2644	176	1,653	0.5215	45	1,776	0.1446
uce-10291	26	1,513	0.5988	34	1,499	0.5412	35	1,496	0.4888
uce-10293	35	1,918	0.5212	68	1,882	0.8990	44	1,892	0.2045
uce-10306	192	2,086	0.7376	275	2,001	0.7167	94	2,154	0.5329
uce-10308	179	1,651	0.9001	288	1,541	0.8539	59	1,751	0.6610
uce-10324	115	1,538	0.6549	174	1,479	0.3018	34	1,617	0.9164
uce-10337	13	802	0.9371	22	746	0.9864	9	627	0.6463
uce-105	33	1,343	0.5911	90	1,279	0.2580	13	1,334	0.8953
uce-106	33	957	0.1612	58	929	0.3718	44	931	0.1070
uce-110	73	1,879	0.1427	199	1,748	0.6055	44	1,896	0.0514
uce-111	41	1,233	0.1140	117	1,153	0.4087	25	1,215	0.0610
uce-1114	65	1,765	0.5713	156	1,654	0.4162	50	1,730	0.5843
uce-113	23	718	0.1832	32	695	0.2108	12	686	0.2294
uce-115	70	2,344	0.4683	185	2,191	0.2970	64	2,221	0.5892
uce-1245	65	1,210	0.0733	114	1,161	0.0911	51	1,224	0.1670
uce-1246	55	725	0.0103	92	687	0.0068	13	763	0.2059
uce-129	15	755	0.0020	24	485	0.1701	17	326	0.0012
uce-1373	743	1,672	0.4136	897	1,517	0.4223	80	2,334	0.3506
uce-1457	40	863	0.3251	74	818	0.2009	13	806	0.5081
uce-149	42	1,037	0.6349	65	978	0.3928	7	948	0.7251
uce-1582	38	1,671	0.1485	80	1,463	0.1641	35	1,341	0.2243
uce-160	39	1,137	0.9796	95	1,025	0.9865	18	936	0.8007
uce-161	45	1,083	0.5576	91	1,035	0.5833	19	1,097	0.4002
uce-1647	150	1,335	0.0003	226	1,259	0.0000	38	1,442	0.8661
uce-1656	44	2,164	0.0745	90	2,089	0.1297	46	1,925	0.1123
uce-1715	79	782	0.1506	115	741	0.0552	21	801	0.6061
uce-1730	1	1,112	0.2480	1	849	0.3160	1	509	0.2280
uce-1736	390	1,888	0.0081	514	1,763	0.0011	97	2,180	0.7588
uce-1956	32	1,507	0.3030	81	1,450	0.3739	54	1,407	0.2327
uce-1959	290	838	0.0969	363	765	0.0635	33	1,095	0.3263

Continued on next page

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-2148	53	893	0.1218	101	843	0.1480	50	882	0.1935
uce-2170	188	988	0.0439	254	922	0.0083	52	1,122	0.7544
uce-2181	167	2,110	0.8058	280	1,996	0.9860	59	2,214	0.4105
uce-2199	57	888	0.1261	105	840	0.1321	21	922	0.2259
uce-2226	166	2,180	0.3789	218	2,122	0.6935	87	2,226	0.1749
uce-2287	69	1,944	0.1167	183	1,792	0.9555	30	1,864	0.0197
uce-2296	21	1,464	0.6525	54	1,361	0.5595	14	1,191	0.5304
uce-231	4	2,139	0.4147	17	2,084	0.8466	19	2,000	0.1535
uce-258	4	1,947	0.6784	24	1,828	0.5440	34	1,674	0.6046
uce-2588	19	969	0.3774	73	900	0.2666	13	903	0.4805
uce-2645	7	1,643	0.9339	21	1,566	0.9705	54	1,456	0.5864
uce-2701	63	678	0.3817	82	657	0.2555	16	712	0.5074
uce-2764	50	1,224	0.0755	84	1,189	0.1238	29	1,238	0.1279
uce-2948	482	1,864	0.7359	663	1,683	0.4954	47	2,298	0.7601
uce-2980	33	1,398	0.8579	73	1,339	0.5865	52	1,336	0.8825
uce-2998	58	1,958	0.5461	96	1,920	0.4571	65	1,947	0.4973
uce-3001	58	1,772	0.9462	101	1,719	0.7614	54	1,760	0.9143
uce-3082	19	842	0.1885	55	804	0.0996	16	824	0.4785
uce-3086	6	1,474	0.7182	22	1,424	0.3307	25	1,347	0.9659
uce-3162	64	1,017	0.8631	147	926	0.7267	31	1,023	0.7440
uce-3313	82	1,627	0.9337	86	1,470	0.8348	45	1,252	0.8061
uce-3342	88	1,741	0.3566	123	1,704	0.1130	77	1,745	0.8836
uce-3423	72	963	0.4730	113	922	0.1410	51	982	0.9907
uce-3800	12	1,146	0.1759	0	1,019	0.1044	93	877	0.4226
uce-3802	260	1,693	0.4459	379	1,573	0.3060	45	1,900	0.5645
uce-387	134	1,695	0.9469	195	1,630	0.7877	58	1,754	0.8916
uce-4286	28	1,915	0.4882	55	1,745	0.4176	55	1,446	0.4560
uce-4295	26	715	0.2147	45	689	0.1522	24	692	0.3840
uce-4318	43	859	0.1544	85	816	0.1576	21	872	0.2449
uce-4543	140	2,138	0.0607	192	2,084	0.0207	55	2,209	0.5130
uce-4577	12	2,003	0.2853	25	1,978	0.4486	58	1,896	0.1910
uce-4628	102	1,173	0.0063	169	1,105	0.0269	41	1,231	0.0322
uce-4753	417	1,294	0.0081	581	1,130	0.0014	44	1,667	0.6204
uce-4863	27	2,080	0.0437	16	1,794	0.1987	51	1,349	0.0402
uce-4864	7	2,006	0.5475	22	1,829	0.3329	10	1,584	0.6676
uce-4877	37	1,978	0.2658	121	1,853	0.0575	33	1,883	0.9874
uce-4882	9	1,617	0.2548	1	516	0.5153	0	184	0.0809
uce-5026	24	1,407	0.6318	37	1,372	0.9572	34	1,256	0.2091
uce-508	46	1,439	0.7144	140	1,333	0.7913	35	1,407	0.4436
uce-5126	0	1,321	0.7204	2	1,254	0.4935	22	1,130	0.7350
uce-5132	126	954	0.0000	147	921	0.0000	54	993	0.8066
uce-5141	31	2,247	0.2480	67	2,186	0.0611	55	2,093	0.9189
uce-5379	58	1,895	0.0356	112	1,834	0.3430	29	1,900	0.0135
uce-538	29	1,006	0.0770	73	962	0.3513	22	1,012	0.0436
uce-5380	82	1,192	0.1850	103	1,169	0.2598	50	1,217	0.1881
uce-5394	191	1,348	0.7980	272	1,266	0.5578	69	1,467	0.7977
uce-5617	44	1,906	0.4934	81	1,843	0.5845	34	1,777	0.2926
uce-5646	157	833	0.0768	197	792	0.0229	63	923	0.6030
uce-5664	0	2,084	0.5212	0	1,041	0.9411	3	485	0.1884
uce-5731	92	2,186	0.2946	134	2,143	0.1096	86	2,187	0.7406
uce-5793	7	2,270	0.6487	15	2,249	0.4031	34	2,132	0.8160
uce-5934	82	1,627	0.7336	161	1,546	0.3888	52	1,647	0.9056
uce-6115	18	1,017	0.5511	22	1,000	0.9361	54	932	0.2102
uce-6262	21	2,388	0.0944	97	2,301	0.0776	50	2,333	0.2635
uce-6284	300	734	0.0055	351	682	0.0098	52	974	0.0750
uce-6304	13	1,201	0.2254	41	1,121	0.0866	11	1,039	0.6597
uce-6385	115	2,030	0.0005	132	2,010	0.0008	101	2,028	0.0625
uce-6443	29	1,507	0.0680	61	1,392	0.0589	38	1,265	0.2444
uce-6448	51	939	0.2699	69	907	0.1877	28	909	0.4242
uce-6482	229	1,255	0.1999	277	1,205	0.0645	49	1,360	0.9766

Continued on next page

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-6485	54	726	0.2615	120	645	0.2615	9	733	nan
uce-6578	1	772	0.2798	4	532	0.2839	22	322	0.2998
uce-6579	2	2,402	0.0338	23	2,123	0.1139	6	1,884	0.0526
uce-6634	246	1,834	0.2429	364	1,714	0.0565	67	2,008	0.9418
uce-6777	142	2,204	0.1585	251	2,085	0.0441	76	2,238	0.7576
uce-694	39	1,236	0.2952	74	1,196	0.1957	30	1,219	0.4587
uce-7156	22	2,117	0.8404	97	1,987	0.4748	32	1,978	0.9708
uce-7157	182	598	0.2261	232	548	0.6381	30	748	0.0907
uce-7286	106	1,910	0.3090	180	1,802	0.1623	51	1,873	0.5731
uce-7352	183	763	0.0084	298	646	0.0028	20	923	0.3616
uce-7414	127	2,083	0.0801	251	1,930	0.0261	57	2,031	0.5674
uce-7458	75	1,410	0.1024	73	1,412	0.0193	78	1,401	0.8821
uce-7690	36	1,290	0.2007	13	1,312	0.1933	114	1,202	0.2810
uce-7758	114	1,597	0.0698	195	1,512	0.0353	64	1,624	0.3798
uce-7775	20	721	0.3263	38	702	0.3520	29	711	0.2991
uce-7805	94	1,559	0.4892	221	1,427	0.4594	19	1,625	0.4159
uce-7838	99	890	0.6794	173	811	0.8807	11	959	0.3460
uce-788	24	877	0.2479	13	800	0.8061	59	662	0.0755
uce-789	27	1,861	0.4304	49	1,802	0.4785	103	1,640	0.3012
uce-8071	90	1,799	0.1633	214	1,665	0.0829	22	1,819	0.4737
uce-8073	141	1,812	0.0379	206	1,740	0.0117	47	1,874	0.5059
uce-813	61	800	0.1535	99	762	0.1656	22	839	0.2317
uce-8168	0	986	0.2801	0	863	0.1559	18	589	0.5915
uce-8230	26	754	0.2401	65	714	0.4285	22	750	0.1575
uce-8323	86	903	0.9568	123	865	0.8039	41	946	0.9075
uce-834	0	1,850	0.4044	14	1,447	0.3187	52	1,205	0.4540
uce-838	49	1,031	0.6270	84	990	0.4443	32	1,027	0.6387
uce-8467	196	1,687	0.8556	290	1,545	0.6149	56	1,609	0.8463
uce-847	0	1,706	0.4074	4	919	0.9125	0	533	0.1320
uce-850	212	1,113	0.3484	356	967	0.1475	22	1,280	0.7163
uce-8640	56	763	0.0219	59	758	0.0870	50	753	0.0415
uce-8739	130	2,215	0.3488	270	2,055	0.2530	25	2,233	0.4528
uce-8854	3	1,464	0.1425	28	846	0.1009	8	552	0.3382
uce-8877	30	750	0.1700	56	721	0.4812	27	745	0.0859
uce-8902	163	578	0.1768	206	534	0.0887	24	715	0.4903
uce-8990	192	1,637	0.0166	270	1,557	0.0035	79	1,724	0.8949
uce-9018	10	934	0.0944	28	776	0.9276	10	493	0.0158
uce-9089	36	707	0.0021	40	191	0.0008	0	167	0.2831
uce-9134	260	1,630	0.0180	405	1,480	0.0020	32	1,835	0.9113
uce-9305	179	1,046	0.1502	269	951	0.2733	55	1,164	0.1357
uce-9340	99	890	0.2073	144	841	0.3362	43	922	0.1674
uce-9459	12	1,261	0.4361	26	1,239	0.1198	20	1,209	0.9972
uce-9683	9	1,016	0.5647	10	840	nan	8	675	nan
uce-9701	104	1,221	0.0451	127	1,195	0.0222	46	1,264	0.3529
uce-9760	33	912	0.8222	86	859	0.6788	21	922	0.7108
uce-9809	35	2,036	0.0961	71	1,690	0.1615	29	1,347	0.1320
uce-9831	24	796	0.8543	16	801	0.9290	35	768	0.5376
uce-9956	85	905	0.0310	130	857	0.0092	33	935	0.5008

Table C.4. PhyloMad substitutional saturation analysis based on matrix 1

Locus	N_ sites	N_ taxa	t_ Entropy	t_ pre- dicted _thre- shold _En- tropy	t_ pre- dicted _TPR_ Entropy	t_ pre- dicted _FPR_ Entropy	Risk_ Entropy	t_ Entro- pyvar	t_ pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR_ Entro- pyvar	t_ pre- dicted _FPR_ Entro- pyvar	Risk_ Entro- pyvar
uce-105	134	53	35.71	9.9	0.64	0.14	low.risk	11.1	13.55	0.92	0.16	high.risk
uce-106	177	45	60.77	12.13	0.68	0.19	low.risk	20.79	13.34	0.89	0.15	low.risk
uce-110	161	63	50.53	10.93	0.66	0.16	low.risk	25.98	15.7	0.91	0.15	low.risk
uce-111	185	51	56.87	12.29	0.68	0.19	low.risk	14.85	13.6	0.91	0.16	low.risk
uce-113	154	39	59	11.3	0.66	0.17	low.risk	19.19	11.62	0.9	0.17	low.risk
uce-115	222	70	62.52	13.31	0.69	0.2	low.risk	26.27	17.33	0.9	0.13	low.risk
uce-129	136	40	28.83	10.4	0.65	0.15	low.risk	NA	11.14	0.91	0.18	low.risk
uce-149	147	47	41.24	10.71	0.66	0.16	low.risk	12.43	12.84	0.91	0.16	high.risk
uce-160	111	49	38.46	8.8	0.63	0.12	low.risk	14.38	12.35	0.93	0.18	low.risk
uce-161	165	48	32.65	11.51	0.67	0.17	low.risk	12.51	13.05	0.91	0.16	high.risk
uce-231	204	66	66.32	12.69	0.68	0.19	low.risk	32.02	16.3	0.91	0.14	low.risk
uce-258	173	63	63.27	11.47	0.67	0.17	low.risk	22.89	15.78	0.91	0.14	low.risk
uce-387	146	61	39.1	10.28	0.65	0.15	low.risk	18.33	15.59	0.91	0.14	low.risk
uce-508	109	55	29.58	8.53	0.62	0.11	low.risk	9.66	14.42	0.91	0.15	high.risk
uce-538	181	46	40.45	12.27	0.68	0.19	low.risk	12.52	13.07	0.9	0.16	high.risk
uce-694	146	51	27.93	10.55	0.65	0.15	low.risk	9.96	13.25	0.92	0.17	high.risk
uce-788	142	43	57.07	10.59	0.65	0.16	low.risk	13.92	12.23	0.91	0.17	low.risk
uce-789	202	62	66.58	12.71	0.68	0.19	low.risk	17.11	15.82	0.9	0.14	low.risk
uce-813	170	42	41.68	11.92	0.67	0.18	low.risk	13.68	11.95	0.91	0.17	low.risk
uce-834	131	62	46.43	9.52	0.64	0.13	low.risk	20.01	15.05	0.92	0.15	low.risk
uce-838	188	47	39.92	12.54	0.68	0.19	low.risk	14.46	13.75	0.89	0.15	low.risk
uce-847	79	61	37.03	6.55	0.59	0.07	low.risk	9.33	14.53	0.93	0.16	high.risk
uce-850	140	52	35.89	10.23	0.65	0.15	low.risk	11.49	13.39	0.92	0.17	high.risk
uce-1017	247	61	54.01	14.44	0.71	0.22	low.risk	23.65	15.93	0.9	0.14	low.risk
uce-1114	147	61	41.04	10.33	0.65	0.15	low.risk	18.73	15.43	0.91	0.15	low.risk
uce-1245	155	51	42.98	10.97	0.66	0.16	low.risk	13.3	14.15	0.9	0.15	high.risk
uce-1246	250	40	54	15.14	0.72	0.24	low.risk	29.3	13.08	0.88	0.15	low.risk
uce-1373	307	70	40.94	16.26	0.73	0.25	low.risk	17.23	17.59	0.89	0.13	high.risk
uce-1457	138	43	36.82	10.4	0.65	0.15	low.risk	9.67	11.91	0.91	0.17	high.risk
uce-1582	132	59	36.04	9.64	0.64	0.13	low.risk	9.12	14.77	0.92	0.15	high.risk
uce-1647	211	55	35.58	13.24	0.69	0.2	low.risk	12.08	15.59	0.89	0.14	high.risk
uce-1656	211	67	58.63	12.95	0.69	0.2	low.risk	16.34	15.54	0.92	0.15	low.risk

Continued on next page

Locus	N_ sites	N_ taxa	t_ Entropy	t_pre- dicted _thre- shold _En- tropy	t_pre- dicted _TPR _Entropy	t_pre- dicted _FPR _Entropy	Risk_ Entropy	t_ Entro- pyvar	t_pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR _Entro- pyvar	t_ pre- dicted _FPR _Entro- pyvar	Risk_ Entro- pyvar
uce-1715	133	42	33.7	10.18	0.65	0.15	low.risk	13.12	12.11	0.9	0.17	low.risk
uce-1730	112	48	34.05	8.89	0.63	0.12	low.risk	15.75	12.26	0.93	0.18	low.risk
uce-1736	293	68	54.53	15.85	0.72	0.25	low.risk	18.53	17.87	0.89	0.12	low.risk
uce-1956	151	56	40.41	10.65	0.66	0.15	low.risk	13.11	13.73	0.93	0.17	high.risk
uce-1959	195	48	40.43	12.79	0.68	0.2	low.risk	11.47	13.21	0.91	0.16	high.risk
uce-2148	141	44	41.82	10.51	0.65	0.16	low.risk	15.6	11.85	0.92	0.18	low.risk
uce-2170	144	49	39.53	10.51	0.65	0.15	low.risk	13.49	13.31	0.91	0.16	low.risk
uce-2181	231	68	52.36	13.69	0.7	0.21	low.risk	25.37	17.19	0.9	0.13	low.risk
uce-2199	145	44	30.06	10.71	0.66	0.16	low.risk	7.77	12.26	0.91	0.17	high.risk
uce-2226	148	69	36.05	10.19	0.65	0.14	low.risk	16.65	16.2	0.92	0.15	low.risk
uce-2287	165	64	62.6	11.09	0.66	0.16	low.risk	29.14	15.86	0.91	0.14	low.risk
uce-2296	192	55	62.34	12.48	0.68	0.19	low.risk	19.2	14.25	0.91	0.16	low.risk
uce-2588	124	45	33.48	9.62	0.64	0.14	low.risk	10.88	12.08	0.92	0.17	high.risk
uce-2645	116	58	28.93	8.83	0.63	0.12	low.risk	11.38	14.02	0.93	0.17	high.risk
uce-2701	172	39	35.48	12.11	0.68	0.19	low.risk	12.53	11.69	0.9	0.17	low.risk
uce-2764	135	51	53.44	10.01	0.65	0.14	low.risk	20.13	14.38	0.9	0.15	low.risk
uce-2948	280	69	43.98	15.4	0.72	0.24	low.risk	14.87	17.78	0.89	0.12	high.risk
uce-2980	116	54	41.67	8.94	0.63	0.12	low.risk	13.86	13.43	0.93	0.17	low.risk
uce-2998	124	64	36.66	9.11	0.63	0.12	low.risk	13.41	15.9	0.91	0.14	high.risk
uce-3001	171	61	33.86	11.43	0.67	0.17	low.risk	17.56	15.45	0.91	0.15	low.risk
uce-3082	147	42	29.93	10.87	0.66	0.16	low.risk	17.27	12.29	0.9	0.16	low.risk
uce-3086	102	55	40.36	8.13	0.62	0.1	low.risk	18.65	13.71	0.92	0.17	low.risk
uce-3162	164	47	66.19	11.5	0.67	0.17	low.risk	19.47	13.24	0.9	0.16	low.risk
uce-3313	191	59	51.08	12.33	0.68	0.19	low.risk	16.97	14.72	0.92	0.16	low.risk
uce-3342	213	61	42.99	13.17	0.69	0.2	low.risk	8.35	15.61	0.91	0.14	high.risk
uce-3423	138	46	37.68	10.3	0.65	0.15	low.risk	17.02	13.25	0.9	0.16	low.risk
uce-3800	184	49	59.84	12.31	0.68	0.19	low.risk	13.79	13.69	0.9	0.16	low.risk
uce-3802	222	63	38.98	13.46	0.69	0.21	low.risk	17.43	15.82	0.91	0.14	low.risk
uce-4286	182	63	59.68	11.86	0.67	0.18	low.risk	15.95	14.8	0.93	0.16	low.risk
uce-4295	113	39	25.97	9.23	0.63	0.13	low.risk	8.89	10.65	0.92	0.19	high.risk
uce-4318	126	43	27.39	9.79	0.64	0.14	low.risk	8.47	12.07	0.91	0.17	high.risk
uce-4543	235	68	62.78	13.84	0.7	0.21	low.risk	31.18	17.69	0.89	0.12	low.risk
uce-4577	133	64	40.8	9.57	0.64	0.13	low.risk	15.3	15.18	0.92	0.16	low.risk
uce-4628	175	51	30.43	11.87	0.67	0.18	low.risk	11.8	13.73	0.91	0.16	high.risk

Continued on next page

Locus	N_ sites	N_ taxa	t_ Entropy	t_pre- dicted _thre- shold _En- tropy	t_pre- dicted _TPR_ Entropy	t_pre- dicted _FPR_ Entropy	Risk_ Entropy	t_ Entro- pyvar	t_pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR_ Entro- pyvar	t_ pre- dicted _FPR_ Entro- pyvar	Risk_ Entro- pyvar
uce-4753	315	59	44.61	16.76	0.73	0.26	low.risk	12.95	16.52	0.88	0.13	high.risk
uce-4863	106	66	41.6	8.09	0.62	0.1	low.risk	19.14	15.39	0.92	0.15	low.risk
uce-4864	104	64	36.85	8.02	0.62	0.1	low.risk	17.26	14.85	0.93	0.16	low.risk
uce-4877	121	64	42.91	8.95	0.63	0.12	low.risk	10.6	15.56	0.92	0.15	high.risk
uce-4882	117	60	38.39	8.84	0.63	0.12	low.risk	NA	13.97	0.94	0.17	low.risk
uce-5026	121	54	28.95	9.21	0.63	0.13	low.risk	13.62	13.5	0.92	0.17	low.risk
uce-5126	130	52	32.42	9.73	0.64	0.14	low.risk	26.47	13.93	0.91	0.16	low.risk
uce-5132	140	47	40.99	10.37	0.65	0.15	low.risk	7.17	12.58	0.92	0.17	high.risk
uce-5141	140	68	37.01	9.83	0.64	0.14	low.risk	11.92	15.5	0.93	0.16	high.risk
uce-5379	207	63	51.6	12.88	0.69	0.2	low.risk	13.5	15.5	0.91	0.15	high.risk
uce-5380	184	51	39.41	12.25	0.68	0.19	low.risk	13.2	13.66	0.91	0.16	high.risk
uce-5394	216	56	35.1	13.41	0.69	0.21	low.risk	16.4	14.89	0.9	0.15	low.risk
uce-5617	145	63	42.86	10.19	0.65	0.14	low.risk	16.01	15.34	0.92	0.15	low.risk
uce-5646	190	45	34.14	12.68	0.68	0.2	low.risk	17.51	12.85	0.9	0.16	low.risk
uce-5664	180	67	67.62	11.68	0.67	0.17	low.risk	NA	15.57	0.92	0.15	low.risk
uce-5731	197	68	36.25	12.37	0.68	0.18	low.risk	18.23	16.63	0.91	0.14	low.risk
uce-5793	157	68	51.51	10.64	0.66	0.15	low.risk	15.22	15.74	0.92	0.15	high.risk
uce-5934	222	59	39.58	13.56	0.69	0.21	low.risk	18.97	16.02	0.89	0.14	low.risk
uce-6115	190	46	37.44	12.65	0.68	0.2	low.risk	13.33	13.73	0.89	0.15	high.risk
uce-6262	178	70	36.35	11.53	0.67	0.17	low.risk	13.9	16.83	0.91	0.14	high.risk
uce-6284	130	46	30.58	9.9	0.64	0.14	low.risk	10.75	12.75	0.91	0.16	high.risk
uce-6304	152	50	54.91	10.86	0.66	0.16	low.risk	30.22	13.77	0.9	0.16	low.risk
uce-6385	213	66	46.82	13.05	0.69	0.2	low.risk	25.58	16.81	0.9	0.13	low.risk
uce-6443	164	56	48.61	11.24	0.66	0.17	low.risk	9.73	14.13	0.92	0.16	high.risk
uce-6448	153	45	42.18	11.05	0.66	0.17	low.risk	16.8	12.6	0.91	0.16	low.risk
uce-6482	181	55	40.25	12.01	0.67	0.18	low.risk	18.85	14.16	0.91	0.16	low.risk
uce-6485	109	40	34.95	8.97	0.63	0.13	low.risk	10.1	10.82	0.92	0.18	high.risk
uce-6578	136	40	53.02	10.4	0.65	0.15	low.risk	10.84	11.41	0.91	0.17	high.risk
uce-6579	166	70	72.3	11	0.66	0.16	low.risk	22.3	16.91	0.91	0.14	low.risk
uce-6634	172	65	48.02	11.38	0.67	0.17	low.risk	27.63	16.41	0.9	0.14	low.risk
uce-6777	234	69	40.1	13.78	0.7	0.21	low.risk	16.04	16.81	0.91	0.14	high.risk
uce-7156	139	66	50.11	9.83	0.64	0.14	low.risk	21.66	15.65	0.92	0.15	low.risk
uce-7157	245	40	34	14.97	0.71	0.24	low.risk	16.26	12.74	0.88	0.15	low.risk
uce-7286	154	64	44.06	10.59	0.65	0.15	low.risk	17.96	15.67	0.91	0.15	low.risk

Continued on next page

Locus	N_ sites	N_ taxa	t_ Entropy	t_pre- dicted _thre- shold _En- tropy	t_pre- dicted _TPR _Entropy	t_pre- dicted _FPR _Entropy	Risk_ Entropy	t_ Entro- pyvar	t_pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR _Entro- pyvar	t_ pre- dicted _FPR _Entro- pyvar	Risk_ Entro- pyvar
uce-7352	293	44	46.41	16.48	0.73	0.26	low.risk	24.04	13.74	0.88	0.15	low.risk
uce-7414	127	67	48.56	9.2	0.63	0.12	low.risk	22.8	16.16	0.91	0.14	low.risk
uce-7458	148	55	36.54	10.53	0.65	0.15	low.risk	22.46	14.29	0.91	0.16	low.risk
uce-7690	87	52	25.34	7.3	0.61	0.09	low.risk	13.68	13.11	0.92	0.17	low.risk
uce-7758	149	59	40.45	10.48	0.65	0.15	low.risk	21.25	15.8	0.9	0.14	low.risk
uce-7775	118	39	35.24	9.5	0.64	0.14	low.risk	15.32	10.99	0.91	0.18	low.risk
uce-7805	153	58	32.85	10.69	0.66	0.15	low.risk	10.56	14.47	0.92	0.16	high.risk
uce-7838	169	45	29.65	11.78	0.67	0.18	low.risk	17.35	12.75	0.9	0.16	low.risk
uce-8071	152	62	33.75	10.54	0.65	0.15	low.risk	15.76	15.16	0.92	0.15	low.risk
uce-8073	198	63	47.34	12.52	0.68	0.19	low.risk	16.33	15.72	0.91	0.14	low.risk
uce-8168	152	45	58.15	11.01	0.66	0.16	low.risk	14.42	12.21	0.92	0.17	low.risk
uce-8230	136	40	30.82	10.4	0.65	0.15	low.risk	11.92	12.05	0.9	0.16	high.risk
uce-8323	160	45	35.75	11.38	0.67	0.17	low.risk	24.93	13.54	0.89	0.15	low.risk
uce-8467	150	62	41.58	10.45	0.65	0.15	low.risk	17.5	15.73	0.91	0.14	low.risk
uce-8640	150	41	34.27	11.04	0.66	0.17	low.risk	13.63	12.12	0.9	0.16	low.risk
uce-8739	267	69	58.42	14.95	0.71	0.23	low.risk	17.65	16.63	0.91	0.14	low.risk
uce-8854	110	56	67.05	8.56	0.63	0.11	low.risk	14.91	13.81	0.92	0.17	low.risk
uce-8877	159	40	34.71	11.49	0.67	0.18	low.risk	12.57	11.52	0.91	0.17	low.risk
uce-8902	174	39	26.5	12.2	0.68	0.19	low.risk	12.75	11.56	0.9	0.17	low.risk
uce-8990	208	61	38.56	12.97	0.69	0.2	low.risk	21.2	16.3	0.89	0.13	low.risk
uce-9018	133	44	42.93	10.12	0.65	0.15	low.risk	13.9	11.75	0.92	0.18	low.risk
uce-9089	123	41	66.36	9.7	0.64	0.14	low.risk	NA	11.46	0.91	0.18	low.risk
uce-9134	165	62	28.96	11.14	0.66	0.16	low.risk	13.16	15.32	0.91	0.15	high.risk
uce-9305	177	50	32.22	11.98	0.67	0.18	low.risk	14.8	13.93	0.9	0.15	low.risk
uce-9340	166	45	34.18	11.65	0.67	0.18	low.risk	13.56	12.42	0.91	0.17	low.risk
uce-9459	152	51	36.24	10.83	0.66	0.16	low.risk	18.6	13.68	0.91	0.16	low.risk
uce-9683	111	46	46.85	8.89	0.63	0.12	low.risk	17.46	12.18	0.92	0.17	low.risk
uce-9701	188	52	39.87	12.39	0.68	0.19	low.risk	21.41	14.54	0.9	0.15	low.risk
uce-9760	170	44	28.93	11.86	0.67	0.18	low.risk	16.8	13.04	0.9	0.16	low.risk
uce-9809	140	65	66.87	9.9	0.64	0.14	low.risk	21.72	16.28	0.9	0.14	low.risk
uce-9831	117	41	37.7	9.38	0.64	0.13	low.risk	14.33	11.46	0.91	0.18	low.risk
uce-9956	140	45	41.63	10.43	0.65	0.15	low.risk	15.21	12.08	0.92	0.17	low.risk
uce-10145	129	67	39.92	9.3	0.64	0.13	low.risk	18.57	15.93	0.92	0.15	low.risk
uce-10291	106	56	28.77	8.33	0.62	0.11	low.risk	13.28	14.5	0.91	0.15	high.risk

Continued on next page

Locus	N_ sites	N_ taxa	t_ Entropy	t_pre- dicted _thre- shold _En- tropy	t_pre- dicted _TPR_ Entropy	t_pre- dicted _FPR_ Entropy	Risk_ Entropy	t_ Entro- pyvar	t_pre- dicted _threshold _Entro- pyvar	t_ pre- dicted _TPR_ Entro- pyvar	t_ pre- dicted _FPR_ Entro- pyvar	Risk_ Entro- pyvar
uce-10293	183	63	40.32	11.9	0.67	0.18	low.risk	18.35	15.76	0.91	0.14	low.risk
uce-10306	160	68	42.35	10.77	0.66	0.15	low.risk	14.81	16.01	0.92	0.15	high.risk
uce-10308	162	61	37.86	11.03	0.66	0.16	low.risk	12.42	15.43	0.91	0.15	high.risk
uce-10324	164	58	41.09	11.19	0.66	0.16	low.risk	19.77	15.47	0.9	0.14	low.risk
uce-10337	164	41	44.13	11.69	0.67	0.18	low.risk	15.46	11.36	0.92	0.18	low.risk

Table C.5. Species excluded from the comparative analysis of leg proportions. Several species were excluded from Fig. 5.20 because the published measurements are probably wrong and we have no specimens available to check them.

Species	Tibia 1	Tibia 2	Reference	Tibia 1/2	Range within genus (N species)
<i>Belisana galeiformis</i> Zhang & Peng, 2011	5.6	2.28	F. Zhang and Peng (2011)	2.46	1.24-1.93 (134)
<i>Belisana mengyang</i> Yao & Li, 2020	1.4	2.19	W. Zhu et al. (2020b)	0.64	1.24-1.93 (134)
<i>Belisana naling</i> Yao & Li, 2020	2.76	2.44	W. Zhu et al. (2020a)	1.13	1.24-1.93 (134)
<i>Belisana tongle</i> Zhang et al. 2008	8.36	3.37	Y. Q. Zhang et al. (2008)	2.48	1.24-1.93 (134)
<i>Meraha shuye</i> (Yao & Li, 2017)	7.95	4.04	Dong et al. (2017)	1.97	1.60-1.71 (6)
<i>Mesabolivar cantharus</i> Machado et al. 2007	5.2	2.5	Machado et al. (2007)	2.08	1.24-1.73 (84)
<i>Pholcus anachoreta</i> Dimitrov & Ribera 2006	12	6	Dimitrov et al. (2006)	2	1.22-1.80 (313)
<i>Pholcus harveyi</i> Zhang & Zhu, 2009	9	8.2	F. Zhang and M.-S. Zhu (2009)	1.1	1.22-1.80 (313)
<i>Pholcus parayichengicus</i> Zhang & Zhu, 2009	10.4	4.6	F. Zhang and M.-S. Zhu (2009)	2.26	1.22-1.80 (313)
<i>Pholcus ping</i> Yao & Li, 2017	21.79	11.03	Dong et al. (2017)	1.98	1.22-1.80 (313)
<i>Psilochorus texanus</i> Slowik 2009	3.71	3.91	Slowik (2009)	0.95	1.28-1.68 (17)
<i>Smeringopous voi</i> Yao & Li, 2019	6.15	8.65	Yao et al. (2019)	0.71	1.31-1.58 (53)

References

- Dimitrov, D. and C. Ribera (2006), *Three new species of Pholcus (Araneae, Pholcidae) from the Canary Islands with notes on the genus Pholcus in the archipelago*, J Arachnol **34** 126.
- Dong, T., G. Zheng, Z. Yao and S. Li (2017), *Ten new species of the spider genus Pholcus Walckenaer, 1805 (Araneae, Pholcidae) from Southeast Asia*, Zootaxa **4306** 339.
- Machado, É. O., F. U. Yamamoto, A. D. Brescovit and B. A. Huber (2007), *Three new ground living pholcid species (Araneae: Pholcidae) from São Paulo, Brazil*, Zootaxa **1582** 35.
- Minh, B. Q., H. A. Schmidt, O. Chernomor et al. (2020), *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era*, Mol Biol Evol **37** 1530.
- Slowik, J. (2009), *A review of the cellar spider genus Psilochorus Simon 1893 in America north of Mexico (Araneae: Pholcidae)*, Zootaxa **2144** 1.
- Yao, Z., G. M. Kioko, E. N. Kioko and S. Li (2019), *Three new species of the spider family Pholcidae (Arachnida: Araneae) from Taita-Taveta, Kenya.*, Zootaxa **4651** zootaxa.
- Zhang, F. and Y.-Q. Peng (2011), *Eleven new species of the genus Belisana Thorell (Araneae: Pholcidae) from south China*, Zootaxa **2989** 51.
- Zhang, F. and M.-S. Zhu (2009), *A review of the genus Pholcus (Araneae: Pholcidae) from China*, Zootaxa **2037** 1.
- Zhang, Y. Q., H. M. Chen and M. S. Zhu (2008), *A new troglophilous Belisana spider from Guangxi, China (Araneae, Pholcidae)*, Acta zootaxonomica sinica **33** 654.
- Zhu, W., Z. Yao, G. Zheng and S. Li (2020a), *Six new species of the spider genus Belisana Thorell, 1898 (Araneae: Pholcidae) from southern China*, Zootaxa **4810** 175.
- Zhu, W., Z. Yao, G. Zheng and S. Li (2020b), *Three new species of the spider genus Belisana Thorell, 1898 (Araneae: Pholcidae) from Xishuangbanna, Yunnan, China*, Zootaxa **4742** 332.

Supplementary information for Chapter 6

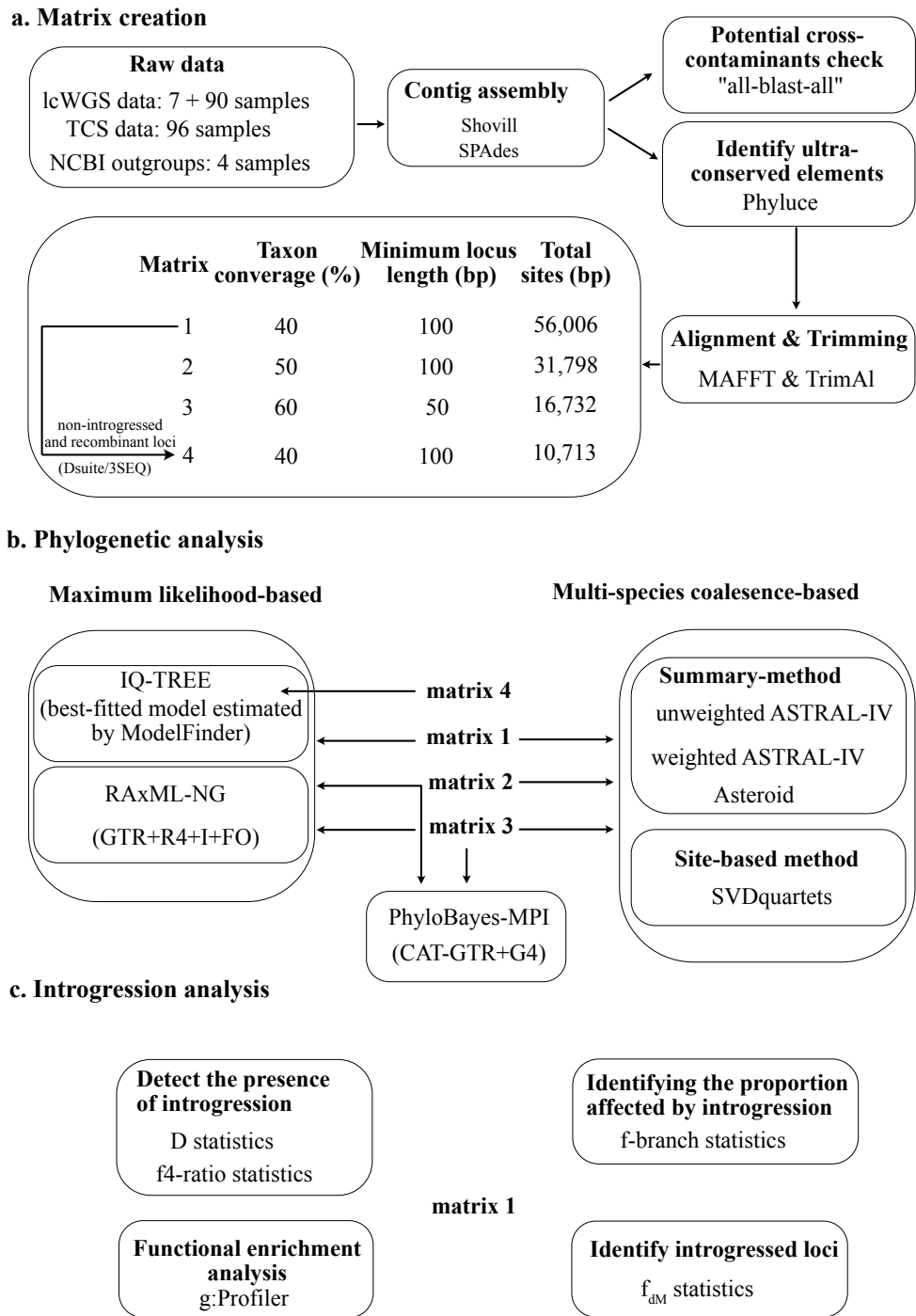


Figure D.1. Workflow for analysis with UCE data. (a) Steps to create the four matrices of UCE. The 7 lcWGS data is from Meng, Podsiadlowski and Huber (2025b) and 90 lcWGS (low-coverage Whole Genome Sequencing) samples were newly sequenced by this study. The 96 TCS (Target Captured Sequencing) data is from Meng, Podsiadlowski, Dimitrov et al. (2025). The TCS data was assembled by SPAdes while the lcWGS data was assembled with Shovill. To detect potential cross-contamination, we blasted the assembly of each sample against the other samples ("all-blast-all"), the hits were filtered out by the method described in Meng, Podsiadlowski, Dimitrov et al. (2025). (b) Phylogenetic analysis with UCE matrices. (c) Steps for introgression analysis.

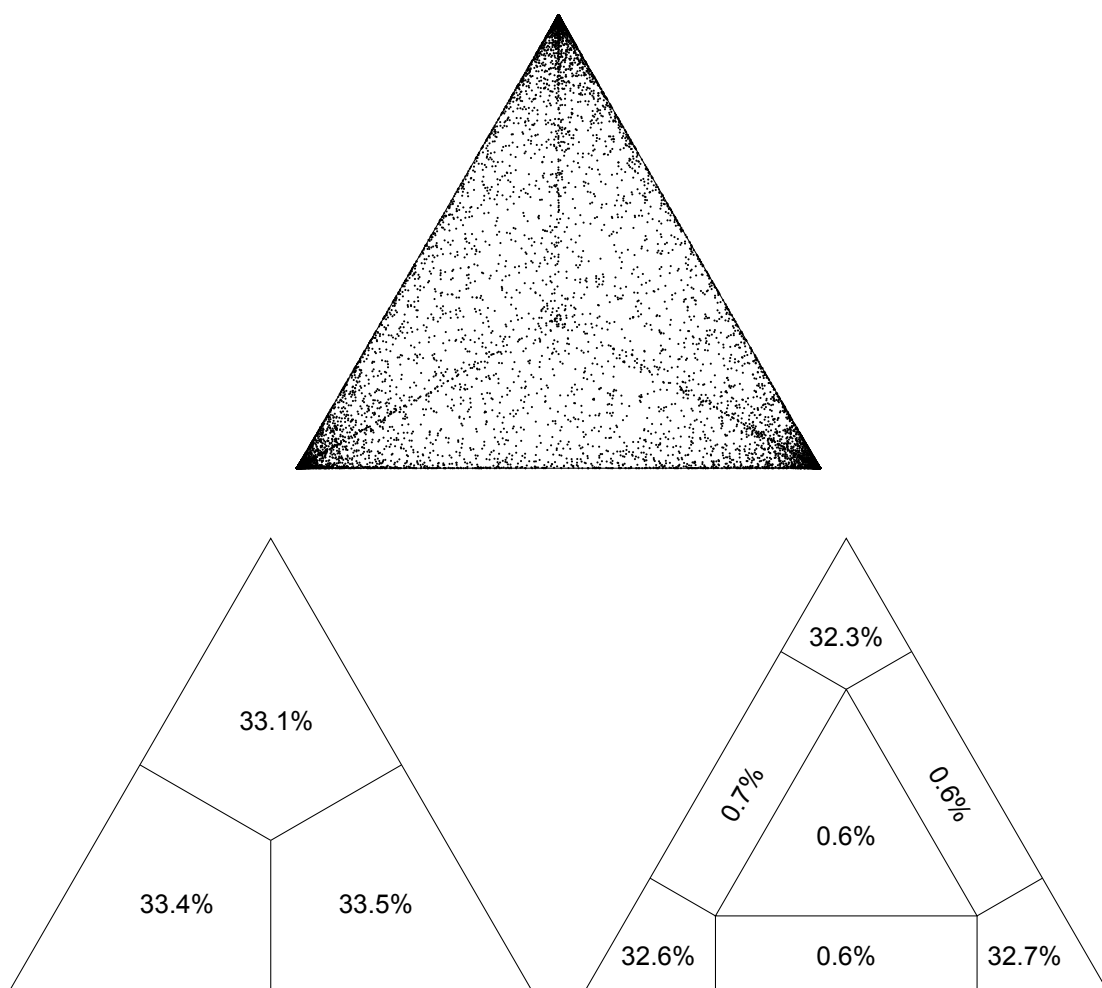


Figure D.2. Likelihood mapping analysis of matrix 1. The analysis was done with IQ-TREE

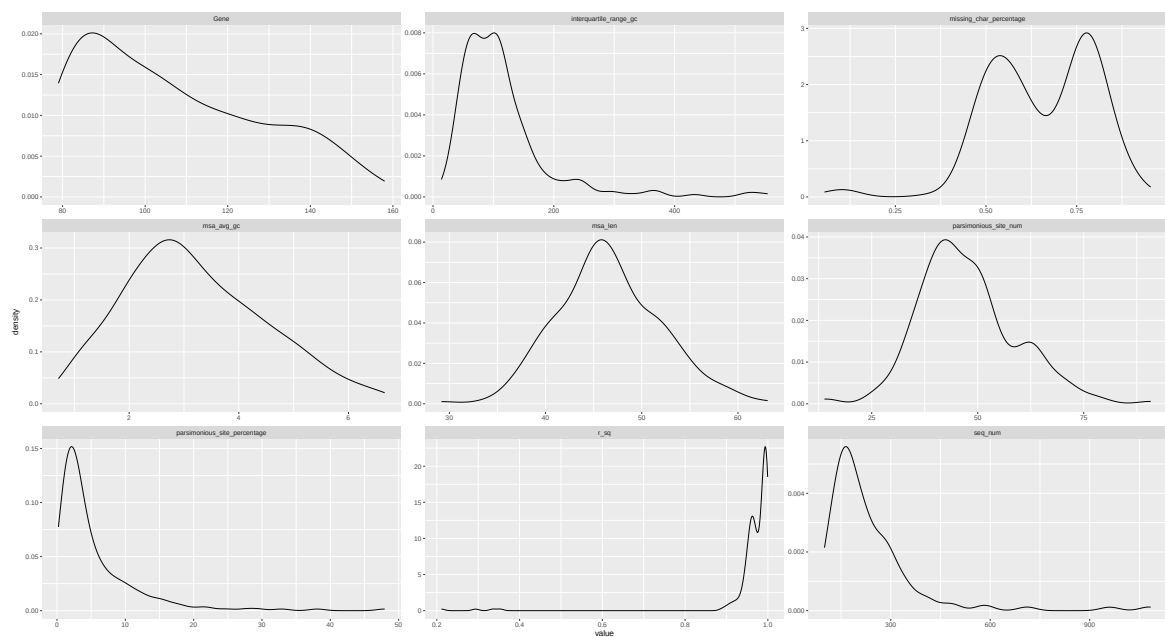


Figure D.3. Density plots of alignment characteristics of matrix 1 . The visualization was finished with ggplot2.

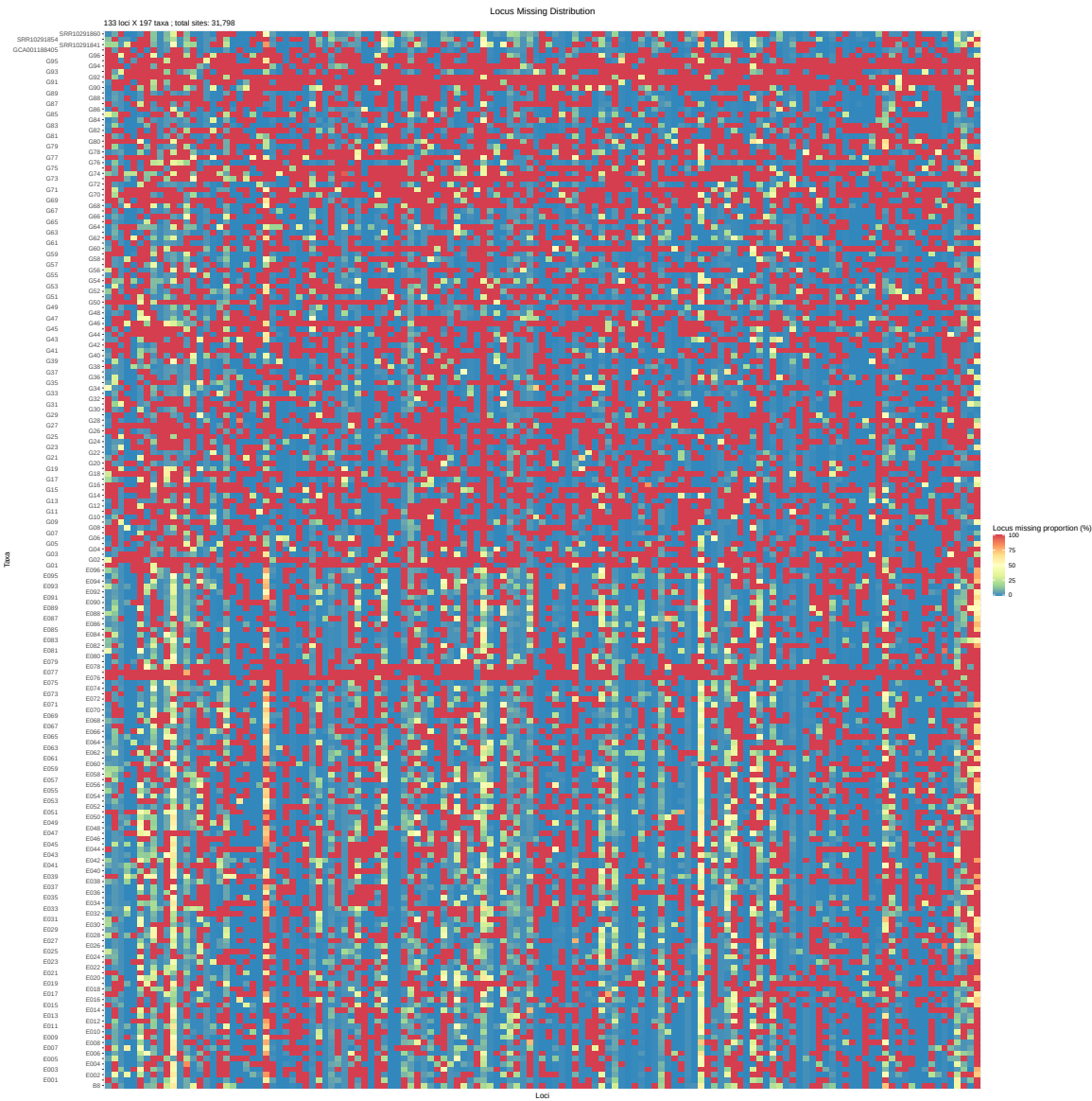


Figure D.5. Missing data distribution of matrix 2. X-axis denotes loci, Y-axis denotes samples.

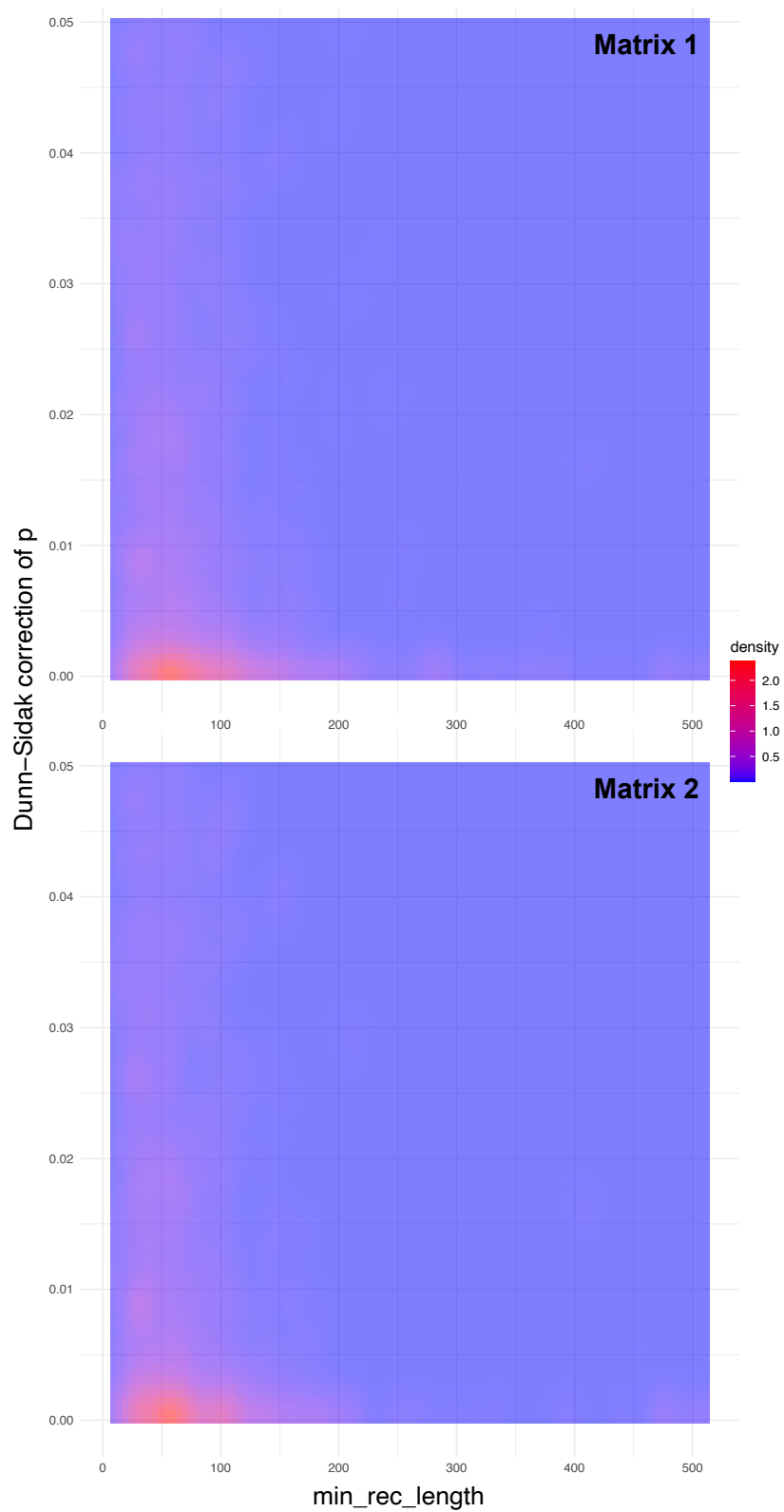


Figure D.6. Intra-locus recombination analysis of matrix 1-2. The analysis was performed with 3SEQ

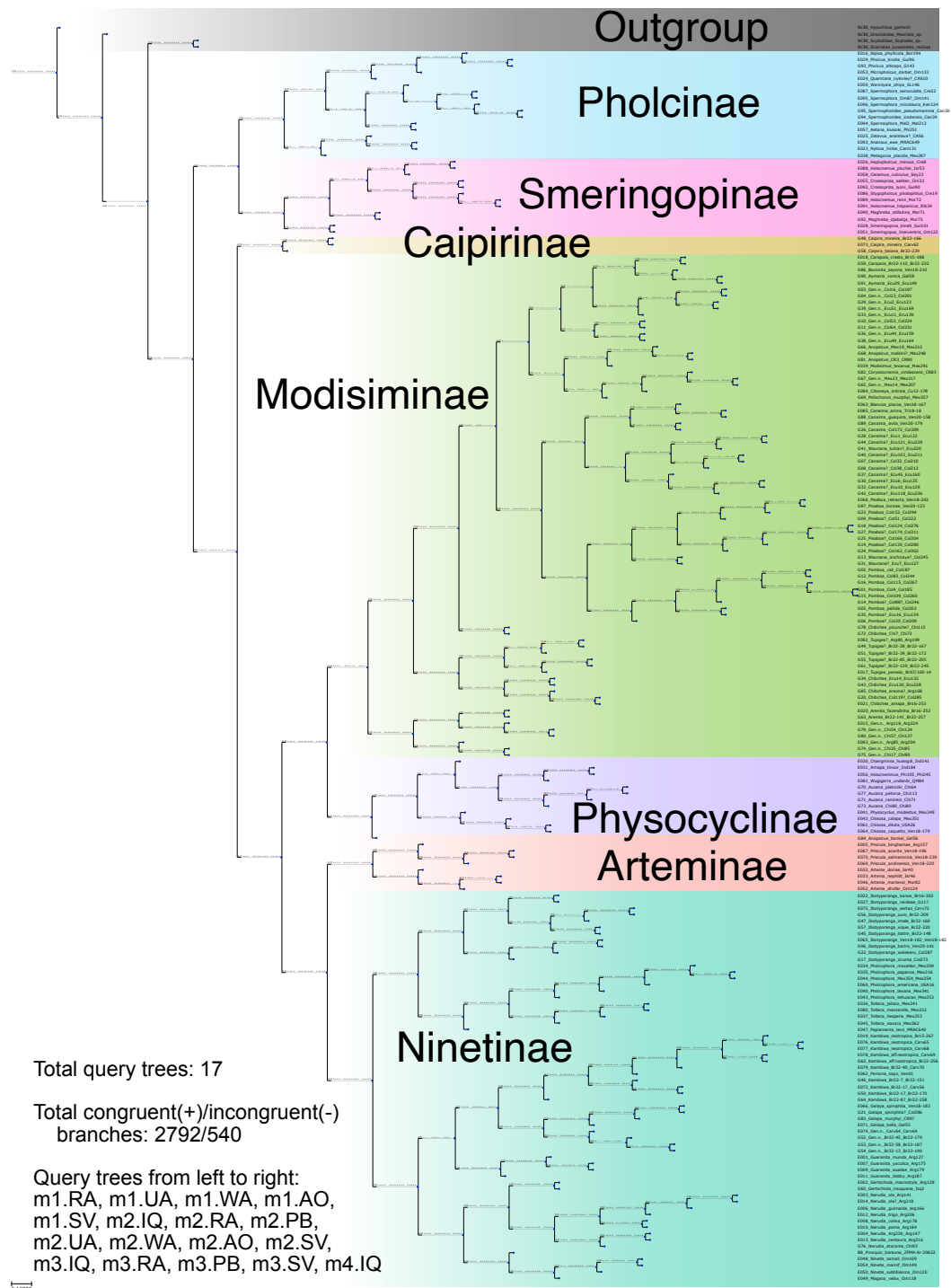


Figure D.7. Summary of phylogenetic trees based on UCE data . The topology (“reference tree”) shown is the same as Figure 6.1, 6.2 and 6.3, and the other trees (“query trees”) are compared with the reference tree. “+” (or “-”) on the branches denotes query tree are congruent (or incongruent) with the reference tree. Data matrix: m1-m4: matrix 1 to matrix 4. Tree construction method: IQ, IQ-TREE; RA, RAxML-NG; UA, unweithed ASTRAL-IV; WA, weithed ASTRAL-IV; AO, Asteroid; SV, SVDquartets; PB, Phylogbayes-MPI.

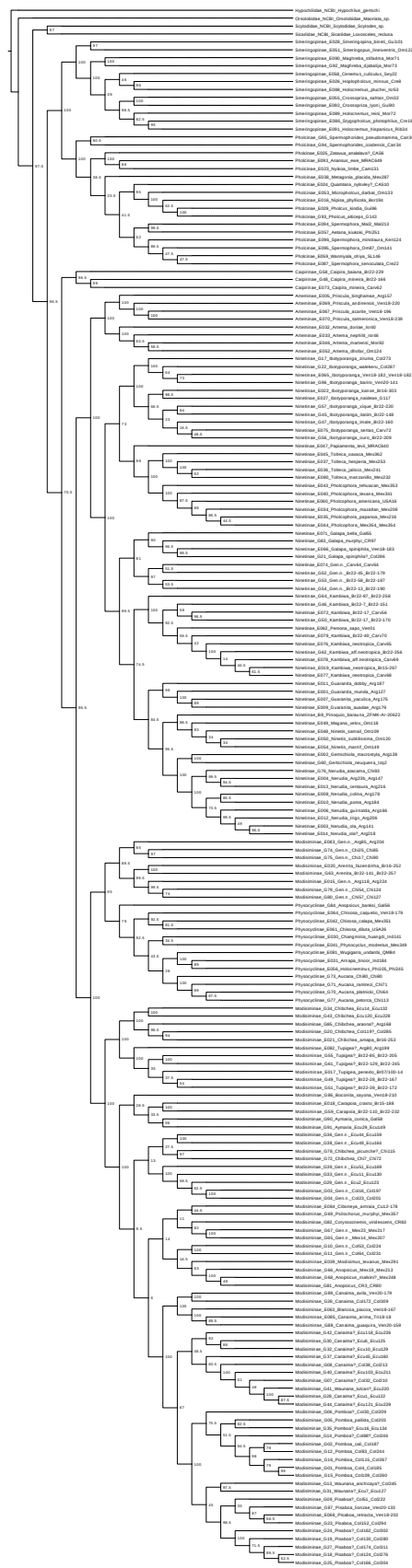


Figure D.8. SVDquartets tree based on matrix 1. Branch numbers are bootstrap supports.



Figure D.9. Unweighted ASTRAL-IV tree based on matrix 2. Branch numbers are local posterior probability/gene concordance factor (gCF)/site concordance factor (sCF), respectively.

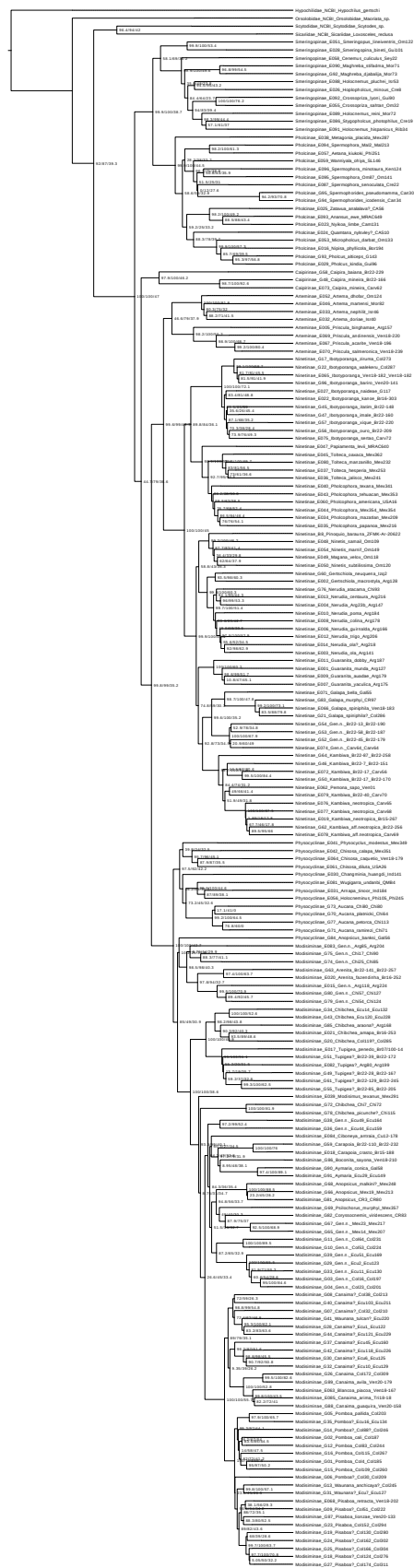


Figure D.10. IQ-TREE tree based on matrix 4. The support values shown on the branches are SH-aLRT support (%) / ultrafast bootstrap support (UFBoot) (%) / site concordance factor (sCF), respectively.

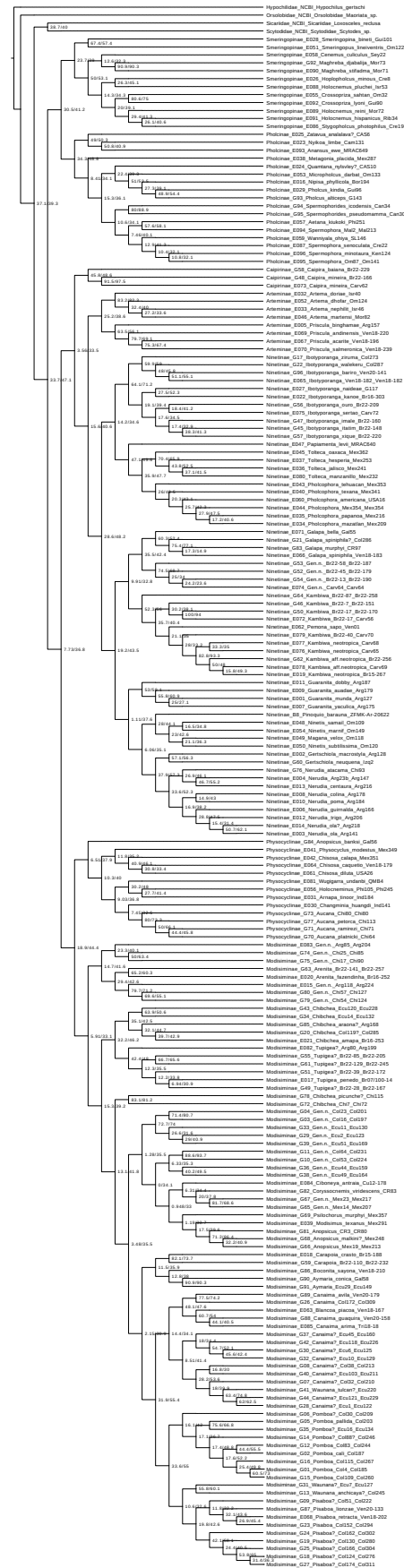


Figure D.11. Asteroid tree based on matrix 1. The support values shown on the branches are gene concordance factor (gCF)/site concordance factor (sCF), respectively.

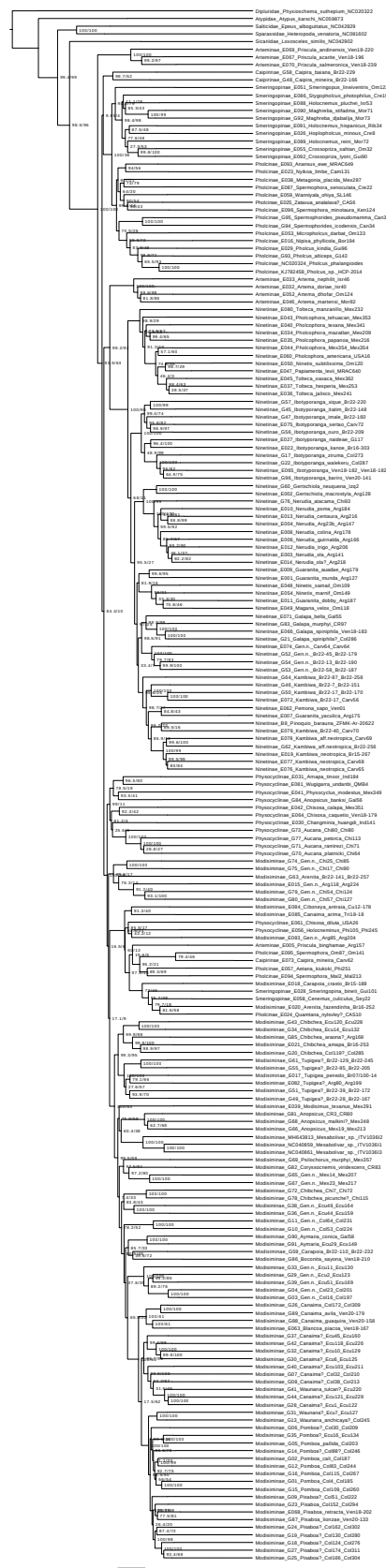


Figure D.12. Mitochondrial phylogeny . The tree was constructed with IQ-TREE and 13 mitochondrial protein-coding genes (PCGs) and 12S + 16S rDNA, which were trimmed with ClipKIT kpi-gappy. The support values shown on the branches are SH-aLRT support (%) /ultrafast bootstrap support (UFBoot) (%), respectively.

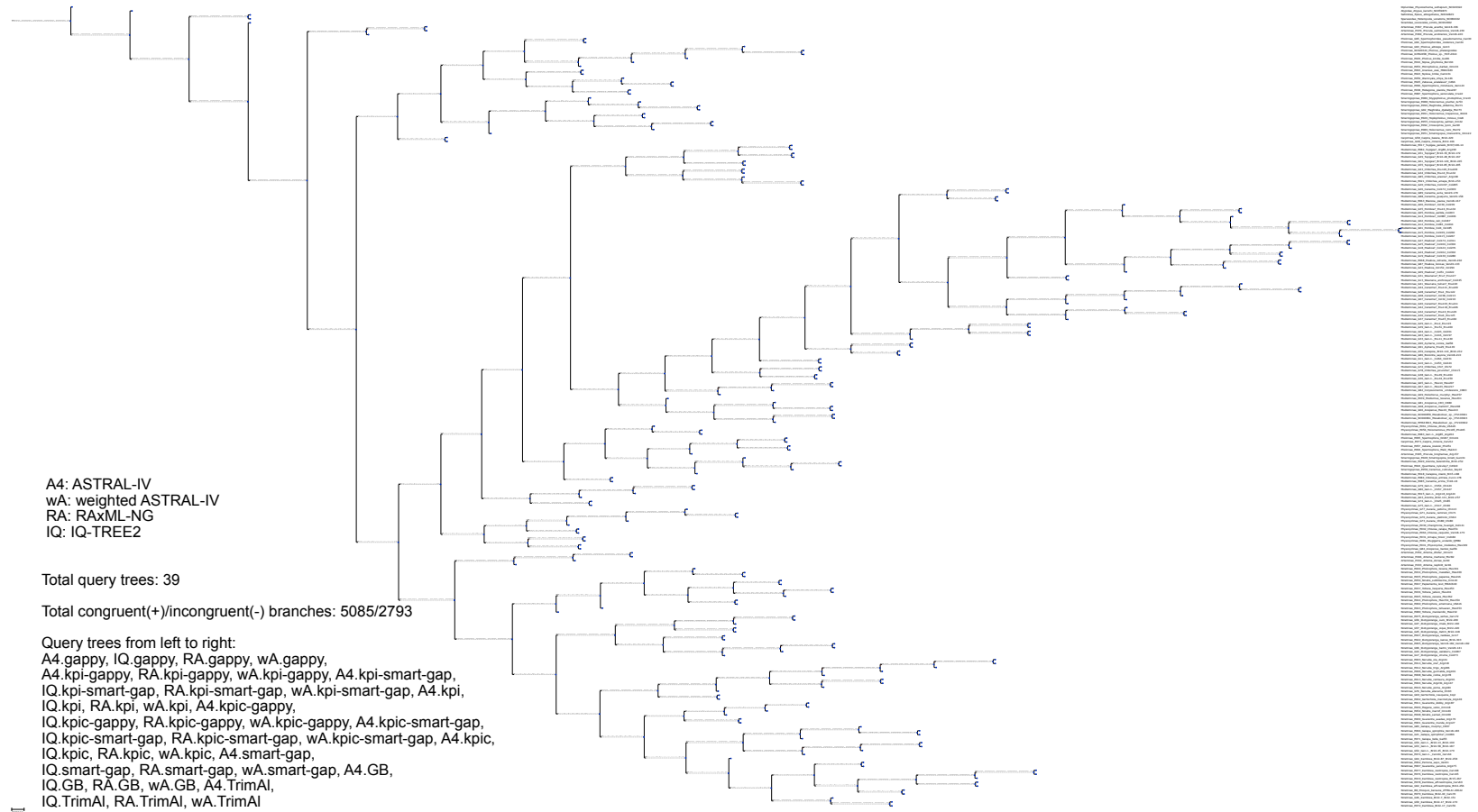


Figure D.13. Summary tree of mitochondrial phylogenies . The topology (“reference tree”) shown is the same as Figure S11, and the other trees (“query trees”) are compared with the reference tree. “+”(or “-”) on the branches denotes query tree are congruent (or incongruent) with the reference tree.

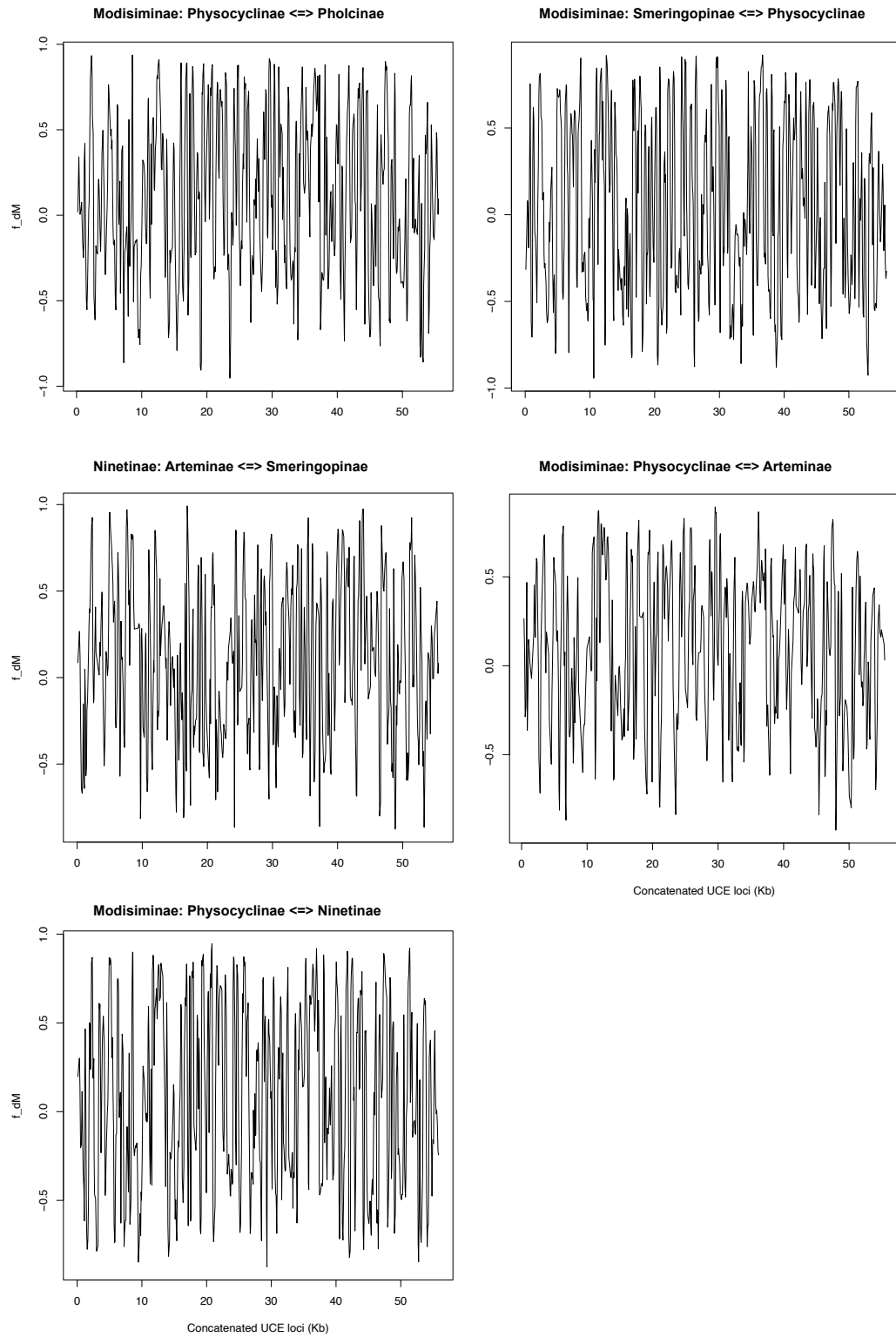


Figure D.14. f_{dM} distribution along untrimmed UCE loci.

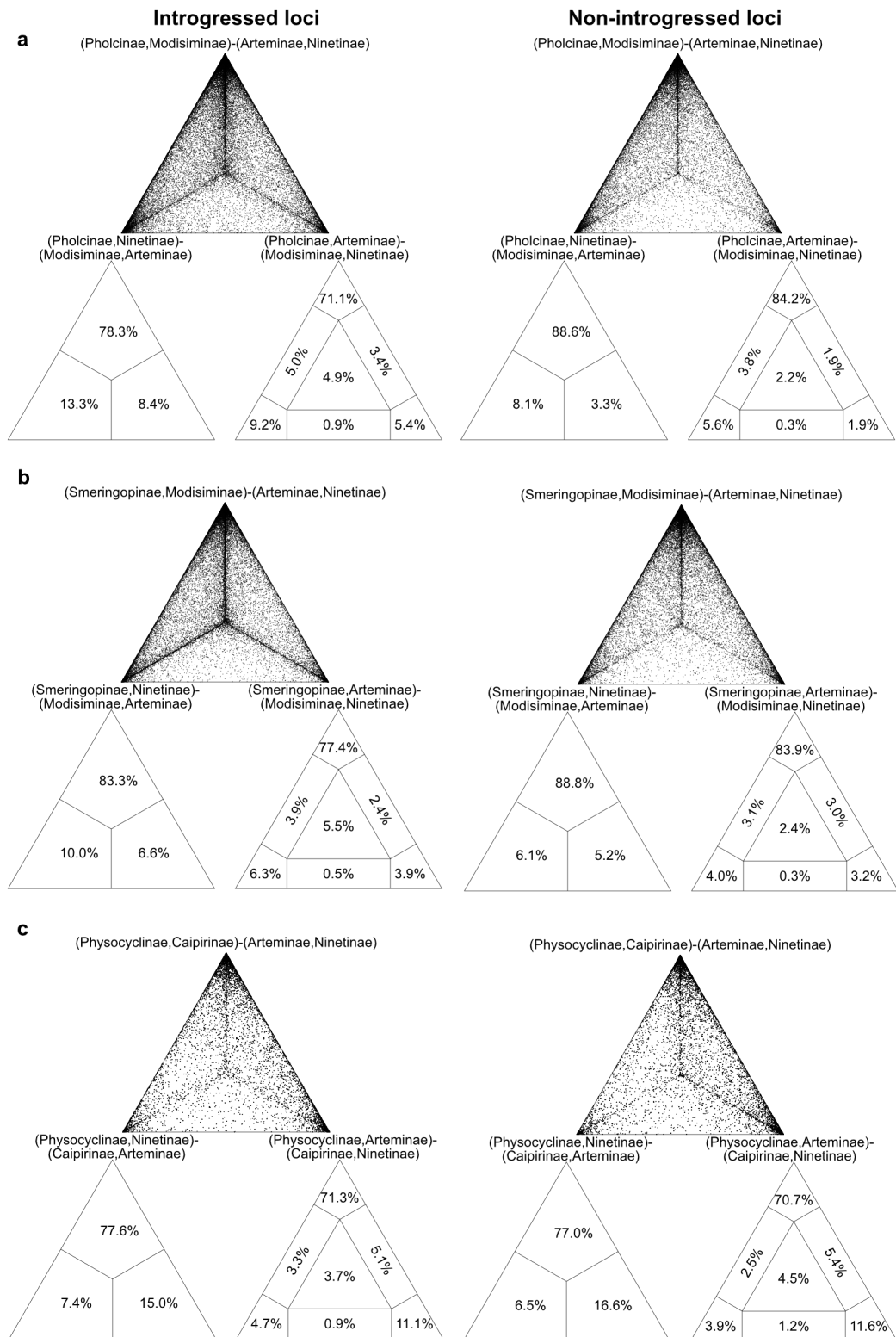


Figure D.15. Likelihood mapping of introgressed and non-introgressed loci..

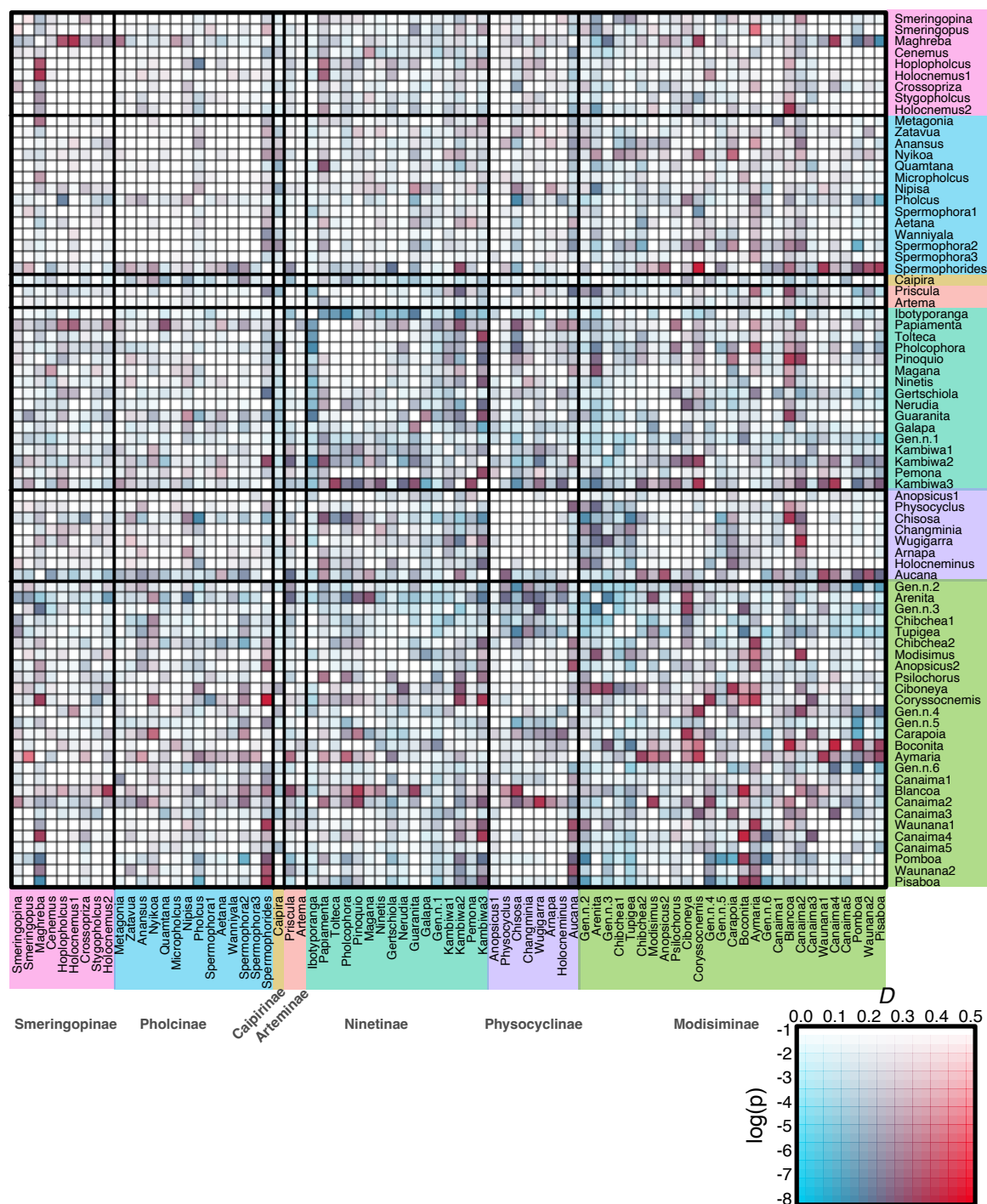


Figure D.16. D statistics showing widespread intra-subfamily introgression..

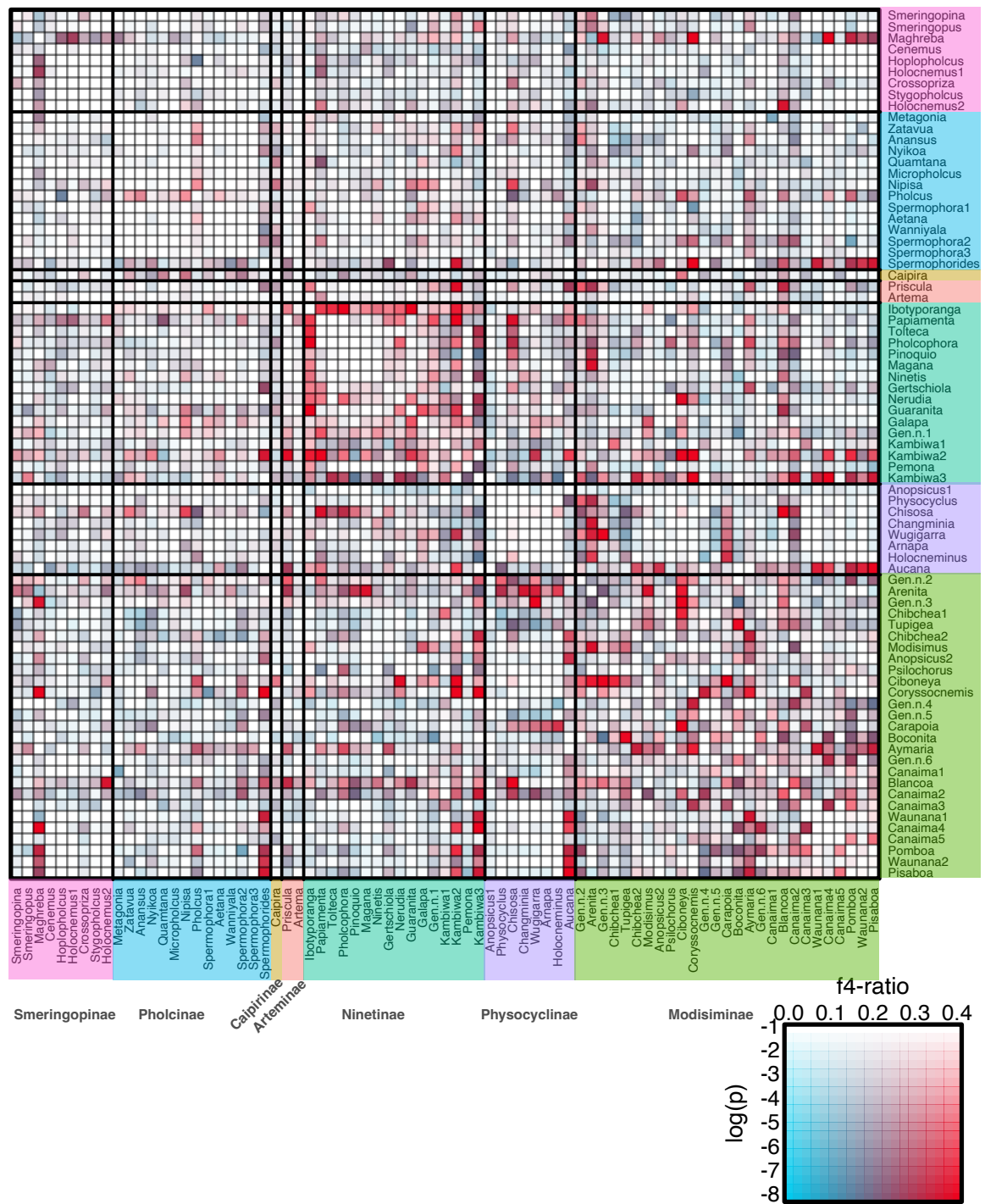


Figure D.17. The f_4 -ratio statistics showing widespread intra-subfamily introgression..

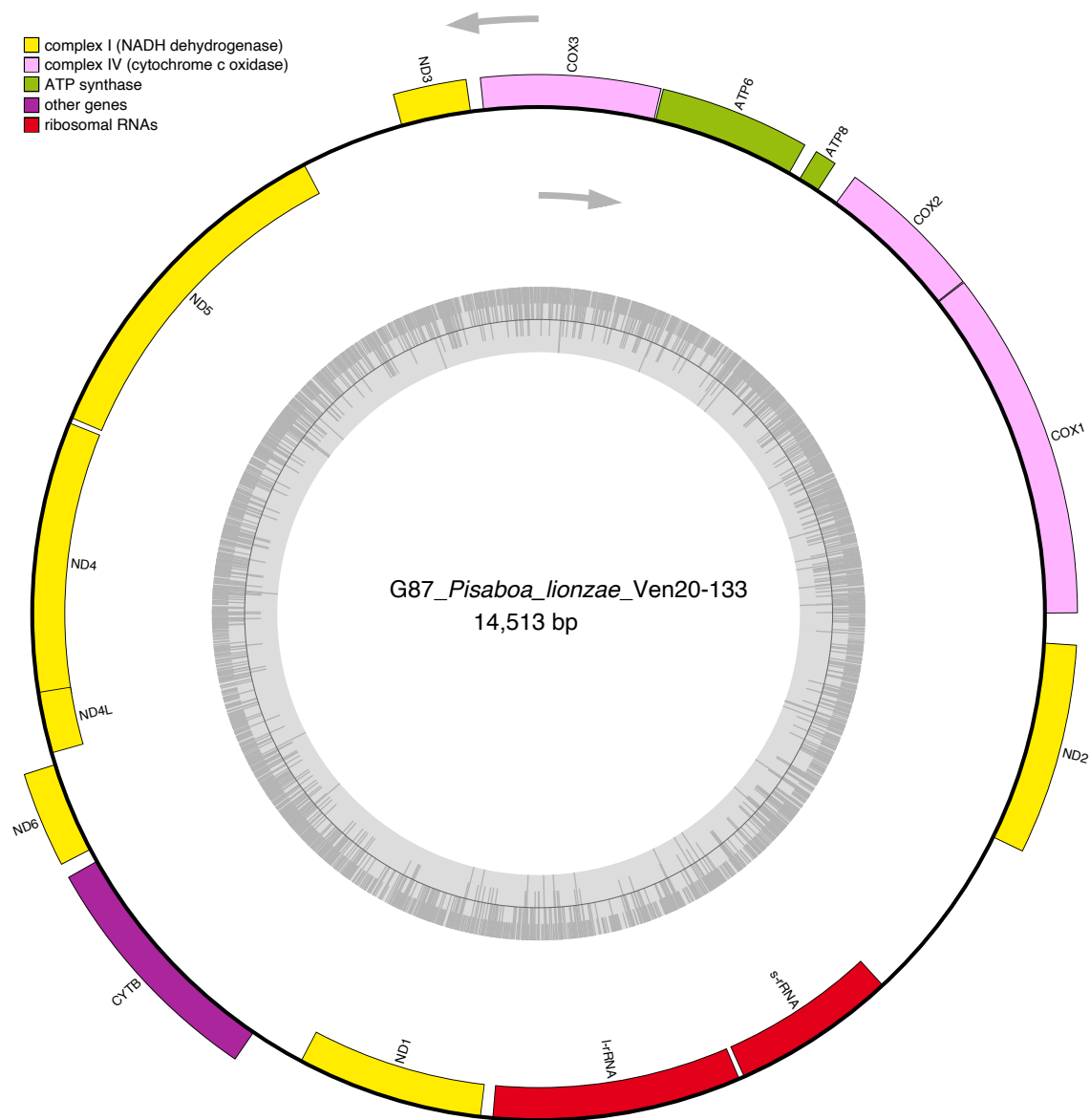


Figure D.18. An example of Pholcidae mitochondrial genomes . Only protein-coding genes and rDNA genes are shown.

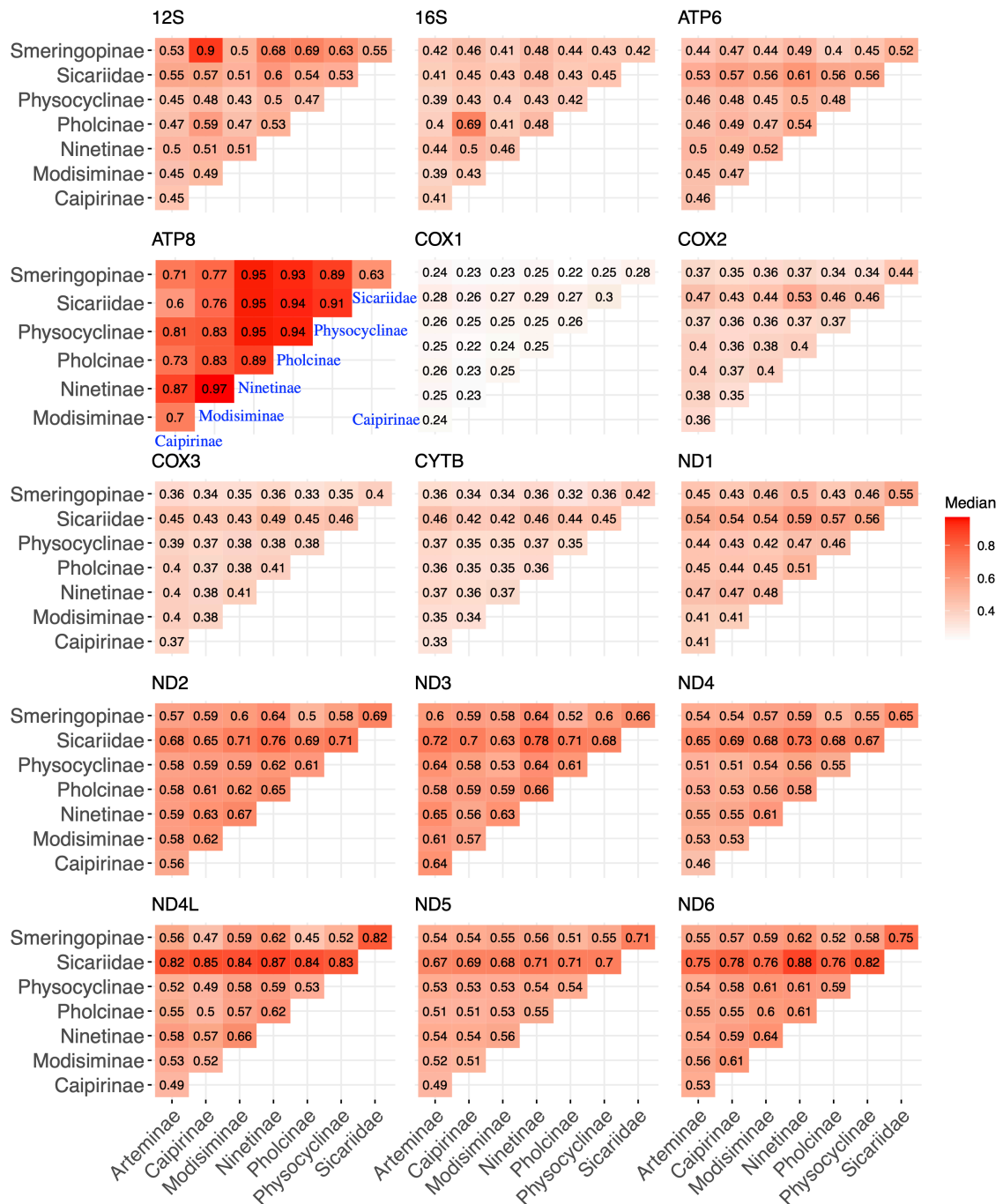


Figure D.19. Inter-subfamily genetic distances. Median values are shown.

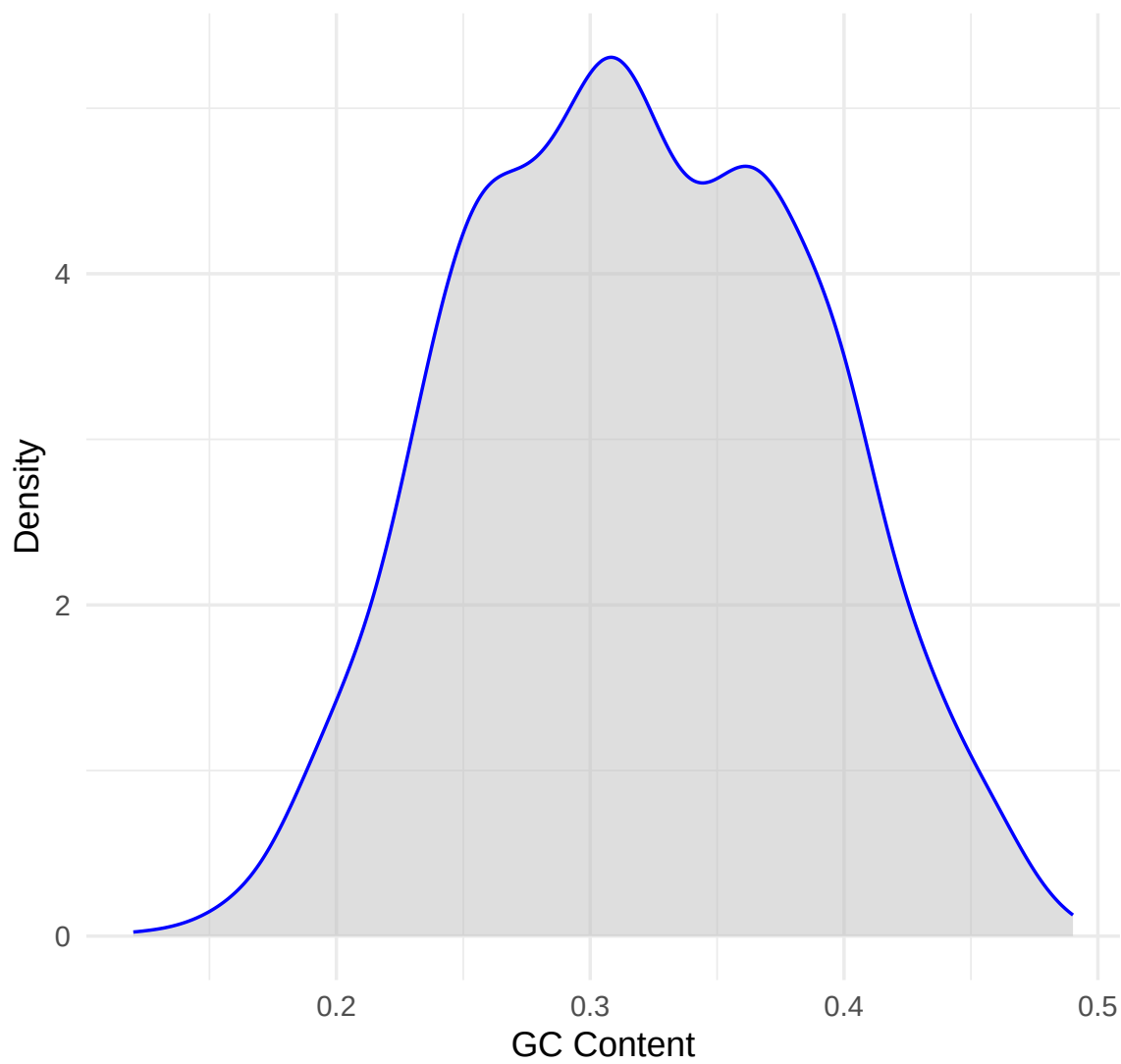


Figure D.20. GC content distribution of Pholcidae mitochondrial PCGs.

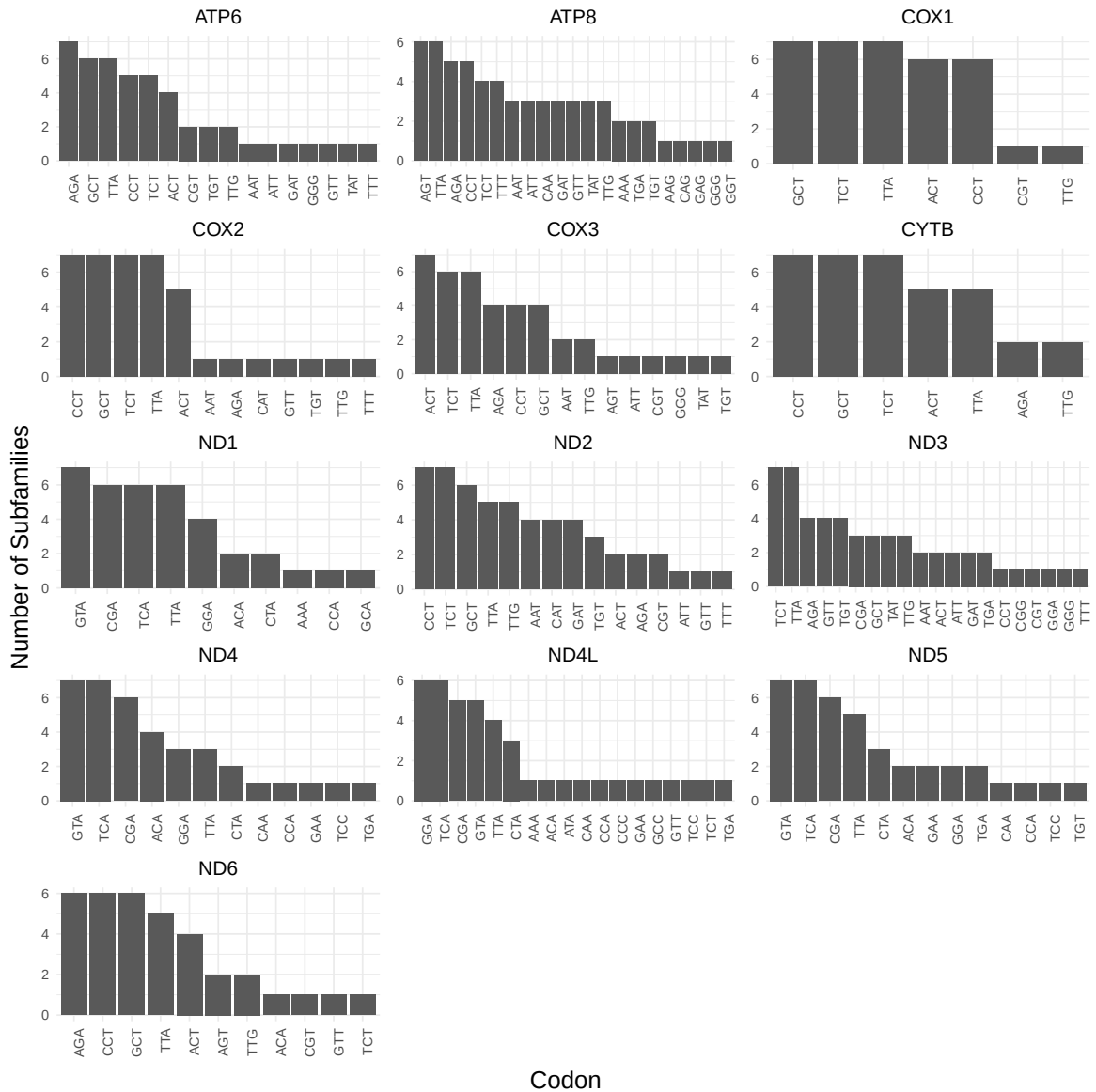


Figure D.21. The number of subfamilies sharing the same codons for each PCG . For each gene, this was done by (1) choosing the top five codons with the highest RSCU values for each subfamily, and then (2) counting the frequency of all codons in the seven Pholcidae subfamilies. Some codons may be presented in all seven subfamilies.

Table D.1. Sample information

Subfamily	Code	Sample name	Source	Mitochondrial genome (circular/linear/fragmented/NA)
Arteminae	E052	E052_Artema_dhofar_Om124	Meng et al. 2025a	fragmented
Arteminae	E032	E032_Artema_doriae_Isr40	Meng et al. 2025a	fragmented
Arteminae	E046	E046_Artema_martensi_Mor82	Meng et al. 2025a	fragmented
Arteminae	E033	E033_Artema_nephilit_Isr46	Meng et al. 2025a	fragmented
Arteminae	E067	E067_Priscula_acarite_Ven18-196	Meng et al. 2025a	fragmented
Arteminae	E069	E069_Priscula_andinensis_Ven18-220	Meng et al. 2025a	fragmented
Arteminae	E005	E005_Priscula_binghamae_Arg157	Meng et al. 2025a	fragmented
Arteminae	E070	E070_Priscula_salmeronica_Ven18-239	Meng et al. 2025a	fragmented
Caj***inae	G58	G58_Gen.n._sp2_Br22-229	Meng et al. 2025b	linear
Caj***inae	E073	E073_Gen.n._sp1_Carv62	Meng et al. 2025b	fragmented
Caj***inae	G48	G48_Gen.n._sp1_Br22-166	Meng et al. 2025b	circular
Modisiminae	G81	G81_Anopsicus_CR3_CR80	this study	circular
Modisiminae	G68	G68_Anopsicus_malkini?_Mex248	this study	circular
Modisiminae	G66	G66_Anopsicus_Mex19_Mex213	this study	circular
Modisiminae	G63	G63_Arenita_Br22-141_Br22-257	this study	circular
Modisiminae	E020	E020_Arenita_fazendinha_Br16-252	Meng et al. 2025a	fragmented
Modisiminae	G90	G90_Aymaria_conica_Gal58	this study	circular
Modisiminae	G91	G91_Aymaria_Ecu29_Ecu149	this study	circular
Modisiminae	E063	E063_Blancoa_piacoa_Ven18-167	Meng et al. 2025a	fragmented
Modisiminae	G86	G86_Boconita_sayona_Ven18-210	this study	circular
Modisiminae	E085	E085_Canaima_arima_Tri18-18	Meng et al. 2025a	fragmented
Modisiminae	G89	G89_Canaima_avila_Ven20-179	this study	circular
Modisiminae	G26	G26_Canaima_Col172_Col309	this study	circular
Modisiminae	G88	G88_Canaima_guaquira_Ven20-158	this study	circular
Modisiminae	G07	G07_Canaima?_Col32_Col210	this study	circular
Modisiminae	G08	G08_Canaima?_Col38_Col213	this study	circular
Modisiminae	G28	G28_Canaima?_Ecu1_Ecu122	this study	circular
Modisiminae	G32	G32_Canaima?_Ecu10_Ecu129	this study	circular
Modisiminae	G40	G40_Canaima?_Ecu103_Ecu211	this study	circular
Modisiminae	G42	G42_Canaima?_Ecu118_Ecu226	this study	circular
Modisiminae	G44	G44_Canaima?_Ecu121_Ecu229	this study	circular
Modisiminae	G37	G37_Canaima?_Ecu45_Ecu160	this study	circular
Modisiminae	G30	G30_Canaima?_Ecu6_Ecu125	this study	circular
Modisiminae	G59	G59_Carapoia_Br22-110_Br22-232	this study	circular
Modisiminae	E018	E018_Carapoia_crasto_Br15-188	Meng et al. 2025a	fragmented
Modisiminae	E021	E021_Chibchea_amapa_Br16-253	Meng et al. 2025a	fragmented
Modisiminae	G85	G85_Chibchea_araona?_Arg168	this study	circular
Modisiminae	G72	G72_Chibchea_Chi7_Chi72	this study	circular
Modisiminae	G20	G20_Chibchea_Col119?_Col285	this study	circular
Modisiminae	G43	G43_Chibchea_Ecu120_Ecu228	this study	circular
Modisiminae	G34	G34_Chibchea_Ecu14_Ecu132	this study	circular
Modisiminae	G78	G78_Chibchea_picunche?_Chi115	this study	circular
Modisiminae	E084	E084_Ciboneya_antraia_Cu12-178	Meng et al. 2025a	fragmented
Modisiminae	G82	G82_Coryssocnemis_viridescens_CR83	this study	circular
Modisiminae	E015	E015_Gen.n._Arg118_Arg224	Meng et al. 2025a	fragmented
Modisiminae	E083	E083_Gen.n._Arg85_Arg204	Meng et al. 2025a	fragmented
Modisiminae	G75	G75_Gen.n._Chi17_Chi90	this study	circular
Modisiminae	G74	G74_Gen.n._Chi25_Chi85	this study	circular
Modisiminae	G79	G79_Gen.n._Chi54_Chi124	this study	circular
Modisiminae	G80	G80_Gen.n._Chi57_Chi127	this study	linear
Modisiminae	G03	G03_Gen.n._Col16_Col197	this study	circular
Modisiminae	G04	G04_Gen.n._Col23_Col201	this study	circular
Modisiminae	G10	G10_Gen.n._Col53_Col224	this study	circular
Modisiminae	G11	G11_Gen.n._Col64_Col231	this study	circular

Continued on next page

Appendix D Supplementary information for Chapter 6

Subfamily	Code	Sample name	Source	Mitochondrial genome (circular/linear/fragmented/NA)
Modisiminae	G33	G33_Gen.n._Ecu11_Ecu130	this study	circular
Modisiminae	G29	G29_Gen.n._Ecu2_Ecu123	this study	circular
Modisiminae	G36	G36_Gen.n._Ecu44_Ecu159	this study	linear
Modisiminae	G38	G38_Gen.n._Ecu49_Ecu164	this study	linear
Modisiminae	G39	G39_Gen.n._Ecu51_Ecu169	this study	circular
Modisiminae	G65	G65_Gen.n._Mex14_Mex207	this study	circular
Modisiminae	G67	G67_Gen.n._Mex23_Mex217	this study	circular
Modisiminae	E039	E039_Modisimus_texanus_Mex291	Meng et al. 2025a	fragmented
Modisiminae	G23	G23_Pisaboa_Col152_Col294	this study	linear
Modisiminae	G87	G87_Pisaboa_lionzae_Ven20-133	this study	circular
Modisiminae	E068	E068_Pisaboa_retracta_Ven18-202	Meng et al. 2025a	fragmented
Modisiminae	G18	G18_Pisaboa?_Col124_Col276	this study	circular
Modisiminae	G19	G19_Pisaboa?_Col130_Col280	this study	circular
Modisiminae	G24	G24_Pisaboa?_Col162_Col302	this study	circular
Modisiminae	G25	G25_Pisaboa?_Col166_Col304	this study	linear
Modisiminae	G27	G27_Pisaboa?_Col174_Col311	this study	linear
Modisiminae	G09	G09_Pisaboa?_Col51_Col222	this study	circular
Modisiminae	G02	G02_Pomboa_cali_Col187	this study	circular
Modisiminae	G15	G15_Pomboa_Col109_Col260	this study	circular
Modisiminae	G16	G16_Pomboa_Col115_Col267	this study	circular
Modisiminae	G01	G01_Pomboa_Col4_Col185	this study	circular
Modisiminae	G12	G12_Pomboa_Col83_Col244	this study	circular
Modisiminae	G05	G05_Pomboa_pallida_Col203	this study	circular
Modisiminae	G06	G06_Pomboa?_Col30_Col209	this study	linear
Modisiminae	G14	G14_Pomboa?_Col88?_Col246	this study	circular
Modisiminae	G35	G35_Pomboa?_Ecu16_Ecu134	this study	circular
Modisiminae	G69	G69_Psilochorus_murphyi_Mex357	this study	circular
Modisiminae	E017	E017_Tupigea_penedo_Br07/100-14	Meng et al. 2025a	fragmented
Modisiminae	E082	E082_Tupigea?_Arg80_Arg199	Meng et al. 2025a	fragmented
Modisiminae	G61	G61_Tupigea?_Br22-129_Br22-245	this study	circular
Modisiminae	G49	G49_Tupigea?_Br22-28_Br22-167	this study	circular
Modisiminae	G51	G51_Tupigea?_Br22-39_Br22-172	this study	circular
Modisiminae	G55	G55_Tupigea?_Br22-85_Br22-205	this study	circular
Modisiminae	G13	G13_Waunana_anchicaya?_Col245	this study	linear
Modisiminae	G41	G41_Waunana_tulcan?_Ecu220	this study	circular
Modisiminae	G31	G31_Waunana?_Ecu7_Ecu127	this study	circular
Ninetinae	E071	E071_Galapa_bella_Gal55	Meng et al. 2025a	fragmented
Ninetinae	G83	G83_Galapa_murphyi_CR97	this study	circular
Ninetinae	E066	E066_Galapa_spiniphila_Ven18-183	Meng et al. 2025a	fragmented
Ninetinae	G21	G21_Galapa_spiniphila?_Col286	this study	circular
Ninetinae	G54	G54_Gen.n._Br22-13_Br22-190	this study	circular
Ninetinae	G52	G52_Gen.n._Br22-45_Br22-179	this study	circular
Ninetinae	G53	G53_Gen.n._Br22-58_Br22-187	this study	circular
Ninetinae	E074	E074_Gen.n._Carv64_Carv64	Meng et al. 2025a	fragmented
Ninetinae	E002	E002_Gertschiola_macrostyla_Arg128	Meng et al. 2025a	fragmented
Ninetinae	G60	G60_Gertschiola_neuquena_Izq2	this study	circular
Ninetinae	E009	E009_Guaranita_auadae_Arg179	Meng et al. 2025a	fragmented
Ninetinae	E011	E011_Guaranita_dobby_Arg187	Meng et al. 2025a	fragmented
Ninetinae	E001	E001_Guaranita_munda_Arg127	Meng et al. 2025a	fragmented
Ninetinae	E007	E007_Guaranita_yaculica_Arg175	Meng et al. 2025a	fragmented
Ninetinae	G96	G96_Ibotyporanga_bariro_Ven20-141	this study	circular
Ninetinae	G47	G47_Ibotyporanga_imale_Br22-160	this study	circular
Ninetinae	G45	G45_Ibotyporanga_itatim_Br22-148	this study	circular
Ninetinae	E022	E022_Ibotyporanga_kanoe_Br16-303	Meng et al. 2025a	fragmented
Ninetinae	E027	E027_Ibotyporanga_naideae_G117	Meng et al. 2025a	fragmented
Ninetinae	G56	G56_Ibotyporanga_ouro_Br22-209	this study	circular
Ninetinae	E075	E075_Ibotyporanga_sertao_Carv72	Meng et al. 2025a	fragmented

Continued on next page

Subfamily	Code	Sample name	Source	Mitochondrial genome (circular/linear/fragmented/NA)
Ninetinae	E065	E065_Ibotyporanga_Ven18-182_Ven18-182	Meng et al. 2025a	fragmented
Ninetinae	G22	G22_Ibotyporanga_walekeru_Col287	this study	linear
Ninetinae	G57	G57_Ibotyporanga_xique_Br22-220	this study	circular
Ninetinae	G17	G17_Ibotyporanga_ziruma_Col273	this study	circular
Ninetinae	E078	E078_Kambiwa_aff.neotropica_Carv69	Meng et al. 2025a	fragmented
Ninetinae	G62	G62_Kambiwa_aff.neotropica_Br22-256	this study	circular
Ninetinae	E072	E072_Kambiwa_Br22-17_Carv56	Meng et al. 2025a	fragmented
Ninetinae	G50	G50_Kambiwa_Br22-17_Br22-170	this study	circular
Ninetinae	E079	E079_Kambiwa_Br22-40_Carv70	Meng et al. 2025a	fragmented
Ninetinae	G46	G46_Kambiwa_Br22-7_Br22-151	this study	circular
Ninetinae	G64	G64_Kambiwa_Br22-87_Br22-258	this study	circular
Ninetinae	E019	E019_Kambiwa_neotropica_Br15-267	Meng et al. 2025a	fragmented
Ninetinae	E076	E076_Kambiwa_neotropica_Carv65	Meng et al. 2025a	fragmented
Ninetinae	E077	E077_Kambiwa_neotropica_Carv68	Meng et al. 2025a	fragmented
Ninetinae	E049	E049_Magana_velox_Om118	Meng et al. 2025a	fragmented
Ninetinae	E004	E004_Nerudia_Arg23b_Arg147	Meng et al. 2025a	fragmented
Ninetinae	G76	G76_Nerudia_atacama_Chi93	this study	circular
Ninetinae	E013	E013_Nerudia_centaura_Arg216	Meng et al. 2025a	fragmented
Ninetinae	E008	E008_Nerudia_colina_Arg178	Meng et al. 2025a	fragmented
Ninetinae	E006	E006_Nerudia_guinalda_Arg166	Meng et al. 2025a	fragmented
Ninetinae	E003	E003_Nerudia_ola_Arg141	Meng et al. 2025a	fragmented
Ninetinae	E014	E014_Nerudia_ola?_Arg218	Meng et al. 2025a	fragmented
Ninetinae	E010	E010_Nerudia_poma_Arg184	Meng et al. 2025a	fragmented
Ninetinae	E012	E012_Nerudia_trigo_Arg206	Meng et al. 2025a	fragmented
Ninetinae	E054	E054_Ninetis_marnif_Om149	Meng et al. 2025a	fragmented
Ninetinae	E048	E048_Ninetis_samail_Om109	Meng et al. 2025a	fragmented
Ninetinae	E050	E050_Ninetis_subtilissima_Om120	Meng et al. 2025a	fragmented
Ninetinae	E047	E047_Papiamenta_levii_MRAC640	Meng et al. 2025a	fragmented
Ninetinae	E062	E062_Pemona_sapo_Ven01	Meng et al. 2025a	fragmented
Ninetinae	E060	E060_Pholcophora_americana_USA16	Meng et al. 2025a	fragmented
Ninetinae	E034	E034_Pholcophora_mazatlan_Mex209	Meng et al. 2025a	fragmented
Ninetinae	E044	E044_Pholcophora_Mex354_Mex354	Meng et al. 2025a	fragmented
Ninetinae	E035	E035_Pholcophora_papanao_Mex216	Meng et al. 2025a	fragmented
Ninetinae	E043	E043_Pholcophora_tehuacan_Mex353	Meng et al. 2025a	fragmented
Ninetinae	E040	E040_Pholcophora_texana_Mex341	Meng et al. 2025a	fragmented
Ninetinae	B8	B8_Pinoquio_baraua_ZFMK-Ar-20622	Meng et al. 2025a	fragmented
Ninetinae	E037	E037_Tolteca_hesperia_Mex253	Meng et al. 2025a	fragmented
Ninetinae	E036	E036_Tolteca_jalisco_Mex241	Meng et al. 2025a	fragmented
Ninetinae	E080	E080_Tolteca_manzanillo_Mex232	Meng et al. 2025a	fragmented
Ninetinae	E045	E045_Tolteca_oaxaca_Mex362	Meng et al. 2025a	fragmented
Pholcinae	E057	E057_Aetana_kiukoki_Phi251	Meng et al. 2025a	fragmented
Pholcinae	E093	E093_Anansus_ewe_MRAC649	Meng et al. 2025a	fragmented
Pholcinae	E038	E038_Metagonia_placida_Mex287	Meng et al. 2025a	fragmented
Pholcinae	E053	E053_Micropholcus_darbat_Om133	Meng et al. 2025a	fragmented
Pholcinae	E016	E016_Nipisa_phyllicola_Bor194	Meng et al. 2025a	fragmented
Pholcinae	E023	E023_Nyikoa_limbe_Cam131	Meng et al. 2025a	fragmented
Pholcinae	G93	G93_Pholcus_alticeps_G143	this study	circular
Pholcinae	E029	E029_Pholcus_kindia_Gui96	Meng et al. 2025a	fragmented
Pholcinae	E024	E024_Quamtana_nylsvey?_CAS10	Meng et al. 2025a	fragmented
Pholcinae	E094	E094_Spermophora_Mal2_Mal213	Meng et al. 2025a	fragmented
Pholcinae	E096	E096_Spermophora_minotaura_Ken124	Meng et al. 2025a	fragmented
Pholcinae	E095	E095_Spermophora_Om87_Om141	Meng et al. 2025a	fragmented
Pholcinae	E087	E087_Spermophora_senoculata_Cre22	Meng et al. 2025a	fragmented
Pholcinae	G94	G94_Spermophorides_icodensis_Can34	this study	circular
Pholcinae	G95	G95_Spermophorides_pseudomamma_Can30	this study	circular
Pholcinae	E059	E059_Wanniyala_ohiya_SL146	Meng et al. 2025a	fragmented

Continued on next page

Appendix D Supplementary information for Chapter 6

Subfamily	Code	Sample name	Source	Mitochondrial genome (circular/linear/fragmented/NA)
Pholcinae	E025	E025_Zatavua_analalava?_CAS6	Meng et al. 2025a	fragmented
Phy***inae	G84	G84_Anopsicus_banksi_Gal56	this study	circular
Phy***inae	E031	E031_Arnapa_tinoor_Ind184	Meng et al. 2025a	fragmented
Phy***inae	G73	G73_Aucana_Chi80_Chi80	Meng et al. 2025b	circular
Phy***inae	G77	G77_Aucana_petorca_Chi113	Meng et al. 2025b	circular
Phy***inae	G70	G70_Aucana_platnicki_Chi64	Meng et al. 2025b	circular
Phy***inae	G71	G71_Aucana_ramirezi_Chi71	Meng et al. 2025b	circular
Phy***inae	E030	E030_Changminia_huangdi_Ind141	Meng et al. 2025a	fragmented
Phy***inae	E042	E042_Chisosa_calapa_Mex351	Meng et al. 2025a	fragmented
Phy***inae	E064	E064_Chisosa_caquetio_Ven18-179	Meng et al. 2025a	fragmented
Phy***inae	E061	E061_Chisosa_diluta_USA26	Meng et al. 2025a	fragmented
Phy***inae	E056	E056_Holocneminus_Phi105_Phi245	Meng et al. 2025a	fragmented
Phy***inae	E041	E041_Physocyclus_modestus_Mex349	Meng et al. 2025a	fragmented
Phy***inae	E081	E081_Wugigarra_undandi_QMB4	Meng et al. 2025a	fragmented
Smeringopinae	E058	E058_Cenemus_culiculus_Sey22	Meng et al. 2025a	fragmented
Smeringopinae	E092	E092_Crossopriza_lyoni_Gui90	Meng et al. 2025a	fragmented
Smeringopinae	E055	E055_Crossopriza_sahtan_Om32	Meng et al. 2025a	fragmented
Smeringopinae	E091	E091_Holocnemus_hispanicus_Rib34	Meng et al. 2025a	fragmented
Smeringopinae	E088	E088_Holocnemus_pluchei_Isr53	Meng et al. 2025a	fragmented
Smeringopinae	E089	E089_Holocnemus_reini_Mor72	Meng et al. 2025a	fragmented
Smeringopinae	E026	E026_Hoplopholcus_minous_Cre8	Meng et al. 2025a	fragmented
Smeringopinae	G92	G92_Maghreba_djabalija_Mor73	this study	circular
Smeringopinae	E090	E090_Maghreba_stifadma_Mor71	Meng et al. 2025a	fragmented
Smeringopinae	E028	E028_Smeringopina_bineti_Gui101	Meng et al. 2025a	fragmented
Smeringopinae	E051	E051_Smeringopus_lineiventris_Om122	Meng et al. 2025a	fragmented
Smeringopinae	E086	E086_Stygopholcus_photophilus_Cre19	Meng et al. 2025a	fragmented
Outgroup for UCE phylogeny	SRR10291841	NCBI_Scytoididae_Scytodes_sp.	Kulkarni et al. 2020	nan
Outgroup for UCE phylogeny	GCA_001188405.1	NCBI_Sicariidae_Loxosceles_reclusa	i5k Consortium 2013	nan
Outgroup for UCE phylogeny	SRR10291854	NCBI_Orsolobidae_Maoriata_sp.	Kulkarni et al. 2020	nan
Outgroup for UCE phylogeny	SRR10291860	NCBI_Hypochilidae_Hypochilus_gertschi	Kulkarni et al. 2020	nan
Modisiminae	NC_040859.1	NC040859_Mesabolivar_sp._ITV103611	Oliveira et al. 2019	circular
Modisiminae	NC_040861.1	NC040861_Mesabolivar_sp._ITV103613	Oliveira et al. 2019	circular
Modisiminae	MH643813.1	MH643813_Mesabolivar_sp._ITV103612	Oliveira et al. 2019	circular
Pholcinae	NC_020324.1	NC020324_Pholcus_phalangioides	Podsiadlowski et al. 2012	circular
Pholcinae	KJ782458.1	KJ782458_Pholcus_sp._HCP-2014	Pan et al. 2016	circular
Outgroup for mitochondrial phylogeny	NC_042902.1	Sicariidae_Loxosceles_similis_NC042902	Kalapothakis et al. 2019	circular
Outgroup for mitochondrial phylogeny	NC_042829.1	Salticidae_Epeus_alboguttatus_NC042829	Yang et al. 2019	circular
Outgroup for mitochondrial phylogeny	NC_081602.1	Sparassidae_Heteropoda_venatoria_NC081602	Jin 2022	circular
Outgroup for mitochondrial phylogeny	NC_059873.1	Atypidae_Atypus_karschi_NC059873	He et al. 2020	circular
Outgroup for mitochondrial phylogeny	NC_020322.1	Dipluridae_Phixioschema_suthepium_NC020322	Podsiadlowski et al. 2012	circular

Reference

- Guanliang Meng, Lars Podsiadlowski, Dimitar Dimitrov and Bernhard A. Huber (2025a), The Complex Interplay between Evolutionary Flexibility and Diversification in a Family of Spiders, submitted (currently under review by Systematic Biology)
- Guanliang Meng, Lars Podsiadlowski and Bernhard A. Huber (2025b), Low coverage genome sequencing reveals a new subfamily of daddy long-legs spiders from Brazilian Caatinga (Araneae, Pholcidae), in preparation
- i5K Consortium. 2013. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered.* 104:595–600.
- Kulkarni S , Wood H , Lloyd M , Hormiga G. 2020. Spider-specific probe set for ultraconserved elements offers new perspectives on the evolutionary history of spiders (Arachnida, Araneae). *Mol Ecol Resour.* 20(1):185–203.
- Yang, D., Yan, X., Xu, K., Yang, W., & Li, C. (2019). The complete mitochondrial genome of *Epeus alboguttatus* (Araneae: Salticidae). *Mitochondrial DNA Part B*, 4(1), 316-317.
- Kalapothakis, Y., Miranda, K. G., Pereira, A. H., Facchin, S., Lucio, N., & Kalapothakis, E. (2019). The brown spider *Loxosceles similis* (Araneae: Sicariidae): complete mitochondrial genome sequence. *Mitochondrial DNA Part B*, 4(1), 1518-1519.
- Pan, W. J., Fang, H. Y., Zhang, P., & Pan, H. C. (2016). The complete mitochondrial genome of long-legs spider *Pholcus* sp.(Araneae: Pholcidae). *Mitochondrial DNA Part A*, 27(2), 1054-1055.
- Podsiadlowski, L., Arabi, J. and Fahrrein, K. (2012) Mitochondrial genomes of three spider species.
- Oliveira, R. R. M., Vasconcelos, S., Pires, E. S., Pietrobon, T., Prous, X., & Oliveira, G. (2019). Complete mitochondrial genomes of three troglophile cave spiders (Mesabolivar, pholcidae). *Mitochondrial DNA Part B*, 4(1), 251-252.
- Jin, X.-Y. 2022. The complete mitochondrial genome of the huntsman spider *Heteropodavenatoria* (Linnaeus, 1767) (Arachnida: Araneae: Sparassidae) and its phylogenetic implication
- He, A. A., Guo, J. J., Huang, Z. Z., Liu, J. J. and Xu, X. X. 2020

Table D.2. Tests of symmetry of matrix 1. IQ-TREE (v2.2.0) (Minh et al. 2020) was used to test the assumptions of stationarity (constant nucleotide/amino-acid frequencies over time) and homogeneity (constant substitution rates over time). The two assumptions are rejected when p-values (SymPval) < 0.01 are found. **Locus**: Partition name; **SymSig**: Number of significant sequence pairs by test of symmetry; **SymNon**: Number of non-significant sequence pairs by test of symmetry; **SymPval**: P-value for maximum test of symmetry; **MarSig**: Number of significant sequence pairs by test of marginal symmetry; **MarNon**: Number of non-significant sequence pairs by test of marginal symmetry; **MarPval**: P-value for maximum test of marginal symmetry; **IntSig**: Number of significant sequence pairs by test of internal symmetry; **IntNon**: Number of non-significant sequence pairs by test of internal symmetry; **IntPval**: P-value for maximum test of internal symmetry

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-1017.fasta	892	8558	0.0594053	1644	7783	0.0595491	217	9162	0.195677
uce-1018.fasta	256	4303	0.313995	373	4167	0.523371	241	4231	0.184608
uce-10291.fasta	144	7114	0.494986	240	6985	0.811366	321	6866	0.218556
uce-10293.fasta	559	10612	0.00142476	724	10364	0.00192719	317	10621	0.0808371
uce-10305.fasta	252	2906	0.51843	403	2733	0.274868	116	2846	0.724072
uce-10306.fasta	1229	9066	0.0166177	1606	8455	0.0120133	445	9186	0.206476
uce-10308.fasta	630	8960	0.63477	1070	8481	0.665079	240	9169	0.434266
uce-10311.fasta	433	3307	1.22e-05	700	3038	4.53e-07	93	3642	0.947513
uce-10316.fasta	1111	4454	0.0489161	1353	4201	0.0393378	320	5194	0.230519

Continued on next page

Appendix D Supplementary information for Chapter 6

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-10324.fasta	219	5767	0.512829	351	5546	0.752463	120	5543	0.256914
uce-10337.fasta	383	6748	0.200795	502	6596	0.109831	328	6699	0.473723
uce-105.fasta	333	6802	0.0807599	746	6323	0.0277833	94	6805	0.543795
uce-1051.fasta	1161	2079	1.42e-05	1295	1943	7.42e-07	269	2958	0.793485
uce-106.fasta	168	5288	0.101554	283	5085	0.0731125	173	5006	0.303324
uce-110.fasta	234	8542	0.189191	600	8153	0.0345902	150	8568	0.991897
uce-111.fasta	3	5230	0.677972	0	893	0.730301	3	185	0.396973
uce-1114.fasta	587	9280	0.210286	957	8751	0.151842	368	9131	0.37488
uce-1119.fasta	63	3096	0.0656797	182	2965	0.0117557	66	3047	0.83901
uce-113.fasta	214	3521	0.00280933	230	3359	0.00122971	50	3329	0.247067
uce-115.fasta	157	7713	0.258652	432	7371	0.232505	115	7485	0.327828
uce-1195.fasta	555	5885	0.000430841	678	5737	0.000161683	224	6126	0.225906
uce-121.fasta	118	3368	0.0405704	186	3274	0.317564	78	3366	0.0219395
uce-1245.fasta	21	3296	0.397549	61	3080	0.198548	61	2752	0.665162
uce-1246.fasta	148	3173	0.0378629	257	2976	0.014489	143	2884	0.422633
uce-1249.fasta	186	3054	0.177369	372	2846	0.250069	62	3086	0.185027
uce-129.fasta	89	4011	0.0782967	31	1424	0.232587	84	792	0.0699314
uce-1337.fasta	238	4414	0.442117	478	4143	0.746907	110	4395	0.202867
uce-1373.fasta	4095	8308	1.55e-06	5166	7228	1.82e-07	309	12054	0.375133
uce-1387.fasta	351	2730	0.00462514	465	2609	0.0348737	111	2932	0.0175287
uce-1457.fasta	471	6549	0.220102	709	6230	0.156067	125	6608	0.387174
uce-149.fasta	181	5703	0.933382	279	5573	0.677109	220	5558	0.955712
uce-1556.fasta	767	2314	0.604378	1128	1935	0.370983	88	2944	0.705478
uce-1582.fasta	322	7669	0.11359	543	6357	0.168438	165	5787	0.155884
uce-160.fasta	328	6673	0.674707	469	6170	0.750481	169	5469	0.289315
uce-161.fasta	183	5812	0.756175	338	5641	0.823602	113	5771	0.475224
uce-1647.fasta	750	5920	0.0177467	1126	5538	0.00551425	232	6406	0.437593
uce-1656.fasta	397	9187	0.435484	940	8432	0.516801	246	8132	0.306336
uce-1710.fasta	109	3042	0.774709	145	2754	0.690762	69	2097	0.614207
uce-1715.fasta	1122	4131	0.587844	1415	3837	0.249109	238	4987	0.908904
uce-1730.fasta	4	6133	0.70372	28	4440	0.502455	5	2932	0.694542
uce-1736.fasta	1068	8248	0.347541	1398	7855	0.180519	414	8702	0.607102
uce-1766.fasta	103	5353	0.133874	297	5041	0.0440626	72	5176	0.63881
uce-1779.fasta	26	3520	0.0768516	34	2863	0.106186	102	2387	0.152212
uce-1881.fasta	565	3994	0.114441	807	3714	0.0204348	119	4359	0.927278
uce-1920.fasta	359	3044	0.445963	534	2854	0.511536	87	3273	0.32143
uce-1947.fasta	711	3944	0.00505467	995	3642	0.00207162	153	4446	0.283889
uce-1956.fasta	350	7648	0.0591117	587	7314	0.0188774	330	7250	0.538435
uce-1959.fasta	696	5737	0.00406821	969	5342	0.00037415	191	5914	0.869284
uce-1964.fasta	676	3419	0.00011978	976	3088	2.23e-05	96	3884	0.361206
uce-2011.fasta	164	2901	7.94e-05	279	2763	4.01e-06	61	2922	0.89764
uce-2033.fasta	946	5158	0.248437	1261	4771	0.106201	193	5652	0.626421
uce-206.fasta	285	3628	0.330861	498	3357	0.0844142	85	3659	0.968084
uce-2074.fasta	326	3077	5.74e-07	439	2960	2e-08	112	3277	0.85306
uce-2110.fasta	571	3989	0.0140119	858	3682	0.0328334	138	4333	0.065666
uce-2170.fasta	603	2967	0.527844	815	2751	0.298054	136	3393	0.695145
uce-2181.fasta	941	9642	0.0823668	1517	9027	0.0517558	334	10144	0.325584
uce-2199.fasta	254	4596	0.0263318	509	4338	0.00999294	125	4704	0.396852
uce-2219.fasta	642	3186	0.794778	924	2884	0.528522	69	3627	0.826954
uce-2222.fasta	23	3134	0.670232	52	3046	0.387922	115	2855	0.795565
uce-2226.fasta	328	9956	0.00521378	757	9301	0.000369293	297	9034	0.994576
uce-2287.fasta	104	7254	0.412607	228	6936	0.175282	105	6846	0.766855
uce-2296.fasta	70	6144	0.314174	90	5954	0.257619	125	5192	0.385974
uce-231.fasta	52	11252	0.261464	106	10716	0.261464	145	9925	nan
uce-2313.fasta	155	3413	0.0704273	362	3184	0.0220326	47	3442	0.568821
uce-2368.fasta	49	3691	0.518681	98	3590	0.205787	91	3396	0.890911
uce-2451.fasta	541	4712	0.111328	857	4388	0.0183811	126	5100	0.958004
uce-258.fasta	35	8028	0.0981235	172	7354	0.164706	123	6617	0.132743
uce-2588.fasta	208	6216	0.372991	555	5386	0.294328	50	5064	0.430963
uce-2634.fasta	250	3320	0.376166	356	3201	0.524292	117	3410	0.240976

Continued on next page

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-2645.fasta	33	9114	0.685107	103	8140	0.517123	188	6549	0.990759
uce-2698.fasta	422	5035	0.0613714	495	4911	0.0168962	229	5026	0.61033
uce-2699.fasta	65	4395	0.391625	111	3988	nan	98	3552	nan
uce-2701.fasta	353	3388	0.0132384	458	3248	0.211573	125	3438	0.00892902
uce-2745.fasta	358	2723	0.05138	603	2464	0.0161808	43	2996	0.704688
uce-2764.fasta	240	5012	0.892842	359	4878	0.676767	104	5099	0.861332
uce-2836.fasta	66	3402	0.112513	221	2930	0.0771564	50	2812	0.325855
uce-2926.fasta	984	2844	0.745737	1291	2527	0.660346	116	3689	0.595351
uce-2948.fasta	2433	8007	0.778986	3241	7174	0.622583	274	10081	0.689495
uce-2980.fasta	71	4091	0.581393	137	3868	0.660531	118	3712	0.374097
uce-2983.fasta	237	3679	0.00190752	414	3477	0.000169868	65	3792	0.823662
uce-2998.fasta	121	7868	0.997701	374	7301	0.963709	69	6889	0.972451
uce-3086.fasta	132	8611	0.00528289	197	7812	0.00047802	144	7007	0.89954
uce-3089.fasta	1350	2745	0.0532834	1674	2421	0.0138253	142	3951	0.620441
uce-3162.fasta	1111	6755	0.19336	1494	6272	0.320516	245	7370	0.160288
uce-3284.fasta	151	3589	0.557113	276	3457	0.850927	144	3557	0.250562
uce-3288.fasta	1054	4406	0.578297	1494	3961	0.251684	156	5288	0.886688
uce-329.fasta	157	3246	0.0446478	239	3116	0.0142371	116	3051	0.508389
uce-3291.fasta	156	3246	0.539307	235	3132	0.375092	97	3184	0.587909
uce-3300.fasta	144	3014	0.287297	300	2722	0.197897	49	2873	0.563703
uce-3313.fasta	70	5451	0.14711	197	4593	0.04227	27	3809	0.725274
uce-3342.fasta	495	5175	0.000148801	787	4879	9.51e-06	174	5485	0.818155
uce-3382.fasta	532	2871	0.000309929	732	2669	7.93e-05	123	3263	0.303582
uce-3423.fasta	441	4709	0.013293	730	4293	0.00796748	68	4518	0.235552
uce-3430.fasta	1004	3001	0.135907	1349	2646	0.255626	111	3858	0.127804
uce-3442.fasta	294	3710	0.521858	551	3429	0.330884	103	3841	0.62594
uce-3687.fasta	166	4985	0.745221	251	4839	0.820988	129	4806	0.462384
uce-3764.fasta	1599	2317	0.0775441	1788	2125	0.0375251	165	3748	0.404283
uce-3800.fasta	166	4684	0.00423985	191	3904	0.000774636	44	3369	0.542059
uce-3802.fasta	1408	8183	3.59e-07	2159	7429	1.03e-08	243	9327	0.93066
uce-3875.fasta	251	3068	0.639822	468	2822	0.40458	68	3054	0.715974
uce-3912.fasta	322	6233	0.0660811	504	5980	0.0125522	215	6091	0.808791
uce-3930.fasta	925	2903	0.0124558	1105	2722	0.0055103	131	3687	0.305027
uce-3945.fasta	146	3255	0.112685	247	3102	0.159759	108	3157	0.162728
uce-3947.fasta	231	3760	0.264865	276	3537	0.15861	186	3281	0.481963
uce-4064.fasta	257	3061	0.0133644	433	2795	0.00392842	50	3046	0.437087
uce-4147.fasta	257	3145	0.426279	436	2955	0.136128	105	3260	0.933948
uce-4246.fasta	1339	2756	0.636576	1571	2522	0.462947	155	3924	0.630769
uce-4286.fasta	450	6562	4.48e-07	596	6353	1.43e-08	245	6401	0.88797
uce-4293.fasta	253	4400	0.912082	409	4227	0.980434	129	4438	0.593602
uce-4295.fasta	97	4653	0.288608	203	4454	0.141511	94	4289	0.591252
uce-4318.fasta	408	5696	0.269755	674	5425	0.085882	162	5910	0.803304
uce-4448.fasta	106	3053	0.385229	168	2943	0.204708	121	2865	0.623071
uce-4495.fasta	185	2975	0.138967	283	2858	0.0551752	140	2935	0.555661
uce-4543.fasta	255	4998	0.649676	551	4697	0.440795	43	5180	0.681578
uce-4577.fasta	300	9707	0.228102	457	9476	0.382265	271	9423	0.166204
uce-4625.fasta	379	2702	0.0518597	521	2554	0.023249	89	2969	0.479948
uce-4628.fasta	472	4087	0.00262146	663	3886	0.000366104	176	4340	0.6264
uce-4661.fasta	903	2418	0.134861	1116	2196	0.0452726	141	3143	0.630309
uce-4753.fasta	2504	7791	0.116091	3207	7078	0.119919	390	9874	0.223736
uce-4756.fasta	129	3523	0.0109943	281	3257	0.0227288	47	3357	0.071451
uce-4787.fasta	55	4585	0.563815	93	3909	0.461392	99	2780	0.518798
uce-4838.fasta	236	4912	9.49e-06	368	4762	5.01e-07	109	4885	0.768497
uce-4846.fasta	196	4363	0.0765453	411	4113	0.0214159	92	4291	0.632347
uce-4863.fasta	131	10993	0.0353582	85	10115	0.0389298	344	8824	0.16058
uce-4864.fasta	127	7232	0.581362	293	6620	0.289787	68	6288	0.810596
uce-4875.fasta	265	3388	0.256851	476	3150	0.0942742	83	3490	0.71352
uce-4877.fasta	147	9394	0.00419592	147	4279	0.00812798	16	1340	0.0661344
uce-4882.fasta	190	7146	0.25493	592	6189	0.127113	73	5946	0.556807
uce-4891.fasta	548	4105	0.991471	880	3737	0.954789	157	4389	0.920083

Continued on next page

Appendix D Supplementary information for Chapter 6

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-4900.fasta	551	6217	0.561025	1166	5275	0.380481	125	5885	0.616351
uce-5026.fasta	92	7282	0.0289349	270	6950	0.0483141	176	6530	0.103523
uce-508.fasta	506	5379	0.0623738	870	4962	0.0180664	156	5576	0.588686
uce-5094.fasta	100	3728	0.749471	199	3613	0.499651	103	3695	0.779306
uce-5126.fasta	9	6981	0.900416	29	6302	0.774393	136	5553	0.779757
uce-5132.fasta	22	6275	0.631665	0	2049	0.542649	61	991	0.534461
uce-5141.fasta	84	9230	0.192235	228	8830	0.17302	180	8378	0.295833
uce-519.fasta	1126	3339	0.345541	1409	3055	0.833576	215	4241	0.117903
uce-530.fasta	218	5129	0.00468093	311	4850	0.000769763	136	4656	0.595284
uce-5342.fasta	952	3704	0.0870062	1379	3276	0.623819	97	4551	0.0257287
uce-5379.fasta	367	8846	0.000968365	663	8197	0.000398895	106	7851	0.228085
uce-5380.fasta	533	4416	0.0295753	684	4254	0.0254665	186	4696	0.195323
uce-5381.fasta	559	4001	0.00416466	852	3671	0.000728483	85	4334	0.559357
uce-5394.fasta	1286	6216	0.0225419	1826	5666	0.00233128	285	7188	0.968841
uce-5617.fasta	182	8190	0.748372	440	7780	0.922795	209	7787	0.393987
uce-5636.fasta	623	2458	0.422894	833	2238	0.263028	85	2919	0.61962
uce-5646.fasta	807	5409	0.00115903	1056	5154	0.0015867	332	5854	0.0779552
uce-5664.fasta	301	7695	0.750409	644	7268	0.989242	134	7590	0.343378
uce-5731.fasta	1098	10530	0.0274801	1482	10097	0.125006	435	11064	0.0373908
uce-575.fasta	2	3600	0.85215	56	3060	0.719012	12	2389	0.729103
uce-5758.fasta	478	2762	0.394567	767	2470	0.465252	58	3171	0.2951
uce-5793.fasta	233	9909	0.000161671	258	9445	1.05e-05	252	8032	0.81573
uce-582.fasta	49	3948	0.286803	139	3660	0.237985	53	3198	0.368052
uce-583.fasta	799	2441	0.401672	1018	2210	0.476834	88	3106	0.295273
uce-5878.fasta	557	3271	3.24e-05	776	3047	6.61e-06	99	3692	0.297714
uce-5934.fasta	249	7873	0.0665854	540	7542	0.174046	188	7804	0.0775048
uce-5940.fasta	378	3808	0.503591	554	3617	0.404078	83	4051	0.493867
uce-6021.fasta	444	2792	0.0117527	673	2547	0.00327523	72	3069	0.44745
uce-6066.fasta	177	3737	0.0267746	183	3707	0.0881937	189	3650	0.0519075
uce-6111.fasta	72	3755	0.199683	148	3665	0.238428	108	3670	0.226951
uce-6115.fasta	307	7196	0.134658	387	7067	0.0359169	219	7071	0.748043
uce-6157.fasta	77	3076	0.0256912	143	2700	0.0113749	37	2386	0.346018
uce-6262.fasta	138	9441	0.0801032	663	8597	0.048238	81	8840	0.507801
uce-628.fasta	228	2931	0.0434318	490	2639	0.0142918	71	2968	0.492719
uce-6304.fasta	120	5515	0.0846463	245	4857	0.0332498	63	4417	0.493124
uce-6385.fasta	480	9390	0.70238	448	9405	0.601665	689	9103	0.583159
uce-6443.fasta	215	8546	0.892042	451	7785	0.700466	140	7475	0.83503
uce-6448.fasta	458	8702	0.0112758	613	8097	0.0550439	284	7803	0.0201368
uce-6482.fasta	464	5529	0.199888	688	5256	0.0645253	143	5519	0.97663
uce-6579.fasta	65	8878	0.0608785	133	7476	0.16967	17	6344	0.0712614
uce-6634.fasta	101	8723	0.069284	231	7885	0.205315	286	6598	0.0596429
uce-6705.fasta	166	4586	0.910665	200	4490	0.709392	189	4394	0.870284
uce-676.fasta	163	3487	0.507276	271	3304	0.269933	65	3410	0.713377
uce-6769.fasta	110	3545	0.0666521	180	3445	0.02891	118	3440	0.428755
uce-6777.fasta	1286	9739	0.166339	1784	9239	0.0411639	413	10604	0.829391
uce-694.fasta	211	8173	0.616162	501	7858	0.753576	176	8097	0.354478
uce-7010.fasta	584	3332	0.0149716	899	3013	0.00220248	79	3824	0.754928
uce-7037.fasta	130	4818	0.228219	220	4653	0.801177	146	4620	0.0676586
uce-7156.fasta	567	10372	0.666044	100	4867	0.562561	165	3260	0.566029
uce-7165.fasta	89	4524	0.540708	251	3994	0.186792	76	3805	0.974114
uce-7167.fasta	205	3363	0.113575	417	3119	0.139842	51	3349	0.187511
uce-7286.fasta	729	8723	0.200789	1355	8030	0.0730361	165	9154	0.663688
uce-7287.fasta	281	4569	8.83e-12	387	4427	4.44e-13	112	4646	0.406863
uce-73.fasta	81	3921	0.198753	170	3766	0.277459	68	3648	0.193326
uce-7352.fasta	926	4429	0.000626034	1493	3836	3.33e-05	137	5154	0.982219
uce-7393.fasta	314	2926	0.379129	448	2790	0.556771	92	3135	0.227835
uce-7414.fasta	134	10969	0.229902	605	7836	0.0769799	301	5392	0.736794
uce-7600.fasta	1641	3210	0.000127823	1917	2933	2.95e-05	199	4645	0.303865
uce-7620.fasta	350	3830	0.319382	719	3410	0.249908	91	3903	0.406244
uce-7690.fasta	231	5440	0.00878437	265	5387	0.0120792	332	5261	0.102105

Continued on next page

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-7758.fasta	1110	8746	0.0159932	1456	7810	0.00145633	57	7200	0.986251
uce-776.fasta	198	3987	0.376568	214	3906	0.729775	124	3867	0.132919
uce-7775.fasta	268	5948	0.345584	413	5794	0.221262	231	5950	0.505196
uce-7805.fasta	649	7864	0.000480343	968	7492	4.92e-05	135	8186	0.65733
uce-7838.fasta	288	4174	0.199519	505	3897	0.510434	61	4212	0.0998531
uce-788.fasta	219	6219	0.454564	301	3721	0.789912	28	1708	0.196852
uce-7884.fasta	203	3710	0.00200343	325	3519	0.00180831	114	3535	0.14273
uce-789.fasta	469	8702	0.869207	667	8393	0.585344	469	8374	0.906598
uce-8071.fasta	280	9581	0.367576	800	8750	0.40635	161	8681	0.306337
uce-8073.fasta	236	7762	0.125258	571	7280	0.0368232	145	7349	0.93966
uce-8098.fasta	245	3076	0.211951	322	2985	0.0981107	138	3150	0.55612
uce-8168.fasta	12	7253	0.813613	79	5979	0.836326	41	4102	0.550471
uce-8222.fasta	447	2874	0.0670198	689	2627	0.0254216	128	3162	0.480733
uce-8230.fasta	593	5182	0.41588	819	4932	0.297829	177	5480	0.517548
uce-8327.fasta	608	5720	0.00114014	961	5344	0.0371398	195	6000	0.00339284
uce-834.fasta	78	8674	0.32219	86	7226	0.237174	213	5838	0.431724
uce-838.fasta	483	5183	0.152767	806	4811	0.199839	88	5384	0.191273
uce-8404.fasta	175	5495	0.0742453	445	5180	0.0872762	121	5318	0.176758
uce-8467.fasta	539	9855	0.121612	843	8524	0.0202752	99	7084	0.966386
uce-850.fasta	1441	8007	0.00112574	2607	6775	0.000436704	167	8921	0.244749
uce-8640.fasta	506	5707	0.0161091	640	5478	0.00802898	133	5750	0.286863
uce-8695.fasta	427	2733	0.0100094	514	2636	0.000843761	126	2987	0.98014
uce-8734.fasta	627	3465	0.000292064	857	3189	9.83e-05	121	3750	0.238361
uce-8739.fasta	494	11901	0.597059	1033	11219	0.338894	135	11691	0.746118
uce-8744.fasta	500	2739	0.189927	683	2550	0.0811903	143	3075	0.573424
uce-8854.fasta	5	4589	0.667655	169	2272	0.439524	33	1599	0.714342
uce-8866.fasta	2077	3383	0.0634606	2435	3022	0.0220499	222	5216	0.5106
uce-8877.fasta	31	3791	0.43043	85	3624	0.201955	79	3234	0.725065
uce-8902.fasta	1933	4737	1.75e-08	2397	4267	5.24e-09	238	6412	0.127485
uce-8975.fasta	110	3889	0.226615	315	3521	0.0979337	62	3426	0.601762
uce-8976.fasta	1417	2238	0.134991	1716	1938	0.0776474	97	3553	0.401466
uce-8990.fasta	433	7562	1.35e-10	636	7299	4.88e-12	216	7601	0.583337
uce-9018.fasta	46	4865	0.103038	112	3975	0.0302815	51	2398	0.652328
uce-907.fasta	472	4275	4.14e-06	774	3781	1.86e-07	66	4375	0.810175
uce-9089.fasta	152	3333	0.252219	329	3094	0.080384	49	3310	0.785787
uce-9134.fasta	1948	8782	0.0573376	2788	7910	0.0329831	179	10344	0.323754
uce-9305.fasta	1443	7201	0.0223924	2241	6341	0.0298877	201	8283	0.12267
uce-9340.fasta	320	6348	0.266979	553	6077	0.0614726	263	6179	0.965327
uce-9371.fasta	953	3897	0.0308442	1412	3430	0.0115282	56	4755	0.414199
uce-9386.fasta	160	3828	0.346727	278	3622	0.637356	147	3674	0.169614
uce-9458.fasta	138	4419	0.701239	297	4122	0.320475	60	3994	0.95696
uce-9459.fasta	94	7527	0.780894	166	7400	0.948995	145	7276	0.413371
uce-9649.fasta	81	3747	0.760305	229	3511	0.831353	49	3239	0.475075
uce-9683.fasta	14	4714	0.564718	20	2525	nan	62	2013	nan
uce-9701.fasta	270	4778	0.630385	391	4638	0.328171	175	4718	0.825571
uce-9809.fasta	167	7212	0.0251681	226	7110	0.0445755	242	6931	0.0953012
uce-9831.fasta	111	6906	0.994794	197	6721	0.920233	194	6470	0.978897
uce-9889.fasta	379	4472	0.0621245	571	4260	0.0216162	118	4666	0.507618
uce-9956.fasta	1859	3812	6.86e-08	2260	3409	1.82e-09	222	5427	0.908352

Table D.3. Tests of symmetry of matrix 2. IQ-TREE (v2.2.0) (Minh et al. 2020) was used to test the assumptions of stationarity (constant nucleotide/amino-acid frequencies over time) and homogeneity (constant substitution rates over time). The two assumptions are rejected when p-values (SymPval) < 0.01 are found. **Locus:** Partition name; **SymSig:** Number of significant sequence pairs by test of symmetry; **SymNon:** Number of non-significant sequence pairs by test of symmetry; **SymPval:** P-value for maximum test of symmetry; **MarSig:** Number of significant sequence pairs by test of marginal symmetry; **MarNon:** Number of non-significant sequence pairs by test of marginal symmetry; **MarPval:** P-value for maximum test of marginal symmetry; **IntSig:** Number of significant sequence pairs by test of internal symmetry; **IntNon:** Number of non-significant sequence pairs by test of internal symmetry; **IntPval:** P-value for maximum test of internal symmetry

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-1017.fasta	892	8558	0.0594053	1644	7783	0.0595491	217	9162	0.195677
uce-10291.fasta	144	7114	0.494986	240	6985	0.811366	321	6866	0.218556
uce-10293.fasta	559	10612	0.00142476	724	10364	0.00192719	317	10621	0.0808371
uce-10306.fasta	1229	9066	0.0166177	1606	8455	0.0120133	445	9186	0.206476
uce-10308.fasta	630	8960	0.63477	1070	8481	0.665079	240	9169	0.434266
uce-10316.fasta	1111	4454	0.0489161	1353	4201	0.0393378	320	5194	0.230519
uce-10324.fasta	219	5767	0.512829	351	5546	0.752463	120	5543	0.256914
uce-10337.fasta	383	6748	0.200795	502	6596	0.109831	328	6699	0.473723
uce-105.fasta	333	6802	0.0807599	746	6323	0.0277833	94	6805	0.543795
uce-106.fasta	168	5288	0.101554	283	5085	0.0731125	173	5006	0.303324
uce-110.fasta	234	8542	0.189191	600	8153	0.0345902	150	8568	0.991897
uce-111.fasta	3	5230	0.677972	0	893	0.730301	3	185	0.396973
uce-1114.fasta	587	9280	0.210286	957	8751	0.151842	368	9131	0.37488
uce-115.fasta	157	7713	0.258652	432	7371	0.232505	115	7485	0.327828
uce-1195.fasta	555	5885	0.000430841	678	5737	0.000161683	224	6126	0.225906
uce-129.fasta	89	4011	0.0782967	31	1424	0.232587	84	792	0.0699314
uce-1373.fasta	4095	8308	1.55e-06	5166	7228	1.82e-07	309	12054	0.375133
uce-1457.fasta	471	6549	0.220102	709	6230	0.156067	125	6608	0.387174
uce-149.fasta	181	5703	0.933382	279	5573	0.677109	220	5558	0.955712
uce-1582.fasta	322	7669	0.11359	543	6357	0.168438	165	5787	0.155884
uce-160.fasta	328	6673	0.674707	469	6170	0.750481	169	5469	0.289315
uce-161.fasta	183	5812	0.756175	338	5641	0.823602	113	5771	0.475224
uce-1647.fasta	750	5920	0.0177467	1126	5538	0.00551425	232	6406	0.437593
uce-1656.fasta	397	9187	0.435484	940	8432	0.516801	246	8132	0.306336
uce-1715.fasta	1122	4131	0.587844	1415	3837	0.249109	238	4987	0.908904
uce-1730.fasta	4	6133	0.70372	28	4440	0.502455	5	2932	0.694542
uce-1736.fasta	1068	8248	0.347541	1398	7855	0.180519	414	8702	0.607102
uce-1766.fasta	103	5353	0.133874	297	5041	0.0440626	72	5176	0.63881
uce-1956.fasta	350	7648	0.000219828	587	7314	4.84e-05	330	7250	0.333172
uce-1959.fasta	696	5737	0.00174813	969	5342	0.000128143	191	5914	0.912723
uce-2033.fasta	946	5158	0.248437	1261	4771	0.106201	193	5652	0.626421
uce-2181.fasta	941	9642	0.0823668	1517	9027	0.0517558	334	10144	0.325584
uce-2199.fasta	254	4596	0.0263318	509	4338	0.00999294	125	4704	0.396852
uce-2226.fasta	328	9956	0.00521378	757	9301	0.000369293	297	9034	0.994576
uce-2287.fasta	104	7254	0.412607	228	6936	0.175282	105	6846	0.766855
uce-2296.fasta	70	6144	0.314174	90	5954	0.257619	125	5192	0.385974
uce-231.fasta	52	11252	0.261464	106	10716	0.261464	145	9925	nan
uce-2451.fasta	541	4712	0.111328	857	4388	0.0183811	126	5100	0.958004
uce-258.fasta	35	8028	0.0981235	172	7354	0.164706	123	6617	0.132743
uce-2588.fasta	208	6216	0.372991	555	5386	0.294328	50	5064	0.430963
uce-2645.fasta	33	9114	0.278252	103	8140	0.212119	188	6549	0.39413
uce-2698.fasta	422	5035	0.0613714	495	4911	0.0168962	229	5026	0.61033
uce-2764.fasta	240	5012	0.892842	359	4878	0.676767	104	5099	0.861332
uce-2948.fasta	2433	8007	0.778986	3241	7174	0.622583	274	10081	0.689495
uce-2998.fasta	121	7868	0.997701	374	7301	0.963709	69	6889	0.972451
uce-3086.fasta	132	8611	0.00528289	197	7812	0.00047802	144	7007	0.89954
uce-3162.fasta	1111	6755	0.19336	1494	6272	0.320516	245	7370	0.160288

Continued on next page

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-3288.fasta	1054	4406	0.578297	1494	3961	0.251684	156	5288	0.886688
uce-3313.fasta	70	5451	0.14711	197	4593	0.04227	27	3809	0.725274
uce-3342.fasta	495	5175	0.000148801	787	4879	9.51e-06	174	5485	0.818155
uce-3423.fasta	441	4709	0.0100776	730	4293	0.00190194	68	4518	0.595579
uce-3687.fasta	166	4985	0.745221	251	4839	0.820988	129	4806	0.462384
uce-3800.fasta	166	4684	0.00423985	191	3904	0.000774636	44	3369	0.542059
uce-3802.fasta	1408	8183	3.59e-07	2159	7429	1.03e-08	243	9327	0.93066
uce-3912.fasta	322	6233	0.0660811	504	5980	0.0125522	215	6091	0.808791
uce-4286.fasta	450	6562	4.48e-07	596	6353	1.43e-08	245	6401	0.88797
uce-4318.fasta	408	5696	0.269755	674	5425	0.085882	162	5910	0.803304
uce-4543.fasta	255	4998	0.649676	551	4697	0.440795	43	5180	0.681578
uce-4577.fasta	300	9707	0.228102	457	9476	0.382265	271	9423	0.166204
uce-4753.fasta	2504	7791	0.116091	3207	7078	0.119919	390	9874	0.223736
uce-4838.fasta	236	4912	9.49e-06	368	4762	5.01e-07	109	4885	0.768497
uce-4863.fasta	131	10993	0.0353582	85	10115	0.0389298	344	8824	0.16058
uce-4864.fasta	127	7232	0.581362	293	6620	0.289787	68	6288	0.810596
uce-4877.fasta	147	9394	0.00419592	147	4279	0.00812798	16	1340	0.0661344
uce-4882.fasta	190	7146	0.25493	592	6189	0.127113	73	5946	0.556807
uce-4900.fasta	551	6217	0.561025	1166	5275	0.380481	125	5885	0.616351
uce-5026.fasta	92	7282	0.0289349	270	6950	0.0483141	176	6530	0.103523
uce-508.fasta	506	5379	0.0623738	870	4962	0.0180664	156	5576	0.588686
uce-5126.fasta	9	6981	0.900416	29	6302	0.774393	136	5553	0.779757
uce-5132.fasta	22	6275	0.631665	0	2049	0.542649	61	991	0.534461
uce-5141.fasta	84	9230	0.192235	228	8830	0.17302	180	8378	0.295833
uce-530.fasta	218	5129	0.00468093	311	4850	0.000769763	136	4656	0.595284
uce-5379.fasta	367	8846	0.000968365	663	8197	0.000398895	106	7851	0.228085
uce-5380.fasta	533	4416	0.0295753	684	4254	0.0254665	186	4696	0.195323
uce-5394.fasta	1286	6216	0.0225419	1826	5666	0.00233128	285	7188	0.968841
uce-5617.fasta	182	8190	0.748372	440	7780	0.922795	209	7787	0.393987
uce-5646.fasta	807	5409	0.00115903	1056	5154	0.0015867	332	5854	0.0779552
uce-5664.fasta	301	7695	0.750409	644	7268	0.989242	134	7590	0.343378
uce-5731.fasta	1098	10530	0.165842	1482	10097	0.0660291	435	11064	0.583174
uce-5793.fasta	233	9909	0.000161671	258	9445	1.05e-05	252	8032	0.81573
uce-5934.fasta	249	7873	0.0665854	540	7542	0.174046	188	7804	0.0775048
uce-6115.fasta	307	7196	0.134658	387	7067	0.0359169	219	7071	0.748043
uce-6262.fasta	138	9441	0.0801032	663	8597	0.048238	81	8840	0.507801
uce-6304.fasta	120	5515	0.0846463	245	4857	0.0332498	63	4417	0.493124
uce-6385.fasta	480	9390	0.70238	448	9405	0.601665	689	9103	0.583159
uce-6443.fasta	215	8546	0.892042	451	7785	0.700466	140	7475	0.83503
uce-6448.fasta	458	8702	0.0112758	613	8097	0.0550439	284	7803	0.0201368
uce-6482.fasta	464	5529	0.199888	688	5256	0.0645253	143	5519	0.97663
uce-6579.fasta	65	8878	0.0608785	133	7476	0.16967	17	6344	0.0712614
uce-6634.fasta	101	8723	0.069284	231	7885	0.205315	286	6598	0.0596429
uce-6777.fasta	1286	9739	0.166339	1784	9239	0.0411639	413	10604	0.829391
uce-694.fasta	211	8173	0.616162	501	7858	0.753576	176	8097	0.354478
uce-7037.fasta	130	4818	0.228219	220	4653	0.801177	146	4620	0.0676586
uce-7156.fasta	567	10372	0.666044	100	4867	0.562561	165	3260	0.566029
uce-7286.fasta	729	8723	0.200789	1355	8030	0.0730361	165	9154	0.663688
uce-7287.fasta	281	4569	8.83e-12	387	4427	4.44e-13	112	4646	0.406863
uce-7352.fasta	926	4429	0.000626034	1493	3836	3.33e-05	137	5154	0.982219
uce-7414.fasta	134	10969	0.229902	605	7836	0.0769799	301	5392	0.736794
uce-7600.fasta	1641	3210	0.000127823	1917	2933	2.95e-05	199	4645	0.303865
uce-7690.fasta	231	5440	0.00878437	265	5387	0.0120792	332	5261	0.102105
uce-7758.fasta	1110	8746	0.0159932	1456	7810	0.00145633	57	7200	0.986251
uce-7775.fasta	268	5948	0.345584	413	5794	0.221262	231	5950	0.505196
uce-7805.fasta	649	7864	0.000480343	968	7492	4.92e-05	135	8186	0.65733
uce-788.fasta	219	6219	0.454564	301	3721	0.789912	28	1708	0.196852
uce-789.fasta	469	8702	0.869207	667	8393	0.585344	469	8374	0.906598
uce-8071.fasta	280	9581	0.98193	800	8750	0.834956	161	8681	0.972531
uce-8073.fasta	236	7762	0.125258	571	7280	0.0368232	145	7349	0.93966

Continued on next page

Appendix D Supplementary information for Chapter 6

Locus	SymSig	SymNon	SymPval	MarSig	MarNon	MarPval	IntSig	IntNon	IntPval
uce-8168.fasta	12	7253	0.813613	79	5979	0.836326	41	4102	0.550471
uce-8230.fasta	593	5182	0.41588	819	4932	0.297829	177	5480	0.517548
uce-8327.fasta	608	5720	0.00114014	961	5344	0.0371398	195	6000	0.00339284
uce-834.fasta	78	8674	0.32219	86	7226	0.237174	213	5838	0.431724
uce-838.fasta	483	5183	0.152767	806	4811	0.199839	88	5384	0.191273
uce-8404.fasta	175	5495	0.0742453	445	5180	0.0872762	121	5318	0.176758
uce-8467.fasta	539	9855	0.121612	843	8524	0.0202752	99	7084	0.966386
uce-850.fasta	1441	8007	0.00112574	2607	6775	0.000436704	167	8921	0.244749
uce-8640.fasta	506	5707	0.0161091	640	5478	0.00802898	133	5750	0.286863
uce-8739.fasta	494	11901	0.597059	1033	11219	0.338894	135	11691	0.746118
uce-8854.fasta	5	4589	0.667655	169	2272	0.439524	33	1599	0.714342
uce-8866.fasta	2077	3383	0.0634606	2435	3022	0.0220499	222	5216	0.5106
uce-8902.fasta	1933	4737	1.75e-08	2397	4267	5.24e-09	238	6412	0.127485
uce-8990.fasta	433	7562	1.35e-10	636	7299	4.88e-12	216	7601	0.583337
uce-9018.fasta	46	4865	0.103038	112	3975	0.0302815	51	2398	0.652328
uce-9134.fasta	1948	8782	0.0573376	2788	7910	0.0329831	179	10344	0.323754
uce-9305.fasta	1443	7201	0.0223924	2241	6341	0.0298877	201	8283	0.12267
uce-9340.fasta	320	6348	0.266979	553	6077	0.0614726	263	6179	0.965327
uce-9371.fasta	953	3897	0.0308442	1412	3430	0.0115282	56	4755	0.414199
uce-9459.fasta	94	7527	0.780894	166	7400	0.948995	145	7276	0.413371
uce-9683.fasta	14	4714	0.564718	20	2525	nan	62	2013	nan
uce-9701.fasta	270	4778	0.630385	391	4638	0.328171	175	4718	0.825571
uce-9809.fasta	167	7212	0.0251681	226	7110	0.0445755	242	6931	0.0953012
uce-9831.fasta	111	6906	0.994794	197	6721	0.920233	194	6470	0.978897
uce-9889.fasta	379	4472	0.0621245	571	4260	0.0216162	118	4666	0.507618
uce-9956.fasta	1859	3812	6.86e-08	2260	3409	1.82e-09	222	5427	0.908352

Table D.4. The details of GO analysis using g:Profile

source	term_name	term_id	highlighted	adjusted_p_value	negative_log10_of_adjusted_p_value	term_size	query_size	intersection_size	effective_domain_size
GO:MF	DNA-binding transcription factor activity	GO:0003700	False	0.012021353970416	1.920046614798792	313	62	9	10688
GO:MF	DNA-binding transcription factor activity, RNA polymerase II-specific	GO:0000981	True	0.0027691436230097	2.5576545187712107	142	62	7	10688
GO:MF	transcription regulator activity	GO:0140110	False	0.00724301176973	2.140080809080018	367	62	10	10688
GO:BP	biological regulation	GO:0065007	False	0.0022938874622039	2.639427892338285	1505	36	21	6343
GO:BP	biosynthetic process	GO:0009058	False	0.0048796230297968	2.3116137276699917	1718	36	22	6343
GO:BP	DNA-templated transcription	GO:0006351	True	1.90353759284e-05	4.720438541910691	624	36	16	6343
GO:BP	gene expression	GO:0010467	False	9.40921771404e-05	4.026446482481023	1251	36	21	6343
GO:BP	intracellular receptor signaling pathway	GO:0030522	True	0.0083748684468685	2.077022006171452	2	36	2	6343
GO:BP	macromolecule biosynthetic process	GO:0009059	False	0.0002947746724037	3.530509834611596	1335	36	21	6343
GO:BP	nucleic acid biosynthetic process	GO:0141187	False	7.79752371813e-05	4.108043295552433	892	36	18	6343
GO:BP	nucleobase-containing compound biosynthetic process	GO:0034654	False	0.0013404535775226	2.872748221871213	1077	36	18	6343
GO:BP	regulation of biological process	GO:0050789	False	0.0015967104730544	2.796773826099553	1473	36	21	6343
GO:BP	regulation of biosynthetic process	GO:0009889	False	0.0006305474575356	3.200282221112773	606	36	14	6343
GO:BP	regulation of cellular process	GO:0050794	False	0.0058169423767673	2.235305238346368	1453	36	20	6343
GO:BP	regulation of DNA-templated transcription	GO:0006355	False	0.0009314407002285	3.0308447890589627	534	36	13	6343
GO:BP	regulation of gene expression	GO:0010468	False	0.0006305474575356	3.200282221112773	606	36	14	6343
GO:BP	regulation of macromolecule biosynthetic process	GO:0010556	False	0.0006305474575356	3.200282221112773	606	36	14	6343

Continued on next page

source	term_name	term_id	highlighted	adjusted_p_value	negative_log10_of_adjusted_p_value	term_size	query_size	intersection_size	effective_domain_size
GO:BP	regulation of macromolecule metabolic process	GO:0060255	False	0.0011448840436181	2.941238497400018	637	36	14	6343
GO:BP	regulation of metabolic process	GO:0019222	False	0.0012563416274893	2.900892250121515	642	36	14	6343
GO:BP	regulation of nucleobase-containing compound metabolic process	GO:0019219	False	0.0016152236344114	2.7917673392976003	561	36	13	6343
GO:BP	regulation of primary metabolic process	GO:0080090	False	0.0033271723706182	2.4779246990803294	599	36	13	6343
GO:BP	regulation of RNA biosynthetic process	GO:2001141	False	0.0009314407002285	3.0308447890589627	534	36	13	6343
GO:BP	regulation of RNA metabolic process	GO:0051252	False	0.0013493003226498	2.8698913757888915	552	36	13	6343
GO:BP	RNA biosynthetic process	GO:0032774	False	6.78476547266e-05	4.16846515989619	884	36	18	6343
GO:BP	RNA metabolic process	GO:0016070	False	0.0003381712197299	3.470863356065074	982	36	18	6343

Table D.4 (continued)

term_id	intersections
GO:0003700	118183479,118185711,118187135,118187145,118187450,118188858,118189526,118192005,118197242
GO:0000981	118183479,118187145,118187450,118188858,118189526,118192005,118197242
GO:0140110	118183479,118184518,118185711,118187135,118187145,118187450,118188858,118189526,118192005,118197242
GO:0065007	118183479,118183553,118183596,118184275,118185711,118186302,118187135,118187145,118187450,118188858,118189412,118189526,118192005,118192900,118193513,118197242,118197831,118197841,118200526,118202262,118205910
GO:0009058	118179737,118180995,118183479,118185711,118186302,118187135,118187145,118187450,118188402,118188858,118189526,118190304,118192005,118192815,118192900,118194057,118197242,118197831,118197841,118203116,118205300,118205910
GO:0006351	118179737,118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118197242,118197831,118197841,118203116,118205300,118205910
GO:0010467	118179737,118183479,118185711,118186302,118187135,118187145,118187450,118188402,118188858,118189526,118190304,118192005,118192815,118192900,118194057,118197242,118197831,118197841,118203116,118205300,118205910
GO:0030522	118187145,118192900
GO:0009059	118179737,118183479,118185711,118186302,118187135,118187145,118187450,118188402,118188858,118189526,118190304,118192005,118192815,118192900,118194057,118197242,118197831,118197841,118203116,118205300,118205910
GO:0141187	118179737,118183479,118185711,118186302,118187135,118187145,118187450,118188402,118188858,118189526,118190304,118192005,118197242,118197831,118197841,118203116,118205300,118205910
GO:0034654	118179737,118183479,118185711,118186302,118187135,118187145,118187450,118188402,118188858,118189526,118190304,118192005,118197242,118197831,118197841,118203116,118205300,118205910
GO:0050789	118183479,118183553,118183596,118184275,118185711,118186302,118187135,118187145,118187450,118188858,118189412,118189526,118192005,118192900,118193513,118197242,118197831,118197841,118200526,118202262,118205910
GO:0009889	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118192900,118197242,118197831,118197841,118205910
GO:0050794	118183479,118183553,118183596,118184275,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118192900,118193513,118197242,118197831,118197841,118200526,118202262,118205910
GO:0006355	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118197242,118197831,118197841,118205910
GO:0010468	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118192900,118197242,118197831,118197841,118205910
GO:0010556	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118192900,118197242,118197831,118197841,118205910
GO:0060255	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118192900,118197242,118197831,118197841,118205910
GO:0019222	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118192900,118197242,118197831,118197841,118205910
GO:0019219	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118197242,118197831,118197841,118205910
GO:0080090	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118197242,118197831,118197841,118205910
GO:2001141	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118197242,118197831,118197841,118205910
GO:0051252	118183479,118185711,118186302,118187135,118187145,118187450,118188858,118189526,118192005,118197242,118197831,118197841,118205910
GO:0032774	118179737,118183479,118185711,118186302,118187135,118187145,118187450,118188402,118188858,118189526,118190304,118192005,118197242,118197831,118197841,118203116,118205300,118205910
GO:0016070	118179737,118183479,118185711,118186302,118187135,118187145,118187450,118188402,118188858,118189526,118190304,118192005,118197242,118197831,118197841,118203116,118205300,118205910

Table D.5. Intra-sbfamily genetic distance for each mitochondrial gene

Subfamily	Gene	Minimum	Maximum	Average	Median
Arteminae	12S	0.210058669	0.78037087	0.455059327	0.423964377
Cai***inae	12S	0.387618224	1.230665049	0.940629842	1.203606255
Modisiminae	12S	0.013161831	2.300837331	0.459488059	0.404099345
Ninetinae	12S	0.017367849	2.372960219	0.586031482	0.440013231
Pholcinae	12S	0.154147565	2.298103049	0.684518757	0.520880986
Phy***inae	12S	0.165032214	1.422287256	0.562988381	0.39247966
Smeringopinae	12S	0.25014433	2.138109391	0.805084927	0.841770333
Arteminae	16S	0.208059024	0.473045907	0.335524761	0.340444572
Cai***inae	16S	0.358406208	1.181519725	0.883640315	1.110995013
Modisiminae	16S	0.013642678	2.143941711	0.39004617	0.356417655
Ninetinae	16S	0.013581581	2.308172892	0.52046534	0.384775077
Pholcinae	16S	0.133547569	1.816646106	0.638087823	0.601817945
Phy***inae	16S	0.214350757	1.404176041	0.467345305	0.350406302
Smeringopinae	16S	0.192067713	1.140412758	0.488613972	0.391450046
Arteminae	ATP6	0.216978045	0.515372128	0.399801704	0.420034614
Cai***inae	ATP6	0.424501671	0.424501671	0.424501671	0.424501671
Modisiminae	ATP6	0.026431564	0.623124247	0.396080859	0.4018375
Ninetinae	ATP6	0.019910747	0.628752231	0.390429869	0.393838301
Pholcinae	ATP6	0.201481758	0.619344827	0.407020856	0.405792003
Phy***inae	ATP6	0.253333109	0.553749437	0.421504061	0.43701565
Smeringopinae	ATP6	0.133126148	0.538016371	0.357986153	0.338952345
Cai***inae	ATP8	0.022731187	0.741518818	0.456447346	0.605092032
Modisiminae	ATP8	0.013793978	1.478672872	0.716446512	0.706302173
Ninetinae	ATP8	0.016950776	1.161224547	0.684351344	0.694587721
Pholcinae	ATP8	0.381948481	0.951986828	0.692967772	0.744968006
Phy***inae	ATP8	0.485426946	0.954376851	0.817029221	0.897621902
Smeringopinae	ATP8	nan	nan	nan	nan
Arteminae	COX1	0.160336287	0.300582127	0.227741319	0.231245356
Cai***inae	COX1	0.216468848	0.216468848	0.216468848	0.216468848
Modisiminae	COX1	0.022072201	0.386272038	0.218484193	0.226260253
Ninetinae	COX1	0.014682418	0.549306144	0.22484444	0.223388186
Pholcinae	COX1	0.125503596	0.311998616	0.225627051	0.224402676
Phy***inae	COX1	0.121720125	0.406697486	0.249436008	0.244393435
Smeringopinae	COX1	0.150145382	0.261007479	0.204464589	0.203245556
Arteminae	COX2	0.17933598	0.473793158	0.34871494	0.377258089
Cai***inae	COX2	0.001537871	0.335813928	0.223609755	0.333477465
Modisiminae	COX2	0.018797676	0.514909628	0.333483468	0.350509515
Ninetinae	COX2	0.009247011	0.478565122	0.303935793	0.304989452
Pholcinae	COX2	0.118933459	0.541249858	0.356376761	0.353112648
Phy***inae	COX2	0.226034302	0.410536466	0.340291383	0.342454516
Smeringopinae	COX2	0.144945994	0.457754581	0.282337732	0.262336762
Arteminae	COX3	0.230500546	0.492409569	0.363303372	0.38301042
Cai***inae	COX3	0.339313947	0.339313947	0.339313947	0.339313947
Modisiminae	COX3	0.024940786	0.588986449	0.350297369	0.361208922
Ninetinae	COX3	0.010758445	1.026691771	0.326863663	0.328179316
Pholcinae	COX3	0.138720823	0.437679738	0.336431938	0.338897327
Phy***inae	COX3	0.246793617	0.464698178	0.358757184	0.360441966
Smeringopinae	COX3	0.110904555	1.756384579	0.296722413	0.27811161
Arteminae	CYTB	0.193478091	0.410711671	0.316318935	0.330388966
Cai***inae	CYTB	0.313134403	0.313134403	0.313134403	0.313134403
Modisiminae	CYTB	0.030765851	0.453261923	0.319796652	0.331776318
Ninetinae	CYTB	0.009029591	0.814867305	0.290373959	0.295773674
Pholcinae	CYTB	0.139397998	0.400454111	0.309538798	0.312423064
Phy***inae	CYTB	0.229779393	0.405519141	0.34225062	0.343247908
Smeringopinae	CYTB	0.202742788	0.420933507	0.290655719	0.290940675
Arteminae	ND1	0.208258728	0.447768507	0.337328734	0.360276414
Cai***inae	ND1	0.417190156	0.417190156	0.417190156	0.417190156

Continued on next page

Subfamily	Gene	Minimum	Maximum	Average	Median
Modisiminae	ND1	0.027241785	0.5595324	0.373264103	0.386148255
Ninetinae	ND1	0.020319743	0.517723762	0.384881577	0.390146363
Pholcinae	ND1	0.22179149	0.582431704	0.412532342	0.410221279
Phy***inae	ND1	0.235317211	0.498509914	0.416068742	0.42368464
Smeringopinae	ND1	0.162951574	0.552715959	0.402206525	0.395669036
Arteminae	ND2	0.265252375	0.617246496	0.464815112	0.49455408
Cai***inae	ND2	0.539798213	0.539798213	0.539798213	0.539798213
Modisiminae	ND2	0.023547854	0.754967769	0.503504463	0.529599071
Ninetinae	ND2	0.0	0.734851523	0.470839628	0.475931629
Pholcinae	ND2	0.244428519	0.674347492	0.520110187	0.516129903
Phy***inae	ND2	0.295055993	0.715546742	0.534237472	0.529862262
Smeringopinae	ND2	0.172413233	0.652360132	0.445281665	0.4422577
Arteminae	ND3	0.281167954	0.70609805	0.540993562	0.641723918
Cai***inae	ND3	0.481262696	0.481262696	0.481262696	0.481262696
Modisiminae	ND3	0.012586574	1.013069724	0.467825229	0.458310086
Ninetinae	ND3	0.004854407	1.631501674	0.548853144	0.529638638
Pholcinae	ND3	0.187048649	0.768125692	0.500074906	0.475713028
Phy***inae	ND3	0.263500057	0.74111687	0.56262996	0.572204647
Smeringopinae	ND3	0.268619474	1.124043739	0.495320375	0.471721925
Arteminae	ND4	0.277507521	0.506151716	0.410074303	0.442290193
Cai***inae	ND4	0.430965655	0.430965655	0.430965655	0.430965655
Modisiminae	ND4	0.029696506	0.707153499	0.4659523	0.489215387
Ninetinae	ND4	0.026719872	0.794968311	0.467503201	0.478995902
Pholcinae	ND4	0.250927512	0.594046075	0.463130569	0.469521287
Phy***inae	ND4	0.297934837	0.589303289	0.468401832	0.468956759
Smeringopinae	ND4	0.279666186	0.702887732	0.45765197	0.421621817
Arteminae	ND4L	0.271262033	0.601794616	0.413183233	0.409259578
Cai***inae	ND4L	0.007615498	0.365293412	0.246067441	0.365293412
Modisiminae	ND4L	0.017124962	0.811215209	0.470942112	0.479909266
Ninetinae	ND4L	0.014662933	0.889084051	0.467942108	0.467527208
Pholcinae	ND4L	0.172514136	0.656601553	0.42354919	0.421102253
Phy***inae	ND4L	0.28555279	0.696546383	0.496435513	0.502315421
Smeringopinae	ND4L	0.228070402	0.615678443	0.42658706	0.413744252
Arteminae	ND5	0.294695544	0.541529271	0.431798027	0.471987728
Cai***inae	ND5	0.439719439	0.439719439	0.439719439	0.439719439
Modisiminae	ND5	0.024258214	0.612982299	0.454179552	0.464514081
Ninetinae	ND5	0.024394008	0.829518252	0.43378026	0.438982457
Pholcinae	ND5	0.241801105	0.814103268	0.470078159	0.460318608
Phy***inae	ND5	0.31745171	0.735387475	0.50411917	0.498544734
Smeringopinae	ND5	0.23917507	0.614901827	0.442175538	0.442470512
Arteminae	ND6	0.220018144	0.597988733	0.414557683	0.442039976
Cai***inae	ND6	0.007281611	0.485717047	0.325939361	0.484819426
Modisiminae	ND6	0.031441192	0.786305984	0.53335093	0.559299667
Ninetinae	ND6	0.007142979	0.963863163	0.449192708	0.455153363
Pholcinae	ND6	0.155419431	0.764177051	0.489282895	0.493754311
Phy***inae	ND6	0.257592048	0.691533385	0.540721908	0.55347989
Smeringopinae	ND6	0.207752305	0.700595952	0.471300531	0.476431536

List of Figures

1.1	Spider tree of life	3
1.2	Spiders (Araneae) descriptions over the years	13
1.3	Pholcidae spider descriptions over the years	17
2.1	Procedures for phylogenomic analysis	34
3.1	Analytical workflow of target capture sequencing data	81
3.2	Maximum likelihood phylogenetic tree and divergence times of Pholcidae	85
3.3	Gene relative frequency and likelihood mapping	87
4.1	Speciation rate dynamics	104
4.2	Ancestral state reconstruction	105
4.3	Trait evolution dynamics	106
4.4	Evolutionary allometry of pholcids	108
4.5	Relationships between disparity, species richness, trait evolution, diversification and microhabitat shift	110
5.1	Maximum likelihood tree based on UCE data	133
5.2	Subfamily-level phylogeny and supporting evidence	134
5.3	Live specimens of <i>GENUS-1 sp1</i> Huber sp. n. and <i>GENUS-1 sp2</i> Huber sp. n.	135
5.4	Known geographic distribution of <i>GENUS-1</i> gen. n.	136
5.5	<i>GENUS-1 sp1</i> Huber sp. n. male palp; male from Gruta do Janelão; ZFMK Ar 24704.	137
5.6	<i>GENUS-1 sp1</i> Huber sp. n. – male procursus and genital bulb; male from Gruta do Janelão; CHNUFPI 3505	139
5.7	<i>GENUS-1 sp1</i> Huber sp. n. – male chelicerae and female genitalia; male and female from Gruta do Janelão; CHNUFPI 3505 (male), ZFMK Ar 24704 (female)	140
5.8	<i>GENUS-1 sp1</i> Huber sp. n. – female epigyna and genitalia; two females from Gruta do Janelão, ZFMK Ar 24704	142
5.9	<i>GENUS-1 sp1</i> Huber sp. n. – prosoma, chelicera, male frontal chelicerapophysis, female palp, male gonopore and epiandrous spigots; male and female from Gruta do Janelão, ZFMK Br22-166	143
5.10	<i>GENUS-1 sp1</i> Huber sp. n. – epigynum, spinnerets; male and female from Gruta do Janelão, ZFMK Br22-166	144
5.11	<i>GENUS-1 sp1</i> Huber sp. n. – palp and procursus ; male from Gruta do Janelão, ZFMK Br22-166	145

5.12	<i>GENUS-1 sp1</i> Huber sp. n. – metatarsus, female palpal tibia and male prolateral trichobothrium; male and female from Gruta do Janelão, ZFMK Br22-166	146
5.13	<i>GENUS-1 sp1</i> Huber sp. n. – short ‘vertical’ hair, chemoreceptor, cuticular plate, and tarsus; male from Gruta do Janelão, ZFMK Br22-166	147
5.14	<i>GENUS-1 sp1</i> Huber sp. n. – tarsal organ; male and female from Gruta do Janelão, ZFMK Br22-166	148
5.15	<i>GENUS-1 sp2</i> Huber sp. n. – male palp; male from W of Queimada Nova; ZFMK Ar 24706	149
5.16	<i>GENUS-1 sp2</i> Huber sp. n. – tarsus and procursus, and genital bulb; male from W of Queimada Nova; ZFMK Ar 24706	150
5.17	<i>GENUS-1 sp2</i> Huber sp. n. – male ocular area, chelicerae and female genitalia; male and female from W of Queimada Nova; ZFMK Ar 24706	151
5.18	<i>GENUS-1 sp2</i> Huber sp. n. – abdomen and genitalia; female from W of Queimada Nova; ZFMK Ar 24706	152
5.19	Type localities of newly described species	155
5.20	Distribution of male tibia 1 / male tibia 2 in 1662 species of Pholcidae	156
6.1	Maximum likelihood phylogeny of Pholcidae based on matrix 1 (part 1)	173
6.2	Maximum likelihood phylogeny of Pholcidae based on matrix 1 (part 2)	174
6.3	Maximum likelihood phylogeny of Pholcidae based on matrix 1 (part 3)	175
6.4	Introgression between Pholcidae subfamilies	176
6.5	Intra-subfamily genetic distance between mitochondrial genes of Pholcidae	178
6.6	The <i>Ka/Ks</i> distribution of Pholcidae mitogenomes	179
6.7	The GC content distribution of Pholcidae mitochondrial PCGs	180
6.8	The RSCU distribution of Pholcidae mitochondrial PCGs	181
6.9	The Enc plot for Pholcidae mitochondrial PCGs	182
6.10	The neutrality plot for Pholcidae mitochondrial PCGs	183
A.1	Missing data distribution of matrix 3	213
A.2	Density distribution of statistics of matrix 3	214
A.3	Likelihood mapping of matrix 3	215
A.4	Compositional homogeneity test of matrix 3	216
A.5	Gene-wise RCFV distribution based on BaCoCa analysis	217
A.6	Taxon-wise RCFV distribution based on BaCoCa analysis	218
A.7	Intra-locus recombination analysis based on data matrix 3	219
A.8	Congruence and incongruence between phylogenetic trees	220
A.9	Matrix-3 maximum likelihood tree from RAxML-NG partitioned analysis	221
A.10	Matrix-3 Bayesian tree from phylogbayes-MPI analysis	222
A.11	Matrix-3 coalescence tree from ASTRAL analysis	223
A.12	Matrix-3 coalescence tree from SVDquartets analysis	224
B.1	ClipKIT smart-gap tree	238
B.2	ClipKIT gappy tree	239
B.3	ClipKIT kpic-smart-gap tree	240
B.4	BAMM single best speciation rate shift configuration	241

B.5	Speciation rate dynamics for different groups	242
B.6	Extinction rate dynamics for different groups	243
B.7	Net-diversification rate dynamics for different groups	244
B.8	Phylogenetic signal of different traits	245
B.9	Ancestral states of $\log_2(\text{body size})$	246
B.10	Ancestral states of $\log_2(\text{leg length})$	247
B.11	Ancestral states of leg proportion	248
B.12	Phenogram of $\log_2(\text{body size})$	249
B.13	Phenogram of $\log_2(\text{leg length})$	250
B.14	Phenogram of leg proportion	251
B.15	Trait value density distribution	252
B.16	Trait evolution rate result	253
B.17	Trait distribution among microhabitats	254
B.18	Relationship between different traits	255
B.19	Relationships between disparity, species richness and speciation rate	256
C.1	Statistics of data matrix 1	279
C.2	Missing data distribution of matrix 1	280
C.3	Likelihood mapping analysis of matrix 1	281
C.4	Intralocus recombination analysis with 3SEQ based on matrix 1	282
C.5	Known geographic distribution of <i>GENUS-1</i> gen. nov.	283
D.1	Workflow for analysis with UCE data	296
D.2	Likelihood mapping analysis of matrix 1	297
D.3	Density plots of alignment characteristics of matrix 1	298
D.4	Missing data distribution of matrix 1	299
D.5	Missing data distribution of matrix 2	300
D.6	Intra-locus recombination analysis of matrix 1-2	301
D.7	Summary of phylogenetic trees based on UCE data	302
D.8	SVDquartets tree based on matrix 1	303
D.9	Unweighted ASTRAL-IV tree based on matrix 2	304
D.10	IQ-TREE tree based on matrix 4	305
D.11	Asteroid tree based on matrix 1	306
D.12	Mitochondrial phylogeny	307
D.13	Summary tree of mitochondrial phylogenies	308
D.14	f_{dM} distribution along untrimmed UCE loci	309
D.15	Likelihood mapping of introgressed and non-introgressed loci.	310
D.16	D statistics showing widespread intra-subfamily introgression.	311
D.17	The f_4 -ratio statistics showing widespread intra-subfamily introgression.	312
D.18	An example of Pholcidae mitochondrial genomes	313
D.19	Inter-subfamily genetic distances	314
D.20	GC content distribution of Pholcidae mitochondrial PCGs	315
D.21	The number of subfamilies sharing the same codons for each PCG	316

List of Tables

A.1	Samples for target capture sequencing and outgroups	225
A.2	Fossil records used for divergence times estimation	228
A.3	Saturation test by PhyloMad based on matrix 3	229
A.4	Tests of symmetry of matrix 3	233
A.5	Topology test	236
B.1	Sequences used to create the six-gene matrices	257
B.2	Pairwise comparisons of rates across clades using Kruskal-Wallis H-test	274
B.3	Akaike weights Table for model fitting of traits	275
B.4	Result for the ES-sim analysis	276
B.5	Model fitting for the QuaSSE analysis	277
C.1	Sample information for UCE-based phylogenetic analysis	284
C.2	Data matrix (1–5) statistics	286
C.3	Tests of symmetry of matrix 1	286
C.4	PhyloMAd substitutional saturation analysis based on matrix 1	289
C.5	Species excluded from the comparative analysis of leg proportions	293
D.1	Sample information	317
D.2	Tests of symmetry of matrix 1	321
D.3	Tests of symmetry of matrix 2	326
D.4	The details of GO analysis using g:Profile	329
D.5	Intra-sbufamily genetic distance for each mitochondrial gene	332