Cumulative dissertation submitted in partial fulfillment of the requirements for the doctoral degree
(Dr. rer. Nat.)

Contributions towards a better understanding of millipede phylogenetics and sawfly genomics

by

Jan Philip Oeyen

from

Lenvik, Norway

Bonn, September 2020

Faculty of Mathematics and Natural Sciences,
Rheinische Friedrich-Wilhelms-Universität, Bonn

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

Angefertigt am Zoologischen Forschungsmuseum Alexander Koenig, Bonn.

Erstgutachter:	Prof. Dr. Bernhard Misof
Zweitgutachter:	Prof. Dr. Oliver Niehuis
Fachnahes Kommissionsmitglied:	Prof. Dr. Lukas Schreiber
Fachfremdes Kommissionsmitglied:	Prof. Dr. Thomas Döring

Tag der Promotion: 17.02.2021
Erscheinungsjahr: 2021

Summary

A loss of biodiversity on a scale and rate comparable to the great extinction events, as a result of anthropogenic activity, has been documented for plants, animals, and micro-organisms. In order to foster a sustainable future, it is imperative that we have a well founded understanding and knowledge of biodiversity as well as the factors that generate and sustain it. This thesis presents research from a range of disciplines all aimed at furthering our understanding of arthropod biodiversity with a focus on Diplopoda (millipedes) and Hymenoptera (sawflies, wasps, ants, and bees).

After providing a general introduction on biodiversity research and the current state of knowledge on millipedes and hymenopterans in Chapter 1, the four following chapters (Chapter 2 – Chapter 5) detail research on the taxonomy and systematics of a small but understudied order of volvating millipedes, the Glomerida. Within these chapters, I utilize a combination of scanning electron microscopy based morphology, DNA barcoding, mitochondrial genomics, and phylogenomic analyses of transcriptome based datasets to investigate the intraordinal relationships of the Glomerida as well as the relationships between the Glomerida and its two potential sister-groups: the Sphaerotheriida and Glomeridesmida. In Chapter 6, I take advantage of the well resolved relationships of the major hymenopteran groups and our knowledge of the timing of their divergence to directly study the evolutionary mechanisms that enabled the diversification of the group. For this, I, along with a large international team of researchers, compare the genomes of a phytophagous and a parasitoid sawfly to those of wasps, ants, and bees (Apocrita). Based on these comparisons, we present insights into the composition of the ancestral hymenopteran genome, the dynamics of hymenopteran genome evolution, the transition from phytophagy to parasitoidism, and the factors that enabled the tremendous diversification of the Apocrita.

Chapter 2 *Eupeyerimhoffia archimedis* (Strasser, 1965) is redescribed based on several specimens collected at a number of sites close to the type locality on Sicily, Italy. Scanning electron microscopy is used to illustrate several unusual morphological characters for a member of the Glomerida for the first time. A fragment of the mitochondrial COI gene (668bp) is sequenced for the first time in *Eupeyerimhoffia* to provide a species-specific barcode and to gain first insights into the genetic distances between the genera in the widespread family Protoglomeridae. The novel sequences are compared to representatives of all other genera of the family: *Protoglomeris vasconica* (Brölemann, 1897) from northern Spain, the dwarfed *Glomerellina laurae* Silvestri, 1908 from Italy and *Glomeroides primus* (Silvestri, 1929) from western North America. The addition of COI sequences from the two other families of the Glomerida renders the family Protoglomeridae paraphyletic with

Glomeroides primus being more closely related to *Glomeridella minima* (Latzel, 1884) than to the other genera in the family. The large genetic distances (13.2–16.8%) between *Eupeyerimhoffia* and the other genera in the order, as well as its unusual morphological characters, including unique morphological adaptations to roll into a ball, are probably an indication of the old age of the group.

Chapter 3 The pill millipedes of the order Glomerida are a moderately diverse group with a classical Holarctic distribution pattern. Their classification is based on a typological system utilizing mainly a single character complex, the male telopods. In order to infer the apomorphies of the Glomerida, to elucidate its position in the Pentazonia, and to test the monophyly of its families and subfamilies, we conduct the first phylogenetic analysis of the order. To provide additional characters, we comparatively analyze the mandible using scanning electron microscopy. The final character matrix consists of 69 characters (11 mandible characters) and incorporates 22 species from 20 of the 34 pill millipede genera, representing all families and subfamilies, except the monotypic Mauriesiinae. Two species from each of the two other Pentazonian orders Sphaerotheriida and Glomeridesmida, as well as two Spirobolida, are included as outgroup taxa. The Glomerida are recovered as monophyletic and are supported by five apomorphies. Within the Pentazonia, the Glomeridesmida are recovered as the sister group to the classical Oniscomorpha (Sphaerotheriida + Glomerida) with weak support. The analysis provides little resolution within the Glomerida, resulting in numerous polytomies. Further morphological characters and/or the addition of molecular analyses are needed to produce a robust phylogenetic classification of the Glomerida.

Chapter 4 We present the nucleotide sequence and a detailed annotation of the complete mitochondrial genome of the pill millipede *Glomeris marginata* (Villers, 1789) (Diplopoda, Glomerida). The mitochondrial genome is 16,514 bp long, has a strong AT bias, and contains 13 protein-coding genes, two ribosomal RNA genes, and 22 transfer RNA genes. We compare the gene arrangement of *G. marginata* to that of a species of Sphaerotheriida, the most closely related millipede group.

Chapter 5 The millipede order Glomerida (Myriapoda, Diplopda), placed within the infraclass Pentazonia, comprises over 300 species of medium sized millipedes that are able to roll up into a ball. The phylogenetic relationships of the order are unresolved. Currently, its species are classified in a typological system that is largely based around characters of the modified last male leg-pair, the telopods. However, morphological phylogenetic analyses have revealed the employed characters to be largely uninformative, failing to resolve the proposed groups, and it has been shown that at least one family (Protoglomeridae) is likely paraphyletic. In order to investigate the intraordinal relationships of Glomerida, we phylogenetically analyzed both amino acid and nucleotide datasets derived from transcriptomes of 12 Glomerida species (9

newly sequenced) and three outgroup millipedes. The sampled Glomerida species include representatives of all but one of the currently recognized sub-families. We analyzed the datasets using supermatrix based maximum likelihood, Bayesian inference, and gene-tree based multi-species coalsence approaches and compare the resulting topologies. The order Glomerida is recovered as monophyletic with strong support and we identify two strongly supported monophyletic units within the order. However, the position of three species could not be confidently recovered. Our results show that none of the currently recognized groups that were represented by more than one species in our dataset represent monophyletic units. Lastly, we discuss the implications of our results on the interpretations of the groups biogeography, as well as the evolution of dwarfism and aberrant morphology in Glomerida.

Chapter 6 The tremendous diversity of Hymenoptera is commonly attributed to the evolution of parasitoidism in the last common ancestor of parasitoid sawflies (Orussidae) and wasp-waisted Hymenoptera (Apocrita). However, Apocrita and Orussidae differ dramatically in their species richness, indicating that the diversification of Apocrita was promoted by additional traits. These traits have remained elusive due to a paucity of sawfly genome sequences, in particular those of parasitoid sawflies. Here we present comparative analyses of draft genomes of the primarily phytophagous sawfly *Athalia rosae* and the parasitoid sawfly *Orussus abietinus*. Our analyses revealed that the ancestral hymenopteran genome exhibited traits that were previously considered unique to eusocial Apocrita (*e.g.*, low transposable element content and activity) and a wider gene repertoire than previously thought (*e.g.*, genes for CO_2 detection). Moreover, we discovered that Apocrita evolved a significantly larger array of odorant receptors than sawflies, which could be relevant to the remarkable diversification of Apocrita by enabling efficient detection and reliable identification of hosts.

Contents

Chapter 1

General Introduction

Since its introduction by Lovejoy (1980), the term biological diversity, or biodiversity for short, has become a commonly used buzzword in research articles, popular science articles, and news media (Swingland 2001). The popularization of the term and related research comes, at least in-part, as a consequence of the rapid decline in species numbers observed across the tree of life (e.g. Barnosky et al. 2011, Ceballos et al. 2015, Newbold et al. 2016).

Despite the likely underestimation of biodiversity indexes (Martin et al. 2019), research has revealed a dramatic acceleration in the loss of biodiversity over the past centuries (Ceballos et al. 2015). This not only includes the extinction of entire species, but also a significant reduction in the number and size of populations, which will likely lead to an even further accelerated loss of biodiversity over time (Ceballos et al. 2017). For wild mammals, it has been estimated that over 80% of the biomass has been lost since the rise of human civilization (Bar-Or et al. 2018). Although these species are not extinct, a large part of their genomic diversity has been decimated, making them less likely to be able to adapt to future changes in their habitat (Thompson et al. 2019). Similar losses have also been reported for other groups of organisms, including plants (e.g. Humphreys et al. 2019), insects (e.g. Burkle et al. 2013, Réginer et al. 2015, Thomas 2016, Hallmann et al. 2017), and entire ecosystems (e.g. soil-biodiversity: Tsiafouli et al. 2015). The observed extent and estimated rate of biodiversity loss have lead authors to term it 'the sixth mass extinction event' (e.g. Ceballos et al. 2015) and 'the insect apocalypse' (Goulson 2019).

The severity of the situation is compounded by human societies dependence on nature, especially so for our food-supply. Around 75% of the main food crops grown for human consumption show an increased production with animal-pollination, which in large is provided by arthropods (Klein et al. 2007). This ecosystem-service alone is estimated to be worth over $153 billion per year (Gallai et al. 2009) and a recent report showed that at least half of the global gross domestic product (GDP), totaling about $44 trillion, is directly dependent on nature (World Economic Forum, 2020).

It is not yet known what the exact long-term consequences of this extinction event will be, however, it is known that the loss or gain of even individual species can have large impacts their community and ecosystem (Bairey et al. 2016). Furthermore, although it has not yet been conclusively shown, it is generally believed that the productiveness and resilience of an ecosystem decreases with the loss of its biodiversity (Cameron 2002).

In order to foster a sustainable protection of biodiversity and, thus, the future of the human population, it is necessary to understand the mechanisms that generate and maintain biodiversity. For this, species are the fundamental unit for all research aimed at understanding different aspects of biodiversity (i.e. taxonomy; Wheeler 1995, Claridge et al. 1997). Taxonomy is the science of identifying, describing, naming, and classifying species (de Queiroz and Gauthier 1994). Humans have named species of plants and animals since long before modern science and modern taxonomy has since continuously developed, both in terms of theory and methods. A full recapitulation of the history of taxonomy is outside of the scope of this introduction, but some notable

milestones should be mentioned: Many view Aristotle as the original father of taxonomy as he was the first to systematically describe biodiversity based on their traits (Lloyd 1961, Leroi 2014). However, the birth of modern taxonomy is largely attributed to the works of the Swedish physician, botanist, and zoologist Carl Linnaeus: the father of modern taxonomy (Calisher 2007). Linnaeus formally implemented the still used binomial naming system through his works *Species Plantaru*m (1753) and *Systema Naturae* (1758), which included species names that distinguished each species from others in the same genus.

After Linnaeus, one of the most notable advancement of our understanding of species, and thus also taxonomy, came with Darwin's book *On the Origin of Species* (1859), in which he presented his hypothesis on evolution and the concept of common decent. However, it was not until Willi Henning introduced his concept on phylogenetic systematics, or cladistics, (Henning 1950) that evolution came to play a central role in the field of taxonomy (de Queiroz and Gauthier 1992). In short, Henning stated that species could only be grouped by shared derived characters (synapomorphies) and that all taxa should include all descendants from one single ancestor (the rule of monophyly). This provided a framework that allowed objective testing of taxonomic and systematic hypotheses.

Next to taxonomy, knowledge about the relationships between species (i.e. phylogenetic systematics) is another prerequisite for understanding the processes that shaped and continue to shape the biodiversity we observe today (Cotterill 1995, Pellens and Grandcolas 2016). For example, researching speciation, the evolutionary

process that forms new species, is inseparable from phylogenetics (Barraclough and Nee 2001). Since Linnaeus, Darwin, and Henning, taxonomy and phylogenetic systematics have continued to see rapid development both in terms of theoretical framework and available methods (Wheeler 2004, Will et al. 2005, Lee and Palci 2015, Young and Gillung 2019).

Early phylogenetic analyses were largely restricted to the inclusion of macroscopic and later microscopic characters of adult individuals. Later, the employment of electron microscopy in the 1970s made the generation of submicroscopic data feasible for adults and development stages (Wanninger 2015). The increased availability of detailed morphological data, together with the developments of computer technology that allowed the simultaneous analysis of complex character matrixes, led to a global breakthrough for phylogenetic systematics (Richter and Wirkner 2014, Wanninger 2015). However, the most impactful development came with the utilization of DNA sequence information as characters for the identification of species (e.g. Hebert et al. 2003) and the inference of phylogenetic relationships (Young and Gillung 2019). Through the introduction of PCR and Sanger sequencing, and more recently 'Next-generation' high-throughput sequencing technologies, scientists are able to rapidly, and increasingly affordably, generate large data-sets (Young and Gillung 2019, Kapli et al. 2020). While most of the modern morphological matrixes comprise of only a few hundred characters (Blanke and Wesener 2014, Neumann et al. 2020, but see Legg et al. 2013), phylogenomic-scale matrixes include over 400,000 amino acid characters (e.g. Misof et al. 2014).

In addition to providing an overwhelming amount of data for phylogenetic inferences, the molecular revolution also enabled the comparison of whole genomes, providing ground-braking insights into genome biology, the evolutionary mechanisms and forces shaping biodiversity (Alföldi and Lindbald-Toh 2013), and the genomic basis of speciation (Seehausen et al. 2014, Wolf and Ellegren 2016, Campbell et al. 2018). However, despite these tremendous technological and methodological advancements, we are still far away from describing and truly understanding global biodiversity.

Currently, about 1.2 million species have been described globally, which is a long way from the 8.7 million species that are estimated to exist (Mora et al. 2012). The gap in our knowledge is so large that it is not believed that taxonomists can close it, leading to the coining of the term the 'taxonomic impediment' (Wheeler et al. 2004). This impediment exists in part due to the tedious and time consuming nature of the science of taxonomy, but also due to another extinction event: the extinction of taxonomists through cuts in funding as well as in training of the next generation (Wheeler 2004, Coleman 2015).

1.1 Millipedes (Diplopoda) – understudied ground-dwellers

One group of arthropods that is historically understudied and that is still only investigated by a very small number of scientists are the millipedes (Diplopoda). There are more than 12,000 described species of millipedes (Diplopoda), making it one of the highly diverse groups of arthropods (Sierwald and Bond 2007). The group is estimated to be around 467 million years old (Rodriguez et al. 2018), is distributed across six continents, and occur in almost all of Earths biomes (Sierwald and Bond

2007). As detritivores, millipedes are important within their ecosystems for the development of soil. Millipedes are among the key organisms for the decomposition of leaves, woody debris, and other detritus (Crawford 1992, Cárcamo et al. 2000, Coleman and Hendrix 2000). Their fecal pellets contribute to the structuring of the soil and also act as hotspots for microbial activity (Dangerfield and Milner 1996). Due to their important functions in soil ecosystems and the resilience of some species against environmental pollutants (Read et al. 1998, Grelle et al. 2000), millipedes are also viewed as highly valuable for the restoration of areas that have been negatively impacted by anthropogenic influence (Snyder and Hendrix 2008).

Despite the impressive diversity and ecological importance of millipedes, our knowledge of the group is lacking at all taxonomic and phylogenetic levels (Sierwald and Bond 2007). Some authors assert that millipede research is behind that of insects and crustaceans with over a hundred years (Blanke and Wesener 2014) and others have stated that the taxonomy of the group is still in its infancy (Brewer et al. 2012).

In terms of taxonomy, the estimated number of undescribed millipede species, ranging from 20,000 to 80,000 (Mora et al. 2011; Brewer et al. 2012), show that a substantial part of the groups' diversity has yet to be discovered. The large discrepancy between different estimates can be attributed to the incomplete data for millipedes and, at least the lower end, is likely to be an underestimate of the true diversity of the group (Brewer et al. 2012). This lack of information is especially severe for the Asian and African continents, which remain largely unexplored (Sierwald and Bond 2007, Bewer et al. 2012). However, even in Europe, authors have noted difficulties when

assessing the potential endangerment of species (Voigtländer et al. 2011), despite it being one of the most well researched geographical areas in terms of millipede taxonomy (Sierwald and Bond 2007). Amongst others, they reported a critical deficiency in current distribution records, knowledge about population sizes and trends, as well as the uncertain taxonomic status of several species (Voigtländer et al. 2011).

The lack of knowledge on millipede taxonomy and slow progress compared to other groups of arthropods can in-part be attributed to a historically low number of researchers focusing on the group, but also to the difficulties surrounding older literature (i.e. obtaining and understanding non-digitized books in foreign languages; Sierwald and Bond 2007, Brewer et al. 2012, Blanke and Wesener 2014). Millipede taxonomists have also been slow to adopt phylogenetic and integrative approaches, including new techniques such as scanning electron microscopy or DNA barcoding (Sierwald and Bond 2007, Brewer et al. 2012). Although there are notable exceptions (e.g. Bond and Sierwald 2002, Spelda et al. 2011, Wesener 2015, Decker 2016, Pimvichai et al. 2020, Short et al. 2020), integrative taxonomy is still far from becoming the standard approach to millipede taxonomy.

Similar to the status of millipede taxonomy, the higher level relationships of millipedes, both within and among orders, have long remained largely unresolved. Only four of the 16 currently recognized millipede orders are classified based on phylogenetic analyses: Polydesmida (Simonsen 1990), Julida (Enghoff 1991), Sphaerotheriida (Wesener and VandenSpiegel 2009, Wesener 2014), and Spirobolida

(Wesener et al. 2008, Pitz and Sierwald 2010). Moreover, only four phylogenetic analyses focusing on intraordinal relationships have been conducted based on morphological data and neither have been bale to produce a well resolved tree (Enghoff 1984, Sierwald et al. 2003, Sierwald and Bond 2007, Blanke and Wesener 2014). These studies are in large hampered by a lack of available morphological characters and poor taxon sampling (Sierwald and Bond 2007, Brewer and Bond 2013, Blanke and Wesener 2014).

Molecular studies including millipedes have so far mostly focused on the position of Myriapoda among other arthropod groups or on the relationships among the major groups within Myriapoda and only included a narrow taxon sampling for millipedes (Regier and Schultz 2001, Regier et al. 2005, Brewer and Bond 2013, Brewer et al. 2013, Rehm et al. 2014, Fernández et al. 2016, Fernandéz et al. 2018). Only three millipede orders have so far been comprehensively analyzed using molecular data: Spirobolida (Pitz and Sierwald 2010), Sphaerotheriida (Wesener et al. 2010), and Polyxenida (Short and Vahtera 2017). Similarly, only a few researchers have employed molecular data to phylogenetically analyze relationships within families or tribes (e.g. Marek and Bond 2006, Marek and Bond 2007, Enghoff et al. 2011, Enghoff et al. 2013, Pimvichai et al. 2014). The most recent transcriptome based study, however, expanded the available taxon sampling, including multiple species from within 14 of the 16 millipede orders (Rodriguez et al. 2018). This study confirmed the monophyly of all orders that were represented with more than one species and provided a large step towards a phylogenetic system of the millipede orders. However, many problems still remain, most notably the lack of autapomorphic

characters of the orders and phylogenetic systems for the classification of species within orders. Fewer than 50% of higher millipede taxa are currently defined by apomorphic characters and many orders are organized in non-phylogenetic classification systems (Sierwald and Bond 2007, Brewer et al. 2012).

1.2 Hymenoptera – a mega-diverse insect order

Another group of arthropods that is highly diverse is the insect order Hymenoptera. The Hymenoptera, with its 153,000 currently described and one million estimated species of sawflies, wasps, ants, and, bees, constitutes one of the four mega-diverse insect orders (Aguiar et al. 2013). The Hymenoptera are not only tremendously diverse in terms of species numbers, but especially so in terms of the numerous unique traits and highly specialized life-strategies that have evolved within the group. Hymenopterans are herbivores, predators, parasitoids, and pollinators, and as such play a central role in the majority of terrestrial ecosystems across the globe (Quicke 1997, Grimaldi and Engel 2005). This diversity is especially striking given the comparably short time in which it evolved (e.g. in comparison with millipedes), with hymenopterans only having started to diversify about 280 million years ago (Peters et al. 2017).

Due to their magnificent diversity, as well as their ecological and economical importance, hymenopterans have been the focus of a large body of research on their taxonomy, phylogenetic relationships, and ecology (Quicke 1997, Grimaldi and Engel 2005). Through large efforts in both morphological (e.g. Ronquist et al. 1999, Vilhelmsen 2001, Sharkey and Roy 2002, Schulmeister 2003) and molecular

phylogenetics (e.g. Branstetter et al. 2017, Peters et al. 2017, Sann et al. 2018) we now have well-founded knowledge on the relationships among most major groups within Hymenoptera, the approximate timing of their diversification, and thus also the life style transitions that accompanied their diversification.

In Hymenoptera, the life style transition that is the most well known and likely the most well researched, is the transition from solitary living to living in eusocial colonies with many hundreds or even hundreds of thousands of members. This transition has happened multiple times convergently within Hymenoptera: in wasps (Vespoidea), ants (Formicoidea), and bees (Anthophila) (Peters et al. 2017, Sann et al. 2018). In these eusocial colonies, members are divided into reproductive and non-reproductive castes, with non-reproductive adults of two or more overlapping generations cooperatively caring for the offspring that is not their own (Wilson and Hölldobler 2005). This complex form of social organization is though to have a range of advantages over solitary living. For example, eusocial colonies have an improved defense against predators and pathogens, as well as an increased efficiency in both foraging and the use of resources (Hölldobler and Wilson 1990). These, and likely further advantages, are thought to be the key factors that enabled the dominance of eusocial insects, in particular ants, bees, and termites, within terrestrial ecosystems (Hölldobler and Wilson 1990).

The diversity and evolutionary success of ants and other eusocial Hymenoptera have made them the focal point of a large proportion of research on Hymenoptera (Schultz 2000). It is therefore no surprise that they were among the first insects to have their

genome sequenced (Weinstock et al. 2006) and are now among the most well represented Hymenopterans in terms of genomic resources. As of 2018, there exist a total of 51 published hymenopteran genomes, out of these 19 and 14 were from ants and bees, respectively (Branstetter et al. 2018). These resources have provided exciting new insights into the molecular underpinnings of eusocial systems. For example, comparative genomic analyses have identified a super-gene in the red imported fire ant *Solenopsis invicta* that regulates their complex phenotypic dimorphism, deciding whether colonies have single or multiple reproductive females at once (Keller and Ross 1998, Wang et al. 2013, Stolle et al. 2019).

Although much of the interest and research on Hymenoptera has been centered around their eusocial species, these are not the most diverse in terms of species numbers. The majority of described, and likely also undescribed, Hymenoptera are in fact parasitoids (Mrinalini and Werren 2015). Parasitoidism is a life-style where the females lay their eggs on or within a single host, which the developing larva feeds upon and ultimately kills. The evolution of this life-style is generally considered the most important factor promoting the diversification of Hymenoptera as a whole (Mrinalini and Werren 2015, Peters et al. 2017), and is by some authors even considered a prerequisite for other key innovations, such as eusociality (Davis et al. 2010).

The origin of parasitoidism in Hymenoptera was debated for a number of years, as the relationships among the sawfly lineages could not be conclusively resolved (e.g. Dowton and Austin 1994, Vilhelmsen 1997, Schulmeister 2003, Davis et al. 2010,

Vilhelmsen et al. 2010, Heraty et al. 2011). Recent studies suggest that parasitoidism evolved only once in Hymenoptera, in the last common ancestor of the parasitic sawflies, (Orussidae) and the wasp-waisted wasps, ants, and bees (Apocrita) (Branstetter et al. 2017, Peters et al. 2017). Given the current evidence, the most likely scenario that led to the evolution of parasitoidism was the transition from an endophytophagous life-style to endophytic parasitoidism.

The ancestral Hymenopterans were endo- or ectophytophagous sawflies (Peters et al. 2017). Phytophagy is still retained in the earliest branching group within Hymenoptera, the Eusymphyta, which comprises of about 7,000 species of primarily ectophytophagous sawflies (Peters et al. 2017). The members of Eusymphyta are both valuable pollinators, as well as economically significant pest species (e.g. Sáringer 1974, Abe 1988, Barbir et al. 2019). The sister-group of Eusymphyta, the Unicalcarida, harbors all three groups of endophytic sawflies (Wood wasps: Siricoidea+Xiphydrioidea; Stem sawflies: Cephoidea; Parasitic sawflies: Orussidae) as a paraphyletic assemblage, with the parasitic sawflies, Orussidae, as the sister group to Apocrita, the wasp waisted wasps, ants, and bees (Peters et al. 2017).

The earliest branching members of the Unicalcarida are the ~250 species of wood wasps (Siricoidea+Xiphydrioidea; Aguiar et al. 2013, Peters et al. 2017), whose larvae develop in wood infected by fungi. Female wood wasps, upon oviposition, injects the tree with both a venom and a symbiotic fungus that she carries with her in specialized organs (Francke-Grosmann 1939, Parkin 1941, Kajimura 2000). The phytotoxic venom (Wang et al. 2016) damages the tree and enables the growth of the fungi,

which in turn acts as the sole food-source for the larva once it hatches (Spradbery and Kirk 1978). This highly effective life-strategy has led some species, in particular *Sirex noctilio*, to become a significant pest species (Wang et al. 2016).

Branching after the wood wasps, are the ~160 species of stem wasps (Cephoidea; Peters et al. 2017). Stem sawflies, as the name suggests, use the stems of plants, predominantly those of grasses, as hosts for their endophytophagous larva (Portman et al. 2018). This has led several species to become significant agricultural pests across the globe (e.g. Gol'berg 1986, Portman et al. 2018). One of the most well known is the wheat stem sawfly, *Cephus cinctus*, that annually causes up to $350 million dollars in damages to crops of wheat, barley, and rye (Criddle 1923, Farstad and Platt 1946, Cockrell et al. 2017) across the United States (Beres et al. 2011). Interestingly, the larvae of *C. cinctus* are also opportunistic cannibals that will prey on their conspecifics if two eggs are deposited in close proximity within the same plant (Beres et al. 2011). Due to the position of the Cephoidea as the sistergroup to Orussidae+Apocrita (Peters et al. 2017), it is tempting to hypothesize that this could be a potential earlier, or even independent, origin of parasitoidism. However, as the larvae of Cephoidea can not develop from feeding on a single conspecific larva, the presence of absence of eggs from conspecifics does not affect their oviposition (Buteler et al. 2009), and it is currently not known if cases of cannibalism translate into fitness benefits, they can not be considered parasitoids. Additionally, it is not well documented how wide spread opportunistic cannibalism is in other species of Cephoidea outside of *C. cinctus* and, therefore, if it is characteristic of the group as a whole or not. It is therefore the prevalent hypothesis that parasitoidism evolved only a

single time in Hymenoptera, namely in the last common ancestor of Orussidea and Apocrita (Peters et al. 2017).

The parasitic sawflies (Orussidae), also referred to as the parasitic wood wasps, that display an interesting mosaic of both advanced traits, some of which are shared with Apocrita, and plesiomorphic traits shared with other sawflies (i.e. the absence of a wasp-waist; Vilhelmsen 2003a). Orussids, similar to the hypothesized ancestor of Orussidae and Apocrita, are ectoparasitoids of wood-boring larvae in dead wood (Vilhelmsen 2003a). To detect their hosts, Orussids use unique approach called vibrational sounding (Vilhelmsen et al. 2001). For this, the female generates vibrations by tapping her antennae against the surface of the wood. The forelegs pick up the reflecting vibrations, that contain information on the presence and location of potential host larvae in the wood. The vibrations are then transmitted through the haemolymph to specialized organs that transduce the information into nerve impulses (Vilhelmsen 2001 et al.). If the female detects a potential host, she will use her ovipositor, that is two times the length of the body, to deposit her egg close to or on the host larva (Ahnlund and Ronquist 2002).

The hatching Orussid larva, shows a number of adaptations to the parasitoid lifestyle that are shared with Apocrita. These include simplification of the sensory apparatus (antennae and eyes), mouth-parts, as well as the legs (Rohwer and Cushman 1917, Parker 1935, Vilhelmsen 2003n). Another striking character of Orussids, which is shared with Apocrita, is the presence of elaborate mushroom bodies in the brain (Farris and Schulmeister 2010). These bodies are centers in the brain that are

important for complex cognitive tasks and were originally thought to be associated with the evolution of eusociality (Howse 1975, Farris and Schulmeister 2010). However, it is now clear they evolved much earlier, in the last common ancestor of Orussidae and Aporita, and are likely important for host-finding behavior, in particular associative and spacial learning, in parasitoids (Farris and Schulmeister 2010).

Despite the evolution of parasitoidism likely being the most important factor having promoted the diversification of Hymenoptera (Davis et al. 2010, Mrinalini and Werren 2015, Peters et al. 2017) and the Orussids' numerous adaptations to the parasitoid life-style, there are currently only 82 described species of parasitoid sawflies (Aguiar et al. 2013). It is therefore evident that the tremendous success and diversification of Apocrita, which currently comprises over 144,000 species (Aguiar et al. 2013), was likely promoted by the acquisition of additional traits. These traits, however, remain largely unknown, as the transition from phytophagy to parasitoidism in Hymenoptera is still poorly understood, especially on a genomic level.

1.3 Research aims

The research aims of the present thesis can be divided into two parts, both of which target different aspects of biodiversity and biodiversity evolution. In the first part (Chapters 2–5), I investigated one of the understudied orders of millipedes, the pill-millipedes (Glomerida). I aimed to provide novel knowledge on their morphological diversity, their phylogenetic relationships, and evolution using traditional morphological techniques, DNA barcoding, and genomic as well as transcriptomic

data gathered using using next-generation sequencing technology. In the second part (Chapter 6), I investigate the evolution of hymenopteran genomes. I aimed to make a contribution towards closing the gap in our knowledge on the early evolution of hymenopteran genomes by characterizing the genomes of the phytophagous sawfly *Athalia rosae* (Eusymphyta: Tenthredinidae) and the parasitic wood-wasp *Orussus abietinus* (Orussidea). I further aimed to comparatively analyze these two sawfly genomes along genome sequences of parasitoid and eusocial apocrita to provide first insights into the genomic changes related to the transition from phytophagy to parasitoidism and factors that lead to the tremendous diversification of Apocrita. Lastly, I aimed to provide insights into other aspects of the evolution of Hymenoptera genomes that have thus far only been studied within Apocrita, including their gene repertoire, conservation of synteny, and transposable element content.

Please note that the second part, the last chapter of this thesis, represents a collaborative effort of a large number of scientists from across the globe, whose primary results contributed to the analyses of the addressed research questions, which I coordinated. Besides coordinating and conducting my own research, I used these primary data for secondary analyses, and syntheses of the overall results into the final research article.

1.4 References

Abe M (1988) A biosystematic study of the genus *Athalia* Leach of Japan (Hymenoptera: Tenthredinidae). Esakia. 26: 91–131.

Aguiar AP, Deans AR, Engel MS, Forshage M, Huber JT, et al. (2013) Order hymenoptera. Zootaxa. 3703: 51–62.

Ahnlund H, Ronquist F (2002) Den röda parasitväxtstekelns (Orussus abietinus) biologi och förekomst i Norden. Entomologisk Tidsskrift. 122: 1–10.

Alföldi J, Lindblad-Toh K (2013) Comparative genomics as a tool to understand evolution and disease. Genome Research. 23: 1063–1068.

Barbir J, Aguado Martín LO, Rodriguez Lloveras X (2019) Impact of Climate Change on Sawfly (Suborder: Symphyta) Polinators in Andalusia Region, Spain. In: Leal Filho W, Barbir J, Preziosi R. (Eds.) Handbook of Climate Change and Biodiversity. Climate Change Management. Springer, Cham. 93–100 pp.

Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, et al. (2011) Has the Earth's sixth mass extinction already arrived? Nature. 471: 51–57.

Bar-On YM, Phillips R, Milo R (2018) The biomass distribution on Earth. Proceedings of the National Academy of Sciences of the U.S.A. 115: 6506–6511.

Barraclough TG, Nee S (2001) Phylogenetics and speciation. Trends in Ecology and Evolution. 16: 391–399.

Beres BL, Dosdall LM, Weaver DK, Carcamo HA, Spaner DM (2011) Biology and integrated management of wheat stem sawfly and the need for continuing research. Canadian Entomologist. 143: 105–25.

Blanke A, Wesener T (2014) Revival of forgotten characters and modern imaging techniques help to produce a robust phylogeny of the Diplopoda (Arthropoda, Myriapoda). Arthropod Structure and Development. 43: 63–75.

Bond JE, Sierwald P (2002) Cryptic speciation in the Anadenobolus excisus millipede species complex on the Island of Jamaica. Evolution 56: 1123–1135.

Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, Gates MW, Kula RR, Brady SG (2017) Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. Current Biology. 27: 1019–1025.

Brewer MS, Bond JE (2013) Ordinal-level phylogenomics of the arthropod class Diplopoda (millipedes) based on an analysis of 221 nuclear protein-coding loci generated using next-generation sequence analyses. PLoS One. 8: e79935.

Brewer MS, Sierwald P, Bond JE (2012) Millipede taxonomy after 250 years: classification and taxonomic practices in a mega-diverse yet understudied arthropod group. PLoS One. 7: e37240.

Brewer MS, Swafford L, Spruill CL, Bond JE (2013) Arthropod Phylogenetics in Light of Three Novel Millipede (Myriapoda: Diplopoda) Mitochondrial Genomes with Comments on the Appropriateness of Mitochondrial Genome Sequence Data for Inferring Deep Level Relationships. PLOS ONE. 8: e68005.

Burkle LA, Marlin JC, Knight TM (2013) Plant-Pollinator Interactions over 120 Years: Loss of Species, Co-Occurrence, and Function. Science. 339: 1611–1615.

Buteler M, Weaver DK, Peterson RKD (2009) Exploring the oviposition behavior of the wheat stem sawfly when encountering plants infested with cryptic conspecifics. Environmental Entomology. 38: 1707–1715.

Calisher CH (2007) Taxonomy: what's in a name? Doesn't a rose by any other name smell as sweet? Croatian medical journal. 48: 268–270.

Cameron T (2002) 2002: The year of the “Diversity - Ecosystem function” debate. Trends in Ecology and Evolution. 17: 495–496.

Campbell CR, Poelstra JW, Yoder AD (2018) What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biological Journal of the Linnean Society. 124: 561–583.

Cárcamo HA, Abe TA, Prescott CE, Holl FB, Chanway CP (2000) Influence of millipedes on litter decomposition, N mineralization, and microbial communities in a coastal forest in British Columbia, Canada. Canadian Journal of Forest Research. 30: 817–826.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM (2015) Accelerated modern human-induced species losses: Entering the sixth mass extinction. Science Advances. 1: 9–13.

Ceballos G, Ehrlich PR, Dirzo R (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proceedings of the National Academy of Sciences of the U.S.A. 114: E6089–E6096.

Claridge MF, Dawah HA, Wilson MR (1997) Species: the units of biodiversity. London, Chapman & Hall. 442 pp.

Cockrell DM, Griffin-Nolan RJ, Rand TA, Altilmisani N, Ode PJ, et al. (2017) Host plants of the wheat stem sawfly (Hymenoptera: Cephidae). Environmental Entomology. 46: 847–854.

Coleman CO (2015) Taxonomy in the times of the taxonomic impediment – examples from the community of experts on amphipod crustaceans. Journal of Crustacean Biology. 35: 729–749.

Coleman DC, Hendrix PF (2000) Invertebrates as webmasters in ecosystems. CABI Publishing. 336 pp.

Cotterill FPD (1995) Systematics, biological knowledge and environmental conservation. Biodiversity and Conservation. 4: 183–205.

Crawford CS (1992) Millipedes as model detritivores. Berichte des Naturwissenschaftlich-medizinischen Vereins Innsbruck. 10: 277–88.

Criddle N (1923) The life habits of *Cephus cinctus* Nort. in Manitoba. Cannadian Entomologist. 55: 1–4.

Dangerfield JM, Milner AE (1996) Millipede fecal pellet production in selected natural and managed habitats of southern Africa: implications for litter dynamics. Biotropica. 28: 113–120.

Darwin C (1859) On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life. London, J. Murray. 502 pp.

Davis RB, Baldauf SL, Mayhew PJ (2010) The origins of species richness in the Hymenoptera: Insights from a family-level supertree. BMC Evolutionary Biology. 10: 109.

Decker P (2016) Integrative taxonomic revision of the polymorphic flat-millipede genera *Oncocladosoma* and *Somethus* in South Australia (Diplopoda:Polydesmida:Paradoxosomatidae). Invertebrate Systematics. 30: 201–218.

de Queiroz K, Gauthier J (1992) Phylogenetic taxonomy*. Annual review of ecology and systematics. 23: 449–480.

de Querioz K, Gauthier J (1994) Toward a phylogenetic system of biological nomenclature. Trends in ecology and evolution. 9: 27–31.

Dowton M, Austin AD (1994) Molecular phylogeny of the insect order Hymenoptera: Apocritan relationships. Proceedings of the National Academy of Sciences of the U.S.A. 91: 9911–9915.

Enghoff H (1984) Phylogeny of millipedes – a cladistic analysis. Journal of Zoological Systematics and Evolutionary Research. 22: 8–26.

Enghoff H (1991) A revised cladistic analysis and classification of the millipede order Julida. Journal of Zoological Systematics and Evolutionary Research. 29: 241–263.

Enghoff H, Petersen G, Seberg O (2011) Phylogenetic relationships in the millipede family Julidae. Cladistics 27:606–616.

Enghoff H, Petersen G, Seberg O (2013) The aberrant millipede genus *Pteridoiulus* and its position in a revised molecular phylogeny of the family Julidae (Diplopoda:Julida). Invertebrate Systematics. 27: 515–529.

Farris SM, Schulmeister S (2010) Parasitoidism, not sociality, is associated with the evolution of elaborate mushroom bodies in the brains of hymenopteran

insects. Proceedings of the Royal Society B: Biological Sciences. 278: 940–951.

Farstad CW, Platt AW (1946) The reaction of barley varieties to wheat stem sawfly attack. Scientific Agriculture. 26: 216–224.

Fernández R, Edgecombe GD, Giribet G (2016) Exploring phylogenetic relationships within Myriapoda and the effects of matrix composition and occupancy on phylogenomic reconstruction. Systematic Biology. 65: 871–889.

Fernández R, Edgecombe GD, Giribet G (2018) Phylogenomics illuminates the backbone of the Myriapoda Tree of Life and reconciles morphological and molecular phylogenies. Scientific Reports. 8: 83.

Francke-Grosmann H (1939) Über das Zusammenleben von Holzwespen (Siricinae) mit Pilzen. Journal of Applied Entomology. 25: 647–680.

Gallai N, Salles JM, Settele J, Vaissière BE (2009) Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. Ecological Economics. 68: 810–821.

Gol’berg AM (1986) Biology of the stem sawflies *Trachelus tabidus* and *Cephus pygmaeus* in the Negev of southern Israel. Entomologia Experimentalis et Applicata. 40: 117–121.

Goulson D (2019) The insect apocalypse, and why it matters. Current Biology. 29: R967–R971.

Grelle C, Fabre MC, Leprêtre A, Descamps M (2000) Myriapod and isopod communities in soils contaminated by heavy metals in northern France. European Journal of Soil Science. 51: 425–433.

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, et al. (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS One. 12: e0185809.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B. 270: 313–321.

Hennig W (1950) Grundzüge einer Theorie der phylogenetischen Systematik. Berlin: Deutscher Zentralverlag. 370 pp.

Heraty J, Ronquist F, Carpenter JM, Hawks D, Schulmeister S, et al. (2011) Evolution of the hymenopteran megaradiation. Molecular Phylogenetics and Evolution. 60: 73–88.

Hölldobler B, Wilson EO (1990) The Ants. Springer, Berlin. 732 pp.

Howse PE (1975) Brain structure and behavior in insects. Annual Review of Entomology. 20: 359–379.

Humphreys AM, Govaerts R, Ficinski SZ, Nic Lughadha E, Vorontsova MS (2019) Global dataset shows geography and life form predict modern plant extinction and rediscovery. Nature Ecology and Evolution. 3: 1043–1047.

Kajimura H (2000) Discovery of mycangia and mucus in adult female Xiphydriid woodwasps (Hymenoptera: Xiphydriidae) in Japan. Annals of the Entomological Society of America. 93: 312–317.

Kapli P, Yang Z, Telford MJ (2020) Phylogenetic tree building in the genomic age. Nature Reviews Genetics. 21: 428–444.

Keller L, Ross KG (1998) Selfish genes: a green beard in the red fire ant. Nature. 394: 573–575.

Klein AM, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, et al. (2007) Importance of pollinators in changing landscapes for world crops. Proceedings of the National Academy of Sciences of the U.S.A. 274: 303–313.

Leroi AM (2014) The Lagoon: How Aristotle Invented Science. London, Bloomsbury Publishing. 512 pp.

Linnaeus C (1753) Species Plantarum. Holmiae [= Stockholm], L. Salvii. 392 pp.

Linnaeus C (1758) Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata. Holmiae [= Stockholm], L. Salvii. 824 pp.

Legg DA, Sutton MD, Edgecombe GD (2013) Arthropod fossil data increase congruence of morphological and molecular phylogenies. Nature Communications. 4: 1–7.

Lloyd GER (1961) The Development of Aristotle's Theory of the Classification of Animals. Phronesis. 6: 59–81.

Lovejoy TE (1980) The Global 2000 Report to the President (G. O. Barney, ed.), The Technical Report. 2: 327–332. Penguin, New York.

Marek PE, Bond JE (2006) Phylogenetic systematics of the colorful, cyanide-producing millipedes of Appalachia (Polydesmida, Xystodesmidae, Apheloriini) using a total evidence Bayesian approach. Molecular Phylogenetics and Evolution. 41: 704–729.

Marek PE, Bond JE (2007) A reassessment of apheloriine millipede phylogeny: additional taxa, Bayesian inference, and direct optimization (Polydesmida : Xystodesmidae). Zootaxa. 1610: 27–39.

Martin PA, Green RE, Balmford A (2019) The biodiversity intactness index may underestimate losses. Nature Ecology and Evolution. 3: 862–863.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. Science. 346: 763–767.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? PLoS Biol. 9: e1001127.

Mrinalini, Werren JH (2015) Parasitoid Wasps and Their Venoms. In: Gopalakrishnakone P., Malhotra A. (Eds.) Evolution of Venomous Animals and Their Toxins. Toxinology. Springer, Dordrecht. 1–26 pp.

Neumann JS, Desalle R, Narechania A, Schierwater B, Tessler M (2020) Morphological Characters Can Strongly Influence Early Animal Relationships Inferred from Phylogenomic Data Sets. Systematic Biology. 0: 1–16.

Newbold T, Hudson LN, Arnell AP, Contu S, De Palma A, et al. (2016) Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. Science. 353: 288–291.

Parikn EA (1941) Symbiosis in Larval Siricidæ (Hymenoptera). Nature. 147: 329–329.

Parker HL (1935) Note on the anatomy of tenthredinid larvae with special reference to the head. Bollettino del Laboratorio di Zoologia Generale e Agraria della R. Scuola Superiore D'Agricoltura in Portici. 28: 159–191.

Pellens R., Grandcolas P. (2016) Biodiversity Conservation and Phylogenetic Systematics. Topics in Biodiversity and Conservation, vol 14. Springer, Cham. 390 pp.

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, et al. (2017) Evolutionary History of the Hymenoptera. Current Biology. 27: 1013–1018.

Pimvichai P, Enghoff H, Panha S (2014) Molecular phylogeny of the Thyropygus allevatus group of giant millipedes and some closely related groups. Molecular Phylogenetics and Evolution. 71: 170–183.

Pimvichai P, Enghoff H, Panha S, Backeljau T (2020) Integrative taxonomy of the new millipede genus *Coxobolellus*, gen. nov. (Diplopoda : Spirobolida : Pseudospirobolellidae), with descriptions of ten new species. Invertebrate Systematics. 34: 591–617.

Pitz KM, Sierwald P (2010) Phylogeny of the millipede order Spirobolida (Arthropoda: Diplopoda: Helminthomorpha). Cladistics. 26: 497–525.

Portman SL, Jaronski ST, Weaver DK, Reddy GVP (2018) Advancing biological control of the wheat stem sawfly: New strategies in a 100-yr struggle to manage a costly pest in the northern great plains. Annals of the Entomological Society of America. 111: 1–7.

Quicke, D.L.J. (1997). Parasitic Wasps. Chapman & Hall, London. 470 pp.

Grimaldi DA, Engel MS (2005) Evolution of the Insects. Cambridge University Press, Cambridge. 772 pp.

Read HJ, Martin MH, Rayner JMV (1998) Invertebrates in woodlands polluted by heavy metals—an evaluation using canonical correspondence analysis. Water, Air, and Soil Pollution. 106: 17–42.

Regier J, Shultz J (2001) A phylogenetic analysis of Myriapoda (Arthropoda) using two nuclear protein-encoding genes. Zoological Journal of The Linnean Society-London. 132: 469–486.

Regier JC, Wilson HM, Shultz JW (2005) Phylogenetic analysis of Myriapoda using three nuclear protein-coding genes. Molecular Phylogenetics and Evolution. 34: 147–158.

Régnier C, Achaz G, Lambert A, Cowie RH, Bouchet P, Fontaine, B (2015) Mass extinction in poorly known taxa. Proceedings of the National Academy of Sciences U.S.A. 112: 7761–7766.

Rehm P, Meusemann K, Borner J, Misof B, Burmester T (2014) Phylogenetic position of Myriapoda revealed by 454 transcriptome sequencing. Molecular phylogenetics and evolution. 77: 25–33.

Richter S, Wirkner CS (2014) A research program for evolutionary morphology. Journal of Zoological Systematics and Evolutionary Research. 52: 338–350.

Rodriguez J, Jones TH, Sierwald P, Marek PE, Shear WA, et al. (2018) Step-wise evolution of complex chemical defenses in millipedes: A phylogenomic approach. Scientific Reports. 8: 1–10.

Rohwer SA, Cushman RA (1917) Idiogastra, a new suborder of Hymenoptera with notes on the immature stages of Oryssus. Proceedings of the Entomological Society of Washington. 19: 89–98.

Ronquist F, Rasnitsyn AP, Roy A, Eriksson K, Lindgren M (1999) Phylogeny of the Hymenoptera: a cladistic reanalysis of Rasnitsyn's (1988) data. Zoologica Scripta. 28: 13–50.

Sáringer G (1974) Problems with *Athalia rosae* L. (Hym., Tenthredinidae) in Hungary. In: Proceedings of the 4. International Rapskongress. Giessen. 575–578 pp.

Sann M, Niehuis O, Peters RS, Mayer C, Kozlov A, et al. (2018) Phylogenomic analysis of Apoidea sheds new light on the sister group of bees. BMC Evolutionary Biology. 18: 1–15.

Schulmeister S (2003) Review of morphological evidence on the phylogeny of basal Hymenoptera (Insecta), with a discussion of the ordering of characters. Biological Journal of the Linnean Society. 79: 209–243.

Schultz TR. 2000. In search of ant ancestors. Proceedings of the National Academy of Sciences of the U.S.A. 97: 14028–14029.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, et al. (2014) Genomics and the origin of species. Nature Reviews Genetics. 15: 176–192.

Sharkey MJ, Roy A (2002) Phylogeny of the Hymenoptera: a reanalysis of the Ronquist et al. (1999) reanalysis, with an emphasis on wing venation and apocritan relationships. Zoologica Scripta. 31: 57–66.

Short M, Vahtera V (2017) Phylogenetic relationships of millipedes in the subclass Penicillata (Diplopoda) with a key to the genera. Journal of Natural History. 51: 2443–2461.

Short M, Vahtera V, Wesener T, Golovatch SI (2020) The millipede family Polyxenidae (Diplopoda, Polyxenida) in the faunas of the Crimean Peninsula and Caucasus, with notes on other European Polyxenidae. Zootaxa. 4772: 306–332.

Sierwald P, Bond JE (2007) Current status of the Myriapod class diplopoda (millipedes): taxonomic diversity and phylogeny. Annual Review of Entomology. 52: 401–420.

Sierwald P, Shear WA, Shelley RM, Bond JE (2003) Millipede phylogeny revisited in the light of the enigmatic order Siphoniulida. Journal of Zoological Systematics and Evolutionary Research. 41: 87–99.

Simonsen Å (1990) Phylogeny and biogeography of the millipede order Polydesmida, with special emphasis on the suborder Polydesmidea. University of Bergen, Bergen. Doctoral dissertation, Museum of Zoology.

Snyder BA, Hendrix PF (2008) Current and Potential Roles of Soil Macroinvertebrates (Earthworms, Millipedes, and Isopods) in Ecological Restoration. Restoration Ecology. 16: 629–636.

Spelda J, Reip HS, Oliveira-Biener U, Melzer RR (2011) Barcoding Fauna Bavarica: Myriapoda - a contribution to DNA sequence-based identifications of centipedes and millipedes (Chilopoda, Diplopoda). Zookeys. 139: 123–139.

Spradbery JP, Kirk AA (1978) Aspects of ecology of siricid woodwasps (Hymenoptera Siricidae) in Europa, North-Africa and Trukey with special reference to biological-control of *Sirex noctilio* in Australia. Bulletin of Entomological Research. 68: 341–359.

Stolle E, Pracana R, Howard P, Paris CI, Brown SJ, et al. (2019) Degenerative Expansion of a Young Supergene. Molecular Biology and Evolution. 36: 553–561.

Swingland IR (2001) Biodiversity, definition of. In: Levin SA (Ed.) Encyclopedia of Biodiversity. Volume 1. Academic Press, San Diego. 377–391 pp.

Thomas JA (2016) Butterfly communities under threat. Science. 353: 216–218.

Thompson TQ, Renee Bellinger M, O'Rourke SM, Prince DJ, Stevenson AE, et al. (2019) Anthropogenic habitat alteration leads to rapid loss of adaptive

variation and restoration potential in wild salmon populations. Proceedings of the National Academy of Sciences U.S.A. 116: 177–186.

Tsiafouli MA, Thébault E, Sgardelis SP, de Ruiter PC, van der Putten WH, et al. (2015) Intensive agriculture reduces soil biodiversity across Europe. Global Change Biology. 21: 973–985.

Vilhelmsen L (1997) The phylogeny of the lower Hymenoptera (Insecta), with a summary of the early evolutionary history of the order. Journal of Zoological Systematics and Evolutionary Research. 35: 49–70.

Vilhelmsen L (2001) Phylogeny and classification of the extant basal lineages of the Hymenoptera (Insecta). Zoological Journal of the Linnean Society. 131: 393–442.

Vilhelmsen L (2003a) Phylogeny and classification of the Orussidae (Insecta: Hymenoptera), a basal parasitic wasp taxon. Zoological Journal of the Linnean Society. 139: 337–418.

Vilhelmsen L (2003b) Larval Anatomy of Orussidae (Hymenoptera). Journal of Hymenoptera Research. 12: 346–354.

Vilhelmsen L, Isidore N, Romani R, Basibuyuk HH, Quicke DLJ (2001) Host location and oviposition in a basal group of parasitic wasps: The subgenual organ, ovipositor apparatus and associated structures in the Orussidae (Hymenoptera, Insecta). Zoomorphology. 121: 63–84.

Vilhelmsen L, Mikó I, Krogmann L (2010) Beyond the wasp-waist: structural diversity and phylogenetic significance of the mesosoma in apocritan wasps (Insecta: Hymenoptera). Zoological Journal of the Linnean Society. 159: 22–194.

Voigtländer K, Reip H, Decker P, Spelda J (2011) Critical reflections on German Red Lists of endangered myriapod species (Chilopoda, Diplopoda) (with species list for Germany). International Journal of Myriapodology. 6: 85–105.

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, et al. (2013) A Y-like social chromosome causes alternative colony organization in fire ants. Nature. 493: 664–668.

Wang T, Zhao M, Rotgans BA, Ni G, Dean JFD, et al. (2016) Proteomic analysis of the venom and venom sac of the woodwasp, *Sirex noctilio* - Towards understanding its biological impact. Journal of Proteomics. 146: 195–206.

Wanninger A (2015) Morphology is dead - long live morphology! Integrating MorphoEvoDevo into molecular EvoDevo and phylogenomics. Frontiers in Ecology and Evolution. 3: 1–9.

Weinstock GM, Robinson GE, Gibbs R, Worley KC, Evans JD, et al. (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature. 443: 931–949.

Wesener T (2014) A new phylogenetic analysis of the Sphaerotheriida (Diplopoda) with a revision of the Australian giant pill-millipedes. Invertebrate Systematics. 28: 196–213.

Wesener T (2015) No millipede endemics north of the Alps? DNA-Barcoding reveals *Glomeris malmivaga* Verhoeff, 1912 as a synonym of *G. ornata* Koch, 1847 (Diplopoda, Glomerida, Glomeridae). Zootaxa. 3999: 571–580.

Wesener T, Enghoff H, Wägele JW (2008) Pachybolini - a tribe of giant Afrotropical millipedes: arguments for monophyly and the description of a new genus from

Madagascar (Diplopoda:Spirobolida:Pachybolidae). Invertebrate Systematics. 22: 37–53.

Wesener T, Raupach MJ, Sierwald P (2010) The origins of the giant pill-millipedes from Madagascar (Diplopoda: Sphaerotheriida: Arthrosphaeridae). Molecular phylogenetics and evolution. 57: 1184-1193.

Wesener T, Van den Spiegel D (2009) A first phylogenetic analysis of Giant Pill-Millipedes (Diplopoda: Sphaerotheriida), a new model Gondwanan taxon, with special emphasis on island gigantism. Cladistics. 25: 545–573.

Wheeler QD (1995) Systematics, the scientific basis for inventories of biodiversity. Biodiversity Conservation. 4: 476–489.

Wheeler QD (2004) Taxonomic triage and the poverty of phylogeny. Philosophical Transactions of the Royal Society B. 359:571–583.

Wheeler QD, Raven PH, Wilson EO (2004) Taxonomy: Impediment or Expedient? Science. 303: 285–285.

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology. 54: 844–851.

Wilson EO, Hölldobler B (2005) Eusociality: Origin and consequences. Proceedings of the National Academy of Sciences of the U.S.A. 102: 13367–13371.

Wolf JBW, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics. 18: 87–100.

World Economic Forum (2020) Nature Risk Rising: Why the Crisis Engulfing Nature Matters for Business and the Economy. Accessed on 26.08.20, from: http://www3.weforum.org/docs/WEF_New_Nature_Economy_Report_2020.pdf

Young AD, Gillung JP (2020) Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. Systematic Entomology. 45: 225–247.

Chapter 2

Steps towards a phylogeny of the pill millipedes: non-monophyly of the family Protoglomeridae, with an integrative redescription of *Eupeyerimhoffia archimedis* (Diplopoda, Glomerida)

This chapter is published in:
Oeyen JP, Wesener T (2015) Steps towards a phylogeny of the pill millipedes: non-monophyly of the family Protoglomeridae, with an integrative redescription of *Eupeyerimhoffia archimedis* (Diplopoda, Glomerida). In: Tuf IH, Tajovský K (Eds) Proceedings of the 16th International Congress of Myriapodology, Olomouc, Czech Republic. ZooKeys. 510: 49–64.

Author contributions to the original article:
Conceived the study: JPO, TW; Molecular lab work: JPO; Disections and SEM: JPO; Performed the analyses: JPO; Interpreted the results: JPO; Collected specimens: JPO; Wrote the manuscript: JPO, TW.

Figure and table numbers have been adapted for the inclusion in the present thesis.

2.1 Introduction

The pill millipedes of the order Glomerida comprise about 290 species in 34 genera (Mauriès 2005, Golovatch et al. 2010, Wesener 2010, 2012) and exhibit a Holarctic distribution, with species found in North America, Europe and North Africa, and Asia with the exception of India south of the Himalayas (Shelley and Golovatch 2011). The Glomerida are currently divided into three families (Mauriès 1971, 2005), the two species-poor families Glomeridellidae and Protoglomeridae, and the family Glomeridae, which contains the majority of species and genera (~240 species in 27 genera) (Mauriès 2005, Wesener 2012).

While the two genera of the Glomeridellidae are Mediterranean, the four genera and 20 species of the Protoglomeridae show a disjunct distribution, partly European, in Spain, the eastern Mediterranean, Algeria and Sicily, and partly in the New World from Guatemala to California (Mauriès 2005).

Here we redescribe the little-known species *Eupeyerimhoffia archimedis* (Strasser, 1965), and describe the male telopods for the first time. Additionally, we illustrate several unusual (and potentially apomorphic) morphological characters of a member of the family Protoglomeridae for the first time using scanning electron microscopy. To complete our integrative approach, we also analyze the genetic distances between the four genera of the family using the common barcoding fragment, COI.

2.2 Material and methods

Samples of *Eupeyerimhoffia archimedis* were collected by hand in July 2013. A single male and several females were collected close to the type locality (Ferla; Fig. 2.1C)

and further samples were collected at a new locality (East of Palazzolo Acreide, Sicily). Exact coordinates are provided in Table 2.1. All samples were conserved in 98% EtOH for further analyses and deposited in the collection of the Zoological Research Museum Alexander Koenig (ZFMK, Bonn, Germany).

Figure 2.1. *Eupeyerimhoffia archimedis* (Strasser 1965) female in situ and habitat. **A:** *E. archimedis* female rolled up, in situ; **B:** *E. archimedis* female in situ; **C:** Habitat of *E. archimedis*, close to the type locality Ferla.

2.2.1 Morphological analysis

A female and the single male from the type locality were dissected under an Olympus SZX12 stereomicroscope with Dumont 5 Inox B forceps. Samples were dehydrated in 100% EtOH for 12 hrs, mounted on aluminum stubs, dried for 12 hrs at 45 °C and

sputter coated with 50 nm of pure gold in a Hummer VI sputtering system (Anatech LTD, USA). Samples were observed with a Hitachi S-2460N SEM (Hitachi LTD, Japan) and digital images were captured using DISS5 (point electronic GmbH, Germany).

2.2.2 Molecular analysis

Muscle tissue was removed from specimens of *Onychoglomeris tyrolensis* (Latzel, 1884), *Protoglomeris vasconica* (Brölemann, 1897), *Glomerellina laurae* Silvestri, 1908, and *Eupeyerimhoffia archimedis* (Strasser, 1965). Sequences of *Glomeroides primus* (Silvestri, 1929) were downloaded from GenBank. Additionally, sequences from GenBank of *Glomeridella minima* (Latzel, 1884), a member of the basal family Glomeridellidae, as well as of *Glomeris marginata* (Villers, 1789), *Geoglomeris subterranea* Verhoeff, 1908 and *Trachysphaera* sp. from the family Glomeridae (Table 2.1) were also downloaded. Specimens from which DNA was extracted were stored as vouchers at the ZFMK. Accession numbers, locality data and voucher information for all samples included in the study are displayed in Table 2.1.

Total genomic DNA was extracted using the Qiagen DNAeasy Blood&Tissue kit following the standard protocol. A fragment of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene was amplified via PCR (Saiki et al. 1988) using the Nancy (Simon et al. 1994) and LCO (Folmer et al. 1994) primer pair following previously published protocols (Wesener et al. 2010).

Table 2.1. Sample information with voucher numbers (ZFMK = Zoological Research Museum Alexander Koenig, Bonn, Germany. ZSM = Bavarian State Collection of Zoology. NHMC = Natural History Museum of Crete), GenBank accession numbers (Acc.#) and locality information. Samples where sequences were downloaded from GenBank are marked with an asterisk.

Species	Specimen Voucher	Acc. #	Locality
**Glomeris marginata* (Villers, 1789)	ZFMK MYR0009	FJ409909	Germany, Nordrhein-Westfalen, Bonn, Venusberg, coll. T. Wesener, IX.2007
**Glomeridella minima* (Latzel, 1884)	ZFMK MYR0003	JQ074181	Slovenia, Lower Sava, Brežice, Prilipe, dry creek valley, 45.8773°N, 15.6246°E, 150 m, coll. H. Reip, 17.x.2009.
**Geoglomeris subterranea* Verhoeff, 1908	BC ZSM MYR 00370	JQ350441	Switzerland, Aargau
**Trachysphaera* sp.	ZFMK MYR0006	JQ074180	Italy, Piemonte, Biella, NW Sanctuary of Oropa, Fagus forest with stones, 45.62947°N, 7.98168°E, 1200 m, coll. T. Wesener, 14.iv.2011
**Glomeroides primus* (Silvestri, 1929)	ZFMK MYR0004	JQ074182	U.S.A., California, Mendocino County, between Fort Bragg and Whiskey Springs, 39.3976°N, 123.6946°W, 35 m, coll E. Garcia, C. Richart & A. Schönhofer, 29.iii.2011.
Onychoglomeris tyrolensis (Latzel, 1884)	ZFMK MYR1276	KP205571	Italy, Trentino-Südtirol, Prov. Trient, Madonna di Campiglio, Beech forest, 46.2209528°N, 010.8296250°E, 1553 m, coll. T. Wesener, 04.x.2012.
Protoglomeris vasconica (Brölemann, 1897)	ZFMK MYR0934	KP205572	Spain, Galicia, Ribadeo, Trabada, deep and moist creek valley with deciduous forest, 43.4295°N, 7.2290°E, coll. H. Reip, 29.vii.2012.
Glomerellina laurae Silvestri, 1908	ZFMK MYR2260	KP205573	Europe, Greece, Rhodos, Kapi - Profitis Ilias, coll. NHMC, 01.i.2000.
Eupeyerimhoffia archimedis (Strasser, 1965) 1	ZFMK MYR1876	KP205574	Italy Sicily, Province Syracuse, South of Ferla, Southern slope, deciduous forest, 37.1151333°N, 014.9403667°E, coll. J.P. Oeyen & P. Erkeling, 10.vii.2013.
Eupeyerimhoffia archimedis (Strasser, 1965) 2	ZFMK MYR1965	KP205575	Italy, Sicily, Province Syracuse, East of Palazzolo Acreide, Ravine, deciduous forest, 37.0997667°N, 015.0232000°E, coll. J.P. Oeyen & P. Erkeling, 13.vii.2013.

Both strands were sequenced by Macrogen (Macrogen Europe Laboratory, Amsterdam, The Netherlands), following the Sanger sequencing method (Sanger et al. 1977). Sequencing reads were assembled and aligned by hand with Bioedit 7.1.3 (Hall 1999) and confirmed with BLAST searches (Altschul et al. 1997). Sequences were uploaded to GenBank (Accession numbers: KP205571 to KP205557).

Mean pairwise distances between terminals (transformed into percentages) were determined using MEGA5.2 (Tamura et al. 2011). To better illustrate relationships between genera, a maximum likelihood phylogenetic analysis was conducted in MEGA5.2 (Tamura et al. 2011). The implemented ModelTest selected the HKY+G+I model (Hasegawa et al. 1985) as best-fitting (BIC = 5783.1, -lnL = -2791.2354, freqA = 0.2647, freqC = 0.1366, freqG = 0.2014, freqT = 0.3972, gamma shape = 0.3364). The bootstrap consensus tree (Fig. 2.2), inferred from 1000 replicates (Felsenstein 1985), is used to represent the evolutionary history of the analyzed taxa. All positions containing gaps and missing data were eliminated. There were a total of 668 positions in the final dataset. While the genetic marker used does not allow a study of the phylogeny of the group, first insights into the separation of the genera are provided.

2.3 Results

2.3.1 Distance analysis

The uncorrected pairwise distances between genera included in the present study are relatively high (Table 2.2). The genetic distances are not lower between species within the same family than between species of different families. The distances range from 18.8% between *Glomerellina laurae* (Protoglomeridae) and *Glomeris marginata*

(Glomeridae) to 12.0% between *Protoglomeris vasconica* (Protoglomeridae) and *Glomeridella minima* (Glomeridellidae). The two *Eupeyerimhoffia archimedis* samples show a 0.2% sequence divergence, but also show both the highest (16.8%: *G. laurae*) and lowest distance (13.2 and 13.4%: *P. vasconica*) to other species within the family.

Table 2.2. Pair-wise uncorrected p-distances (%) of the COI-fragment.

#	Species	1	2	3	4	5	6	7	8	9
1	*Glomeris marginata*									
2	*Glomeridella minima*	16.0								
3	*Geoglomeris subterranea*	17.4	15.6							
4	*Trachysphaera sp.*	15.0	13.2	15.6						
5	*Glomeroides primus*	16.3	14.2	16.4	15.0					
6	*Onychoglomeris tyrolensis*	14.3	13.2	16.8	13.2	15.3				
7	*Protoglomeris vasconica*	14.8	12.0	15.3	13.5	15.0	13.2			
8	*Glomerellina laurae*	18.8	16.0	18.3	15.3	16.7	16.3	15.8		
9	*Eupeyerimhoffia archimedis 1*	16.1	15.2	15.0	13.1	16.5	15.0	13.2	16.8	
10	*Eupeyerimhoffia archimedis 2*	16.2	15.3	15.0	13.2	16.7	15.2	13.4	16.8	0.2

2.3.2 Tree description

The maximum likelihood tree receives little to no support, most nodes remain unresolved and all taxa are separated by long branches (Fig. 2.2). The family Protoglomeridae (P) could not be recovered. All members of the family are recovered within a major polytomy together with species from both Glomeridellidae and Glomeridae (Fig. 2.2). Within the polytomy *Glomeroides primus* (Protoglomeridae) groups together with *Glomeridella minima* (Glomeridellidae). *Glomerellina laurae* (P) does not cluster with any species within the polytomy and rests on the longest branch within the tree. *Protoglomeris vasconica* (P) and *Eupeyerimhoffia archimedis* (P) are recovered in a polytomy together with *Geoglomeris subterranea* (Glomeridae). Only the subfamily Glomerinae (*Glomeris* & *Onychoglomeris*) could be resolved as monophyletic (Fig. 2.2) as the sister group to the polytomy, though

with weak support.

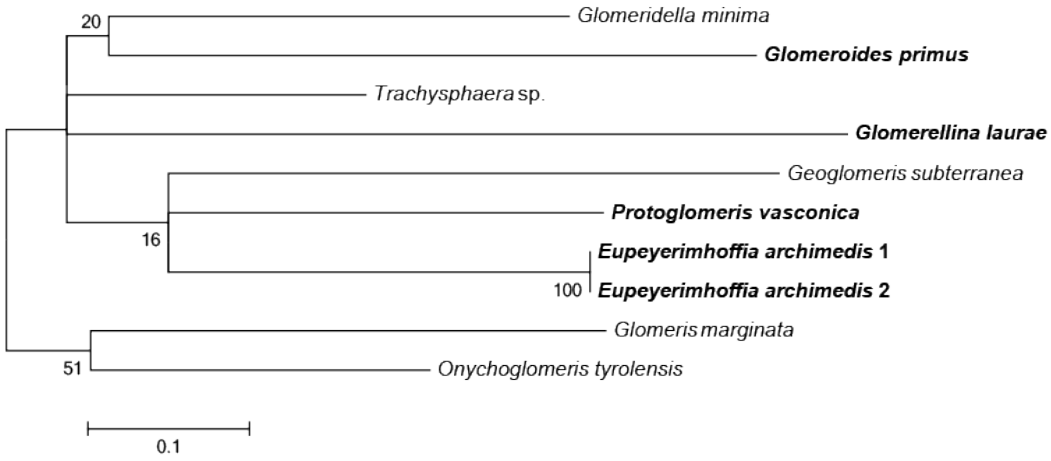

Figure 2.2. Maximum likelihood bootstrap consensus tree. Members of the family Protoglomeridae are marked in bold.

Family Protoglomeridae Brölemann, 1913

Diagnosis. Simple telopods with four podomeres distal to syncoxite, forming pincers. Telopoditomeres 1–3 lacking trichosteles. Telopoditomere 2 with a non-membranous immovable finger located almost parallel to telopoditomere 3. Here, we follow the typological system of Mauriès (2005), despite the fact that no phylogenetic analysis has been undertaken to characterize the families in the order.

Eupeyerimhoffia Brölemann, 1913

Eupeyerimhoffia Brölemann, 1913: 166–174 (first description); Jeekel 1971: 13 (note); Strasser 1976: 581–583 (synonymization *Trinacriomeris*); Hoffann 1980: 67 (list); Foddai et al. 1995: 11 (list); Shelley et al. 2000: 11 (list); Mauriès 2005: 242 (classification); Kime and Enghoff 2011: 39 (atlas).

Trinacriomeris Strasser, 1965: 10–14. syn.

Diagnosis. Tergite 11 fused to anal shield. Telopod simple with four podomeres. Process of telopoditomere 2 of telopod short and stout. Male leg-pair 17 with four podomeres. Coxa of male leg-pair 18 not fused to syncoxite. Lateral palpi of gnathochilarium as large as inner palpi. One of the largest Glomerida, 18–22 mm long. Mandible with large condylus. Members might be mistaken in habitus, size and color with the species of the genus *Onychoglomeris* Verhoeff, 1909, whose species differ greatly in the telopods and many other characteristics.

Type species. *Eupeyerimhoffia algerina* Brölemann, 1913 from Algeria

Other species included. *Eupeyerimhoffia archimedis* (Strasser, 1965) from southern Sicily.

Eupeyerimhoffia archimedis **(Strasser, 1965)**

Trinacriomeris archimedis Strasser 1965: 10–14 (first description);

Trinacriomeris archimedis Strasser 1970: 153 (list);

Trinacriomeris archimedis Strasser 1976: 581–583 (synonymization *Trinacriomeris*);

Eupeyerimhoffia archimedis Foddai et al. 1995: 11 (list);

Trinacriomeris archimedis Shelley et al. 2000: 11 (list).

Material examined. 1 F, **MHNG**, lectotype (designated herewith), labeled paratype, in 70% ethanol, Italie (Sicile): Siracuse: Avola pr. Siracuse. 1 F, **MHNG 3460**, dried and mounted, Italie (Sicile): Siracuse: Avola pr. Siracuse; 1 F, **ZFMK MYR01879**, 1 M, **ZFMK MYR01875**, Italy, Sicily at type locality, south of Ferla, 37.1151333°N, 014.9403667°E, coll. J.P. Oeyen & P. Erkeling, 10.vii.2013; 1 F, **ZFMK MYR 1965**,

Italy, Sicily, Province Syracuse, East of Palazzolo Acreide, Ravine, deciduous forest, 37.0997667°N, 015.0232000°E, coll. J.P. Oeyen & P. Erkeling, 13.vii.2013.

Comment. A second female type specimen from Ferla, Sicily was, according to the first description, stored at the University of Catania, Institute of Zoology, Italy.

Re-diagnosis. Can easily be distinguished from the other Sicilian Glomerida species by size and color. It is the largest and only light brown species on the island. It can be distinguished from its congener *E. algerina* in having: (1) Single continuous anterior stria on collum, posterior stria divided in lateral parts; (2) thoracic shield with single continuous stria reaching the lateral lobes on both sides.

Description. General coloration (living specimen) light brown, almost copper. Collum, head, antennae, posterior margin and lateral speckled fields of tergites lighter, almost golden cream color (Fig. 2.1A, B).

Head sparsely covered with minute setae, >10 supralabral setae (Fig. 2.3A, C). Incisura lateralis (IL) directed slightly laterally, not reaching height of organ of Tömösváry (TO) or antennal basis (Fig. 2.3A–C). Lateral marginal bulge thickest at IL, decreasing gradually dorsally until terminating at height of dorsal-most ocellus (Fig. 2.3C). A furrow running laterally between ventral-most ocellus and TO, circumventing antennal fossa and terminating at height of IL (Fig. 2.3C).

Labrum wide, with 19 marginal setae (Fig. 2.3A, B). Central labral tooth projecting beyond lateral margin.

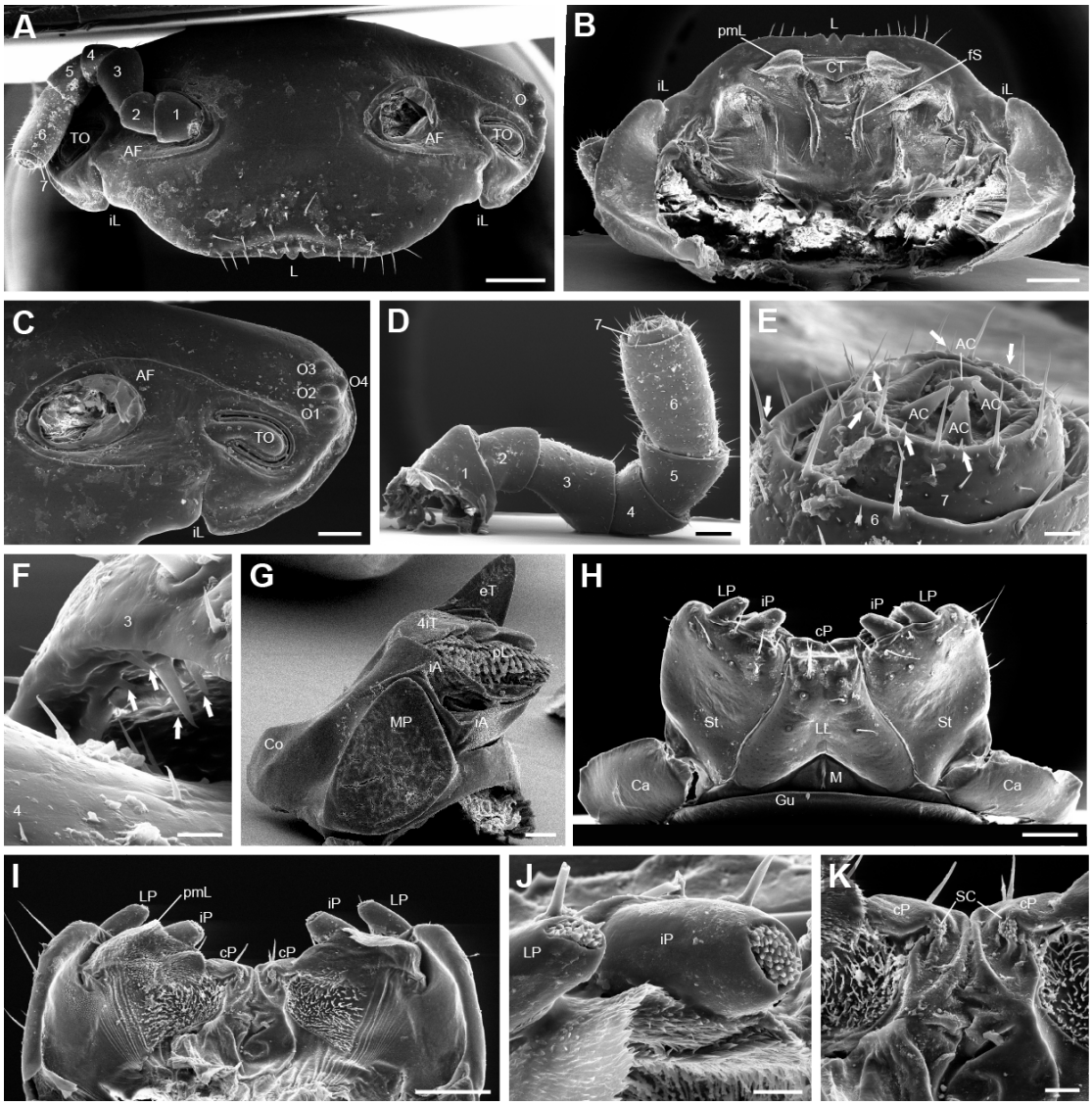

Figure 2.3. *Eupeyerimhoffia archimedis* (Strasser 1965) male, SEM. **A:** Head, frontal view; **B:** Head, ventral view; **C:** Head, detail of lateral area; **D:** Antenna, posterior view; **E** Antenna, antennomere 6 and 7; **F:** Antenna, apical edge of antennomere 3; **G:** Mandible, mesal view; **H:** Gnathochilarium, ventral view; **I:** Endochilarium, dorsal view; **J:** Gnathochilarium, lateral and inner palpi, dorsal view; **K:** Endochilarium, detail of median area, dorsal view. **Abbreviations:** 1-7 = Antennal segments number 1-7; 4iT = 4-combed inner tooth; AC = Api- cal cone; AF = Antennal fossa; Ca = Cardine; Co = Condylus; cP = Central pads; CT = Central tooth; eT = External tooth; fS = fringed seam; Gu = Gula; iA = Intermediate area; iL = Incisura lateralis; iP = Inner palpi; L = Labrum; LL = Lamella linguales; LP = Lateral palpi; M = Mentum; MP = Molar plate; O = Ocellaria; O1-O4 = Ocelli 1-4; pL = Pectinate lamellae; pmL = paramedian lobe; TO = Organ of Tömösváry; SC = Sensory clusters; St = Stipites. Arrows mark *sensilla basiconica*. Scale bar: 400 µm (**A, B**); 200 µm (**C**); 150 µm (**D**); 25 µm (**E**); 10 µm (**F**); 100 µm (**G**); 250 µm (**H, I**); 40 µm (**J**); 50 µm (**K**).

Epipharynx with pronounced central tooth and two lateral membranous lobes, covered densely in cuticular scales (Fig. 2.3B). Incisura lateralis clearly visible, reaching margin of head capsule. Two paramedian fringed seams stretching from central tooth posteriorly towards hypopharynx.

Ocellaria black, 3+1 convex lenses (Fig. 2.3C).

Antennae with four apical cones (Fig. 2.3E). Antennomere 3 approximately as long as 1 and 2 combined (Fig. 2.3D). Antennomere 6 approximately 1.9 times longer than wide. Antennomeres 1–5 only sparely setose, 6th more densely setose. Multiple sensilla basiconica on proximal apical edge of antennomere 3 (Fig. 2.3F) as well as at apical edge of antennomere 7 (Fig. 2.3E).

Organ of Tömösváry recessed, elongate, curved ventrally (Fig. 2.3C). 1.9 times longer than wide. Bulging cone and slit margins smooth. Cone narrower at midpoint. No internal structures visible in SEM.

Gnathochilarium ventrally with 8 large setae on lamella linguales, 12 large setae on each stipites (Fig. 2.3H). Remaining ventral surface glabrous. Cardines large. Inner palpi slightly larger than lateral palpi (Fig. 2.3H–J). Inner palpi with >40 sensory cones standing in single field (Fig. 2.3J). Lateral palpi also with field of >20 sensory cones (Fig. 2.3J).

Endochilarium with large anterior membranous paramedian lobes (pmL), densely covered with cuticular scales (Fig. 2.3I). Fields of long setae posterior to membranous lobes. Central pads with single cluster of sensilla directed towards median furrow (Fig. 2.3I, K).

Mandible with single large outer tooth and four-combed inner tooth (Fig. 2.3G). Proximal comb of inner tooth slightly ovoid. Six rows of pectinate lamellae. Lateral

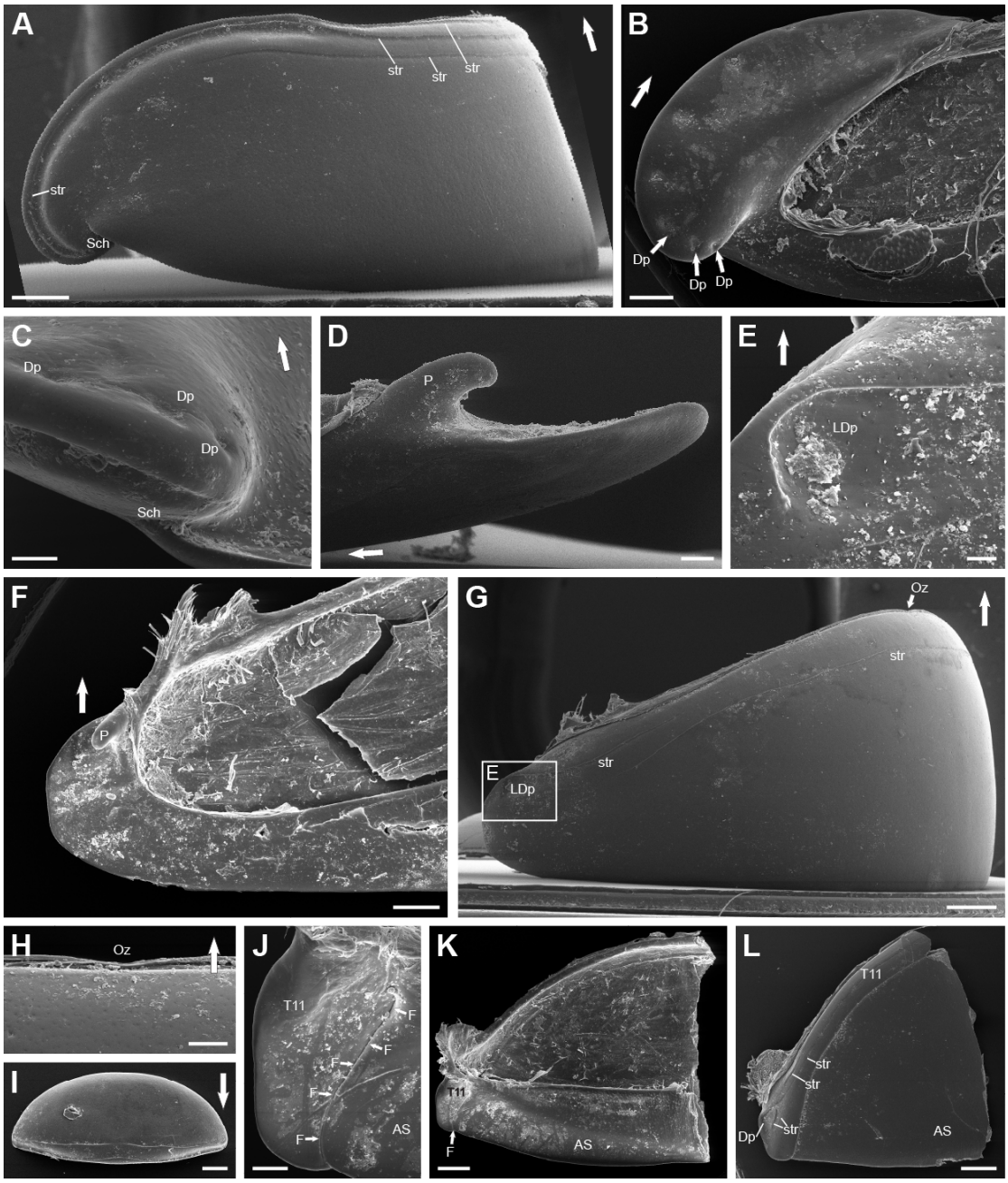

Figure 2.4. *Eupeyerimhoffia archimedis* (Strasser 1965) male, SEM. **A:** Thoracic shield, dorso-lateral view; **B:** Thoracic shield, meso-lateral; **C:** Thoracic shield, schism detail, ventro-lateral view; **D:** Tergite, detail of peg, antero-lateral view; **E:** Tergite, detail of depression, lateral view; **F:** Tergite, ventral view; **G:** Tergite, dorso-lateral view; **H:** Tergite, ozopore, dorsal view; **I:** Collum, dorsal view; **J:** Tergite 11 and anal shield, detail of furrow, ventro-lateral view; **K:** Tergite 11 and anal shield, left side, anterior view; **L:** Tergite 11 and anal shield, right side, posterior view. **Abbreviations:** AS = Anal shield; Dp = Depression; F = Furrow; Oz = Ozopore; LDp = Lateral depression; P = Peg; Sch = Schisma; stri = striae; T11 = Tergite 11. Arrows point anteriorly. Scale bar: 400 µm (**A, H, I, K, L**); 300 µm (**B, F**); 100 µm (**C, D, J**); 50 µm (**E**); 500 µm (**G**).

areas of intermediate area covered with small cuticular scales, central part smooth with possible pore. Molar plate almost triangular, marginal bulge at anterior edge, no anterior depression and posterior tip slightly curved towards mandibular basis. Condylus pronounced (Fig. 2.3G).

Collum with one continuous anterior and two posterior lateral striae (Fig. 2.4I). Uniformly covered with minute setae, recessed into small pits.

Thoracic shield with very small schism (Fig. 2.4A). 3 median striae. Marginal furrow widest laterally, narrowing medially. Uniformly covered with minute recessed setae. Ventral area of lobe with seven anterior marginal depressions on lobe and a single depression at the posterior margin (Fig. 2.3B, C; see below for function of depressions).

Tergites 3–10 covered with minute recessed setae, with single complete transverse anterior stria and short lateral striae anteriorly circumventing a depression (Fig. 2.4E, G). Lateral edges not projecting posteriorly. Stout pegs on ventral areas projecting posteroventrally from lateral most part of anterior edge (Fig. 2.4D, F).

Ozopore simple, neither with special sutures nor other structures (Fig. 2.4H).

Tergite 11 and anal shield completely fused but both dorsally and ventrally distinguishable by a pronounced furrow (Fig. 2.4J–L). Tergite 11 with 3 short lateral striae and a single stria circumventing a lateral depression and stretching across whole tergite (Fig. 2.4L). Tergite 11 and anal shield dorsally evenly covered with minute setae, with neither any special notches nor structures.

Pleurites evenly covered with small setae, bulge at anterior edge widest medially narrowing towards proximal edge. Pleurite 1.2 times wider than long.

Stigmatic plates reaching around coxa on both anterior and posterior sides. 1.5 times

wider than long, almost pentagonal in shape. Plate with regular margin, lacking any projections. Spiracle inconspicuous, protected by small knob.

Midbody legs sparsely covered with minute setae (Fig. 2.5F). Coxa almost triangular, much wider at base than apically. Coxa mesally elongated to process carrying a single spine. Two coxal furrows originating at center of coxal basis, one stretches apically around coxa, the second terminates after 2/3 of coxal height in a meso-apical direction. Tibia, pre- and postfemur with a single mesal spine, femur with two. Apical margin of prefemur with a single small apical protrusion. Femur almost 3 times longer than wide. Tarsus with no apical, 11 dorsal and 8–11 ventral spines. Tarsus 4.5 times longer than wide. Claw elongated.

Male sexual characters.

Male tergite 11 and anal shield do not show any special structures (Fig. 2.4J–L). See further and more detailed descriptions above.

Male first leg-pair sparsely covered with minute setae (Fig. 2.5A). Coxa not widened at basis, but mesally elongated to a process carrying two spines. Postfemur and tibia each with single mesal apical spine, prefemur and femur with two. Apical margin of prefemur with a single small protrusion. Tarsus with 7–10 dorsal and 8 ventral spines. Claw elongated but stout at basis. Tarsus almost 4 times longer than wide.

Male second leg-pair similar to midbody legs, but with a bulbous medial coxal protrusion carrying two spines (Fig. 2.5C), similar to leg 1. Tarsus approximately 3.8 times longer than wide.

Male gonopore clam-shaped and mesally protruding from posterior side of coxa 2 (Fig. 2.5E). Single elongate membranous opening surrounded by 9 or 10 apical and 4 basal setae. No division into separate plates.

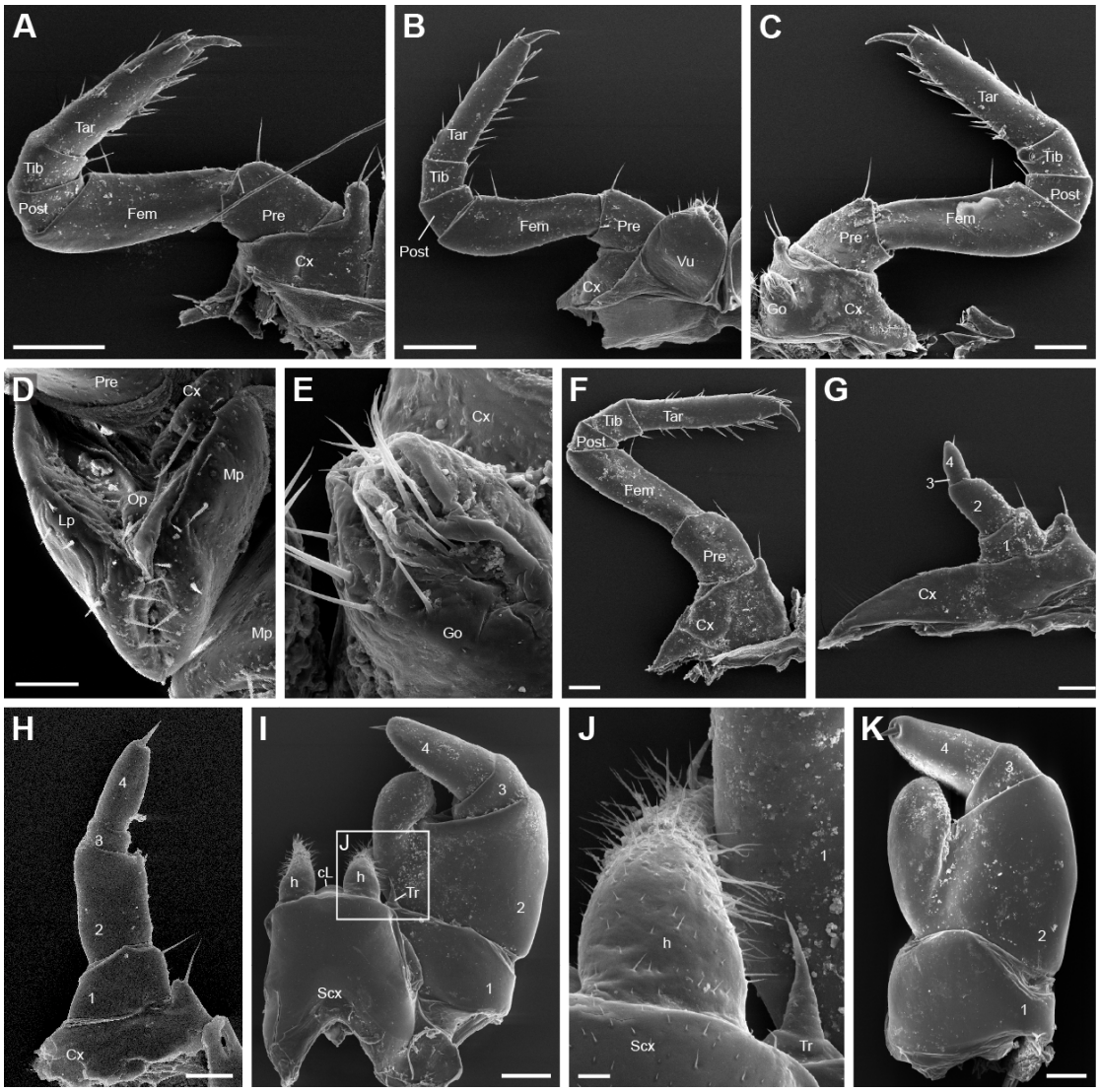

Figure 2.5. *Eupeyerimhoffia archimedis* (Strasser 1965) male and female, SEM. **A:** Leg-pair 1, male, right side, posterior view; **B:** Leg-pair 2, female, right side, posterior view; **C:** Leg-pair 2, male, left side, posterior view; **D:** Leg-pair 2, female, right vulva, ventral view; **E:** Leg-pair 2, male gonopore, posterior view; **F:** Leg- pair 9, male, right side, posterior view; **G:** Leg-pair 17, left side, anterior view; **H:** Leg-pair 18, right side, posterior view; **I:** Telopod with syncoxite, anterior view; **J:** Telopod, inner horn of syncoxite; **K:** Telopod, posterior view. **Abbreviations:** Cx = Coxa; Pre = Prefemur; Fem = Femur; Post = Postfemur; Tib = Tibia; Tar = Tarsus; Vu = Vulva; Go = Gonopore; Mp = Median plate; Lp = Lateral plate; Op = Operculum; 1-4 = Podomere 1-4; cL = Central lobe; h = Inner horn; Scx = Syncoxite; Tr = Trichostele. Scale bar: 400 µm (**A,B**); 250 µm (**C**); 100 µm (**D,H**); 25 µm (**E**); 200 µm (**F,I**); 150 µm (**G,K**); 50 µm (**J**).

Male leg 17 reduced with 4 podomeres (Fig. 2.5G). Coxa with small medial process bearing a subapical spine and a wide but narrow, almost triangular, coxal lobe. Apical edge of coxa with small protrusion. Podomere 1 with mesal spine. Second podomere approximately 1.8 times wider than podomere 3. Podomere 3 very short and inconspicuous. Podomere 4 with subapical spine. Complete leg sparsely covered with minute setae.

Male leg 18 reduced, but to a lesser extent than leg-pair 17 (Fig. 2.5H). Coxa slightly damaged during dissection, but apparently without widened coxal lobe and not fused to syncoxite. Small mesal coxal process with single subapical spine. Single, well-developed medial spine on podomere 1. Apical edge of podomere 1 with apical protrusion. Podomere 2 approximately 1.5 times as wide as podomere 3. Podomere 3 very short, no spines and with very inconspicuous borders to podomere 4. Podomere 4 with apical spine.

Telopod (male leg 19) stout, syncoxite likewise (Fig. 2.5I–K). Syncoxal lobe small and rounded. Inner horns of syncoxite with numerous hairs of varying length and well-developed subapical spine, which is curved almost 90° (Fig. 2.5J). Telopoditomere 1 with mesal, highly reduced trichostele (Fig. 2.5I, J). Telopoditomere 2 mesally elongated into large bulbous process (immovable finger) with knobbed proximal surface. Telopoditomere 3 short, approximately 2 times wider than long, devoid of any peculiarities. Telopoditomere 4 with medial field of knobs juxtaposed to process of telopoditomere 2, and a large posteriorly oriented spine. Telopoditomere 4 forms chela (pincer) against medial process of second telopoditomere.

Female sexual characters.

Female second leg-pair similar to midbody legs, but coxa with two spines on

separate medial protrusions which are fused basally (Fig. 2.5B). Tarsus 4 times longer than wide.

Female vulva large, attached to posterior side of coxa via membranes (Fig. 2.5D). Operculum recessed between vulva plates. Posterior end of operculum narrower than anterior one, with two spines. Vulva with ventrally symmetrical mesal and lateral plates, carrying altogether 7 or 8 spines. Lateral plate overlaps mesal one apically on posterior side, but both are fused together at their base via a membranous connection.

Intraspecific variation. Not enough samples present to describe morphological variation. The populations from Ferla and Palazzolo Acreide have two different haplotypes, differing at one base pair position.

Volvation. As described by Strasser (1965), the genus has a volvation strategy, which differs from what is known from most other Glomerida. When rolling up into a ball the ventral ends of the tergites are not inserted in the schisma of the thoracic shield, unlike in *Glomeris*, but rest on top of it (Fig. 2.1A). The pegs on the ventral side of the tergites (Fig. 2.4D, F) rest within the depressions on the ventral side of the thoracic shield (Fig. 2.4B, C).

2.4 Discussion

2.4.1 Problems during the morphological analysis

The described position of the vulva operculum might be an artifact, as the structural integrity of membranous structures was not preserved by critical point drying. This should be considered for future studies of glomerid vulvae.

The sampling within the present study did not allow for a description of the morphological variation within the species. However, the 0.2% sequence divergence

between the two reported localities shows that there is variability within the species, with at least two haplotypes present on the island.

2.4.2 Unique morphological characters of *Eupeyerimhoffia*

Eupeyerimhoffia archimedis shows several interesting characters. The mandible with a large condylus and flat molar plate lacking a groove (Fig. 2.3G) is very atypical of Glomerida. Glomerida are generally described as lacking a condylus and always possessing a molar plate with a distinct deep groove (Köhler and Alberti 1990). Furthermore, contrary to previous descriptions, the Protoglomeridae-like telopods possess a trichostele on the first podomere, which represents another special character of the species, if not of the genus. The presence of this trichostele violates the diagnosis of the family as proposed by Mauriès (2005).

2.4.3 Volvation strategy

The volvation strategy of *Eupeyerimhoffia* is another striking and possibly unique feature of the genus inside the order Glomerida. Similar pegs on the tergites have been reported for members of the genera *Epiromeris* (Thaler and Knoflach-Thaler 1998) and *Trachysphaera* (Strasser, 1965). These do not, however, possess the herein described thoracic shield lobe with a reduced schisma in combination with ventral depressions (Fig. 2.4A–C). Both modifications allow *Eupeyerimhoffia* a unique method of rolling into a ball. To understand this phenomenon, further inquiries into the origin and diversification of glomerid volvation are necessary and jointly represent a very interesting future research topic on its own.

2.4.4 Relationships of the four genera of the Protoglomeridae and impact on Glomerida phylogeny

As stated above, the COI fragment is not well suited to study the group's phylogeny. Therefore, it is not surprising that the COI tree lacks resolution and receives little statistical support. Nonetheless, together with the distance analysis, it is sufficient to observe that the members of the family Protoglomeridae are not each other's closest relatives (e.g. *Glomeroides primus* grouping with *Glomeridella minima* from the separate suborder Glomeridelloidea) and possibly that the family does not constitute a monophyletic unit. Similar results have also been reported by Wesener (2012) in a study that did not include all members of the family. This supports the notion that characters based mainly on the telopods are not sufficient to infer relationships within the order Glomerida. This is especially true when considering the close relationship between *Eupeyerimhoffia archimedis* and *Protoglomeris vasconica*, despite the fact that *E. archimedis* does not conform to the diagnosis of the family. Therefore, a phylogenetic analysis based on a much broader dataset, including further molecular markers and morphological characters, is required to illuminate the evolutionary history of the pill millipedes.

2.5 Acknowledgements

We thank K. Ulmen (ZFMK: SEM) and C. Etzbauer (ZFMK: Molecular lab) for help in the lab, H. Reip (Jena) for the specimens of *Protoglomeris vasconica*, S. Simaiakis (NHMC) for the specimens of *Glomerellina laurae*, and A. Schönhofer for the specimens of *Glomeroides primus*. We also thank P. Schwendinger (MHNG) for his advice and for hosting one of us (TW) during a stay at Geneva in the summer of 2012.

We thank the Alexander-Koenig-Gesellschaft (AKG) for funding the trip to re-collect *Eupeyerimhoffia archimedis* in Sicily and P. Erkeling for his help in the field. Sergei Golovatch and a second reviewer provided useful comments that enhanced the quality of the here presented article. Lastly, we thank P.B. Frandsen (Rutgers University, USA) for proofreading and giving constructive comments on the manuscript.

2.6 References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 25: 3389–402.

Brölemann HW (1913) Eupeyerimhoffia algerina nouvelle forme de Gloméride (Myriapodes). Bulletin de la Société d'Histoire naturelle de l'Afrique du Nord. 5: 166–174.

Felsenstein J (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution. 39: 783.

Foddai D, Minelli A, Scheller U, Zapparoli M (1995) Chilopoda, Diplopoda, Pauropoda, Symphyla. In: Minelli A, Ruffo S, La Posta S (Eds.) Checklist delle Specie della Fauna Italiana. Bologna. 32/33: 1–35 pp.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular marine biology and biotechnology. 3: 294–9.

Golovatch SI, Mikhaljova EV, Chang HW (2010) Pill-millipedes (Glomerida, Diplopoda) in Taiwan. Zootaxa. 2477: 1–20.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor. Nucleic Acids Symposium Series. 41: 95–98.

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution. 22: 160–174.

Hoffman R (1980) Classification of the Diplopoda. Museum d'Histoire Naturelle, Genève. 1–237 pp.

Jeekel CAW (1971) Nomenclator generum et familiarum Diplopodorum: a list of the genus and family-group names in the class Diplopoda from the 10th edition of Linnaeus, 1758, to the end of 1957. Monografieen van de Nederlandse Entomologische Vereniging. 5: 1–412.

Kime RD, Enghoff H (2011) Atlas of European millipedes (Class Diplopoda): Volume 1, orders Polyxenida, Glomerida, Platydesmida, Siphonocryptidae, Polyzoniida, Callipodida, Polydesmida. Pensoft, Sofia-Moscow. 282 pp.

Köhler HR, Alberti G (1990) Morphology of the mandibles in the millipedes (Diplopoda, Arthropoda). Zoologica Scripta. 19: 195–202.

Mauriès JP (1971) Diplopodes épigés et cavernicoles de Pyrénées espagnoles et des monts cantabriques: VII. Glomérides. Essai de classification des Glomeroidea. Bulletin de la Société d'Histoire naturelle de Toulouse. 107: 423–453.

Mauriès JP (2005) Attempt at a classification of Glomerida (Diplopoda), with descriptions of two new genera from the northwest of the Iberian Peninsula. Arthropoda Selecta. 14: 241–249.

Saiki R, Gelfand D, Stoffel S, Scharf S, Higuchi R, et al. (1988) Primer- directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 239: 487–491.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences. 74: 5463–5467.

Shelley RM, Golovatch SI (2011) Atlas of Myriapod Biogeography. I. Indigenous Ordinal and Supra-Ordinal Distributions in the Diplopoda: Perspectives on Taxon Origins and Ages, and a Hypothesis on the Origin and Early Evolution of the Class. Insecta Mundi. 0158: 1–134.

Shelley RM, Sierwald P, Kiser SB, Golovatch SI (2000) Nomenclator generum et familiarum Diplopodorum II. Sofia.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, et al. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America. 87: 651–701.

Strasser K (1965) I diplopodi della fauna siciliana. Atti della Accademia gioenia di Scienze naturali in Catania. 7: 1–35.

Strasser K (1970) Diplopodi della Sicilia e della Calabria. Memorie del Museo civico di Storia naturale di Verona. 17: 151–200.

Strasser K (1976) Über Diplopoda-Chilognatha Griechenlands, II. Revue suisse de Zoologie. 83: 579–645.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood,

evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution. 28: 2731–9.

Thaler K, Knoflach-Thaler B (1998) Das Männchen von *Epiromeris aelleni* Strasser (Diplopoda, Glomerida). Entomologische Nachrichten und Berichte. 42: 1–4.

Wesener T (2010) Revision of the American Pill Millipedes I: *Onomeris* and *Trichomeris*. Zootaxa. 2725: 28–40.

Wesener T (2012) *Nearctomeris*, a new genus of Pill Millipedes from North America, with a comparison of genetic distances of American Pill Millipede Genera (Glomerida, Glomeridae). Zootaxa. 68: 58–68.

Wesener T, Raupach MJ, Sierwald P (2010) The origins of the giant pill-millipedes from Madagascar (Diplopoda: Sphaerotheriida: Arthrosphaeridae). Molecular Phylogenetics and Evolution. 57: 1184–93.

Chapter 3

A first phylogenetic analysis of the pill millipedes of the order Glomerida, with a special assessment of mandible characters (Myriapoda, Diplopoda, Pentazonia)

This chapter is published in:
Oeyen JP, Wesener T (2018) A first phylogenetic analysis of the pill millipedes of the order Glomerida, with a special assessment of mandible characters (Myriapoda, Diplopoda, Pentazonia). Arthropod Structure and Development. 47: 214–228. doi: 10.1016/j.asd.2018.02.005

Author contributions to the original article:
Conceived the study: JPO, TW; Collected specimens: JPO, TW; Disections and SEM: JPO; Compiled the character matrix: JPO; Performed the analyses: JPO; Interpreted the results: JPO; Wrote the manuscript: JPO, TW.

Figure and table numbers have been adapted for the inclusion in the present thesis.

3.1 Introduction

The millipedes (Diplopoda) constitute the third largest class of terrestrial arthropods with more than 11,000 described species and are the most species-rich group of the Myriapoda (Enghoff et al. 2015). Most species are soil-dwelling and play a crucial role as macroinvertebrates in soil decomposition and nutrient cycling (Ashwini and Sridhar 2002, Curry 1994, Crawford 1992, Wolters and Ekschmitt 1997). The Diplopoda, especially in comparison to centipedes (e.g. Murienne et al. 2010) and insects (e.g. Beutel et al. 2011), are poorly known in terms of phylogenetically informative morphological characters (Blanke and Wesener 2014). To date, only four interordinal phylogenetic analyses have been conducted based on morphological data (Enghoff 1984, Sierwald et al. 2003, Sierwald and Bond 2007, Blanke and Wesener 2014) and merely four of the 16 Diplopoda orders have been partially classified based on phylogenetic analyses: Polydesmida (Simonsen 1990), Julida (Enghoff 1991), Sphaerotheriida (Wesener and VandenSpiegel 2009, Wesener 2014), and Spirobolida (Wesener et al. 2008, Pitz and Sierwald 2010). Consequently, the relationships within the Diplopoda are still highly disputed and many orders, including the Glomerida, remain poorly defined, as their apomorphic characters have not been inferred.

The order Glomerida, commonly called pill millipedes, are a Holarctic group distributed in North America, Europe, and of Asia (Shelley and Golovatch 2011). The pill millipedes are moderately diverse, currently comprising more than 300 species and 34 genera (Enghoff et al. 2015). The first comprehensive classification of the order, including all known genera, was presented by Mauriés (1971). The proposed typological system is based largely on characters of the modified last male leg-pair,

the telopods, which were studied and illustrated for most known genera. Although the system was taxonomically comprehensive, it did not distinguish between apomorphies and plesiomorphies. A vastly different system was proposed by Hoffman (1980), who argued that telopod characters were subject to convergent reductions and therefore not informative. However, Hoffman did not provide any morphological characters to support his proposed system, making it impossible to verify. The original classification by Mauriés has since been updated twice (Mauriés 2005, Enghoff et al. 2015) and currently separates the three families Glomeridellidae, Protoglomeridae, and Glomeridae (Enghoff et al. 2015). The family Glomeridae is further subdivided into the four subfamilies Glomerinae, Haploglomerinae, Doderiinae, and Mauriesiinae (Enghoff et al. 2015). However, the underlying character selection, as well as the character discussion, remains unchanged. First molecular analyses have shown that at least one of the families (Protoglomeridae) likely does not represent a monophyletic unit (Wesener 2012, Oeyen and Wesener 2015). These results support the view that telopod characters alone might not be sufficient to establish a phylogenetic system for the order and that additional characters are sorely needed to further our understanding of the group.

Mandible characters have long been and still are of great importance in higher-level classifications of arthropods, especially concerning the position of the Myriapoda (Snodgrass 1950, Manton 1964, Lauterbach 1972, Wägele 1993, Edgecombe et al. 2003). Despite its historical prominence and some authors arguing strongly for its usefulness in the classification of the Diplopoda (Ishii and Tamura 1995), the mandible has only been utilized in few phylogenetic or taxonomic studies within the

class. Strong differences between orders have been noted in the structure of both the basal joint (Blanke and Wesener 2014) and the gnathal lobe (Köhler and Alberti 1990, Ishii and Tamura 1995). The mandibles show family-specific characters in Polydesmida (Ishii and Tamura 1996) and have been successfully exploited as taxonomic characters in the orders Julida (Enghoff 1979) and Polyxenida (Ishii 1988, Ishii and Liang 1990). However, little variation was found in the orders Spirobolida (Wesener et al. 2008, Wesener et al. 2009) and Sphaerotheriida (Wesener and VandenSpiegel 2009, Wesener 2014, Wesener 2016). In the order Glomerida, the mandible has generally been ignored as a character in both taxonomic and systematic studies. Published SEM images only exist for six species from two families (Glomeridae and Protoglomeridae). These include two species of *Glomeris* (Glomeridae: Glomerinae; Köhler and Alberti 1990), and a single species of *Hyleoglomeris* (Glomeridae: Doderiinae; Ishii and Tamura 1995), *Onomeris* (Glomeridae: Doderiinae; Wesener 2010), *Nearctomeris* (Glomeridae: Doderiinae; Wesener 2012), and *Eupeyerimhoffia* (Protoglomeridae; Oeyen and Wesener 2015). Hence, the mandible of the Glomerida still represents an unexploited and potentially valuable source of informative characters.

In the present study we search for mandibular characters within the order Glomerida and examine their variation based on scanning electron microscopy images. Our analysis incorporates 22 species of 20 genera, covering all but one of the currently proposed subfamilies. We combine the novel mandibular characters with characters from the published literature to conduct the first provisional phylogenetic analysis for the Glomerida. Furthermore, by incorporating the two other pentazonian taxa,

Glomeridesmida and Sphaerotheriida, we are for the first time able to infer order-level apomorphies for the Glomerida.

3.2 Material and methods

3.2.1 Taxon selection

In order to provide insight into the phylogeny of the Glomerida, species from as many different suborders, families, subfamilies, and genera as possible were included in the present study. The sampling includes the two genera of the species-poor suborder Glomeridelloidea, all four genera of the basal family Protoglomeridae, and three out of the four Glomeridae subfamilies, only excluding the monogeneric Mauriesiinae. The two mega-diverse genera *Glomeris* (see Golovatch et al. 2009) and *Hyleoglomeris* (see Golovatch et al. 2006, Golovatch et al. 2013), which together encompass more than half of the species diversity of the order, are each present with two species. In total, the analysis encompasses 22 species from 20 out of the 34 known genera (Table 3.1; Enghoff et al. 2015).

In order to identify potential apomorphies of the Glomerida, representatives of the two other orders in the subclass Pentazonia, Sphaerotheriida and Glomeridesmida (Blanke and Wesener 2014), were included as near-outgroups. The sampling includes both Glomeridesmida families, Glomeridesmidae and Termitodesmidae, as well as two of the five Sphaerotheriida families (Wesener 2014), Procyliosomatidae and Arthrosphaeridae (Table 3.1). In order to polarize the characters within the Pentazonia, two representatives of the subclass Helminthomorpha, both belonging to the order Spirobolida, were included as far-outgroups (Table 3.1). The Spirobolida were selected as the far-outgroup as high quality SEM images are available for all

examined characters, which minimizes the amount of missing data in the character matrix.

Table 3.1: Material examined, with classification and voucher depository. Abbreviations: FMNH-INS - Field Museum of Natural History, Invertebrate collection; MNHN - Museum nationale d'histoire naturelle, Paris; SWUNM - Srinakharinwirot University Natural History Museum, Bangkok; ZMB - Zoologisches Museum Berlin; ZFMK - Zoological Research Museum A. Koenig, Bonn.

Classification	**Taxon**	**#**
Far outgroup		
Spirobolida, Pachybolidae, Pachybolini	*Madabolus maximus* Wesener & Enghoff, 2008	FMNH-INS 5466
Spirobolida, Pachybolidae, Atopochetini	*Pseudocentrobolus aureus* Wesener, 2009	FMNH-INS 5495
Outgroup		
Sphaerotheriida, Procyliosomatidae	*Procyliosoma leae* Silvestri, 1917	QVMAG 23:45802
Sphaerotheriida, Arthrosphaeridae	*Sphaeromimus vatovavy* Wesener, 2014	MNHN
Glomeridesmida, Glomeridesmidae	*Glomeridesmus* sp.	FMNH-INS 11916
Glomeridesmida, Termitodesmidae	*Termitodesmus ceylonicus* Silvestri, 1911	ZMB 5117
Ingroup		
Glomeridelloidea, Glomeridellidae	*Glomeridella minima* Latzel, 1884	ZFMK MYR 856
Glomeridelloidea, Glomeridellidae	*Typhloglomeris* sp.	ZFMK MYR 2261
Glomeroidea, Protoglomeridae	*Protoglomeris vasconica* Brölemann, 1897	ZFMK MYR 934
Glomeroidea, Protoglomeridae	*Eupeyerimhoffia archimedis* (Strasser, 1965)	ZFMK MYR 1875
Glomeroidea, Protoglomeridae	*Glomerellina laurae* Silvestri, 1908	ZFMK MYR 2260
Glomeroidea, Protoglomeridae	*Glomeroides prima* Silvestri, 1929	ZFMK MYR 2440
Glomeroidea, Glomeridae, Doderiinae	*Adenomeris* cf. *hispida* Ribaut, 1909	ZFMK MYR 82
Glomeroidea, Glomeridae, Doderiinae	*Epiromeris aelleni* Strasser, 1976	ZFMK MYR 163
Glomeroidea, Glomeridae, Doderiinae	*Hyleoglomeris* sp. 1	SWUNM-MYR-D0001
Glomeroidea, Glomeridae, Doderiinae	*Hyleoglomeris* sp. 2	ZFMK MYR 820
Glomeroidea, Glomeridae, Doderiinae	*Rhopalomeris* sp	ZFMK MYR 5553
Glomeroidea, Glomeridae, Doderiinae	*Nearctomeris inexpectata* Wesener, 2012	ZFMK MYR 11
Glomeroidea, Glomeridae, Doderiinae	*Onomeris sinuata* Loomis, 1943	ZFMK MYR 1564
Glomeroidea, Glomeridae, Doderiinae	*Trachysphaera* sp.	ZFMK MYR 895
Glomeroidea, Glomeridae, Doderiinae	*Geoglomeris subterranea* Verhoeff, 1908	ZFMK MYR 742
Glomeroidea, Glomeridae, Haploglomerinae	*Haploglomeris multistriata* (Koch, 1844)	ZFMK MYR 245
Glomeroidea, Glomeridae, Haploglomerinae	*Simplomeris montivaga* (Faes, 1902)	ZFMK MYR 2626
Glomeroidea, Glomeridae, Haploglomerinae	*Schismaglomeris occultocolorata* (Verhoeff, 1892)	ZFMk MYR 4953
Glomeroidea, Glomeridae, Glomerinae	*Glomeris distichella* Berlese, 1887	ZFMK MYR 1895
Glomeroidea, Glomeridae, Glomerinae	*Glomeris marginata* (Viller, 1789)	ZFMK MYR 2255
Glomeroidea, Glomeridae, Glomerinae	*Loboglomeris* sp.	ZFMK MYR 239
Glomeroidea, Glomeridae, Glomerinae	*Onychoglomeris fagi* Verhoeff, 1930	ZFMK MYR 624

3.2.2 Dissection and imaging

Samples were processed for the scanning electron microscope following the procedure previously published in Oeyen and Wesener (2015).

3.2.3 Micro-CT scan and three-dimensional visualization

In order to observe the position of the mandible in situ, single specimens of *Eupeyerimhoffia archimedis* (ZFMK MYR5100) and *Rhopalomeris* sp (ZFMK MYR5552) were scanned in a Skyscan 1272 system (Bruker microCT, Kontich, Belgium). The specimens were scanned in 95% ethanol without prior dissection. Settings were: source voltage = 60 kV, source current = 160 mA, 180° rotation, angular step size = 0.1°, exposure time = 3417 ms, no binning, filter = Aluminum 0.25 mm, averaging = 7, random movement = 15, 1896 projections + reference scans, pixel size = 2.5 mm, flat field correction = ON. Thermal drift correction, ring artifact reduction and digital section reconstruction was done in NRecon 1.7.0.4 (Bruker microCT, Kontich, Belgium), reducing the pixel size to 10 mm. The resulting image stack was cropped in Data Viewer 1.5.2.4 (Bruker microCT, Kontich, Belgium) and converted to 8bit TIFF in ImageJ (Abràmoff et al. 2004). Subsequent volume renderings were done in Drishti 2.6 (Limaye 2012). The resulting volume was clipped to show the right mandible in mesal view (see Fig. 3.1C). All utilized software and packages are published under licenses that allow free use in non-commercial organizations and research. Original mCT scans of both species are deposited in the MorphoBank database (O'Leary and Kaufman 2012; Project number 3105).

3.2.4 Character selection

A total of 69 characters was included in the present study. Characters of the Pentazonia were taken from Blanke and Wesener (2014). Other characters, especially concerning the Sphaerotheriida, were taken from Wesener (2014, 2016). Numerous characters were taken from the published literature, especially from previous classifications of the Glomerida (e.g. Verhoeff 1928, Hoffman 1982. Mauriés 1971, Mauriés 2005). Eleven mandible characters were directly scored from images. All characters are discussed in detail (Table 3.2).

Table 3.2: Character discussion. BW# = character numbers in Blanke and Wesener (2014).

#	**Description:**
Head characters:	
1	*Head capsule*: (0) with occipital phragma; (1) lacking a occipital phragma (BW2). See Blanke & Wesener (2014) for a discussion of this character. We have unpublished data on Glomeridesmida and Sphaerotheriida, that show its absence. For the Glomerida taxa investigated here, the head was dissected in all specimens, without the discovery of a occipital phragma. However, in all pentazonians, markedly in Glomerida, paired triangular endoskeletal processes arise from the occiput that are structurally and functionally comparable to the internal pseudoccipital processes of the occipital phragma in helminthomorph millipedes (see, e.g., Silvestri, 1903, his Figs. 21-22; Verhoeff, 1928: 771; Dohle, 1964: 270).
2	*Head capsule, parietal sclerite (genae)*: (0) present; (1) absent (BW1).
3	*Tentorium, transverse bar*: (0) present; (1) reduced. A transverse bar is completely reduced in the Sphaerotheriida, despite the presence of an incisura lateralis (Koch, 2015; Sagorny and Wesener, 2017).
4	*Labrum*: (0) with three teeth; (1) with a single, triangular, anteriomedian tooth (BW3).
5	*Head, large lateral antennal grooves*: (0) absent; (1) present. Large lateral antennal grooves are only present in the order Sphaerotheriida (Wesener and VandenSpiegel, 2009; Wesener, 2016). Head grooves also exist as shallow invaginations in some Glomerida and Spirobolida, but never as special lateral pits on the head.
6	*Antennae placed on head*: (0) laterally, laterally of organ of Tömösváry; (1) centrally, always mesally of organ of Tömösváry. Only in the Glomerida and Glomeridesmida the antennae are located centrally, mesally of the organ of Tömösváry (Wesener, 2016).
7	*Antenna, apical disc, number of apical cones*: (0) 4; (1) 10–200. The number of antennal cones in Diplopoda is usually 4, but this varies greatly between species and genera in numerous orders, especially in the order Sphaerotheriida (Wesener and VandenSpiegel 2009; Wesener, 2016).

8	*Antenna, distal segment*: (0) normal; (1) swollen. In our data set, inside the Glomerida only the genus *Rhopalomeris* has an apically swollen distal antennomere (Mauriès, 2005).
9	*Epipharynx*: (0) loosely attached to head capsule; (1) fixed to head capsule. See Koch (2015). This character was investigated in all analyzed taxa.
10	*Eyes*: (0) present; (1) absent (BW4). All Glomeridesmida are blind, while all Sphaerotheriida have numerous ommatidia. In Glomerida, blind genera and species exist, usually those adapted to cave life or deeper soil layers.
11	*Eyes, if present*: (0) with numerous, 50–90 ommatidia; (1) with less than 10 ommatidia arranged in 1-2 rows (BW5). All Sphaerotheriida and many Helminthomorpha groups feature eyes consisting of numerous, often more than 50 ommatidia. The Glomerida with eyes usually feature 5–7, rarely up to 9 (7+2) ommatidia (Enghoff et al., 2015).
12	*Tömösváry organ*: (0) absent; (1) present (BW6).
13	*Tömösváry organ*: (0) small, at most twice as large as single ocellus; (1) very large (BW7).
14	*Gula of gnathochilarium*: (0) normal-sized; (1) very large. The gula of the gnathochilarium is greatly enlarged in species of the orders Glomerida and Glomeridesmida (Wesener and VandenSpiegel, 2009, own observations).
15	*Gnathochilarium, lateral palpi*: (0) present; (1) rudimentary. The lateral palpi are reduced in the Sphaerotheriida (Wesener and VandenSpiegel, 2009).
16	*Gnathochilarium, lingual palps*: (0) present, with sensory cones; (1) replaced by central pads, devoid of sensory cones. Lingual palps are transformed to pads without any sensory cones in the orders Glomerida and Sphaerotheriida (Wesener and VandenSpiegel, 2009; Koch, 2015).
17	*Gnathochilarium, cardines*: (0) small, (1) large. The cardines are small in the Glomeridesmida and Sphaerotheriida as well as in most Helminthomorpha, but greatly enlarged in the Glomerida (Wesener and VandenSpiegel, 2009).
18	*Gnathochilarium, mentum and lamella linguales*: (0) separate; (1) fused, with suture still visible. The lamellae linguales are fused to the mentum in the Glomeridesmida and Sphaerotheriida (Wesener and VandenSpiegel, 2009).
19	*Gnathochilarium, parts of mentum*: (0) always located basally to apical half of lamella linguales; (1) located laterally to apical half of lamella linguales. Only in the Sphaerotheriida parts of the mentum are also surrounding the lamella linguales laterally (Wesener and VandenSpiegel, 2009).
20	*Gnathochilarium, mentum*: (0) undivided; (1) divided. Only in the Glomerida the mentum is divided, a unique character for the Diplopoda (Enghoff, 1990; Koch, 2015).
21	*Gnathochilarium, endochilarium, sensory cones*: (0) two fields with normal cones; (1) one continuous field with cones and small granules. Only in the Sphaerotheriida the sensory cones of the endochilarium are arranged in a continuous field (Wesener and VandenSpiegel, 2009).
22	*Gnathochilarium, maxillary nephridia*: (0) open on suture between stipes and lamellae lingulaes; (1) open on the endochilarium. The maxillary nephridia open on the endochilarium only in the order Sphaerotheriida

	(Verhoeff, 1928; Koch, 2015, own observations).
23	*Gnathochilarium, lamellae linguales*: (0) free; (1) fused (BW12).
24	*Mandible consisting of*: (0) single basal joint plus gnathal lobe; (1) two basal joints, cardo and stipes (BW13).
25	*Mandible, basal article and gnathal lobe*: (0) connected via two joints; (1) lacking the condylic joints, only syndetic joint is present (BW14). Slightly rephrased utilizing the terminology of Koch (2015).
26	*Mandible, inner tooth, number of cusps*: (1) 3; (2) 4; (3) 5. Outgroup data taken from the literature: Spirobolida (Wesener et al., 2008, 2009); Sphaerotheriida (Wesener and VandenSpiegel, 2009; Wesener et al., 2014b); Glomeridesmida (Iniesta et al., 2012).
27	*Mandible, apical (lateral) tooth*: (0) present; (1) absent. A free lateral tooth is present in the Spirobolida (Wesener et al., 2008, 2009). Outgroup data taken from the literature: Spirobolida (Wesener et al., 2008, 2009); Sphaerotheriida (Wesener and VandenSpiegel, 2009; Wesener et al., 2014b); Glomeridesmida (Iniesta et al., 2012).
28	*Mandible, molar plate*: (0) without groove; (1) with groove. Outgroup data taken from the literature: Spirobolida (Wesener et al., 2008, 2009); Sphaerotheriida (Wesener and VandenSpiegel, 2009; Wesener et al., 2014b); Glomeridesmida (Iniesta et al., 2012). For Glomerida characters, see Figs 2-5.
29	*Mandible, molar plate*: (0) without membranous fringe; (1) with a short membranous fringe; (2) with a very long membranous fringe. Outgroup data taken from the literature: Spirobolida (Wesener et al., 2008, 2009); Sphaerotheriida (Wesener and VandenSpiegel, 2009; Wesener et al., 2014b); Glomeridesmida (Iniesta et al., 2012). Only Glomeridesmida have a strongly elongated fringe. For Glomerida characters, see Figs. 2-5. The membranous fringe is often 'rubbed down' in older individuals. The fringe was scored as present if small remains were visible at part of the mandible.
30	*Mandible, molar plate, surface*: (0) 'riffled' into several teeth; (1) relatively even and flat. A 'riffled' mandible is present in several Helminthomorpha (e.g. Ishii and Tamura, 1996), including the species of Spirobolida utilized here. Outgroup data taken from the literature: Spirobolida (Wesener et al., 2008, 2009); Sphaerotheriida (Wesener and VandenSpiegel, 2009; Wesener, 2014); Glomeridesmida (Iniesta et al., 2012).
31	*Mandible, molar plate, condylus*: (0) absent; (1) large and well-developed; (2) small, only a bump. A mandible condylus is a potential apomorphy of the Pentazonia (Wesener and VandenSpiegel, 2009; Blanke and Wesener, 2014; Koch, 2015). However, in most Glomerida, the condylus is only present as a small bump (Figs. 2-5).
32	*Mandible, rows of pectinate lamellae*: (0) 7 or more; (1) 6 or less. Outgroup data taken from the literature: Spirobolida (Wesener et al., 2008, 2009); Sphaerotheriida (Wesener and VandenSpiegel, 2009; Wesener, 2014); Glomeridesmida (Iniesta et al., 2012). For Glomerida characters, see Figs. 2-5.
33	*Mandible, intermediate area*: (0) large spines; (1) scale-like. Scale-like structures are apparently only present in species of the orders Glomeridesmida and Glomerida (Figs. 2-5). Outgroup data taken from the literature: Spirobolida (Wesener et a., 2008, 2009); Sphaerotheriida (Wesener and VandenSpiegel, 2009; Wesener, 2014); Glomeridesmida (Iniesta et al., 2012).
34	*Mandible, external tooth*: (0) single, sharp-edged tip; (1) serrated tip; (2) bilobed tip. The tip of the external tooth is modified in several species of the Glomerida (Figs. 2-5).

Body characters:	
35	*Body, number of tergites*: (0) 10; (1) 11; (2) 12; (3) 20; (4) more than 38. Our Helminthomorpha outgroup taxa of the order Spirobolida always have more than 38 tergites (Wesener et al., 2008). Fully grown Glomeridesmida always have 20 tergites (Enghoff et al., 2015), while all Sphaerotheriida have always 12 (Wesener and VandenSpiegel, 2009; Wesener, 2014). In Glomerida, species have either 10 or 11 tergites. As Palaeozoic taxa associated with the "Oniscomorpha" have 14 or 15 tergites (e.g. Hannibal, 1984; Racheboeuf et al., 2004), a reduction of tergites within the group is possible. However, we do not know if the state of the Glomeridesmida is the ancestral state, as the basal-most Diplopoda, the Polyxenida have less tergites than the Glomeridesmida, therefore multiple independent reductions are also possible. We code the number of tergites as independent characters, rather than as a transformation series. Data taken from the literature (Mauriès, 2005; Enghoff et al., 2015).
36	*Tergites and pleurites*: (0) free; (1) fused (BW30).
37	*Collum (tergite 1)*: (0) much smaller than following tergites; (1) slightly smaller than following tergites; (2) enlarged, covering posterior part of head, larger than following tergites (BW31).
38	*Tergite 2*: (0) almost as large as following tergites; (1) much larger than following tergites, called thoracic shield (BW32).
39	*Tergites, thoracic shield*: (0) smooth; (1) with pits. Large pits of unknown sensory function are present of the thoracic shield of the genera *Trachysphaera* (Wilbrandt et al., 2015) and *Adenomeris* (Blower, 1985). Such pits were also described in the genus *Tectosphaera*, which could not be incorporated into this analysis (Mauriès, 2005).
40	*Tergites*: (0) without transversal ridges of coagulated secretions; (1) with transversal ridges of coagulated secretions. Transversal ridges on the tergite with secretions are present in the genera *Trachysphaera* (Wilbrandt et al., 2015) and *Adenomeris* (Blower, 1985). Such ridges were also described in the genus *Tectosphaera*, which could not be incorporated into this analysis (Mauriès, 2005).
41	*Stigmatic plates*: (0) 'divided' or not fused; (1) fused to pleurites (BW19).
42	*Sternites at first leg pair*: (0) present; (1) absent. In the orders Glomerida and Sphaerotheriida the sternites are completely reduced, without any remnants being visible (Wesener et al., 2014a).
43	*Sternites at midbody segments*: (0) present; (1) completely reduced or at most only rudimentary. Real sternites are completely reduced in all Glomerida and Sphaerotheriida. In Glomeridesmida, remnants are still visible at the anterior and posterior leg pairs (Wesener et al., 2014a).
44	*Legs, coxal pouches*: (0) absent; (1) present (BW35). Coxal pouches are absent in all species of the Sphaerotheriida and Glomerida.
45	*Legs, tarsi, paronychium*: (0) absent; (1) present. In the dataset presented here, paronychia are only present in all species of the Glomeridesmida (Iniesta et al., 2012; Enghoff et al., 2015).
46	*8th leg in adult males*: (0) = walking leg; (1) accessory gonopod (BW45).
47	*9th leg in adult males*: (0) = walking leg; (1) functional gonopod (BW46).
48	*Development of male copulation legs (gonopods/telopods)*: (0) gradual; (1) abrupt (BW44). Males are still unknown for the family Termitodesmidae of the order Glomeridesmida,

	therefore their telopod development remains unknown.
49	*Spiracles on body segments 1 and 2*: (0) present; (1) absent (BW22).
50	*Stigma opening, location in mid-body segments*: (0) on a separate stigmatic plate; (1) on a plate fused with the coxa; (2) on a plate fused to the pleurites. In our Spirobolida outgroup, the stigma openings are located on stigmatic plates fused with the pleurites and tergites (Wesener et al., 2008, 2009), while the stigmatic plates are fused to the coxa in the Glomeridesmida, and free in the Glomerida and Sphaerotheriida (Wesener et al., 2014a).
51	*Tracheae*: (0) in tracheal bush; (1) dichotomous (BW23).
52	*Intestine*: (0) straight; (1) curved, N-shaped (BW43).
53	*Lateral defence glands*: (0) absent; (1) present (BW57).
54	*Unpaired mid-dorsal defence glands*: (0) absent; (1) present (BW58). Some genera of the Glomerida have mid-dorsal defence glands (Shear et al., 2011). Data about the presence or absence of defence glands was taken from the literature (Enghoff et al., 2015), or has been checked on actual specimens for T*yphloglomeris* sp. and *Glomerellina laurae*, which both have defence glands.
55	*Defence fluid*: (0) not containing benzochinone; (1) containing benzochinone (BW61). Here we extrapolate as our two analysed Spirobolida taxa did not have their defence fluids analyzed yet. Currently all analyzed defence fluids of Juliformia contain benzochinone in their defence fluids (Shear, 2015).
56	*Defence fluid*: (0) not containing quinazolinone alkaloids; (1) containing quinazolinone alkaloids (BW62). Data on this character is only available for two taxa of the Glomerida: *Glomeris marginata* (Schildknecht et al., 1966; Meinwald et al., 1966) and *Onomeris sinuata* (Shear et al., 2011). Defence secretions of other Glomerida have not been analyzed yet. Currently we assume that the defence fluids of all Glomerida with defence glands contain quinazolinone alkaloids. For Glomerida species without ozopores, this character was scored as inapplicable.
Sexual characters:	
57	*Male gonopore position*: (0) behind coxa 2, eversible; (1) "in" coxa 2, non-eversible (BW49). This character was slightly rephrased to be congruent with the observations made by Koch (2015).
58	*Female sexual opening location*: (0) attached posteriorly to coxa 2; (1) part of coxa 2, opening directly through the coxa. The female sexual opening is located attached the second coxa in most millipedes, in the Sphaerotheriida it opens directly through the coxa (Wesener and VandenSpiegel, 2009). This character was slightly rephrased to be congruent with the observations made by Koch (2015).
59	*Penultimate leg pair in females*: (0) normal walking leg; (1) weakly modified, with a narrow coxosternite and a spiracle opening located mesally instead of laterally. Only in the Glomeridesmida is the penultimate female leg specially modified (Iniesta et al., 2012; Enghoff et al., 2015).
60	*Last leg pair in females*: (0) normal walking leg; (1) heavily modified, separated from one another by a large undivided sternite, protruding above anal shield through special indentations of the last pleurite and tergite, fulfilling a sensory function. Only in the Glomeridesmida is the ultimate female leg pair modified and protruding posteriorly (Iniesta et al., 2012; Enghoff et al., 2015).
61	*Last male leg pair*: (0) unmodified; (1) modified into telopods, consisting of pincer-shaped,

	three- or four-jointed structures formed by the telopodites and inner horns formed by the fused syncoxite (BW48).
62	*Modified male penultimate leg pair into clasping organ, anterior telopod*: (0) absent; (1) present. An anterior telopod is present in all Sphaerotheriida (Wesener and VandenSpiegel, 2009), as well as in some members of the Glomerida suborder Glomeridelloidea (Mauriès 2005).
63	*Modified male penultimate leg pair*: (0) normal walking leg but shorter in size; (1) modified into anterior telopods, femur long; (2) modified into anterior telopods, femur wide or globular. In the Glomerida suborder Glomeridelloidea, the subfamily Glomeridellinae has anterior telopods with an elongated femur, while in the subfamily Typhloglomerinae, the femur is globular (Mauriès, 2005).
64	*Male telopod, podomere 1, trichostele*: (0) absent; (1) present. The protoglomerid genus *Eupeyerimhoffia* features a trichostele on the first podomere (Oeyen et al. 2015), as do all members of the family Glomeridae except the subfamily Haploglomerinae (Mauriès, 2005).
65	*Male telopod, shape*: (0) very long; (1) of 'usual' shape. Character taken from the literature. The defining characters of the subfamily Haploglomerinae are the elongated telopods (Mauriès, 2005).
66	*Male telopod, podomere 2, lamellar outgrowth*: (0) absent; (1) present. Character taken from the literature. The defining character of the family Glomeridae is the lamellar outgrowth at the telopod (Mauriès, 2005).
67	*Male telopod, podomere 2, lamellar outgrowth*: (0) 'normal' shaped; (1) strongly differentiated distocaudally. Character taken from the literature. The defining character of the family Doderiinae is a strongly differentiated lamellar outgrowth at the telopod (Mauriès, 2005), which is absent in the subfamilies Glomerinae and Haploglomerinae. This character could not be scored for the non-Glomeridae.
68	*Male telopod, podomere 3, trichostele*: (0) absent; (1) present. Character taken from the literature. Only members of the family Glomeridae have a trichostele at the third podomere of the telopod (Mauriès, 2005).
69	*Male telopod, podomere number*: (0) 3; (1) 4. Only the Glomerida genus *Glomeridella* has a telopod with 3 podomeres above the syncoxite, as does *Glomeridesmus* of the outgroup. All other Glomerida have 4 podomeres (Mauriès, 2005; Enghoff et al., 2015). In the Sphaerotheriida this character varies even inside some genera (Wesener and VandenSpiegel 2009; Wesener, 2014).

3.2.5 Phylogenetic analysis

The character matrix (Table 3.3) was built utilizing the software Mesquite 3.04 (Maddison and Maddison 2015). The final character set comprises 28 terminals and 69 characters (Table 3.3). All characters were scored as “unordered”. The complete character matrix is given in Supplementary File 3.1.

The maximum parsimony branch-and-bound tree search conducted in PAUP* 4.0b10

(Swofford 2002) recovered 53 minimum-length trees with a length of 106. A strict consensus tree was constructed from the minimum-length trees. A heuristic tree search under the tree-bisection-reconnection (TBR) branch-swapping algorithm, with an unlimited 'MaxTrees' setting, conducted in PAUP* 4.0b10 (Swofford 2002) recovered an identical topology. Accelerated transformation was used as the character optimization criterion. All but one of the characters were parsimony informative. To verify that local optima did not affect the obtained results, a ratchet 'island hop' analysis (Nixon 1999) was conducted in Winclada/Asado 1.7 (Nixon 2002) with slightly modified settings (20,000 replicates, one tree held on each step, with a random constraint level set to 10). The 'island-hop' analysis yielded 6 trees, of which the consensus did not differ from the tree obtained by PAUP*, except that the subfamily Haploglomerinae was recovered. Furthermore, a bootstrap analysis (Felsenstein 1985), incorporating 1000 pseudoreplicates, was undertaken in PAUP 4.0b10 under the TBR branch-swapping algorithm, with an unlimited setting for each replicate. To examine the dependence of the topology on single characters, a 1000 replicate Jackknife analysis, with 10 search repetitions, one starting tree per hold, and a maximum trees setting of 10,000 (Farris et al. 1996) was also undertaken in Winclada/Asado 1.7 (Nixon 2002). Bootstrap and Jackknife support values were plotted on the strict consensus tree. Characters were mapped to the inferred strict consensus tree using Winclada/Asado 1.7 (Nixon 2002; see Supplementary File 3.2). Unsupported nodes were collapsed before character optimization. Due to the unstable topology of the inferred tree, we only consider unambiguous changes, only discuss potential apomorphic characters, and unless stated otherwise, all described results are derived from the inferred strict consensus tree.

Table 3.3. Character matrix. Inapplicable characters coded as '-'. Data also available in Nexus file (Supplementary File 3.1).

Species \ Character #	1-10	11-20	21-30	31-40	41-50	51-60	61-69
Madabolus maximus	0000000001	00-0000000	0001010000	0100412000	1000011112	0010100000	0000-----
Pseudocentrobolus aureus	0000000001	00-0000000	0001010000	0000412000	1000011112	0010100000	0000-----
Procyliosoma leae	1111100011	0100110110	1110111101	1000200100	0110000000	1100--1100	110010-01
Sphaeromimus vatovavy	1111101111	0100110110	1110111101	1000200100	0110000000	1100--1100	110010-01
Glomeridesmus sp.	1101010010	-111000100	0010131021	1010301000	0011100001	1100--0011	100010-00
Termitodesmus ceylonicus	1101010010	-111000100	0010131021	1010302000	0011100001	1100--0011	-00010-0-
Glomeridella minima	1101010011	1111011001	0010121111	2012000100	0110000000	1101010000	111010-00
Typhloglomeris sp.	1101010011	1111011001	0010121111	2012100100	0110000000	1101010000	112010-01
Protoglomeris vasconica	1101010011	1111011001	0010121111	2011000100	0110000000	1101010000	100010-01
Eupeyerimhoffia archimedis	1101010011	1111011001	0010121001	1010000100	0110000000	1101010000	100110-01
Glomerellina laurae	1101010011	1111011001	0010131101	2112000110	0110000000	1101010000	100010-01
Glomeroides prima	1101010011	1111011001	0010121111	2012100100	0110000000	1101010000	100010-01
Adenomeris cf. *hispida*	1101010010	-111011001	0010131111	2112100111	0110000000	1100--0000	100111111
Epiromeris aelleni	1101010011	1111011001	0010131101	2010100100	0110000000	1101010000	100111111
Haploglomeris multistriata	1101010011	1111011001	0010121111	2010100100	0110000000	1101010000	100001011
Simplomeris montivaga	1101010011	1111011001	0010121111	2010100100	0110000000	1101010000	100001011
Schismaglomeris occultocolorata	1101010011	1111011001	0010121111	2012100100	0110000000	1101010000	100001011
Glomeris distichella	1101010011	1111011001	0010121111	2010100100	0110000000	1101010000	100111011
Glomeris marginata	1101010011	1111011001	0010111111	2010100100	0110000000	1101010000	100111011
Loboglomeris sp.	1101010011	1111011001	0010121111	2010100100	0110000000	1101010000	100111011
Onychoglomeris fagi	1101010011	1111011001	0010121111	2010100100	0110000000	1101010000	100111011
Hyleoglomeris sp. 1	1101010011	1111011001	0010121011	2012100100	0110000000	1101010000	100111111
Hyleoglomeris sp. 2	1101010011	1111011001	0010121111	2010100100	0110000000	1101010000	100111111
Rhopalomeris sp.	1101011111	1111011001	0010121111	2010100100	0110000000	1101010000	100111111
Nearctomeris inexpectata	1101010011	1111011001	0010121111	2012100100	0110000000	1101010000	100111111
Onomeris sinuata	1101010011	1111011001	0010121111	2012100100	0110000000	1101010000	100111111
Trachysphaera sp.	1101010011	1111011001	0010121111	2112000111	0110000000	1100--0000	100111111
Geoglomeris subterranea	1101010010	-111011001	0010121111	2112100100	0110000000	1101010000	100111111

3.3 Results

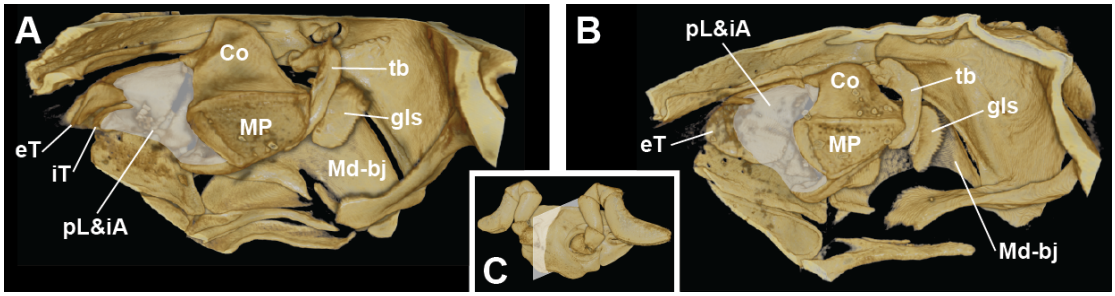

Figure 3.1. Glomerida mandible 3D reconstruction based on CT data. A: *Eupeyerimhoffia archimedis* (Strasser 1965); B: *Rhopalomeris* sp. C: *Rhopalomeris* sp., representation of sectional plane as applied in A and B. Not to scale. Abbreviations: Co = Condylus; eT = external tooth; gls = gnathal lobe sclerite; iA = intermediate area; iT = inner teeth; Md-bj = basal joint of mandible; pL = pectinate lamellae; tb = transverse bar of tentorium.

3.3.1 Mandible diversity in the Glomerida

The mandible of the Glomerida consists of an undivided basal joint and an apical gnathal lobe (Fig. 3.1A and B). A gnathal lobe sclerite, which acts as an attachment point for muscle tissue, is present basally (Fig. 3.1A and B). The gnathal lobe is divided into the molar plate and a distal part, separated by the intermediate area (Fig. 3.1A and B). The molar plate is a strongly sclerotized structure often with an apical groove, a membranous fringe, and dorsally with a condylus which seems to provide an additional point of articulation against the head capsule (Fig. 3.1A and B). The intermediate area is covered by cuticular scale-like projections (Figs. 3.2–3.5). A perimolar pore was observed in the intermediate area of some species (Figs. 3.3C, 3.4A, and 3.4C). The distal area consists of an external tooth, an internal tooth and a number of pectinate lamellae (Figs. 3.1–3.5). A considerable intraordinal variation of the mandibular structures was observed amongst the analyzed Glomerida genera. The external tooth can have a serrated (Fig. 3.2C), simple (Fig. 3.2D), or bilobed tip (Fig. 3.3A). The internal tooth can have 5 (Fig. 3.3A) and (Fig. 3.4A) or 4 cusps (Fig.

3.4C). The condylus on the molar plate can be well-developed (Figs. 3.1A and 3.2D) or less pronounced (Figs. 3.1B and 3.4B). The molar plate itself can be with (Fig. 3.3F) or without (Figs. 3.2D and 3.4D) a groove. A membranous fringe is present on the molar plate in several genera (Fig. 3.4E and F) and is absent in others (Figs. 3.2D–E and 3.3B).

3.3.2 Phylogeny of the Pentazonia

We recover the Pentazonia as a monophyletic group with unambiguous statistical support (Bootstrap = 100/Jackknife = 100; Fig. 3.6). The group is recovered based on 22 characters already listed in previous work (see Blanke and Wesener 2014). Inside the Pentazonia, the Glomeridesmida are recovered as the sister group to the classical Oniscomorpha, uniting the Sphaerotheriida and Glomerida with weak statistical support (62/52; Fig. 3.6). The far-outgroup, Spirobolida, is not recovered as monophyletic (see discussion).

The order Glomeridesmida, represented by both known genera (Termitodesmus and Glomeridesmus), is recovered as monophyletic with strong statistical support (100/99; Fig. 3.6). Four characters were recovered as potential apomorphies of the Glomeridesmida: (1) strongly elongated membranous fringe on the mandible molar plate (Table 3.2, c29); (2) body with 20 tergites (c35); (3) penultimate leg pair of females slightly modified, with a narrow coxosternite and a mesal stigma opening (c59), and (4) ultimate leg pair in females modified to fulfill a sensory function (c60). An additional two characters are recovered as unique for the Glomeridesmida, but are also present in other millipedes not incorporated in the present analysis (Blanke and

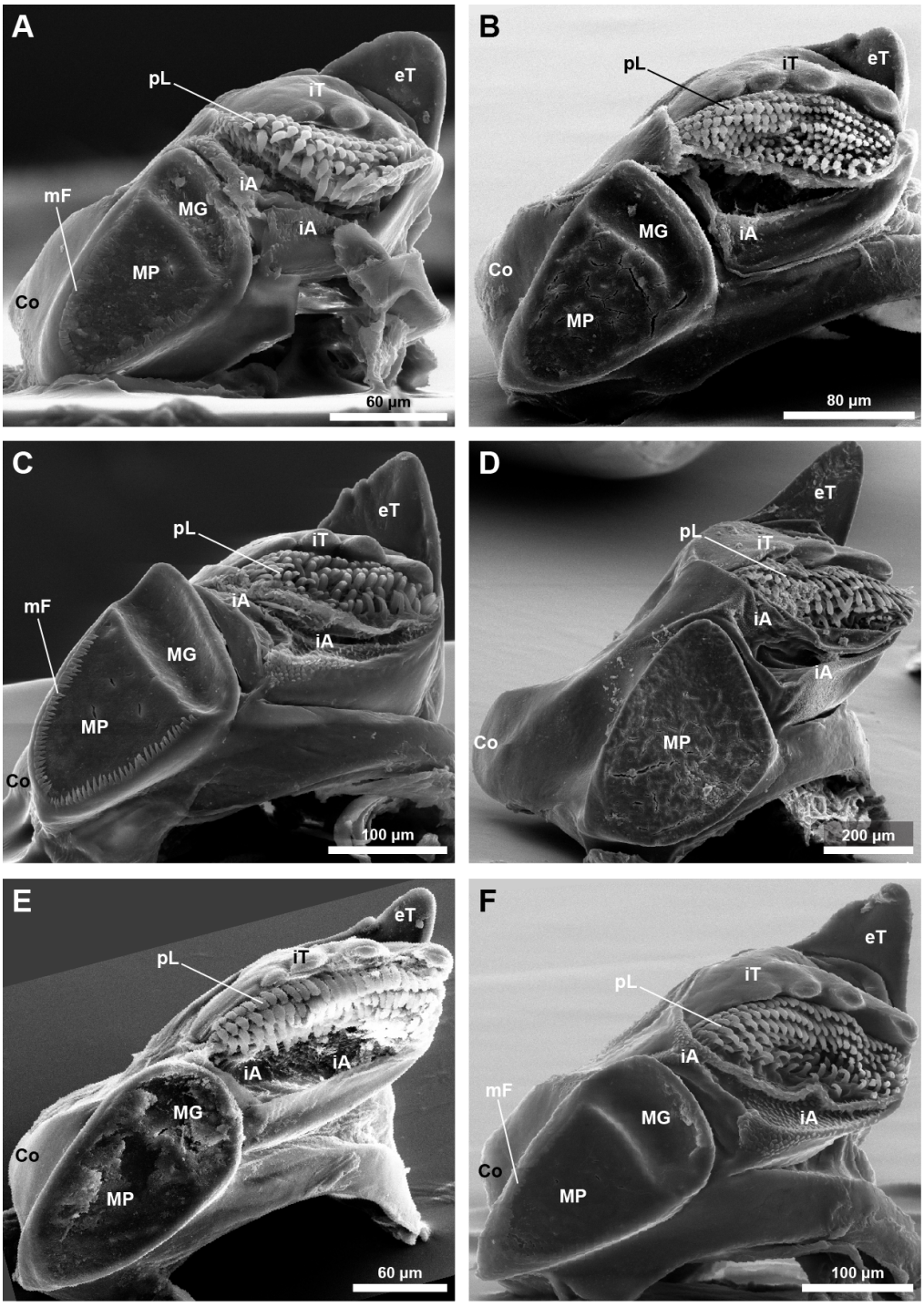

Figure 3.2. SEM images of Glomerida mandibles, mesal view. **A:** *Glomeridella minima* Latzel, 1884, left mandible; **B:** *Typhloglomeris* sp., left mandible; **C:** *Protoglomeris vasconica* Brölemann 1897, right mandible; **D**: *Eupeyerimhoffia archimedis* (Strasser 1965), left mandible, modified from Oeyen and Wesener (2015); **E:** *Glomerellina laurae* Silvestri 1908, left mandible; **F:** *Glomeroides prima* Silvestri 1929, left mandible. **Abbreviations:** Co = condylus; eT = external tooth; iA = intermediate area; iT = inner tooth; mF = membranous fringe; MG = molar groove; MP = molar plate; pL = pectinate lamellae.

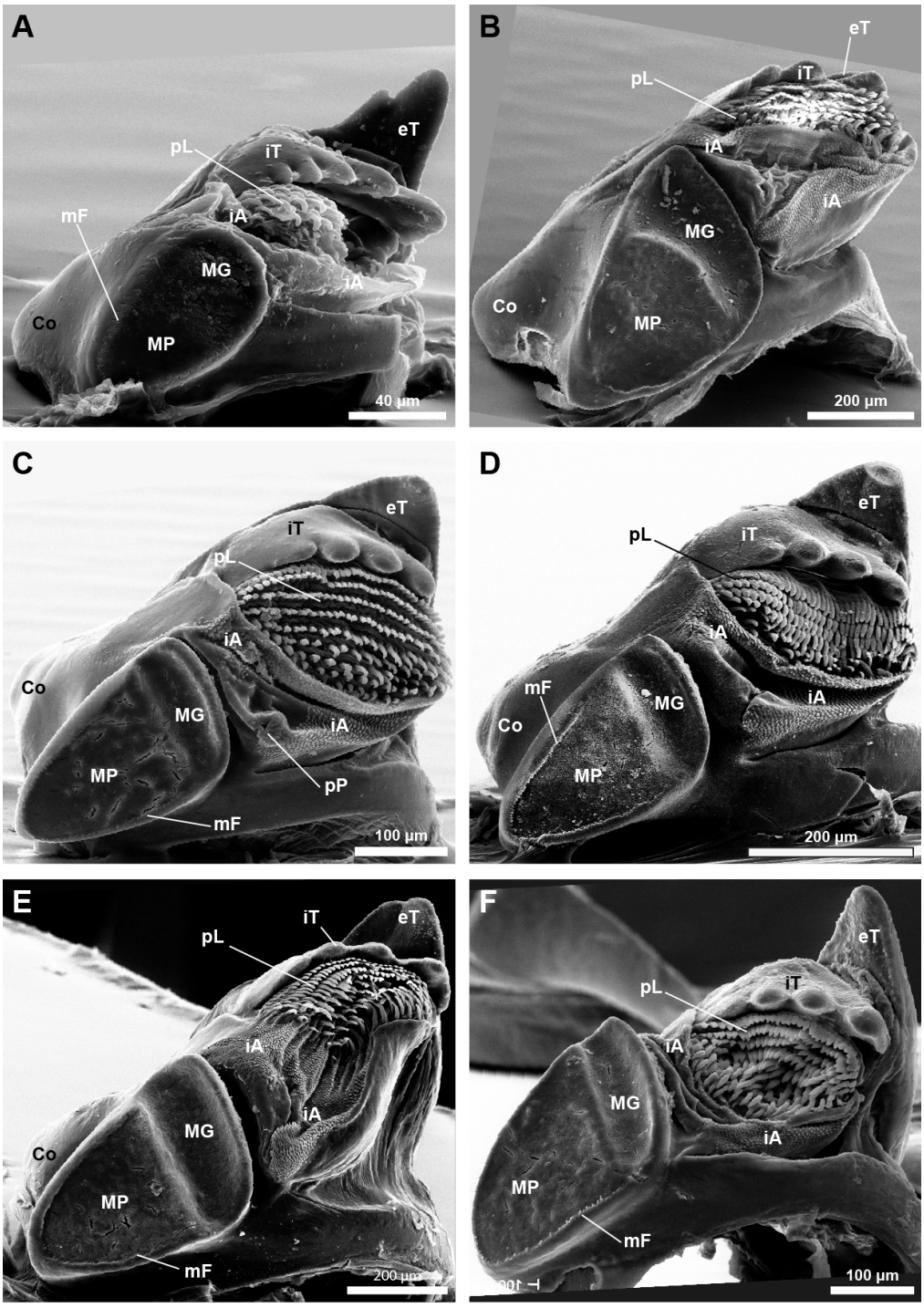

Figure 3.3. SEM images of Glomerida mandibles, mesal view. **A:** *Adenomeris* cf. *hispida* Ribaut 1909, left mandible; **B:** *Epiromeris aelleni* Strasser 1976, left mandible; **C:** *Haploglomeris multistriata* (Koch 1844), right mandible; **D:** *Simplomeris montivaga* (Faes 1902), left mandible; **E:** *Schismaglomeris occultocolorata* (Verhoeff 1892), left mandible; **F:** *Glomeris distichella* Berlese 1887, right mandible. **Abbreviations:** Co = condylus; eT = external tooth; iA = intermediate area; iT = inner tooth; mF = membranous fringe; MG = molar groove; MP = molar plate; pL = pectinate lamellae; pP = perimolar pore.

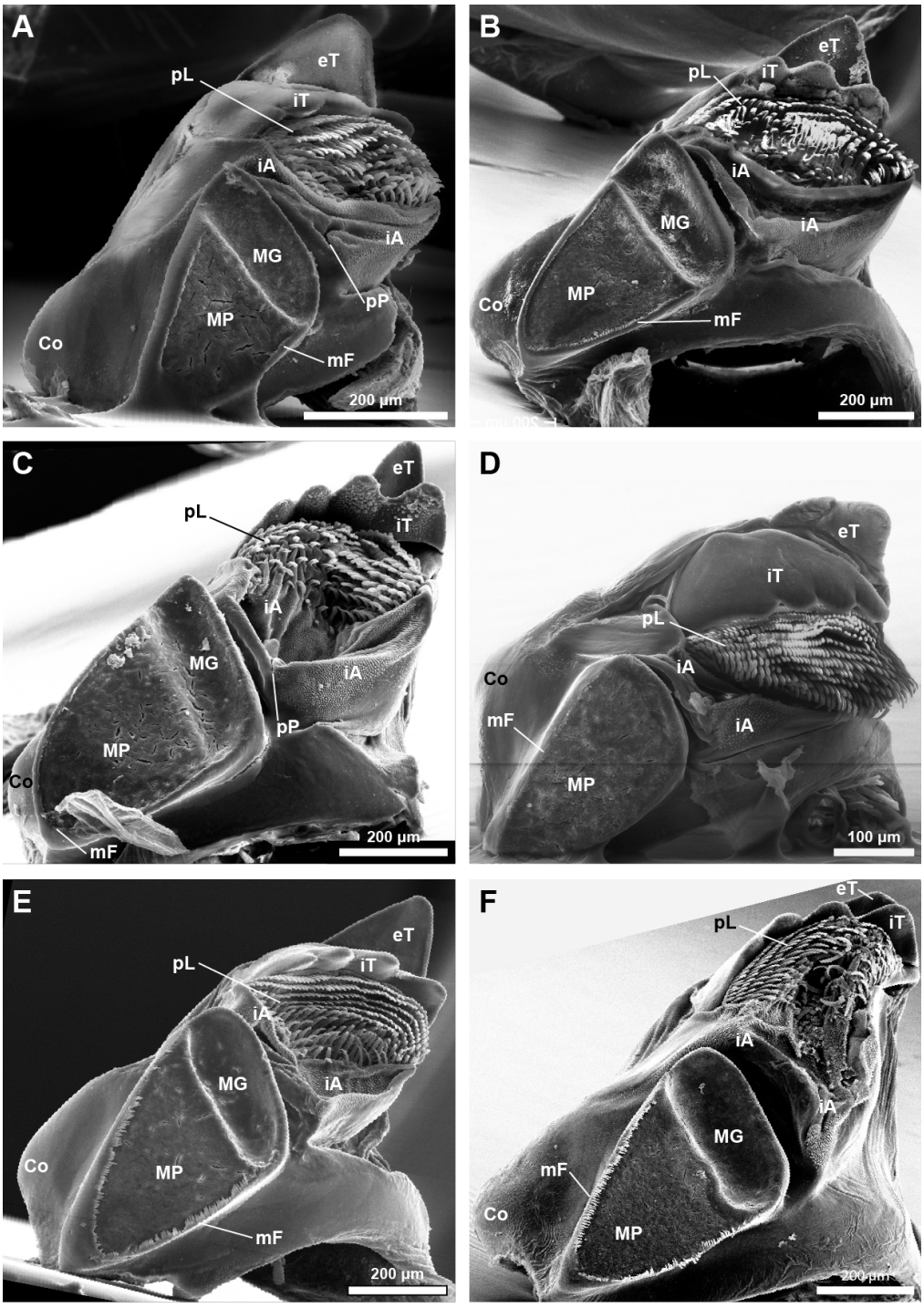

Figure 3.4. SEM images of Glomerida mandibles, mesal view. **A:** *Glomeris marginata* (Viller 1789), left mandible; **B:** *Loboglomeris* sp., right mandible; **C:** *Onychoglomeris fagi* Verhoeff, 1930, left mandible; **D:** *Hyleoglomeris* sp. 1, right mandible; **E:** *Hyleoglomeris* sp. 2, left mandible; **F:** *Rhopalomeris* sp., left mandible. **Abbreviations:** Co = condylus; eT = external tooth; iA = intermediate area; iT = inner tooth; mF = membranous fringe; MG = molar groove; MP = molar plate; pL = pectinate lamellae; pP = perimolar pore.

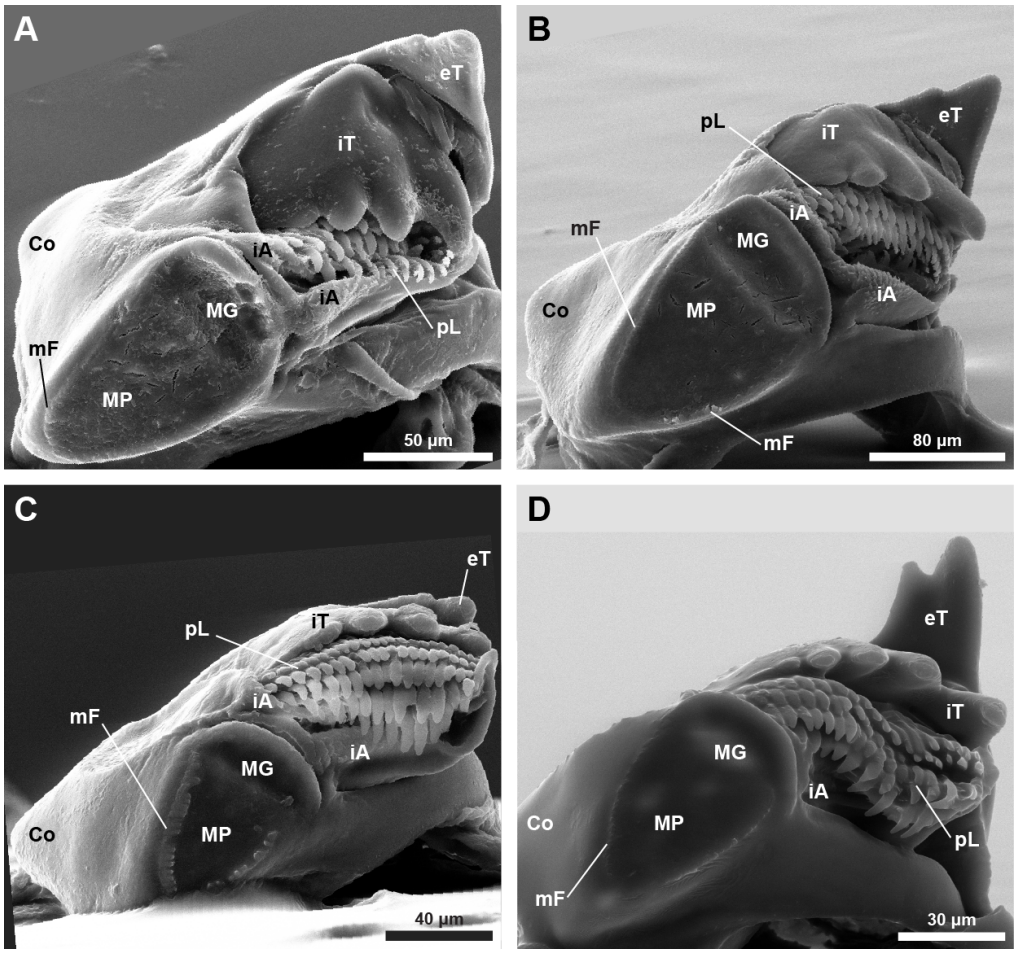

Figure 3.5. SEM images of Glomerida mandibles, mesal view. **A:** *Nearctomeris inexpectata* Wesener 2012, right mandible; **B:** *Onomeris sinuata* Loomis 1943, right mandible; **C:** *Trachysphaera* sp., right mandible; **D:** *Geoglomeris subterranea* Verhoeff 1908, right mandible; **Abbreviations:** Co = condylus; eT = external tooth; iA = intermediate area; iT = inner tooth; mF = membranous fringe; MG = molar groove; MP = molar plate; pL = pectinate lamellae.

Wesener 2014, Enghoff et al. 2015). These characters are the presence of coxal pouches (c44) and the presence of a paronychium on the tarsus (c45).

The superorder Oniscomorpha is supported by four potential apomorphies: (1) lingual palps of the gnathochilarium devoid of sensory cones (c16); (2) small tergite 1, the collum (c37); (3) enlarged tergite 2, the thoracic shield (c38); (4) completely reduced sternites at leg pair 1 (c42).

The sister-order of the Glomerida, the Sphaerotheriida, is recovered as monophyletic with unambiguous statistical support (100/100; Fig. 3.6) and seven potential

apomorphies: (1) reduction of the transverse bar of the tentorium (c3); (2) head with large lateral antennal grooves (c5); (3) reduction of the lateral palpi of the gnathochilarium (c15); (4) parts of the gnathochilarium mentum surrounding the lamellae linguales (c19); (5) central pads of the endochilarium with a field of special sensory cones (c21); (6) maxillary nephridia opening on the endochilarium (c22); (7) body consisting of 12 tergites (c35).

3.3.3 Monophyly of the Glomerida

The order Glomerida is recovered as monophyletic with high statistical support (94/92; Fig. 3.6) and is supported by five potential apomorphies: (1) eyes with less than 10 ommatidia arranged in one or two rows (c11); (2) gnathochilarium with greatly enlarged cardines (c17); (3) gnathochilarium with a divided mentum (c20); and (4) tergites with unpaired mid-dorsal defense-glands (c54), which (5) contain quinazolinone alkaloids (c56).

3.3.4 Intraordinal relationships of the Glomerida

The phylogenetic analysis provided little resolution of the relationships within the Glomerida (Fig. 3.6). The genus *Eupeyerimhoffia* (Protoglomeridae) is recovered in a basal position with moderate statistical support (70/60; Fig. 3.6). *Eupeyerimhoffia* is the only genus of the Glomerida in our dataset with a well-developed condylus on the mandible (c31:1), all other analyzed genera have a rudimentary condylus only resembling a bump (c31:2). The genus *Glomerellina* (Protoglomeridae) is also recovered in a basal position, but without any statistical or clear character support. The remaining genera of the Glomerida are recovered as a group with weak statistical

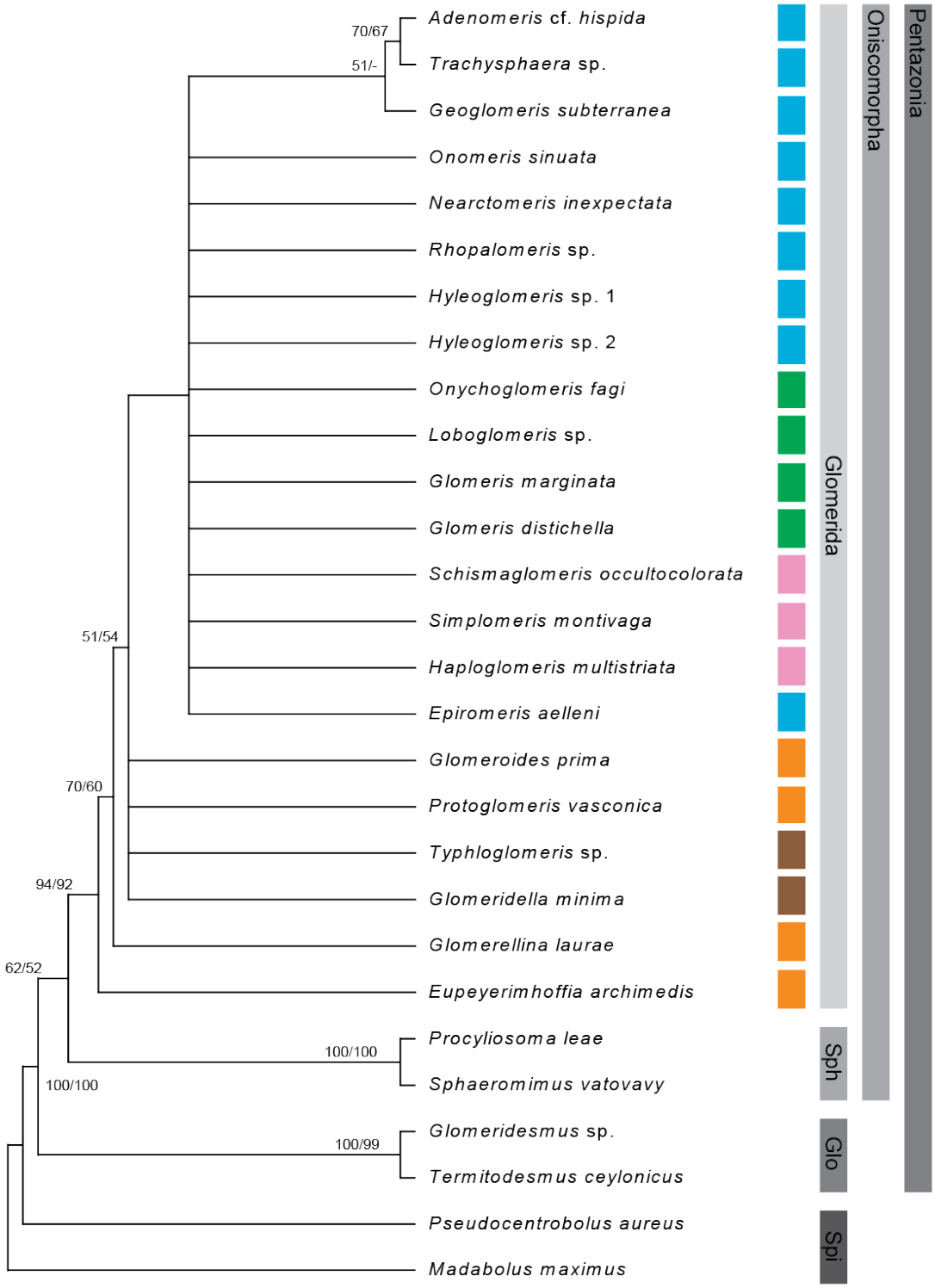

Figure 3.6 Maximum parsimony strict consensus tree. Colored bars correspond to groups as defined in Enghoff et al. (2015): Brown = Glomeridellidae; Orange = Protoglomeridae; Green = Glomerinae; Pink = Haploglomerinae; Blue = Doderiinae. **Abbreviations:** Sph = Sphaerotheriida; Glo = Glomeridesmida; Spi = Spirostreptida. Nodal support values represent bootstrap and jackknife scores.

support (51/54; Fig. 3.6) and without any character support. Within this group, a polytomy unites four genera from the families Glomeridellidae (*Glomeridella* and *Typhloglomeris*) and Protoglomeridae (*Protoglomeris* and *Glomeroides*) with a group containing all investigated species of the family Glomeridae. The monophyly of the family Glomeridae receives no statistical support, but it is supported by two potential apomorphies: presence of a lamellar outgrowth on the podomere 2 of the posterior telopod (c66), and podomere 3 of the posterior telopod with a trichostele (c68). Within the Glomeridae, none of the three analyzed subfamilies (Glomerinae, Haploglomerinae, and Doderiinae) are recovered (Fig. 3.6). Within the family, 11 of the 14 genera are recovered in a single polytomy (Fig. 3.6). A group uniting three of the genera containing only dwarf species (species <5 mm in length; *Geoglomeris*, *Adenomeris* + *Trachysphaera*) is recovered with weak to no statistical support (51/-; Fig. 3.6) and no supporting character. *Adenomeris* and *Trachysphaera* are recovered as sister-taxa with moderate support (70/67; Fig. 3.6) and a single potentially apomorphic character: the tergites are covered with transversal ridges of coagulated secretions (c40). Despite the mostly unresolved strict consensus tree, some groups were present in some of the 53 shortest trees. This is true for both the families Glomeridae and Glomeridellidae, while the family Protoglomeridae was never recovered. Within the Glomeridae, the three subfamilies Glomerinae, Haploglomerinae and Doderiinae are also present as monophyletic amongst some of the shortest trees.

3.4 Discussion

3.4.1 Mandible character diversity and phylogenetic influence in the Glomerida

The diversity of mandible characters and their variation within the order Glomerida surpasses what was previously reported for the group (Köhler and Alberti 1990), and is comparable to the morphological diversity described for the Polydesmida (Ishii and Tamura 1996). This is surprising, as the Polydesmida show a much higher diversity at the species and family level compared to the Glomerida (Enghoff et al. 2015). The mandible diversity of the Glomerida also stands in contrast to the conserved, almost uniform, mandible morphology in all of the analyzed species and genera of the putative sister-group Sphaerotheriida (Wesener and VandenSpiegel 2009, Wesener 2009, Wesener 2014, Wesener 2016). However, the lack of resolution within the Glomerida in the phylogenetic analysis shows that this apparent variation does not contain a strong phylogenetic signal (Fig. 3.6). The great variety observed in the different mandibular characters in the Glomerida shows the value of the mandible as a source of important taxonomic characters on the genus- and species-level. This is underlined by the large differences shown between the two analyzed species of the exceptionally species-rich genus *Hyleoglomeris* (compare Fig. 3.4D and 3.4E).

The perimolar pore, which was only observed in three of the 22 analyzed Glomerida species (Figs. 3.3C, 3.4A and 3.4C), has to our knowledge not previously been described in the literature. The pore might be related to the salivary ostioles present on the mandibles in Polyxenida (Ishii and Tamura 1995). However, due to the structure of the intermediate area, which is easily damaged or distorted during dissection, the perimolar pore described here could possibly be an artifact, such as a cuticular fold.

Further investigations employing histological sectioning are needed to clarify the interpretation of this structure.

The apparent lack of a phylogenetic signal in mandibular characters on the higher level of the Glomerida could be the result of convergent adaptations to specific food sources. Convergent morphological adaptations in the structure of the mandible to similar food sources, such as bat guano in cave environments, have been described in millipedes (Enghoff 1985, Liu et al. 2017). However, a recent study found no correlation between the structure of mandibles and trophic niche differentiation in soil and leaf litter dwelling millipedes (Semenyuk et al. 2011).

3.4.2 Monophyly of the Oniscomorpha

The monophyly of the superorder Oniscomorpha, uniting the Glomerida and Sphaerotheriida, has always been only weakly supported in morphological analyses (Enghoff 1984, Sierwald et al. 2003, Blanke and Wesener 2014) and is supported almost exclusively by characters connected to their ability to volvate, which could have evolved convergently (Blanke and Wesener 2014). However, molecular (Regier et al. 2005, Fernández et al. 2016) and total-evidence (Sierwald and Bond 2007) analyses consistently recover the Glomeridesmida and Glomerida as sister taxa, rather than the Sphaerotheriida and Glomerida. Despite the introduction of several characters, which have previously not been considered for phylogenetic analyses, we again recovered the Oniscomorpha with weak statistical support (Fig. 3.6). Out of the four characters supporting the classical Oniscomorpha in our dataset, three characters are directly related to the ability to volvate (small collum (c37), enlarged tergite 2

(c38); and completely reduced sternites at leg pair 1 (c42)), while a single character is not (lingual palps of the gnathochilarium devoid of sensory cones (c16)). Both the reduction of sternites (derived from Wesener et al. 2014a) and the lingual palp character (derived from Koch 2015) have not been included in previous analyses. The alternate topology, uniting the Glomerida and Glomeridesmida, is also supported by four characters in our analysis. These are: (1) the central position of the antennae (c6); (2) the presence of a large, horse-shoe shaped organ of Tömösváry (c13); (3) an enlarged gula of the gnathochilarium (c14); (4) the intermediate area of the mandible being covered by scale-like spines (c33). Three of these characters (c6, c14, and c33) are utilized for the first time in a phylogenetic context and all four seem to be independent characters. Therefore, although the Oniscomorpha is recovered with four supporting characters in the present analysis, the alternate sister group relationship between Glomerida and Glomeridesmida appears more likely given the equal number of independent supporting characters.

The failure to recover a monophyletic Spirobolida can be attributed to the fact that no characters applying only to the Spirobolida were included in the matrix, as they were not the focus of the present study. The monophyly of the group has been shown to be robust in previous studies (Wesener et al. 2008, Pitz and Sierwald 2010).

3.4.3 Monophyly and autapomorphies of the Glomerida

Recovering the Glomerida as a strongly supported monophyletic group was expected as it has never been questioned in the myriapod literature (Verhoeff 1928, Hoffman 1982, Enghoff 1984, Sierwald et al. 2003, Enghoff et al. 2015). While generally the

Diplopoda orders are supported by numerous characters (e.g. Hoffman 1982, Enghoff et al. 2015), the lack of phylogenetic studies within the class currently prevents any distinction between plesiomorphies and apomorphies. The monophyly of the Glomerida can now, for the first time, be confirmed and is supported by five inferred autapomorphies. This represents an important step towards a more robust classification of the Diplopoda.

3.4.4 Intraordinal relationships of the Glomerida

The present dataset does not contain a sufficient phylogenetic signal to resolve the intraordinal relationships within the Glomerida. It does, however, provide additional evidence that the characters employed in the current non-phylogenetic classification of the Glomerida (Enghoff et al. 2015) are not sufficient to define any meaningful groups. None of the suborders, families, subfamilies, or any alternative groups, could be recovered as monophyletic in the strict consensus tree (Fig. 3.6). The monophyly of the family Protoglomeridae has already been questioned based on both molecular (Wesener 2012, Oeyen and Wesener 2015) and morphological data (Oeyen and Wesener 2015). The present analysis further substantiates this by not resolving the four genera of the family as each other's closest relatives in any of the shortest trees. However, the families Glomeridae and Glomeridellidae, as well as the Glomeridae sub-families were present as monophyletic groups within some of the shortest trees and thus could possibly be recovered as monophyletic groups with the addition of further evidence in future.

In conclusion, the relationships within the Pentazonia remain uncertain, but new

morphological support is found for the Glomerida + Glomeridesmida relationship. The order Glomerida is confirmed as a monophyletic unit and is supported as such by five autapomorphic characters. The mandible of the Glomerida shows too much variation on the genus- and in some cases even species-level to currently provide characters for the phylogenetic classification of the order. This flaw, however, makes the mandible a worthy taxonomic character that should be included in future species descriptions. The novel mandible characters, combined with all currently known characters, are not sufficient to clarify the relationships within the Glomerida. Therefore, additional morphological and/or molecular characters are needed to resolve the phylogeny of the pill millipedes.

3.5 Acknowledgments

This work would not have been possible without the providers of numerous rare specimens by Robert Mesibov, Stelios Simaiakis, Axel Schönhofer, Jason Dunlop, Hans Reip, Paul Lee, Karin Voigtländer, Nattarin Wongthamwanich, and Peter Kautt, for which we are very thankful.

Mandibles of the order Glomeridesmida and Spirobolida could be studied during a postdoctoral stay of TW at the Field Museum, funded by NSF grant DEB 05-29715 to Petra Sierwald, Jason Bond and William Shear. We are thankful for the continuous advice of Betty Strack (Field Museum) at the scanning electron microscope. The material of *Sphaeromimus vatovavy* at the Muséum national d'Histoire naturelle, Paris (MNHN), was studied by TW during a stay supported by the SYNTHESYS European Commission's (FP 6) Integrated Infrastructure Initiative program grant FR-TAF-1166 to TW. Jean-Jacques Geoffroy, Monique Nguyen-Duy Jacquemin and Mark Judson

were great hosts during the stay at the MNHN. Karin Ulmen advised at the SEM in Bonn. We thank Thorsten Klug for his general assistance. We also thank Leif Moritz for the conduction of the CT scans and the reconstructions. The Alexander Koenig Society (AKG) funded a collection trip to Sicily and northern Italy during which several of the here studied specimens could be collected. We also thank the two anonymous reviewers for their constructive criticism and numerous suggestions. Lastly, we thank native English speaker T.O. Burt for reviewing the language of the manuscript.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.asd.2018.02.005.

3.6 References

Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. Biophotonics international. 11: 36–42.

Ashwini KM, Sridhar KR (2002) Towards organic farming with millipede *Arthrosphaera magna*. Current Science. 82: 20–22.

Beutel RG, Friedrich F, Hörnschemeyer T, Pohl H, Hünefeld F, et al. (2011) Morphological and molecular evidence converge upon a robust phylogeny of the megadiverse Holometabola. Cladistics. 27: 341–355.

Blanke A, Wesener T (2014) Revival of forgotten characters and modern imaging techniques help to produce a robust phylogeny of the Diplopoda (Arthropoda, Myriapoda). Arthropod Structure and Development. 43: 63–75.

Blower JG (1985) Millipedes. Synopses of the British Fauna, No. 35. Field Studies Council, Shrewsbury. 242 pp.

Crawford CS (1992) Millipedes as model detritivores. Berichte des Naturwissenschaftlich-Medizinischen Verein Innsbruck. 10: 277–288.

Curry JP (1994) Grassland Invertebrates. Chapman and Hall, London. 438 pp.

Edgecombe GD, Richter S, Wilson GD (2003) The mandibular gnathal edges: homologous structures throughout Mandibulata? African Invertebrates. 44: 115–135.

Enghoff H (1979) Taxonomic significance of the mandibles in the millipede order Julida. In: Camatini M (Ed.), Myriapod Biology. Academic Press, London. 27–38 pp.

Enghoff H (1984) Phylogeny of millipedes–a cladistic analysis. Journal of Zoological Systematics and Evolutionary Research. 22: 8–26.

Enghoff H (1985) Modified mouthparts in hydrophilous cave millipedes (Diplopoda). Bijdragen tot de Dierkunde. 55: 67–77.

Enghoff H (1990) The ground-plan of chilognathan millipedes (external morphology). In: Minelli A (Ed.), Proceedings of the 7th International Congress of Myriapodology. EJ Brill, Leiden. 1–21 pp.

Enghoff H (1991) A revised cladistic analysis and classification of the millipede order Julida. Journal of Zoological Systematics and Evolutionary Research. 29: 241–263.

Enghoff H, Golovatch S, Short M, Stoev P, Wesener T (2015) Diplopoda—taxonomic overview. In: Minelli, A. (Ed.). Treatise on Zoology-Anatomy, Taxonomy, Biology. The Myriapoda, Vol. 2. Brill, Leiden. 363–453 pp.

Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. Cladistics. 12: 99–124.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39: 783–791.

Fernández R, Edgecombe GD, Giribet G (2016) Exploring phylogenetic relationships within Myriapoda and the effects of matrix composition and occupancy on phylogenomic reconstruction. Systematic Biology. 65: 871–889.

Golovatch SI (2013) Three new species of the millipede genus *Hyleoglomeris* Verhoeff, 1910 from the Aegean region of Greece (Diplopoda, Glomerida, Glomeridae). Biodiversity Data Journal. 1: e1000.

Golovatch SI, Geoffroy J, Mauriès J (2006) Review of the millipede genus *Hyleoglomeris* Verhoeff, 1910 (Diplopoda, Glomerida, Glomeridae), with descriptions of new species from caves in Southeast Asia. Zoosystema-Paris. 28: 887–915.

Golovatch SI, Mauriès JP, Akkari N, Stoev P, Geoffroy JJ (2009) The millipede genus *Glomeri*s Latreille, 1802 (Diplopoda, Glomerida, Glomeridae) in North Africa. ZooKeys. 12: 47–86.

Hannibal J (1984) Pill millipedes from the Coal Age. Field Museum of Natural History Bulletin. 55: 12–16.

Hoffman RL (1980) Classification of the Diplopoda. Museum d'Histoire naturelle, Genève. 237 pp.

Hoffman RL (1982) Diplopoda. In: Parker SB (Ed.), Synopsis and Classification of Living Organisms. McGraw-Hill Book Company, New York. 689–724 pp.

Iniesta LFM, Ferreira RL, Wesener T (2012) The first troglobitic *Glomeridesmus* from Brazil, and a template for a modern taxonomic description of Glomeridesmida (Diplopoda). Zootaxa. 3550: 26–42.

Ishii K (1988) On the significance of the mandible as a diagnostic character in the taxonomy of penicillate diplopods (Diplopoda: Polyxenidae). The Canadian Entomologist. 120: 955–963.

Ishii K, Liang L (1990) Two new species of penicillate diplopods of the family Polyxenidae (Diplopoda: Penicillata) from China. The Canadian Entomologist. 122: 1239–1246.

Ishii K, Tamura H (1995) Comparative study of the mandibular structure as a diagnostic character in taxonomy of diplopods. Acta Zoologica Fennica. 196: 232–235.

Ishii K, Tamura H (1996) A taxonomic study of polydesmoid millipedes (Diplopoda) based on their mandibular structures. Mémoires du Muséum National d'Histoire Naturelle. 169: 101–111.

Koch M (2015) Diplopoda – General Morphology. In: Minelli A (Ed.), Treatise on Zoology – Anatomy, Taxonomy, Biology. The Myriapoda, vol. 2. Brill, Leiden. 7–76 pp.

Köhler HR, Alberti G (1990) Morphology of the mandibles in the millipedes (Diplopoda, Arthropoda). Zoologica Scripta. 19: 195–202.

Lauterbach KE (1972). Über die sogenannte Ganzbein-Mandibel der Tracheata, insbesondere der Myriapoda. Zoologischer Anzeiger. 188: 145–154.

Limaye A (2012) Drishti: a volume exploration and presentation tool. In: Stock SR (Ed.), Proc Spie, vol. 8506. Developments in X-Ray Tomography VIII. 85060X pp.

Liu W, Golovatch SI, Wesener T, Tian M (2017) Convergent evolution of unique morphological adaptations to a subterranean environment in cave millipedes (Diplopoda).PloS one. 12: e0170717.

Maddison WP, Maddison DR (2015) Mesquite: a modular system for evolutionary analysis. Version 3.02. 2015. Avaliable from: http://mesquiteproject.org (Accessed 5 Jul. 2016).

Manton SM (1964) Mandibular mechanisms and the evolution of arthropods. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 247: 1–183.

Mauriès JP (1971) Diplopodes épigés et cavernicoles de Pyrénées espagnoles et des monts cantabriques: VII. Glomérides. Essai de classification des Glomeroidea. Bulletin de La Société d'Histoire Naturelle de Toulouse. 107: 423–453.

Mauriès JP (2005) Attempt at a classification of Glomerida (Diplopoda), with descriptions of two new genera from the northwest of the Iberian Peninsula. Arthropoda Selecta. 14: 241–249.

Meinwald YC, Meinwald J, Eisner O (1966) 1, 2-Dialkyl-4(3H)-quinazolinones in the defensive secretion of a millipede (*Glomeris marginata*). Science. 154: 390–391.

Murienne J, Edgecombe GD, Giribet G (2010) Including secondary structure, fossils and molecular dating in the centipede tree of life. Molecular Phylogenetics and Evolution. 57: 301–313.

Nixon KC (1999) The Parsimony Ratchet, a new method for rapid parsimony analysis. Cladistics. 15: 407–414.

Nixon KC (2002) WinClada. Version 1.00.08. Published by the author, Ithaca, New York.

Oeyen JP, Wesener T (2015) Steps towards a phylogeny of the pill millipedes: non-monophyly of the family Protoglomeridae, with an integrative redescription of *Eupeyerimhoffia archimedis* (Diplopoda, Glomerida). ZooKeys. 510: 49–64.

O'Leary MA, Kaufman SG (2012) MorphoBank 3.0: Web application for morphological phylogenetics and taxonomy. http://morphobank.org.

Pitz KM, Sierwald P (2010) Phylogeny of the millipede order Spirobolida (Arthropoda: Diplopoda: Helminthomorpha). Cladistics. 26: 497–525.

Racheboeuf PR, Hannibal JT, Vannier J (2004) A new species of the diplopod *Amynilyspes* (Oniscomorpha) from the Stephanian Lagerstätte of Montceau-les-mines, France. Journal of Paleontology. 78: 221–229.

Regier JC, Wilson HM, Shultz JW (2005) Phylogenetic analysis of Myriapoda using three nuclear protein-coding genes. Molecular Phylogenetics and Evolution. 34: 147–158.

Sagorny C, Wesener T (2017) Two new giant pill-millipede species of the genus *Zoosphaerium* endemic to the Bemanevika area in northern Madagascar (Diplopoda, Sphaerotheriida, Arthrosphaeridae). Zootaxa. 4263: 273–294.

Schildknecht H, Wenneis WF, Weis KH, Maschwitz U (1966) Glomerin, ein neues Arthropoden-alkaloid. Zeitschrift für Naturforschung. 21: 121–127.

Semenyuk II, Tiunov AV, Golovatch SI (2011) Structure of mandibles in relation to trophic niche differentiation in a tropical millipede community. International Journal of Myriapodology. 6: 37–49.

Shear WA (2015) The chemical defenses of millipedes (Diplopoda): biochemistry, physiology and ecology. Biochemical Systematics and Ecology. 61: 78–117.

Shear WA, Jones TH, Wesener T (2011) Glomerin and homoglomerin from the North American pill millipede *Onomeris sinuata* (Loomis, 1943) (Diplopoda, Pentazonia, Glomeridae). International Journal of Myriapodology. 4: 1–10.

Shelley RM, Golovatch SI (2011) Atlas of Myriapod Biogeography. I. Indigenous Ordinal and Supra-Ordinal Distributions in the Diplopoda: Perspectives on Taxon Origins and Ages, and a Hypothesis on the Origin and Early Evolution of the Class. Insecta Mundi. 158: 1–134.

Sierwald P, Bond JE (2007) Current status of the myriapod class Diplopoda (millipedes): taxonomic diversity and phylogeny. Annual Review of Entomology. 52: 401–420.

Sierwald P, Shear WA, Shelley RM, Bond JE (2003) Millipede phylogeny revisited in the light of the enigmatic order Siphoniulida. Journal of Zoological Systematics and Evolutionary Research. 41: 87–99.

Simonsen Å (1990) Phylogeny and biogeography of the millipede order Polydesmida, with special emphasis on the suborder Polydesmidea. Doctoral dissertation, Museum of Zoology, University of Bergen, Bergen.

Snodgrass RE (1950) Comparative studies on the jaws of mandibulate arthropods. Smithsonian Miscellaneous Collections. 116: 1–85.

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods). Sunderland, MA.

Verhoeff KW (1928) Klasse Diplopoda, 1. Teil. In: Bronn's Klassen und Ordnungen des Tier-Reichs. Fünfter Band, II. Abteilung: Gliederfüßler: Arthropoda. Akad. Verlagsgesselschaft Leipzig, Leipzig. 1–1071 pp.

Wägele JW (1993) Rejection of the “Uniramia” hypothesis and implications of the Mandibulata concept. Zoologische Jahrbücher Abbteilung für Systematik. 120: 253–288.

Wesener T (2009) Unexplored richness: discovery of 31 new species of Giant Pill-Millipedes endemic to Madagascar, with a special emphasis on microendemism (Diplopoda, Sphaerotheriida). Zootaxa. 2097: 1–134.

Wesener T (2010) Revision of the Amercican Pill Millipedes I: *Onomeris* and *Trichomeris*. Zootaxa. 2725: 28–40.

Wesener T (2012) *Nearctomeris*, a new genus of Pill Millipedes from North America, with a comparison of genetic distances of American Pill Millipede Genera (Glomerida, Glomeridae). Zootaxa. 3258: 58–68.

Wesener T (2014) A new phylogenetic analysis of the Sphaerotheriida (Diplopoda) with a revision of the Australian giant pill-millipedes. Invertebrate Systematics. 28: 196–213.

Wesener T (2016) The Giant Pill-Millipedes, order Sphaerotheriida–An annotated species catalogue with morphological atlas and list of apomorphies (Arthropoda: Diplopoda). Bonn zoological Bulletin–Supplementum. 63: 1–104.

Wesener T, VandenSpiegel D (2009) A first phylogenetic analysis of Giant Pill Millipedes (Diplopoda: Sphaerotheriida), a new model Gondwanan taxon, with special emphasis on island gigantism. Cladistics. 25: 545–573.

Wesener T, Enghoff H, Wägele JW (2008) Pachybolini–a tribe of giant Afrotropical millipedes: arguments for monophyly and the description of a new genus from Madagascar (Diplopoda: Spirobolida: Pachybolidae). Invertebrate Systematics. 22: 37–53.

Wesener T, Enghoff H, Hoffman RL, Wägele JW, Sierwald P (2009) Revision of the endemic giant fire millipedes of Madagascar, genus *Aphistogoniulus* (Diplopoda: Spirobolida: Pachybolidae). International Journal of Myriapodology. 2: 15–52.

Wesener T, Le D, Loria S (2014b) Integrative revision of the giant pill-millipede genus *Sphaeromimus* from Madagascar, with the description of seven new species (Diplopoda, Sphaerotheriida, Arthrosphaeridae). ZooKeys. 414: 67–107.

Wesener T, Sierwald P, Wägele JW (2014a) Sternites and spiracles–The unclear homology of ventral sclerites in the basal millipede order Glomeridesmida (Myriapoda, Diplopoda). Arthropod Structure and Development. 43: 87–95.

Wilbrandt J, Lee P, Read H, Wesener T (2015) A first integrative study of the identity and origins of the British Dwarf Pill Millipede populations, *Trachysphaera* cf. *lobata* (Diplopoda, Glomerida, Glomeridae). Biodiversity Data Journal. 3: e5176.

Wolters V, Ekschmitt K (1997) Gastropods, isopods, diplopods, and chilopods: neglected groups of the decomposer food web. In: Benckiser G (Ed.), Fauna in

Soil Ecosystems. Dekker, New York. 265–306 pp.

Chapter 4

The complete mitochondrial genome of *Glomeris marginata* (Diplopoda, Glomerida)

This chapter is intended for publication in Mitochondrial DNA part B.
Authors: Oeyen JP, Wesener T, Niehuis O, Misof B.

Author contributions to the article:
Conceived the study: JPO, ON; Collected specimens: JPO, TW; Performed the analyses: JPO; Interpreted the results: JPO; Wrote the manuscript: JPO, TW, ON, BM

4.1 Introduction

Pill millipedes (Myriapoda: Diplopoda: Glomerida) are small to medium sized millipedes that are widely known for and easily recognized by their ability to roll-up into a ball (Figure 4.1A). The order Glomerida comprises more than 300 described species in 35 genera that occur throughout the Holarctic region, with a distinct species-diversity hot spot in Europe (Enghoff et al. 2015). Within Diplopoda, the order Glomerida is placed in the infraclass Pentazonia among the giant pill-millipedes (Sphaerotheriida) and the Glomeridesmida (Blanke and Wesener 2014). The relationships among these three lineages have remained contested, however. Traditionally, Glomerida and Sphaerotheriida have been considered sister-groups (superorder Oniscomorpha) based on results from studying morphological data (e.g., Blanke and Wesener 2014). More recent analyses that also included molecular data suggested Glomeridesmida to represent the sister-group of the Glomerida (e.g., Fernández et al. 2016, Rodriguez et al. 2018). So far, only comparatively few molecular data for studying phylogenetic relationships of millipedes have been compiled, especially so from species of the three orders of the Pentazonia. Expanding the available molecular resources is crucial in order to confidently resolve the phylogenetic relationships among and within these millipede orders in the future. To this end, we here present the entire DNA sequence of a mitochondrial genome from species of the order Glomerida — that of *Glomeris marginata* — and compare its gene arrangement with that published for a mitochondrial genome of a species of the millipede order Sphaerotheriida.

4.2 Materials and methods

Total genomic DNA was extracted from heads of 45 *Glomeris marginata* males collected at the Landskrone (Heppingen/Gimmingen, Rhineland-Palatinate, Germany; 50°33'03.0" N / 007°10'31.8" E; November 14, 2014) and used to constructed a paired end 250-bp insert library that was sequenced on an Illumina HiSeq2000 by BGI-Shenzhen following standard Illumina protocols. Raw reads were trimmed using Trimmomatic (version 0.32; Bolger et al. 2014) with the Illumina clip option (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10) and the supplied Illumina adapter file. Quality trimming used a sliding window of 4 bp with a minimum phred score of 28, retaining only paired reads with a minimum length of 75 bp. The cleaned genomic DNA reads of *G. marginata* were assembled with the software MITObim version 1.8 (Hahn et al. 2013), which employs the assembler MIRA version 4.0 (Chevreux et al. 2004) and requires are reference genome for initiating the assembly process. We procided the software for this purpose that published mitochondrial genome of *Prionobelum* sp. (Sphaerotheriida) (GenBank Accession: NC_018361; Dong et al. 2012, Dong et al. 2016). To increase the specificity of MIRA, an initial mapping assembly was created using the full library (option: -quick) before reducing the readpool to an expected 100x coverage. The final assembly was inferred using the reduced dataset following the two-step procedure recommended by the MITObim developers (Hahn et al. 2013). A majority rule consensus assembly was created using the miraconvert script supplied with MIRA version 4.0 (options: miraconvert -f maf -t fasta -A "SOLEXA_SETTINGS -CO:fnicpst=yes" sample-ref_out.maf majority_rule_consensus), and the resulting assembly was manually checked for errors and circularized by mapping the reads to the assembly using the map to

reference function in Geneious version 7.0 (Kearse et al. 2012). Genes were annotated using the MITOS web-server (http://mitos.bioinf.uni-leipzig.de/index; Bernt et al. 2013) and manually checking *E* values of all annotations to confirm the robustness of the results. The order of genes in the mitochondrial genomes of *G. marginata* and *Prionobelum* sp. (Sphaerotheriida) were compared and potential rearrangements scenarios were estimated using CREx (Common Interval Rearragement Explorer; Bernt et al. 2007).

4.3 Results

The assembly of the mitochondrial genome of *G. marginata* is 16,514 bp long and contains a total of 13 protein-coding genes, two ribosomal RNA genes, and 22 transfer RNA genes (Figures 4.1B and C). Twenty-two of the genes are encoded on the positive strand, with only a single block of fifteen genes encoded on the negative strand (Figure 4.1B). The DNA sequence of the genome has a strong AT bias, with GC content of only 29.9 %. Except for fourteen genes spanning from *cox1* to *trnE*, the gene order of *G. marginata* differs from that reported for *Prionobelum* sp. (Sphaerotheriida). The analysis of potential rearrangement scenarios resulted in a single scenario with two rearrangements (Figure 4.1C): (1) a transposition of *trnQ* and (2) a tandem duplication followed by random loss (Figure 4.1C). The gene order of *G. marginata* is almost identical to those of *Narceus annularis* (Diplopoda, Spirobolida) and *Thyropygus* sp. (Diplopoda, Spirostreptida) (Lavrov et al. 2002) of the superorder Juliformia, only differing by the position of two RNA genes (*trnT* and *trnC*).

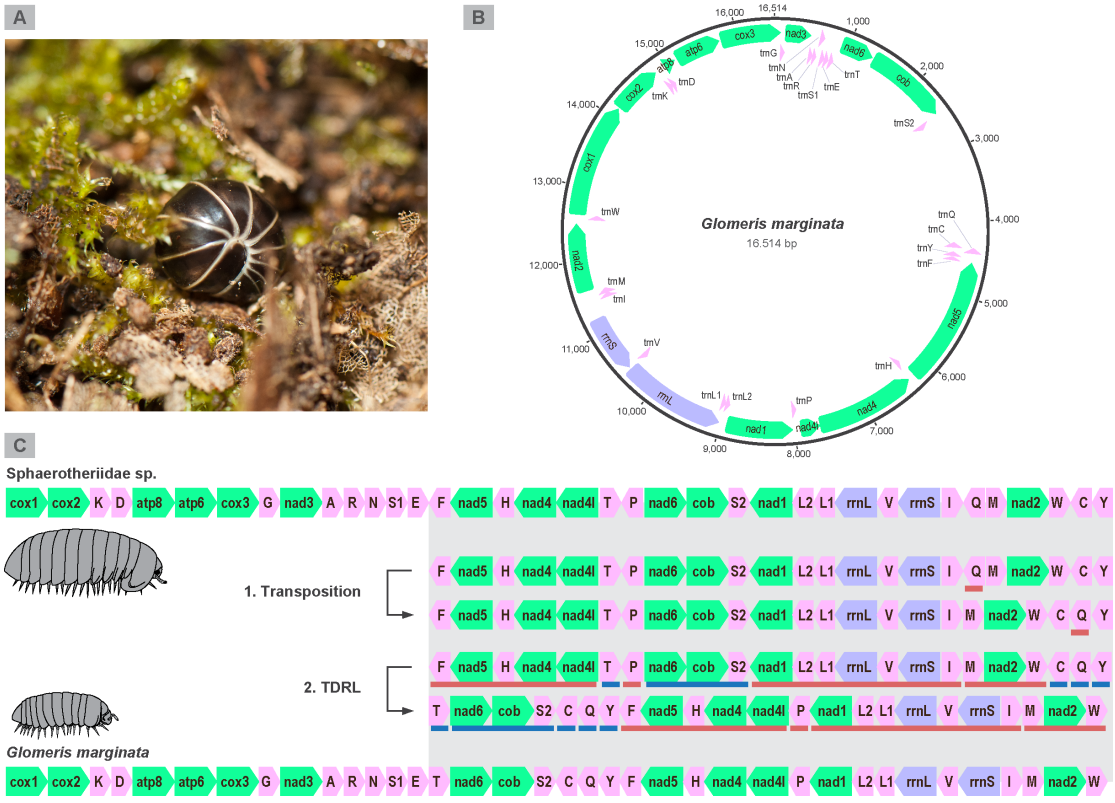

Figure 4.1. A: *Glomeris marginata*, photographed *ex-situ*. **B:** Complete mitochondrial genome of *G. marginata*. Protein-coding genes are indicated by green annotations, ribosomal RNA genes by purple annotations, and tRNA genes by pink annotations. **C:** Gene rearrangement scenario from the ancestral myriapod mitochondrion, as reported for Sphaerotheriida, to the here reported mitochondrion of *G. marginata* inferred using CREx.

4.4 Discussion

The size of the *G. marginata* mitochondrial genome and its strong AT bias is similar to what is known from other Diplopoda (Dong et al. 2012; Dong et al. 2016). Furthermore, the gene content is identical to that reported for other Myriapoda (e.g. Lavrov et al. 2002; Lavrov et al. 2002; Brewer et al. 2013), including the proposed sister taxon of the Glomerida, the Sphaerotheriida (Dong et al. 2012). However, the gene order differs from that of Sphaerotheriida (Dong et al. 2012) in numerous aspects and to those of other Diplopoda in few rearrangements (Dong et al. 2016). As the mitochondrial genome organization of Sphaerotheriida is thought to represent the

ancestral state of all Myriapoda (Dong et al. 2012), the rearrangements of *G. marginata* likely represent a derived state. The near-identical genome arrangement of *G. marginata* with *N. annularis* and *Thyropygus* sp. is likely the result of convergent evolution within Diplopoda, as both *Narceus* and *Thyropygus* belong to the Juliformia, while the Glomerida belong to the Pentazonia. Pentazonia and Juliformia were never recovered as sister-groups in any morphological (Blanke & Wesener 2014) or molecular study (Brewer & Bond 2013; Brewer et al. 2013; Rodriguez et al. 2018) of the Diplopoda. This convergently evolved near-identical genome arrangement is consistent with the idea of a non-random occurrence of mitochondrial rearrangements due to functional constraints (Lavrov et al. 2002).

Despite problems associated with inferring phylogenetic relationships among myriapods from mitochondrial protein-coding sequences (Brewer et al. 2013), these relationships can, with the expansion of available mitochondrial genomes, in the future be addressed using genome rearrangements as characters.

As far as the phylogenetic relationships within Pentazonia are concerned, these could potentially be clarified with the addition of further mitochondrial genomes. Most importantly a mitochondrial genome has not been published for the third order of the Pentazonia, the order Glomeridesmida. The here presented mitochondrial genome of *G. marginata* thus paves the road for a thorough analysis of myriapod phylogenetic relationships that is based on genomic metacharacters rather than primary sequence information.

Disclosure statement

The authors report no conflict of interest.

Data availability statement

Voucher specimens were deposited in the collection of the Zoological Research Museum Alexander Koenig (Bonn, Germany) under the voucher number **ZFMK-MYR-3757**. The mitochondrial genome sequence has been deposited in the NCBI GenBank and is available under the accession number **MT881677**.

4.5 Acknowledgments

We acknowledge the help of Thorsten Klug in collecting and sorting samples, Dr. Christoph Hahn for his help and suggestions regarding the assembly strategy. We also thank the Leibniz Graduate School on Genomic Biodiversity research, BGI, and Xin Zhou for granting access to the original sequencing data.

4.6 References

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, et a. (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. Molecular phylogenetics and evolution. 69: 313–319.

Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, et al. (2007) CREx: inferring genomic rearrangements based on common intervals. Bioinformatics. 23: 2957–2958.

Blanke A, Wesener T (2014) Revival of forgotten characters and modern imaging techniques help to produce a robust phylogeny of the Diplopoda (Arthropoda, Myriapoda). Arthropod Structure and Development. 43: 63–75.

Bolger AM, Lohse M, Usadel, B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30: 2114–20.

Brewer MS, Bond JE (2013) Ordinal-level phylogenomics of the arthropod class Diplopoda (millipedes) based on an analysis of 221 nuclear protein-coding loci generated using next-generation sequence analyses. PloS one. 8: e79935.

Brewer MS, Swafford L, Spruill CL, Bond JE (2013) Arthropod phylogenetics in light of three novel millipede (myriapoda: diplopoda) mitochondrial genomes with comments on the appropriateness of mitochondrial genome sequence data for inferring deep level relationships. PloS one. 8: e68005.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, et al. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Research. 14: 1147–1159.

Dong Y, Xu JJ, Hao SJ, Sun HY (2012) The complete mitochondrial genome of the giant pill millipede, Sphaerotheriidae sp. (Myriapoda: Diplopoda: Sphaerotheriida). Mitochondrial DNA. 23: 333–335.

Dong Y, Zhu L, Bai Y, Ou Y, Wang C (2016) Complete mitochondrial genomes of two flat-backed millipedes by next-generation sequencing (Diplopoda, Polydesmida). Zookeys. 637: 1–20.

Enghoff H, Golovatch S, Short M, Stoev P, Wesener T (2015) Diplopoda – Taxonomic overview. In: Minelli A. (Ed.), Treatise on Zoology-Anatomy, Taxonomy, Biology. The Myriapoda, vol. 2. Brill, Leiden. 363–453 pp.

Fernández R, Edgecombe GD, Giribet G (2016) Exploring Phylogenetic Relationships within Myriapoda and the Effects of Matrix Composition and Occupancy on Phylogenomic Reconstruction. Systematic Biology. 65: 871–889.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 29: 644–52.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. Nucleic Acids Research. 41: e129.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28: 1647–1649.

Lavrov DV, Boore JL, Brown WM. 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. Molecular Biology and Evolution. 19: 163–169.

Rodriguez J, Jones TH, Sierwald P, Marek PE, Shear WA, et al. (2018) Step-wise evolution of complex chemical defenses in millipedes: a phylogenomic approach. Scientific Reports. 8: 3209.

Chapter 5

Phylogenomic analyses reveal paraphyly across Pill-Millipede families and subfamilies (Diplopoda: Glomerida)

This chapter is intended for publication in Molecular Phylogenetics and Evolution.

Authors: Oeyen JP, Vasilikopoulos A, Szucsich N, Bartel D, Makris C, Read H, Richards P, Rühr P, Donath A, Meusemann K, Podsiadlowski L, Niehuis O, Misof B, Wesener T.

Author contributions: Conceived the study: JPO; Collected specimens: JPO, TW, NS, DB, CM, HR, PRichards, PRühr; Data processing: JPO, AV, AD, KM, LP; Analyzed the data: JPO, AV; Interpreted the results: JPO, AV, BM; Wrote the manuscript: JPO, TW, BM, ON.

5.1 Introduction

The order Glomerida (Myriapoda, Diplopoda), commonly referred to as Pill-Millipedes, comprises of more than 300 species and 34 genera of small to medium sized (2–20 mm long) millipedes that are able to roll up (volvate) into a sphere (Figures 5.1A, C, and I). Glomerids occur throughout most of the Holarctic and parts of the Oriental sub-region, with a prominent center of biodiversity in Europe (23 genera; Enghoff et al. 2015). Among the European species are also a number of dwarves (< 5 mm in length), some of which show a unique aberrant morphology (Figures 5.1F and G). Within their distribution, glomerids mainly inhabit deciduous forests where they predominantly feed on decaying leaf litter and play an essential role in the soil ecosystem, contributing to litter decomposition, soil formation, nutrient cycling, and diagenesis (e.g. Rawlins et al. 2006, Bonkowski et al. 1998). Despite their ecological importance and having been established as model organisms for studying arthropod development (e.g. Dohle 1964, Janssen et al. 2008), the order as a whole has received little attention in terms of systematic research (Oeyen and Wesener 2018).

Phylogenetically, Glomerida is currently placed in the infraclass Pentazonia together with Sphaerotheriida (giant pill-millipedes) and Glomeridesmida. Based on morphological evidence, Glomerida has been considered the sister group of Sphaerotheriida (superorder Oniscomorpha; Enghoff 1984, Sierwald et al. 2003, Blanke and Wesener 2014). However, only a small number of characters that are all related to their ability volvate supported this group (Wesener and Blanke 2014). Molecular (Regier et al. 2005, Fernandez et al. 2016, Rodriguez et al. 2018) as well as

total evidence (Sierwald and Bond 2007) studies have, on the other hand, consistently recovered Glomeridesmida as the sister group to Glomerida. This alternative relationship has also found some, although weak, support based on a morphological analysis including new characters unrelated to their ability to volvate (Oeyen and Wesener 2018). In addition to the unclear position of Glomerida among the orders of Pentazonia, there is currently no phylogenetic system for the intraordinal relationships in Glomerida.

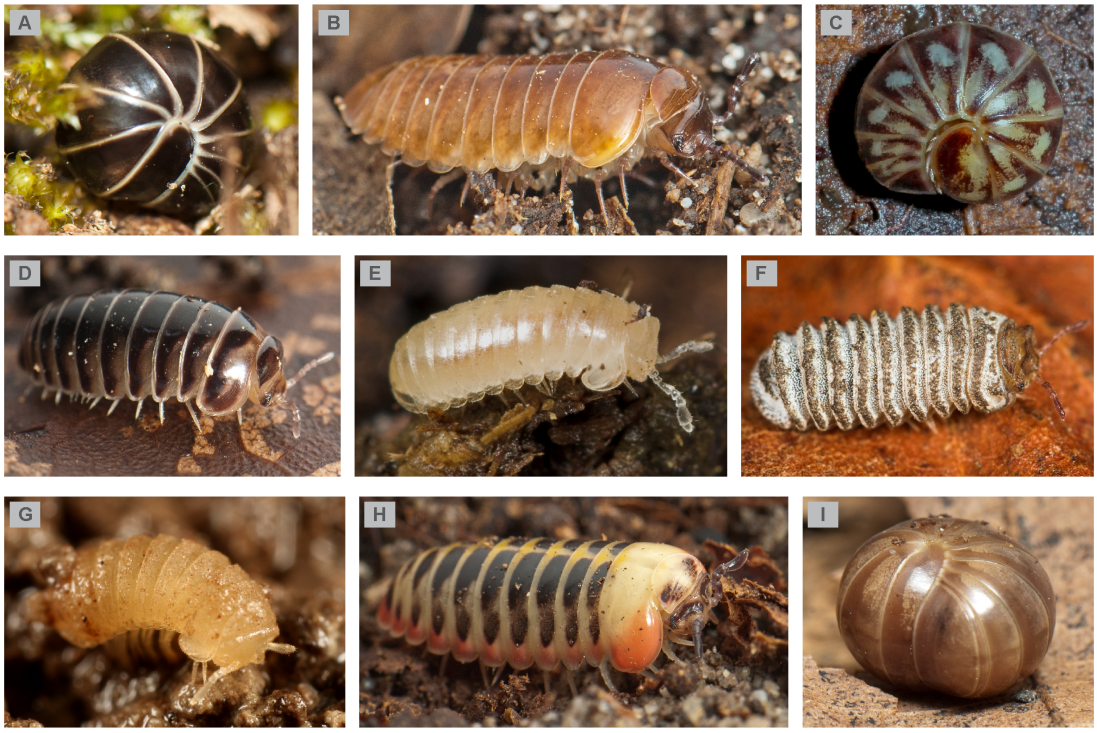

Figure 5.1: Habitus of Glomerida species. **A:** *Glomeris marginata*. **B:** *Protoglomeris vasconica*. **C:** *Typhloglomeris martensi*. **D:** *Hyleoglomeris* sp. Japan. **E:** *Geoglomeris subterranea*. **F:** *Trachysphaera lobata*. **G:** *Adenomeris gibbosa*. **H:** *Rhopalomeris carnifex*. **I:** *Eupeyerimhoffia archimedis*. Not to scale. Photographs E, F, and G courtesy of P. Richards.

Currently, the order is classified in a typological system that was originally proposed by Mauriès (1971) and that has since been revised twice (Mauriès 2005, Enghoff et al. 2015). The system separates three families: Glomderidellidae, Protoglomeridae, and

Glomeridae. Glomeridae is further subdivided into the five subfamilies Glomerinae, Haploglomerinae, Doderiinae, and Mauriesiinae. This classification is predominantly based on characters of the telopods. The telopods are the last male leg pairs that are modified into claspers. The telopods hold the vulva during mating, while the inner horns (coxal processes) of the syncoxite, on which the seminal fluid has been deposited, are inserted into it (Haacker 1969). Arguing that the telopods were uninformative due to potential convergent reductions, an alternative system was proposed by Hoffman (1980). This alternative system divided the Glomerids into the three families Glomeridellidae, Trachysphaeridae, and Glomeridae (Hoffman 1980). The latter of which is further subdivided into the two subfamilies Protoglomerinae and Glomerinae. One of the main differences to the currently recognized system was the placement of the American genera (*Onomeris* and *Glomeroides*) into the same subfamily (Protoglomerinae; Hoffman 1980) rather than separate families (*Glomeroides* in Protoglomeridae and *Onomeris* in Glomeridae; Enghoff et al. 2015). However, the system proposed by Mauries remains accepted, as Hoffman failed to present morphological characters supporting his hypothesis, rendering it unverifiable.

Although the system proposed by Mauries remains accepted, the concerns about its validity have remained. Initial molecular studies utilizing a single mitochondrial gene (COI) and a small taxonomic sample have suggested that at least one family, Protoglomeridae, is paraphyletic (Wesener 2012, Oeyen and Wesener 2015). This was further substantiated by a recent study that analyzed all known morphological characters, including a set of novel mandible characters, from a comprehensive set of taxa (Oeyen and Wesener 2018). The morphological characters were found to be

largely uninformative, Protoglomeridae was resolved as paraphyletic, and none of the groups above genus level received any support (Oeyen and Wesener 2018).

The uncertain relationships of the Glomerida, especially among genera within the same sub-family, not only impedes the work of taxonomists, but has also resulted in open questions regarding the evolution and biogeography of the group. For instance, close relationships of geographically distant taxa habe been suggested based on morphological similarities, but have not been further evaluated. Most notably, two American genera, *Nearctomeris* and *Onomeris,* have been suggested to be more closely related to the predominantly Asian genus *Hyleoglomeris* than to the third American genus Glomeroides (Mauries 1984, Wesener 2010, Wesener 2012).

Another intriguing enigma is the unique evolution of dwarfism and aberrant morphology in Europe. The currently described dwarf species are assigned to seven genera that exclusively contain dwarf species (Enghoff et al. 2015). Three genera contain species that show a Glomerida-typical morphology (*Glomerellina*, *Geoglomeris*, and *Tectosphaera*) while while the morphologically abberant species are assigned to four separate genera (*Adenomeris*, *Doderia*, *Strasseria*, and *Trachysphaera*). With the exception of *Glomerellina,* these genera are all placed in the subfamily Doderiinae (Enghoff et al. 2015) and have also been suggested to be more closely related to *Hyleoglomeris* than to other European genera (Mauries 1984). As dwarfism is currently thought to be a derived trait within Glomerida (Oeyen and Wesener 2018), this hypothesis implies that dwarfism in Glomerida evolved independently two or, depending on the still unresolved relationships within Doderiinae, even more times. Some authors have also hypothesized a potential third

case, suggesting a common origin of the three genera that contain morphologically aberrant dwarf species, separating them into their own family, the Trachysphaeridae (Attems 1926, Verhoeff 1932, Schubart 1934, Hoffman 1980). The separation of *Glomerellina* from the other genera of dwarf species, as well as a close relationship of the aberrant species, has found some, although weak, support based on the analysis morphological characters (Oeyen and Wesener 2018). However, due to the unresolved internal relationships of the group, it remains unclear what the closest relatives of the dwarf species are and also how often dwarfism and the aberrant morphology of some dwarf species evolved.

In order to gain insights into the phylogenetic relationships and evolution of Glomerida, we compiled and analyzed a transcriptome-based dataset with representative species of all but one of the currently recognized subfamilies. For this, we sequenced the transcriptomes of nine Glomerida and combined them with previously published data. We analyzed both amino acid and nucleotide datasets using maximum likelihood, bayesian inference, and multi-species coalescence based approaches, compare the results, and discuss their implications and potential future directions of Glomerida systematics.

5.2 Materials and methods

5.2.1 Taxon sampling, sequencing, and assembly

We compiled a dataset of transcriptomes covering all but one (Mauriesiinae) of the currently recognized Glomerida subfamilies (Table 5.1). In addition to the previously published transcriptomes of *Glomeris marginata* (Fernandez et al. 2016) and

Glomeris pustulata (Misof et al. 2014) and the transcriptomes of *Haploglomeris multistriata* and *Glomeridella minima,* which were kindly provided by Szucsich et al. (in review), we sequenced the transcriptomes of nine further Glomerida species from separate genera.

Table 5.1. Overview of taxa included in the transcriptome analysis ordered alphabetically by species name.

Order	**Family**	**Subfamily**	**Species**
Glomerida	Glomeridae	Doderiinae	*Adenomeris gibbosa* Mauriès, 1960
Glomerida	Protoglomeridae	Protoglomerinae	*Eupeyerimhoffia archimedis* (Strasser, 1965)
Glomerida	Glomeridae	Doderiinae	*Geoglomeris subterranea* Verhoeff, 1908
Glomerida	Glomeridellidae	Glomeridellinae	*Glomeridella minima* (Latzel, 1884)
Glomerida	Glomeridae	Glomerinae	*Glomeris pustulata* (Fabricius, 1781)
Glomerida	Glomeridae	Glomerinae	*Glomeris marginata* (Villiers, 1789)
Glomerida	Glomeridae	Haploglomerinae	*Haploglomeris multistriata* (Koch, 1844)
Glomerida	Glomeridae	Doderiinae	*Hyleoglomeris* sp. Japan
Glomerida	Glomeridae	Doderiinae	*Onomeris underwoodii* Cook, 1896
Glomerida	Protoglomeridae	Protoglomerinae	*Protoglomeris vasconica* (Brölemann, 1897)
Glomerida	Glomeridae	Doderiinae	*Rhopalomeris carnifex* (Pocock, 1889)
Glomerida	Glomeridae	Doderiinae	*Trachysphaera* sp.
Glomerida	Glomeridellidae	Typhloglomerinae	*Typhloglomeris martensi* (Golovatch, 1981)
Outgroup			
Platydesmida	Andrognathidae	Bazillozoniinae	*Brachycybe* sp.
Sphaerotheriida	Cyliosomatidae	n/a	*Cyliosoma* sp.
Polyxenida	Polyxenidae	n/a	*Eudigraphis taiwanensis* Ishii, 1990
Glomeridesmida	Glomeridesmidae	Glomeridesminae	*Glomeridesmus* sp.
Spirobolida	Spirobolidae	Spirobolinae	*Narceus americanus* (Palisot de Beauvouis, 1817)

These included three dwarf species from separate genera (*Adenomeris*, *Geoglomeris*, and *Trachysphaera*), two of which show aberrant morphology (*Trachysphaera* and *Adenomeris*) as well as the American species *Onomeris underwoodi.* As outgroups, we included the already published transcriptomes of *Brachycybe* sp. (Platydesmida), *Cyliosoma* sp. (Sphaerotheriida), *Eudigraphis taiwanensis* (Polyxenida), *Glomeridesmus* sp. (Glomeridesmida), and *Narceus americanus* (Spirobolida) (Brewer and Bond 2013, Fernandez et al. 2016). With the exception of *Glomeris*

pustulata, where an assembly was available, raw sequencing reads of all previously sequenced species were downloaded and processed along those of the newly sequenced species (Table 5.1 and 5.2). Raw sequencing reads were filtered, trimmed, and *de-novo* assembled, before screening for potential contamination (see Supplementary Text). Raw reads and assemblies of the newly sequenced transcriptomes were submitted to the NCBI Short Read Archive (SRA) and the NCBI Transcriptome Shotgun Assembly Sequence Database (TSA), respectively (see Table 5.2 for individual accession numbers). A detailed description of all methods applied can be found in the Supplementary Text and descriptive statistics of the assemblies are shown in Table 5.3.

5.2.2 Orthology assignment and alignment

We searched the each transcriptome for single copy genes (COGs) using Orthograph (version 0.6.3; Petersen et al. 2017) and a set containing 1,066 arthropod single copy orthologs based on BUSCO v2 (Simao et al. 2015) (see Supplementary Text). The reference set included 16 species that represent the diversity of arthropods and includes *Strigamia maritima*, the only high quality myriapod genome that is currently available (Chipman et al. 2014). Orthograph outputs both amino acid and nucleotide sequences of all identified COGs for downstream analyses. All amino acid sequences identified for each gene, including sequences of reference species, were aligned using MAFFT L-INS-I (version 7.310; Katoh and Standley 2013). Potential outliers (miss-identified or erroneously aligned sequences) were identified in the amino acid sequence alignments and subsequently removed from both the amino acid and nucleotide datasets (see Misof et al. 2014 and Supplementary Text for details) before

removing all sequences of reference taxa from both the amino acid alignments and the unaligned nucleotide sequences. Next, all positions that contained only gaps or ambiguous characters were removed from the amino acid alignments. Finally, we used a modified version of Pal2Nal (version 14; Suyama et al. 2006), which is able to handle ambiguous codons (Misof et al. 2014), to generate nucleotide multiple sequence alignments using the corresponding amino acid multiple sequence alignments as a guide.

5.2.3 Alignment processing

Both the nucleotide and amino acid multiple sequence alignments were analyzed using Alicscore (version 2.0; Kück et al. 2010, Misof and Misof 2009) in order to identify ambiguous and randomly similar sites across all possible pairwise comparisons between species. Identified sites were subsequently masked using Alicut (version 2.3; github.com/PatrickKueck/AliCUT). The resulting alignments were concatenated using FASconCAT-G (version 1.02; Kück and Longo 2014), creating amino acid and nucleotide supermatrixes. We then calculated the phylogenetic information content (IC) of each gene partition using MARE (version 0.1.2-rc; Misof et al. 2013) and removed all genes without phylogenetic information (IC = 0) from each of the supermatrixes. Next, we extracted four additional nucleotide supermatrixes that included only codon positions 1+2, 1, 2, and 3, respectively. An additional copy of each supermatrix was also filtered using the BMGE filtering algorithm with the BLOSUM62 matrix for the amino acid datasets and the PAM-100 matrix for the nucleotide dataset, setting the entropy cut-off to 0.5 (version 1.12;

Table 5.2. Overview of transcriptome data analyzed in the present study. Names of newly sequenced species in bold. Lengths given in base-pairs.

Species	TSA Accession	BioSample	BioProject	SRR Accession	Source	No. of contigs	After local VecScreen	Final no. of contigs	Mean length	Median length	N50 length
Adenomeris aibbosa	GHHY00000000	SAMN10983292	PRJNA523521	SRR8707657	This study	92,826	92,811	92,803	1,273	541	2,732
Eupeverimhoffia archimedis	GHIC00000000	SAMN10983309	PRJNA523528	SRR8707658	This study	73,632	73,638	73,624	2,242	1,333	4,306
Geoglomeris subterranea	GHHX00000000	SAMN10983323	PRJNA523531	SRR8707644	This study	114,077	114,061	114,050	1,504	598	3,424
Glomeridella minima	GERY00000000	SAMN04604763	PRJNA316405	SRR3485983	Szucsich et al. in review	39,296	39,296	39,283	903	581	1,424
Glomeris marginata	n/a	SAMN0455850	PRJNA315427	SRR3233211	Fernandez et al. 2016	152,067	149,579	149,579	638	367	1,459
Glomeris pustulata	GAKW00000000	SAMN01907469	PRJNA208707	SRR768330	Misof et al. 2014	0	0	40,834	375	330	373
Haploglomeris multistriata	GESA00000000	SAMN04604765	PRJNA316407	SRR3485985	Szucsich et al. in review	40,790	40,787	40,776	1,175	552	2,364
***Hyleoglomeris* sp. Japan**	GHHT00000000	SAMN10983325	PRJNA523532	SRR8713551	This study	331,916	33,912	33,750	1,286	631	2,523
Onomeris underwoodi	GHHZ00000000	SAMN10983326	PRJNA523533	SRR8713553	This study	74,520	74,509	74,447	1,878	890	3,857
Protoglomeris vasconica	GHIA00000000	SAMN10983328	PRJNA523535	SRR8713537	This study	32,573	32,571	32,254	1,341	641	2,713
Rhopalomeris carnifex	GHIE00000000	SAMN10983386	PRJNA523536	SRR8713586	This study	100,041	100,009	99,953	1,552	618	3,503
***Trachysphaera* sp.**	GHID00000000	SAMN10983388	PRJNA523541	SRR8713587	This study	100,804	100,797	100,762	1,575	623	3,526
Typhloglomeris martensi	GHIB00000000	SAMN10983397	PRJNA523542	SRR8713588	This study	95,185	95,173	95,165	1,612	690	3,411
Outgroup											
Brachycybe sp.	n/a	SAMN02205327	PRJNA209355	SRR945430	Brewer and Bond 2013	31,734	31,727	31,727	928	532	1,599
Cyliosoma sp.	n/a	SAMN04558501	PRJNA315428	SRR3458641	Fernandez et al. 2016	124,043	121,490	121,490	352	272	348
Eudiagraphis taiwanensis	n/a	SAMN04558465	PRJNA315426	SRR3458640	Fernandez et al. 2016	216,860	214,611	214,611	588	349	815
Glomeridesmus sp.	n/a	SAMN02205326	PRJNA209355	SRR941771	Brewer and Bond 2013	34,947	34,940	34,940	790	474	1,244
Narceus americanus	n/a	SAMN04558508	PRJNA315433	SRR3233222	Fernandez et al. 2016	150,738	148,041	148,041	417	279	441

Criscuolo and Gribaldo 2010). Genes that were under 150 characters long were excluded at this point. The supermatrixes were then filtered to insure a high overall completeness. The filtering criteria applied were: (1) sequences cover at least 75 % of the length of the partition, not including gaps and (2) partitions have sequence information from at least four in-group taxa.

Finally, in order to test for potential model violations in our data sets, we used Symtest (version 2.0.47; github.com/ottmi/symtest) to calculate the overall deviation from stationarity, reversibility, and homogeneity (SRH) between all species in each supermatrix (Jermiin et al. 2008). Following Misof et al. (2014) and Vasilikopoulos et al. (2019), we applied the Bowker's test of symmetry (Bowker 1948) to explore the assumption of evolution under SRH processes. The heat-maps generated by SymTest visualize the overall compositional homogeneity of each supermatrix (Figure 5.2A–D; Figure S1). Due to the large compoistional heterogeneity observed in the nucleotide datasets containing codon positions 1+2+3, 1+2, 1, and 3 (Figure S1), only the second codon position and the amino acid supermatrixes were further analyzed (Table 5.4).

5.2.4 Maximum likelihood phylogenetic analyses

The optimal partitioning scheme for the each supermatrix and corresponding substitution models were identified using ModelFinder (Chernomor et al. 2016, Kalyaanamoorthy et al. 2017) as implemented in IQTREE (version 1.6.10; Nguyen et al. 2015). The optimal partitioning scheme was identified using the relaxed clustering algorithm (options: -rcluster 50 -rcluster-max 10000). For the amino acid supermatrixes, the following substitution models were considered while allowing for

all possible rate heterogeneity types (option -mrate E,I,G,I+G,R): DAYHOFF (Dayhoff et al. 1978), DCMUT (Kosiol and Goldman 2005), JTT (Jones et al. 1992), JTTDCMUT (Kosiol and Goldman 2005), LG (Le and Gascuel 2008), WAG (Whelan and Goldman 2001), LG4X (Le et al. 2012), and LG4M (Le et al. 2012). All implemented models were considered for the nucleotide supermatrixes. The median approximation for Gamma rate heterogeneity was used for all supermatrixes and the best model fitting was selected based on the corrected Akaike information citerion (options: -gmedian -merit AICc). Maximum likelihood tree searches were conducted in IQTREE (version 1.6.10) using the edge-linked-proportional partition model (option: -spp). We conducted ten independent tree searches using the default 100 parsimony tress and an additional ten searches using a random starting tree (option: -t RANDOM). To account for potential effects of heterotachy, we additionally analyzed both nucleotide supermatrixes under the ghost model with unlinked parameters and base frequencies among classes (option: -m GTR+FO*H4; Crotty et al. 2019). Statistical branch support values for all trees were assessed with 500 non-parametric bootstrap replicates.

5.2.5 Coalescence based phylogenetic analyses

As an alternative to the supermatrix approach, we conducted a coalescent based species tree analysis in ASTRAL III (version 5.6.3; Mirarab and Warnow 2015, Zhang et al. 2018). For this, individual gene trees of all genes included in the supermatrixes were first inferred under the maximum likelihood optimality criterion using IQTREE (version 1.6.10), selecting the best fitting model using the implemented ModelFinder, and considering the same set of models and rates that were used in the respective

supermatrix analyses (see section 5.2.4). We conducted 10 gene tree searches for each gene and retained the best scoring tree for the species tree reconstruction. The species trees were then inferred from the gene trees in ASTRAL III, branch lengths were estimated in coalescence units, and the resulting tree was annotated with quartet support values (option: -t 2). For each branch, the three resulting quartet support values (q1, q2, and q3) show the percentage of quartets in the gene trees which support the species tree topolgy (q1) and the two alternative topologies (q2 and q3). Additionally, we scored the maximum likelihood tree based on the amino acid supermatrix to produce quartet scores as an alternative measure of branch support (option: -q).

5.2.6 Bayesian phylogenetic analysis

As an alternative to the two maximum likelihood based approaches, we conducted a bayesian based tree analysis using PhyloBayes (mpi-version 1.8; Lartillot and Philippe 2004, Lartillot and Philippe 2006, Lartillot et al. 2007). Due to computational constraints, only the BMGE filtered amino acid supermatrix was analyzed. Two separate chains were run using the CAT-GTR model with 4 discrete categories for the gamma distribution (options: -cat -gtr -dgam 4). Testing for convergence among the chains in both continuous parameters and the tree space was done using the included tracecomp and bcomp programs with a burn-in of 1000 and sub-sampling every 10 trees. At the point of stopping the tree-search, the maximal discrepancy between bi-partitions (max-diff) was 0.014, the effective size of all parameters were >50, and the two chains contained 1042 and 1031 trees, respectively (Table S3).

5.2.7 Topology testing, rogue taxon identification, and root placement

As we uncovered conflicting results based on different datasets and reconstruction methods, we tested the six alternative hypotheses (Figure 5.4) using the approximate unbiased (AU) test (Shimodaira 2002) implemented in IQTREE (version 1.6.12) on all four of the analyzed datasets. Furthermore, we conducted a rogue taxon identification analysis using RogueNaRok (version 1.0; Aberer et al. 2013) on the 500 non-parametric bootstrap replicates which were generated for each of the IQTREE analyses (section 2.4). Lastly, in order to test for potential effects of the outgroups on the tree reconstruction and root placement, we re-analyzed both nucleotide datasets without the outgroup taxa using IQTREE, following the same steps as in the previous analyses (section 2.4). The inferred tree was then rooted using the exhaustive search (option --exhaustive) in RootDigger (version 1.1.4; Bettisworth and Stamatakis 2020).

5.3 Results

5.3.1 Orthology assignment and alignment processing

We identified between 660 and 1,019 (median = 1002.5) of the 1,066 COGs with the total number of amino acids in each species ranging from 310,607 to 73,181 (median = 302,360 AA; Table 5.3). Due to the low number and short genes identified in the transcriptomes of *Cyliosoma* sp. (660 COGs, 73,181 AA), *Glomeris pustulata* (849 COGs, 111,669 AA), and *Narceus americanus* (895 COGs, 164,631 AA), these species were excluded from further analysis to prevent them from negatively impacting the alignments and matrix completeness. After the removal of 35 sequences, which were identified as outliers, the dataset consisted of 1,046 genes. The search for potentially ambiguous or erronusly aligned sites using ALISCORE and

ALICUT masked 55,226 and 164,678 in the amino acid and nucleotide datasets, respectively. Removing genes without phylogenetic information (IC = 0) using MARE further reduced the datasets to 1033 genes. After applying the BMGE filtering and filtering for completeness the final supermatrixes contained 504–991 genes spanning 153,799–318,796 aligned positions, of which 10.44–19.76% were parsimony informative (Table 5.4). The bowker's test of symmetry shows that there are significant deviations from the SRH-conditions in across our datasets. Due to the particularly strong deviations observed in the nucleotide data sets containing all three codon positions, positions 1+2, 1, and 3 (Figure S1), only the two amino acid supermatrixes and the two supermatrixes derived from the second codon position were further analyzed. In these remaining datasets, the most significant deviations can be seen in respect to the outgroup taxa and overall the scores are improved in the BMGE-filtered datasets with the nucleotide datasets performing slightly better than the amino acid datasets (Figure 5.2A–D). Similarly, the BMGE-filtering also improved the completeness scores compared to the respective unfiltered datasets (Figure 5.2E–H).

Table 5.3. Summary statistics of the orthology assignment at the amino acid level calculated using the helper script provided with the Orthograph pipeline. Asterisk indicates taxa not included in subsequent analyses. Protein lengths given as number of amino acid positions.

Species	No. orthologous hits	Proportion of COGs (%)	Total no. of amino acids	No. of X residues	No. of stop codons	N50 of protein lengths	Mean protein length	Median protein length	Max. protein length	Min. protein length
Adenomeris gibbosa	1,006	94.37	308,199	0	13	363	306	250	1,761	32
Brachycybe sp.	978	91.74	273,739	0	3	323	279	235	1,594	30
**Cyliosoma sp.*	660	61.91	73,181	0	6	120	110	100	473	27
Eudigraphis taiwanensis	984	92.31	259,958	0	7	306	264	220	1,755	29
Eupeyerimhoffia archimedis	1,007	94.47	308,623	0	10	360	306	254	1,756	32
Geoglomeris subterranea	1,003	94.09	303,587	0	10	363	302	248	2,101	33
Glomeridella minima	977	91.65	273,504	6	10	324	279	236	1,371	33
Glomeridesmus sp.	986	92.50	271,703	0	4	322	275	233	1,760	14
Glomeris marginata	995	93.34	258,711	0	14	303	260	226	1,756	32
**Glomeris pustulata*	849	79.64	111,669	28	43	144	131	118	483	8
Haploglomeris multistriata	1,007	94.47	302,910	2	6	357	300	246	1,756	10
Hyleoglomeris sp. Japan	1,011	94.84	306,747	15	7	361	303	253	1,744	25
**Narceus americanus*	895	83.96	164,631	0	13	213	183	157	797	31
Onomeris underwoodi	1,017	95.40	310,607	0	5	358	305	251	2,052	32
Protoglomeris vasconica	1,002	94.00	301,810	9	9	359	301	251	1,756	18
Rhopalomeris carnifex	1,010	94.75	306,622	0	7	354	303	253	1,811	25
Trachysphaera sp.	1,019	95.59	309,804	0	8	358	304	252	1,814	22
Typhloglomeris martensi	1,010	94.75	308,522	0	12	355	305	252	2,517	33

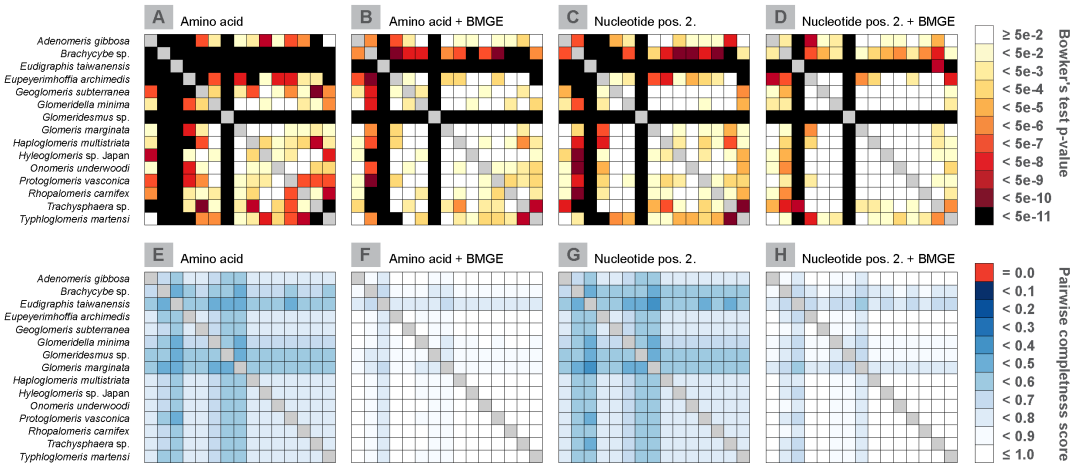

Figure 5.2: Heat-maps showing symmetry and completeness of the analyzed datasets. **A–D:** Pairwise p-values for Bowker's test of symmetry generated with SymTest. **E–H:** Pairwise completeness scores computed using AliStat.

Table 5.4. Summary statistics of the four analyzed datasets.

Matrix	No. of genes	No. of alignment sites	No. Parsimony informative sites	Percentage of parsimony informative sites	Overall alignment completeness score (Ca)	Minimum completeness score for pairs of sequences (Cr ij)	Average p-distance between sequences	Information content (IC)	Median pairwise p-value to the Bowker's test	Percent of pairwise p-values < 0.05 to the Bowker's test
Amino acid	991	318,762	62,994	19.76%	0.772	0.402	0.173	0.527	3.39E-08	87.62%
Amino acid + BMGE filtering	504	153,799	29,906	19.44%	0.941	0.641	0.141	0.643	5.59E-04	68.57%
Nucleotide (2. codon pos.)	844	318,796	33,335	10.46%	0.772	0.402	0.157	0.721	9.29E-05	75.24%
Nucleotide (2. codon pos.) + BMGE filtering	597	188,571	19,682	10.44%	0.927	0.594	0.134	0.747	6.59E-03	59.05%

5.3.2 Phylogenetic analyses

Glomerida were recovered as the monophyletic sister group to Glomeridesmida with strong support across all datasets and analyses (Figures 5.3, 5.4, and S3–6). However, none of the currently proposed groups, that were represented with more than one species, could be recovered as monophyletic. Within Glomerida, only two groups, respectively comprising of seven and two species, are consistently recovered and supported as monophyletic. Within the first monophyletic group (MG1),

Trachypshaera sp. is recovered as the earliest branching member, branching before *Glomeridella minima* (Figures 5.3 and S3–6). Within the clade that is sister to *G. minimia*, *Rhopalomeris carnifex* + *Hyleoglomeris* sp. Japan were recovered as the sister group to a clade where *Protoglomeris vasconica* is the sister to *Glomeris marginata* + *Haploglomeris multistriata* (Figures 5.3 and S3–6). The second stable monophyletic group (MG2) comprises of *Adenomeris gibbosa* and *Geoglomeris subterranea* (Figures 5.3 and S3–6). The relationships between these two groups and the remaining three Glomerida, however, are unstable across analyses and datasets, with a total of six different inferred topologies (Figures 5.4 and S3–6).

Outside of the two monophyletic groups, the analyses of the two amino acid datasets resulted in three different topologies depending on the applied reconstruction method (Figures 5.4A–C, S3, and S4). ASTRAL recovered MG2 as the earliest branching Glomerida, followed by the consecutively branching *T. martensi*, *E. archimedis*, *O. underwoodi*, and MG1 (Figure 5.4A). IQTREE recovered *T. martensi* as the earliest branching Glomerida, branching before MG2 (Figure 5.4B). MG2 was recovered as the sister group to a clade uniting MG1 as the sister group to *O. underwoodi* + *E. archimedis* (Figure 5.4B). PhyloBayes recovered *O. underwoodi* + MG1 as the sister group to a clade in which *E. archimedis* + *T. mart*ensi was recovered as the sister group to MG2 (Figure 5.4C). The topologies recovered by PhyloBayes and IQTREE were overall strongly supported by posterior probabilities (all > 90; Figures 5.4C, S3 and S4) and bootstrap scores (only two branches < 90 %; Figure 5.4B, S3 and S4), respectively.

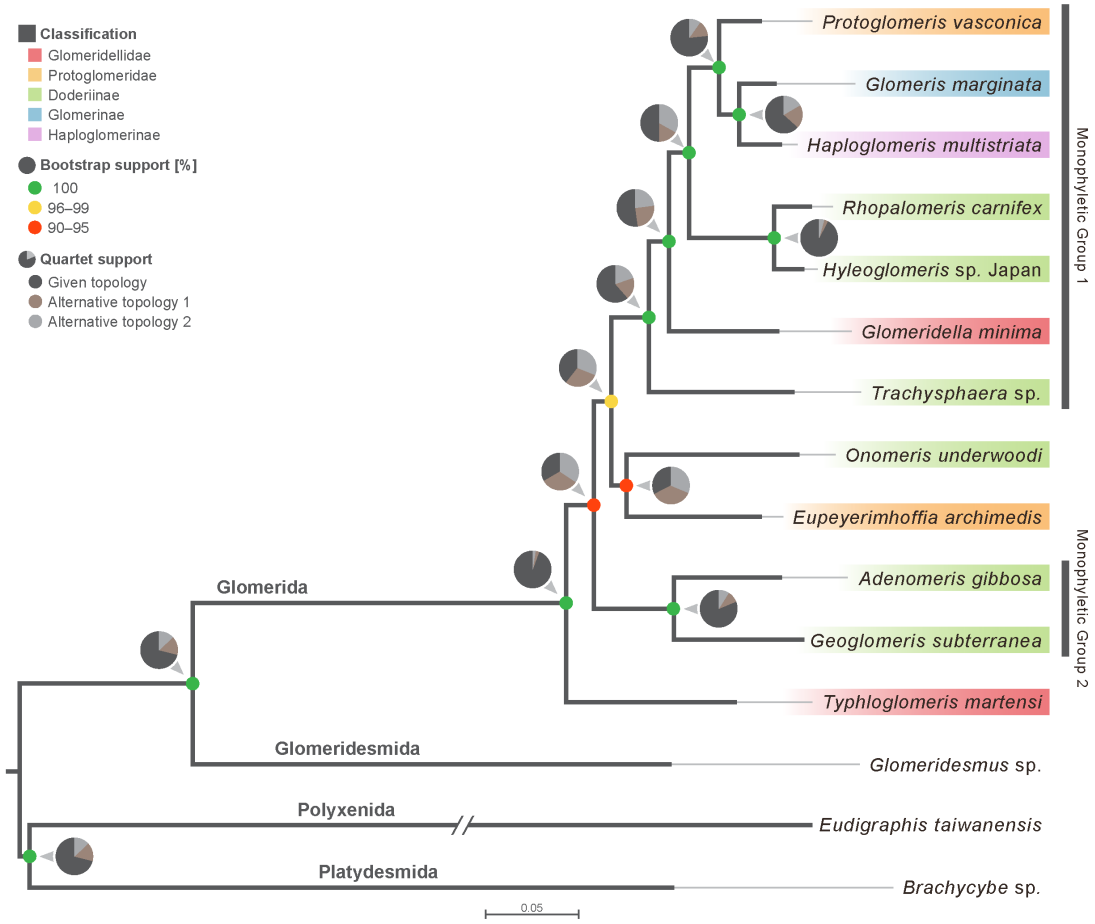

Figure 5.3: Maximum likelihood tree obtained from the analysis of the unfiltered amino acid dataset. Branch support values were generated with 500 non-parametric bootstrap replicates and the quartet support (pie charts) was generated by scoring the shown topology using all gene-trees from the same dataset using ASTRAL. Classification and colors refer to families and subfamilies as defined by Enghoff et al. (2015).

The posterior probabilities of the topology recovered by the ASTRAL analyses were comparably low, only giving strong support (> 90) to the monophyly of Glomerida and the two monophyletic group (Figures 5.4A, S3, and S4). The quartet scores of the ASTRAL analyses show a close to equal distribution (~ 30 %) among the three possible topologies at the branches placing *O. underwoodi, E. archimedis, and T. martensi* in both the topologies inferred using ASTRAL (Figures S3 and S4) and the scoring of the IQTREE topology derived from the amino acid dataset (Figure 5.3).

The analyses of the two nucleotide datasets resulted in three new topologies (Figures 5.4D–F and S5–6) and one which was identical to the topology inferred by the ASTRAL analyses of the amino acid datasets (Figures 5.4A and S5). The latter was inferred by the IQTREE analysis of the non-BMGE-filtered dataset using the ghost model. The three remaining analyses of the nucleotide datasets using IQTREE, irrespective of the model applied, resulted in an identical topology (Figures 5.4D and S5–6). Here, the position of MG2 and *T. martensi* are exchanged relative to topology 1 (Figures 5.4D and S5–6). The ASTRAL analyses of the two nucleotide datasets both recovered MG2 as the earliest branching Glomerida, but differ in the placement of the remaining taxa (Figures 5.4E–F and S5–6). The ASTRAL analysis of the non-BMGE-filtered dataset recovered *T. martensi*, *O. underwoodi*, *E. archimedis*, and MG1 as consecutively branching after MG2 (Figures 5.4E and S5). The ASTRAL analysis of the BMGE-filtered dataset recovered *E. archimedis* + *T. martensi*, as the sistergroup to *O. underwoodi* + MG1 (Figures 5.4D and S6). All of the topologies inferred by IQTREE received strong branch support across the tree (100 %; Figures 5.4A, 5.4D and S5–6). The two topologies inferred using ASTRAL, similar to those of the amino acid datasets, only receive strong branch support for the monophyly of Glomerida, MG1, and MG2 (Figures 5.4E–F and S5–6).

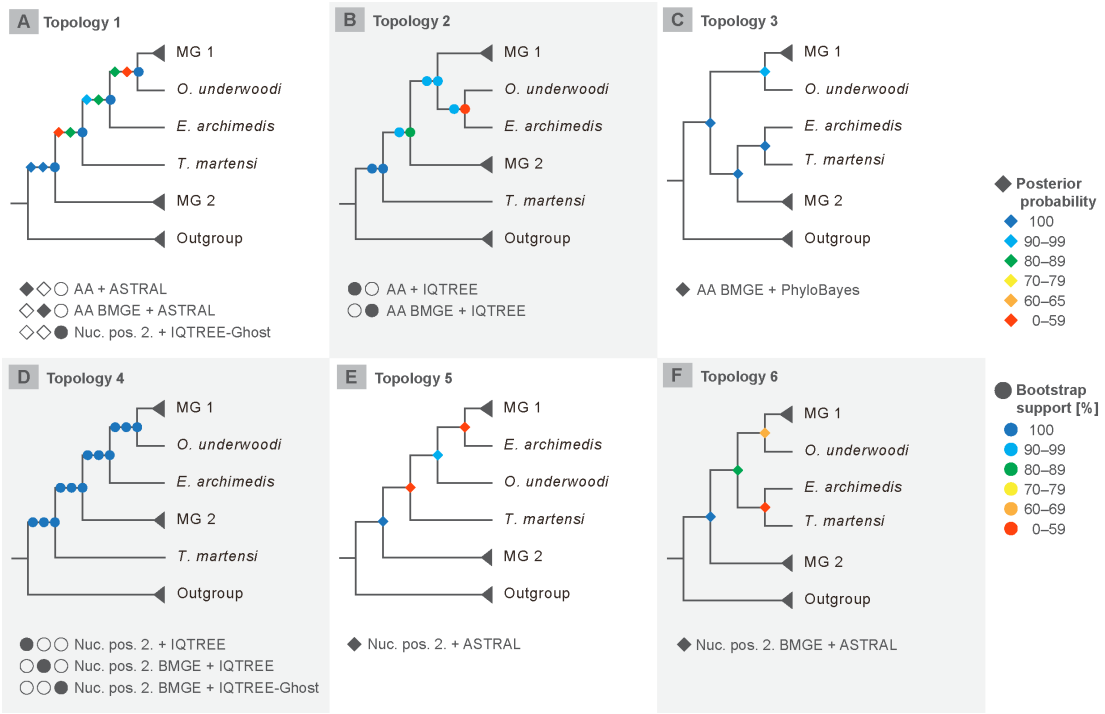

Figure 5.4: Summary of topologies inferred from all conducted analyses on all investigated datasets. Branch support values derived from 500 non-parametric bootstrap replicates (IQTREE) and posterior probabilities (ASTRAL and PhyloBayes). Abbreviations: AA = Amino acid; Nuc = Nucleotide; pos. 2 = second codon position; MG1 = Monophyletic group 1; MG2 = Monophyletic group 2.

5.3.3 Topology testing, rogue taxon identification, and root placement

The topology test significantly rejected topology 3 (AA BMGE + PhyloBayes) and 5 (Nuc. pos. 2. + ASTRAL) across all datasets (Table 5.5). Topology 1 was rejected based on both the amino acid datasets, while topology 6 was rejected based on both the BMGE-filtered datasets (Table 5.5). Topologies 2 and 4 were not rejected based on any dataset (Table 5.5).

Table 5.5. p-values of approximately unbiased (AU) test (Shimodaira 2002) on all four analyzed datasets for six alternative topologies (see Figure 5.4). Minus sign (-) denotes significantly rejected topologies.

Dataset	AA		AA BMGE		Nuc. pos. 2.		Nuc. pos. 2. BMGE	
Topology 1	0.043	-	0.0287	-	0.194		0.294	
Topology 2	0.612		0.942		0.547		0.328	
Topology 3	2.88E-07	-	1.40E-44	-	2.68E-05	-	2.96E-05	-
Topology 4	0.592		0.121		0.683		0.805	
Topology 5	2.27E-05	-	0.000144	-	0.00128	-	0.000377	-
Topology 6	0.142		0.0235	-	0.194		0.0388	-

No rogue taxa were identified in any of the datasets. The reanalysis of the two nucleotide datasets without the outgroup taxa, that were the source of the most significant model violations (Figure 5.2A–D), and subsequent automatic rooting using RootDigger resulted in an identical topology to the original analyses (Figure 5.5). However, the likelihood weight ratio values (LWR) values indicate that alternative root placements are almost equally likely (Figure 5.5).

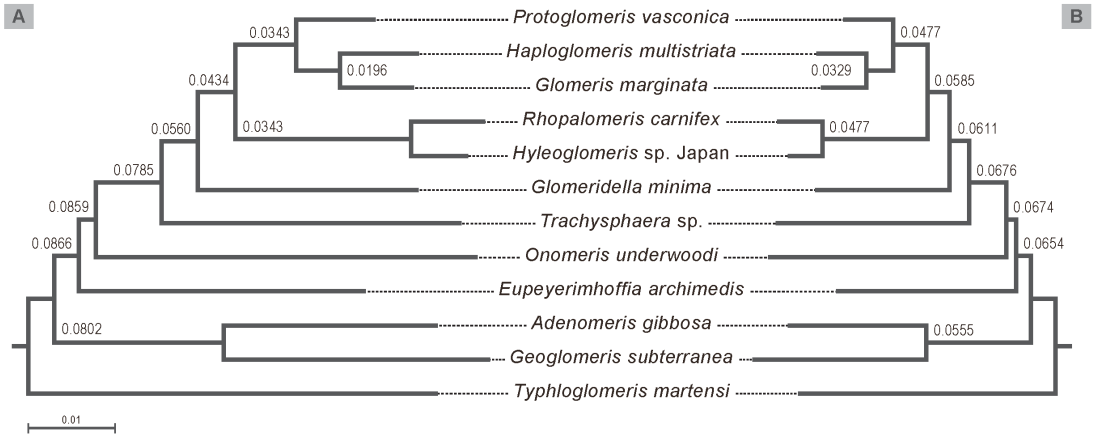

Figure 5.5: Root placement and likelihood weight ratios (LWR) for alternative placement as inferred using RootDigger. **A:** Topology inferred from the nucleotide dataset without additional filtering. **B:** Topology inferred from the BMGE filtered nucleotide dataset.

5.4 Discussion

5.4.1 Discordance among reconstruction methods and datasets

Although a majority of the relationships within Glomerida are well resolved, we could not robustly resolve the position of three species: *O. underwoodi, E. archimedis,* and *T. martensi*. Conflicting results between datasets and reconstruction methods have been attributed to model violations (e.g. Feuda et al. 2017). However, although the analyzed datasets show some violations of model assumtions (SRH-citerion), they do not seem to be a large driving factor of discordance in our analyses. The reduction of model violations in our datasets using the BMGE filtering only slightly changed the topologies of the inferred trees in two analyses of the nucleotide data (IQTREE with the ghost model and ASTRAL), in one of which neither topology received statistical support. Furthermore, the tree reconstruction after the removal of the outgroup taxa, which introduced the most substantial model violations, and subsequent automatic root placement did not lead to changes in the inferred topologies.

Other often cited potential sources of conflict, that can be hard to distinguish from each other, are ancient rapid speciation events (Whitfield and Lockhart 2007, Whitfield and Kjer 2008, Sayyari and Mirarab 2016), sometimes accompanied by subsequent incomplete lineage sorting (ILS; Maddison 1997), and confounding factors in the dataset (Whitfield and Kjer 2008). Speciation processes and divergence rates have yet to be studied in detail within Myriapoda. However, given the relatively high estimated age of Glomerida (200 mya; Rodriguez et al. 2018), with the most diverse extant genus *Hyleoglomeris* being at least 28-40 mya (Wesener 2019), and that speciation in millipedes can in some cases occur over a relatively short time

(Spelda et al. 2011, Wesener and Conrad 2016), it is conceivable that ancient rapid speciation could have taken place during the groups evolutionary history. Both ancient rapid speciation events, with or without ILS, and confounding can result in trees with short internal branches. We observe relatively short branches along the backbone of all the inferred trees, particularly those branches close to the three problematic species (Figures 5.3 and S3–S6). Furthermore, we observe strong gene tree and species tree discordance along these branches, as can be characteristic of ILS (Maddison 1997). Although the numerous inferred topologies prevent conclusive testing for the presence of confounding factors, initial tests utilizing the hypotheses derived from the maximum likelihood and coalescence based analyses of the amino acid data indicate that there are likely no strong confounding factors affecting the reconstructions based on this dataset (see Supplementary Text). Ultimately, a wider taxon sampling in combination with a dating and diversification analysis will be necessary to disentangle these effects in future studies. If ILS is the root cause of the conflict observed in our analyses, it is likely that alternative methods are required to confidently resolve the relationships of Glomerida. One such approach is using structural characters derived from whole genome sequencing (e.g. Niehuis et al. 2012, Suh et al. 2015).

Alternatively, the reference set of orthologous genes could be a factor contributing to the short branch lengths and lack of a clear phylogenetic signal at some branches. Due to the dearth of available myriapod reference genomes for the identification of group-specific single copy genes, the set was based on a wide taxonomic sampling of reference species, resulting in a potentially overly conservative gene set, indicated by the low percentage of parsimony informative sites in our datasets (10–20 %, Table

5.4). Other studies on myriapods have used *de-novo* identification of orthologous groups from transcriptomes in order to increase the number of genes available for analysis (e.g. > 2,000 genes; Rodriguez et al. 2018). However, without genomic evidence, this approach runs the risk of inaccurate orthology prediction due to the inherent incompleteness of transcriptomes and the difficulty involved in distinguishing between isoforms and paralogs of genes (Petersen et al. 2017). Improving the genome availability within Myriapoda is therefore a crucial step for furthering gene-based phylogenomics of the group, especially for addressing higher-level relationships.

5.4.2 Intraordinal relationships of Glomerida

Our results corroborate the previously indicated shortcomings of the current classification (Hoffman 1980, Oeyen and Wesener 2015, Oeyen and Wesener 2018) and show that issues are prevalent throughout the current system. Irrespective of the true position of *O. underwoodi*, *E. archimedis*, and *T. martensi*, the Glomeridellidae, Protoglomeridae, and Doderiinae are rendered non-monophyletic. These results affirm the reported lack of phylogenetic signal in the morphological characters currently used to define groups within the order (Oeyen and Wesener 2018). This includes characters that were previously considered strong potential apomorphies, unrelated to modifications of the primary telopods. For example, the species of the polyphyletic Glomeridellidae are characterized by a transformation of the second to last leg pair to anterior telopods (Enghoff et al. 2015).

The monophyly of the sub-families, Glomerinae and Haploglomerinae, could not be tested in our analyses, as they are both only represented by a single species. The two

subfamilies are currently differentiated on a comparably simple character, the presence or absence of an internal femoral process on the telopods (Enghoff et al. 2015). Therefore, increasing the sampling of these two subfamilies should be an aim of future studies.

5.4.3 Biogeography and the evolution of dwarfism

The hypothesized close relationship of the American Glomerida, represented by *Onomeris* in our dataset, and the European genera comprising only of dwarf species to the predominantly Asian genus *Hyleoglomeris* (Wesener 2010, Wesener 2012) are not corroborated by our results. Although the exact placement of *Onomeris* remains unresolved, *Onomeris* was never recovered as a close relative of *Hyleoglomeris* in either of the inferred topologies. Furthermore, the group uniting *Hyleoglomeris* and the Asian genus *Rhopalomeris* was consistently recovered within the first monophyletic group along European genera with strong support across all analyses. However, our analyses only included one out of three American genera and a single species of *Hyleoglomeris*, which is one of the largest and geographically widest spread Glomerida genera (Golovatch et al. 2006). Therefore, a more inclusive sampling is needed to fully unravel the origins of the American Glomerida.

Similarly, we did not find a close affinity of any of the included dwarf genera to either of the Asian genera (*Hyleoglomeris* and *Rhopalomeris*). We inferred *Geoglomeris* as the sister-taxon to the aberrant *Adenomeris*, separate from the aberrant *Trachysphaera* with strong support across analyses. Contrary to the monophyletic grouping of three dwarf genera suggested by morphological evidence (Oeyen and Wesener 2018) and the hypothesized single origin of aberrant species (Attems 1926, Verhoeff 1932,

Schubart 1934, Hoffman 1970), this indicates that both dwarfism and the aberrant morphology evolved at least twice independently within Glomerida. However, as our dataset only includes a small sample of the diversity of dwarf and aberrant species, further cases of convergent evolution could potentially be revealed with a wider taxonomic sampling. Currently, a third independent origin seems likely if the distant relationship of Glomerellina to the other European dwarves (Hoffman 1980; Enghoff et al. 2015; Oeyen and Wesener 2018) can be confirmed.

5.5 Conclusions and future outlook

In conclusion, we found (1) that at least three out of the five included groups of the current system do not represent monophyletic units, that (2) neither the American genus *Onomeris* nor the European dwarf species are closely related to the Asian genera, and that (3) dwarfism and aberrant morphology evolved at least twice within the order.

As a consequence of the wide spread paraphyly in the current system, a representative approach to resolving the phylogenetic relationships within the order is not feasible and future studies will require a comprehensive sampling, both in terms of taxonomic and geographic diversity. An increased sampling of *Glomeris* and *Hyleoglomeris* is of particular importance, as they together contain over half the currently described species of Glomerida (ca 80 and 90, respectively) and also show the widest geographic distribution (Enghoff et al. 2015). Because Glomerida are rarely collected and can also be difficult to collect despite having precise locality information, this could be achieved by applying methods that allow for the usage of museum

specimens. Such methods include anchored hybrid enrichment (Mayer et al. 2016; Sann et al. 2018) or alternatively low coverage whole genome sequencing (Johnson 2019). Lastly, our results corroborate the previously reported need for new, telopod-unrelated, morphological characters for accurate taxonomy.

5.6 Acknowledgements

We thank the 1KITE project for use of their infrastructure and pipelines. We also thank Claudia Etzbauer and Sandra Kukowka for their support in the molecular lab, Shibo-Kun for his general support, and Dr. Christoph Mayer for his suggestions and productive discussions. Lastly, JPO thanks the Zoological Research Museum Alexander Koenig as well as the Leibniz Graduate School of Genomic Biodiversity Research for funding.

5.7 References

Aberer AJ, Krompass D, Stamatakis A (2013) Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. Systematic Biology. 62: 162–166.

Attems C (1926) Myriopoda. In: Kükenthal W, Krumbach T (Eds) Handbuch der Zoologie. Eine Naturgeschichte der Stämme des Tierreiches. 4(1). Progoneata, Chilopoda, Insecta I: 1–402. Walter de Gruyter, Berlin.

Bettisworth B, Stamatakis A (2020) RootDigger: a root placement program for phylogenetic trees. BioRxiv. Doi: 10.1101/2020.02.13.935304.

Blanke A, Wesener T (2014) Revival of forgotten characters and modern imaging techniques help to produce a robust phylogeny of the Diplopoda (Arthropoda, Myriapoda). Arthropod Structure and Development. 43: 63–75.

Bonkowski M, Scheu S, Schaefer M (1998) Interactions of earthworms (*Octolasion lacteum*), millipedes (*Glomeris marginata*) and plants (*Hordelymus europaeus*) in a beechwood on a basalt hill: Implications for litter decomposition and soil formation. Applied Soil Ecology. 9: 161–166.

Bowker AH (1948) A test for symmetry in contingency tables. Journal of the American Statistical Association. 43: 572–574.

Brewer MS, Bond JE (2013) Ordinal-Level Phylogenomics of the Arthropod Class Diplopoda (Millipedes) Based on an Analysis of 221 Nuclear Protein-Coding Loci Generated Using Next-Generation Sequence Analyses. PLoS One. 8:e79935.

Chernomor O, von Haeseler A, Minh BQ (2016) Terrace aware data structure for phylogenomic inference from supermatrices. Systematic Biology. 65: 997–1008.

Chipman AD, Ferrier DEK, Brena C, Qu J, Hughes DST, et al. (2014) The First Myriapod Genome Sequence Reveals Conservative Arthropod Gene Content and Genome Organisation in the Centipede *Strigamia maritima*. PLoS Biology. 12: e1002005.

Criscuolo A, Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evolutionary Biology. 10: 210.

Crotty SM, Minh BQ, Bean NG, Holland BR, Tuke J, et al. (2019) GHOST: Recovering historical signal from heterotachously-evolved sequence alignments. Systematic Biology. 69: 249–264.

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. Atlas of Protein Sequence Structure. 5: 345–351.

Dohle W (1964) Die Embryonalentwicklung von *Glomeris marginata* (Villers) im Vergleich zur Entwicklung anderer Diplopoden. Zoologische Jahrbücher Abteilung für Anatomie und Ontogenie der Tiere. 81: 241–310.

Enghoff H (1984) Phylogeny of millipedes–a cladistic analysis. Journal of Zoological Systematics and Evolutionary Research. 22: 8–26.

Enghoff H, Golovatch S, Short M, Stoev P, Wesener T (2015) Diplopoda—taxonomic overview. In: Minelli, A. (Ed.). Treatise on Zoology-Anatomy, Taxonomy, Biology. The Myriapoda, Vol. 2. Brill, Leiden. 363–453 pp.

Fernández R, Edgecombe GD, Giribet G (2016) Exploring phylogenetic relationships within Myriapoda and the effects of matrix composition and occupancy on phylogenomic reconstruction. Systematic Biology. 65: 871–889.

Feuda R, Dohrmann M, Pett W, Philippe H, Rota-Stabelli O, et al. (2017) Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals. Current Biology. 27: 3864–3870.

Golovatch SI, Geoffroy J, Mauriès JP. Review of the millipede genus *Hyleoglomeris* Verhoeff, 1910 (Diplopoda, Glomerida, Glomeridae), with descriptions of new species from caves in Southeast Asia. Zoosystema-Paris. 28: 887–915.

Haacker U (1969) Sperma übertragung von *Glomeris* (Diplopoda). Naturwissenschaften. 56: 467.

Hoffman RL (1980) Classification of the Diplopoda. Museum d'Histoire naturelle, Genève. 237 pp.

Janssen R, Budd GE, Prpic NM, Damen WGM (2011) Expression of myriapod pair rule gene orthologs. EvoDevo. 2: 5.

Jermiin LS, Jayaswal V, Ababneh F, Robinson, J (2008) Phylogenetic model evaluation. In: In: Keith J (Ed.) Bioinformatics, Data, Sequences Analysis and Evolution, vol. I. Humana Press, Totowa. 331–363 pp.

Johnson KP (2019) Putting the genome in insect phylogenomics. Current Opinion in Insect Science. 36: 111–117.

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences. 8: 275–282.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. Nature Methods. 14: 587–589.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution. 30: 772–780.

Kosiol C, Goldman N (2005) Different versions of the dayhoff rate matrix. Molecular Biology and Evolution.. 22: 193–199.

Kück P, Longo GC (2014) FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Frontiers in Zoology. 11: 81.

Kück P, Meusemann K, DambachJ, Thormann B, von Reumont BM, et al. (2010) Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Frontiers in Zoology,. 7: 10.

Lartillot N, Brinkmann H, Philippe H (2007) Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evolutionary Biology. 7: S4.

Lartillot N, Philippe H (2004) A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. Molecular Biology and Evolution. 21: 1095–1109.

Lartillot N, Philippe H (2006) Computing Bayes factors using thermodynamic integration. Systematic Biology. 55: 195–207.

Le SQ, Dang CC, Gascuel O (2012) Modeling protein evolution with several amino acid replacement matrices depending on site rates. Molecular Biology and Evolution. 29: 2921–2936.

Le SQ, Gascuel O (2008) An Improved General Amino Acid Replacement Matrix. Molecular Biology and Evolution. 25: 1307–1320.

Maddison WP (1997) Gene trees in species trees. Systematic Biology. 46: 523–536.

Mauriès JP (1984) Diplopodes cavernicoles de Toscane (Materiel du Musée Zoologique, Florence): Nouveautes chorologiques et taxonomiques sur les genres *Geoglomeris*, Devillea et *Orphanoiulus*. Redia. 67: 401–416.

Mauriès JP (1971) Diplopodes épigés et cavernicoles de Pyrénées espagnoles et des monts cantabriques: VII. Glomérides. Essai de classification des Glomeroidea. Bulletin de La Société d'Histoire Naturelle de Toulouse. 107: 423–453.

Mauriès JP (2005) Attempt at a classification of Glomerida (Diplopoda), with descriptions of two new genera from the northwest of the Iberian Peninsula. Arthropoda Selecta. 14: 241–249.

Mayer C, Sann M, Donath A, Meixner M, Podsiadlowski L, et al. (2016) BaitFisher: A Software Package for Multispecies Target DNA Enrichment Probe Design. Molecular Biology and Evolution. 33: 1875–1886.

Mirarab S, Warnow T (2015) ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 31: i44–i52.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. Science. 346: 763–767.

Misof B, Meyer B, von Reumont BM, Kück P, Misof K, et al. (2013) Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. BMC Bioinformatics. 14: 348.

Misof B, Misof K (2009) A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Systematic Biology. 58: 21–34.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Molecular Biology and Evolution. 32: 268–274.

Niehuis O, Hartig G, Grath S, Pohl H, Lehmann J, et al. (2012) Genomic and morphological evidence converge to resolve the enigma of Strepsiptera. Current Biology. 22: 1309–1313.

Oeyen JP, Wesener T (2015) Steps towards a phylogeny of the pill millipedes: non-monophyly of the family Protoglomeridae, with an integrative redescription of *Eupeyerimhoffia archimedis* (Diplopoda, Glomerida). ZooKeys. 510: 49–64.

Oeyen JP, Wesener T (2018) A first phylogenetic analysis of the pill millipedes of the order Glomerida, with a special assessment of mandible characters (Myriapoda, Diplopoda, Pentazonia). Arthropod Structure and Development. 47: 214–228.

Petersen M, Meusemann K, Donath A, Dowling D, Liu S, et al. (2017) Orthograph: A versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. BMC Bioinformatics. 18: 1–10.

Rawlins AJ, Bull ID, Poirier N, Ineson P, Evershed RP (2006) The biochemical transformation of oak (*Quercus robur*) leaf litter consumed by the pill millipede (*Glomeris marginata*). Soil Biology and Biochemistry. 38: 1063–1076.

Regier JC, Wilson HM, Shultz JW (2005) Phylogenetic analysis of Myriapoda using three nuclear protein-coding genes. Molecular Phylogenetics and Evolution. 34: 147–158.

Rodriguez J, Jones TH, Sierwald P, Marek PE, Shear WA, et al. (2018) Step-wise evolution of complex chemical defenses in millipedes: A phylogenomic approach. Scientific Reports. 8: 1–10.

Sann M, Niehuis O, Peters RS, Mayer C, Kozlov A, et al. (2018) Phylogenomic analysis of Apoidea sheds new light on the sister group of bees. BMC Evolutionary Biology. 18: 71.

Sayyari E, Mirarab S (2016) Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution. 33: 1654–1668.

Schubart O (1934) Tausendfüssler oder Myriapoda I: Diplopoda. Die Tierwelt Deutschlands und der angrenzenden Meeresteile nach ihren Merkmalen und nach ihrer Lebensweise (Begr. von F. Dahl, weitergef. v. M. Dahl & H. Bischoff), 28. Teil: vii. G. Fischer, Jena. 318 pp.

Shimodaira H (2002) An Approximately Unbiased Test of Phylogenetic Tree Selection, Systematic Biology. 51: 492–508.

Sierwald P, Bond JE (2007) Current status of the myriapod class Diplopoda (millipedes): taxonomic diversity and phylogeny. Annual Review of Entomology. 52: 401–420.

Sierwald P, Shear WA, Shelley RM, Bond JE (2003) Millipede phylogeny revisited in the light of the enigmatic order Siphoniulida. Journal of Zoological Systematics and Evolutionary Research. 41: 87–99.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31: 3210–3212.

Spelda J, Reip HS, Oliveira-Biener U, Melzer RR (2011) Barcoding Fauna Bavarica: Myriapoda - a contribution to DNA sequence-based identifications of centipedes and millipedes (Chilopoda, Diplopoda). Zookeys. 139: 123–139.

Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. PLoS Biology. 13: 1–18.

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research. 34: W609–W612.

Szucsich N, Bartel D, Blanke A, Böhm A, Donath A, et al. (in review) Four myriapod relatives – but who are sisters? No end to debates on relationships among the four major myriapod subgroups. BMC Evolutionary Biology.

Vasilikopoulos A, Balke M, Beutel RG, Donath A, Podsiadlowski L, et al. (2019) Phylogenomics of the superfamily Dytiscoidea (Coleoptera: Adephaga) with an evaluation of phylogenetic conflict and systematic error. Molecular Phylogenetics and Evolution. 135: 270–285.

Verhoeff KW (1932) Diplopoda 2. Bronn's Klassen und Ordnungen des Tier-Reichs. 5 (II, 7-13). Akademischer Verlag, Leipzig. 1073–2084 pp.

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution. 18: 691–699.

Whitfield JB, Kjer KM (2008) Ancient Rapid Radiations of Insects: Challenges for Phylogenetic Analysis. Annual Review of Entomology. 53: 449–472.

Whitfield JB, Lockhart PJ (2007) Deciphering ancient rapid radiations. Trends in Ecology and Evolution. 22: 258–265.

Wesener T (2010) Revision of the Amercican Pill Millipedes I: *Onomeris* and *Trichomeris*. Zootaxa. 2725: 28–40.

Wesener T (2012) *Nearctomeris*, a new genus of Pill Millipedes from North America, with a comparison of genetic distances of American Pill Millipede Genera (Glomerida, Glomeridae). Zootaxa. 3258: 58–68.

Wesener T (2019) The oldest pill millipede fossil: A species of the Asiatic pill millipede genus *Hyleoglomeris* in Baltic amber (Diplopoda: Glomerida: Glomeridae). Zoologischer Anzeiger. 283: 40–45.

Wesener T, Conrad C (2016) Local Hotspots of Endemism or Artifacts of Incorrect Taxonomy? The Status of Microendemic Pill Millipede Species of the Genus *Glomeris* in Northern Italy (Diplopoda, Glomerida). PLoS One. 11: e0162284.

Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 19: 153.

Chapter 6

Sawfly genomes reveal evolutionary acquisitions that fostered the mega-radiation of parasitoid and eusocial Hymenoptera

This chapter is published in:
Oeyen JP, Baa-Puyouletet P, Benoit JB, Beukeboom LW, Bornberg-Bauer E, et al. (2020) Sawfly genomes reveal evolutionary acquisitions that fostered the mega-radiation of parasitoid and eusocial Hymenoptera. Genome Biology and Evolution. 12: 1099–1188.

Authors: Oeyen JP (JPO), Baa-Puyoulet P (PBP), Benoit JB (JBB), Beukeboom LW (LWB), Bornberg-Bauer E (EBB), Buttstedt A (AB), Calevro F (FC), Cash EI (EC), Chao H (HsC), Charles (HuC), Chen MM (MC), Childers C (CC), Cridge AG (AGC), Dearden P (PD), Dinh H (HD), Doddapaneni HV (HVD), Dolan A (AmD), Donath A (AlD), Dowling D (DD), Dugan S (SD), Duncan E (ED), Elpidina EN (ENE), Friedrich M (MF), Geuverink E (EG), Gibson JD (JDG), Grath S (SG), Grimmelikhuijzen CJP (CJPG), Große-Wilde E (EGW), Gudobba C (CG), Han Y (YH), Hansson BS (BH), Hauser F (FH), Hughes DST (DH), Ioannidis P (PI), Jacquin-Joly E (EJJ), Jennings EC (ECJ), Jones JW (JWJ), Klasberg S (SK), Lee SL (SLL), Lesný P (PL), Lovegrove M (ML), Martin S (SM), Martynov AG (AGM), Mayer C (CM), Montagné N (NM), Moris VC (VCM), Munoz-Torres M (MMT), Murali SC (SCM), Muzny DM (DMM), Oppert B (BO), Parisot N (NP), Pauli T (TP), Peters RS (RSP), Petersen

M (MaP), Pick C (CP), Persyn E (EP), Podsiadlowski L (LP), Poelchau MF (MFP), Provataris P (PP), Qu J (JQ), Reijnders MJMF (MR), von Reumont BM (BMvR), Rosendale AJ (AJR), Simao FA (FAS), Skelly J (JS), Sotiropoulos AG (AGS), Stahl AL (ALS), Sumitani M (MS), Szuter EM (EMS), Tidswell O (OT), Tsitlakidis E (ET), Vedder L (LV), Waterhouse RM (RMW), Werren JH (JHW), Wilbrandt J (JW), Worley KC (KCW), Yamamoto DS (DSY), Zande Lvd (LvdZ), Zdobnov EM (EZ), Ziesmann T (TZ), Gibbs RA (RAG), Richards S (SR), Hatakeyama M (MH), Misof B (BM), Niehuis O (ON).

Author contributions **Conceived the project:** BM, JPO, MH, ON, RSP, SR. **Lead authors:** BM, JPO, ON. **Major contributors:** BM, BO, JPO, MH, ON, SR. **Project coordinator:** JPO. **Project management:** JPO, MFP, RAG, SD, SR. **Antioxidants:** ECJ, JBB. **Aquaporins:** ALS, JBB. **Automated annotation:** DH, SR. **Autophagy genes:** EMS, JBB. **BUSCO:** EZ, FAS, PI, RMW. **CEGMA:** JPO, ON. **Chemoreceptor repertoire:** BH, EGW, EJJ, EP, NM. **Cuticular proteins:** AJR, JBB. **Metabolism:** FC, HuC, JPO, NP, PBP. **Cysteine peptidases:** AGM, BO, ENE. **Desaturases:** EC. **Dosage compensation:** DSY, MH, MS. **Developmental genes:** AGC, ED, JS, ML, OT, PD. **DNA methylation:** BM, PP, SG. **Elongases:** ON, VCM. **Facilitated manual gene curation:** MMT. **Gene family evolution:** BM, JPO. **Gene structure comparison:** BM, JPO, JW, ON. **GO-term:** RMW, MR. **Heat shock proteins:** ECJ, EMS, JBB. **Hexamerins:** CP. **Immune genes:** LP, PL. **Insulator proteins:** LV, ON, TP. **Lateral gene transfer:** AmD, JHW, MH, ON. **Major royal jelly proteins:** AB. **Metabolism of host plant toxins:** DSY, MH, MS. **Metallopeptidases:** LP, SM. **Microsynteny:** BM, JPO, ON. **Mitochondrial genomes:** AlD, JPO, ON. **Neuropeptides and their G protein-coupled receptors:** AGS, CJPG, ET, FH. **Non-coding RNA:** AlD, TZ. **Orthology:** EZ, PI, RMW. **OXPHOS genes:** JDG. **Protein domains:** EBB, SK. **Quality control and official gene set generation:** CC, MC, MFP. **Repeats and transposable elements:** CM, MaP, JPO. **RNAi:** AlD, DD, ON. **Sequencing and assembly:** DH, DMM, HD, HsC, HVD, JPO, JQ, KCW, ON, MH, SCM, SLL, YH . **Sex determination:** DSY, EG, LWB, LvdZ, MH, MS. **Silk proteins:** DSY, MH, MS. **Variable sites:** LP. **Venom proteins:** BMvR. **Vision genes:** CG, JWJ, MF. **Web Apollo:** CC.

Figure numbers have been adapted for the inclusion in the present thesis.

6.1 Introduction

Hymenoptera (sawflies, wasps, ants, and bees) represent one of the four mega-diverse insect orders. It is estimated to comprise over one million species and currently includes over 153,000 described species (Aguiar et al. 2013). The transition from an ancestral ectophytophagous lifestyle, retained by the majority of sawflies ("Symphyta"), to parasitoidism, a lifestyle in which a larva develops by feeding upon and killing a single host specimen, is generally considered the most important factor that promoted the diversification of Hymenoptera (Peters et al. 2017, Mrinalini and Werren 2017). Results from phylogenetic analyses imply that this transition occurred only once during the evolution of Hymenoptera: in the stem lineage of the parasitoid sawfly family Orussidae and the wasp-waisted Hymenoptera (Apocrita) (Peters et al. 2017). The transition was associated with the evolution of numerous adaptations in behavior, morphology, and physiology to a parasitoid lifestyle (Whitfield 1998). For example, parasitoids critically depend on their ability to locate hosts, to successfully lay eggs on or in their hosts, to inject venom to immobilize their host and/or to antagonize their hosts' immune response, and to metabolize a nitrogen-rich animal-based diet (as compared to a nitrogen-poor plant-based diet). Intriguingly, however, wasp-waisted Hymenoptera diversified far more (144,593 described species, > 90 % of the extant species of Hymenoptera) than parasitoid sawflies (82 described species), indicating that the diversification of the Apocrita was likely promoted by the evolutionary acquisition of traits that parasitoid sawflies lack. Yet, the transition from phytophagy to parasitoidism and the factors contributing to the massive speciation of Apocrita have remained largely unstudied. The tremendous diversity, as well as the ecological and economical importance of Hymenoptera, have led the order to be the

focus of a wealth of taxonomic, evolutionary, and ecological research (Quicke 1997, Grimaldi and Engel 2005, Sharkey 2007, Peters et al. 2017). However, most of the comparative genomic research on Hymenoptera has been focused on Apocrita and especially on the multiple origins of eusociality within this clade. As a result, all but one of the published draft genomes of Hymenoptera refer to species of Apocrita (Branstetter et al. 2018). The only published draft genome of a sawfly is that of the wheat stem sawfly, *Cephus cinctus* (Cephoidea) (Robertson et al. 2018). The larvae of Cephoidea are endophytophagous, feeding on a wide range of large-stemmed grasses, including economically important crops, and show an opportunistic cannibalistic behavior (Beres et al. 2011). As the sister group to Orussidae + Apocrita (Peters et al. 2017), the superfamily Cephoidea represents an important lineage in the hymenopteran tree of life for understanding the possible onsets of parasitoidism. At the same time, the derived ecology of Cephoidea, whose larvae are neither strictly phytophagous nor parasitoid, and its specific systematic position prevent the drawing of major conclusions on the composition of the ancestral genome of (phytophagous) Hymenoptera or on factors contributing to the disparate diversification of the parasitoid Orussidae and Apocrita.

Knowledge of the composition of the ancestral genome of Hymenoptera is fundamental for tracing the evolution of traits within Hymenoptera. In addition, due to the phylogenetic position of Hymenoptera as the sister group of all remaining holometabolous insects (Savard et al. 2006, Peters et al. 2014), the composition of the ancestral genome of Hymenoptera has major implications for understanding the evolution of holometabolous insects and their genomes. Previous studies on Apocrita

have shown that the repertoire of immune response genes (Evans et al. 2006, Gadau et al. 2012, Barribeau et al. 2015), of vision genes (opsins) (Henze and Oakley 2015), and the GC content (Standage et al. 2016) of Hymenoptera genomes are reduced compared to genomes of other insects. A reduction has also been found in the diversity and abundance of transposable elements, which are key drivers of genome size evolution in insects (Petersen et al. 2019), in social Apocrita (Kapheim et al. 2015). It remains to be investigated, however, whether these traits are characteristic of all Hymenoptera or whether they are specific to Apocrita. Also of interest are the origin and diversification of major royal jelly proteins (MRJPs), which were first discovered in the eponymous royal jelly (Hanes and Šimuth 1992), a honeybee gland secretion fed by young worker bees to developing larvae and triggering queen development (Snodgrass 1925). These proteins are encoded by a varying number of genes (*mrjp* and *mrjp*-like) that are exclusive to Hymenoptera and have been found in all but one of their genomes sequenced thus far (Werren et al. 2010, Bonasio et al. 2010, Nygaard et al. 2011, Kapheim et al. 2015, Sadd et al. 2015, Chris R Smith et al. 2011, Christopher D Smith et al. 2011, Buttstedt et al. 2014, Kupke et al. 2012). The *mrjp-l* genes likely originated from yellow genes (Hanes and Šimuth 1992), which are found across insects, but it is unknown when they originated and started to diversify in Hymenoptera. The current taxonomically biased distribution of genome sequencing data prevents the reliable inference of the ancestral features of Hymenoptera genomes and genomic traits that likely fostered the evolution of parasitoidism.

Here we present comparative analyses of draft genomes of the ectophytophagous sawfly *Athalia rosae* and the parasitoid sawfly *Orussus abietinus*. *Athalia rosae*

(Tenthredinoidea) is a representative of Eusymphyta, which a recent phylogenetic analysis suggests to be the sister lineage of all remaining Hymenoptera (Peters et al. 2017). *Athalia rosae* has retained the ancestral ectophytophagous lifestyle of Hymenoptera and feeds on crucifers (Brassicaceae), of which it is also an important agricultural pest (Sáringer 1974, Abe 1988). The species is readily bred under lab conditions, is currently being established as a model species, and is studied for a wide range of research questions (*e.g.*, in developmental biology (Yamamoto et al. 2004, Sekine et al. 2015), on sex determination (Mine et al. 2017), and on chemical defense (Abdalsamee and Müller 2012)). *Orussus abietinus* is a representative of the relatively species-poor group of parasitoid sawflies (also referred to as parasitic wood wasps), consisting exclusively of the family Orussidae. Like other orussids, *O. abietinus* is an ectoparasitoid of xylophagous larvae (beetles and wood wasps) developing in dead wood, a lifestyle considered to likely mirror the ancestral state of parasitoids (Peters et al. 2017). Orussids detect their hosts via vibrational sounding: the female wasps generate vibrations via frequent tapping of the antennae against the wood. The reflecting vibrations (containing information on the presence of host larvae in the wood) are in turn picked up by the forelegs and transmitted through the haemolymph to specialized organs, where they are transduced into nerve impulses (Vilhelmsen et al. 2001). If a host larva is detected, the female orussid lays an egg on or close to the host larva, which the orussid larva feeds upon when hatched (Ahnlund and Ronquist 2002). The anatomy of the orussid larva is simplified compared to those of other sawflies and is more similar to those of Apocrita (Vilhelmsen 2003). For example, orussid larvae lack eyes and legs (as do the larvae of Apocrita and in contrast to the larvae of sawflies) and their antennae and mouthparts are strongly simplified

(Vilhelmsen 2003). These morphological characteristics are considered adaptations to a parasitoid lifestyle. Our analyses of the draft genomes of *A. rosae* and *O. abietinus*, including comparisons with those of other Hymenoptera, provide first insights into (1) the composition of the ancestral genome of Hymenoptera, (2) traits related to the transition from phytophagy to parasitoidism, and (3) features that enabled the massive speciation of Apocrita. We also revisit multiple long-standing ideas on hymenopteran genome evolution, the results of which highlight the importance of comprehensive taxonomic sampling in comparative genomics.

6.2 Results and discussions

We sequenced and assembled the genome of the turnip sawfly, *Athalia rosae*, (Tenthredinidae; a representative of the phytophagous "Symphyta"; Fig. 6.1A–C) and the parasitoid sawfly *Orussus abietinus* (Orussidae; a representative of the parasitoid "Symphyta"; Fig. 6.1A–C) at a base coverage depth of 525 x and 255 x, respectively, from Illumina paired-end and mate-pair libraries using DNA of haploid males (SI II.1). After assembling the reads with ALLPATHS-LG (Gnerre et al. 2011) and scaffolding the resulting contigs using Atlas-Link and Atlas-Fill, the draft genome assemblies of *A. rosae* and *O. abietinus* span 164 Mbp and 201 Mbp, respectively (Fig. 6.1D). The assembly sizes closely match *in silico* genome size estimates (170 Mbp and 247 Mbp) inferred from the 17-mer distribution in the Illumina paired-end reads. The two genome assemblies are of high contiguity (522 and 936 scaffolds with N50 of 1.37 Mbp and 2.37 Mbp, respectively) compared to other Hymenoptera draft genome assemblies (File S1). Assessments of gene space coverage using the Arthropoda gene set of Benchmarking Universal Single-Copy Orthologs (BUSCO;

Simão et al. 2015) further revealed that the genome assemblies encompass the majority (96 % and 93 %) of the expected protein-coding genes (SI II.2.1).

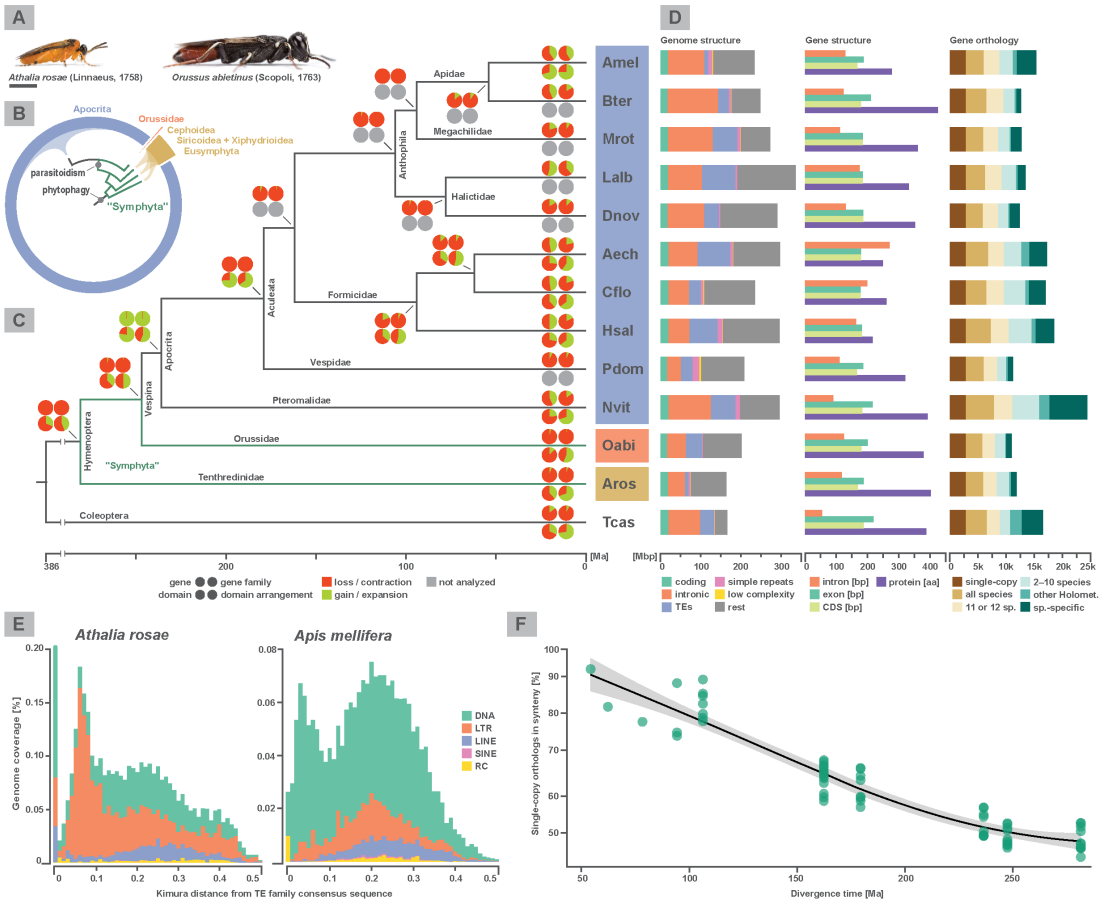

Figure 6.1. Hymenoptera genome evolution. Adult males of *Athalia rosae*[1] and *Orussus abietinus*. Scale bar: 2.5 mm. **(B)** Number of described species (Apocrita: 144,593; Orussidae: 82; "Symphyta" excl. Orussidae: 7,983) of, relationships of, and ecological transitions in Hymenoptera (Peters et al. 2017, Aguiar et al. 2013). **(C)** Ratio of gain and loss of genes, domains, and domain arrangements, as well as ratio of gene families that experienced expansions or contractions. Gene and gene family evolution were analyzed by applying the maximum likelihood optimality criterion, a single coupled birth and death rate, and using the divergence time estimates and phylogenetic relationships inferred by Peters et al. (2017). Domain and domain arrangement evolution were analyzed by applying the maximum parsimony optimality criterion. **(D)** Absolute number of nucleotides occupied by genomic components (left column), median length of various gene structure parameters (center column), and gene orthology in the genome of each species (right column; unit = number of genes). **(E)** Divergence distribution of transposable element (TE) copies in the genome of *Athalia rosae* and that of *Apis*

mellifera, estimated from the Kimura distance of the nucleotide sequence of each TE copy to the TE family nucleotide consensus sequence. **(F)** Loss of synteny over time in the genomes of twelve Hymenoptera, inferred from the proportion of 3,983 shared single-copy orthologs (SCOs) retaining the same neighboring SCO, relative to the divergence time, in all possible pairwise comparisons. The curve represents the smoothed conditional mean. **Abbreviations:** aa = amino acids; bp = base pairs; CDS = coding sequence; LINE = long interspersed nuclear element; LTR = long terminal repeats; Ma = million years ago; RC = rolling circle transposons; SINE = short interspersed nuclear element; TE = transposable elements; Aech = *Acromyrmex echinatior*; Amel = *Apis mellifera*; Aros = *A. rosae*; Bter = *Bombus terrestris*; Cflo = *Camponotus floridanus*; Dnov = *Dufourea novaeangliae*; Hsal = *Harpegnathos saltator*; Lalb = *Lasioglossum albipes*; Mrot = *Megachile rotundata*; Nvit = *Nasonia vitripennis*; Oabi = *Orussus abietinus*; Pdom = *Polistes dominula*; Tcas = *Tribolium castaneum.* All photographs by Oliver Niehuis, with assistance from Thomas Pauli and Ralph S. Peters. [1] Note that while the photograph shows a male of the nominate form, we sequenced and report the genome of the Eastern Palearctic subspecies *A. rosae ruficornis.*

The assemblies are close in size to that of the wheat stem sawfly, *Cephus cinctus* (Cephidae; 205 Mbp; Hanrahan and Johnston 2011, Robertson et al. 2018), and fall within the lower range of the known genome sizes of Hymenoptera (98 Mbp to 1.3 Gbp; Ardila-Garcia et al. 2010, Tavares et al. 2010, Gregory 2018). In fact, the genome of *A. rosae* is the smallest of all Hymenoptera sequenced so far. The two sawfly genomes have a higher GC content than most apocritan genomes (sawflies: 41 % and 45 %; Apocrita: median 37 %; SI II.4.2). This is consistent with the hypothesis that the low GC content of Apocrita genomes represents a derived state, possibly caused by high recombination rates associated with GC-biased gene conversion (Wilfert et al. 2007, Niehuis et al. 2010, Kent et al. 2012). However, the cause and effect relationship of recombination rate and GC content remains to be disentangled.

6.2.1 Copy number and amino acid sequence of conserved genes and gene families substantiate the high quality of the sawfly draft genomes

The evolution of the hymenopteran gene repertoire was studied in detail by manually annotating more than 1,000 protein-coding genes and non-coding (nc) RNAs in each of the two sawfly genomes. We found a wide range of genes and gene families to be conserved in amino acid sequence and copy number across Hymenoptera, consistent with *a priori* expectations, and confirming the high coverage of the sawfly genomes by the draft assemblies. Manually annotated and studied genes and gene families include ncRNAs, potentially laterally transferred genes, major royal jelly proteins, storage proteins, developmental genes, insulator proteins, DNA methyltransferases, silk proteins, elongases, desaturases, opsins, metallopeptidases, heat shock proteins, aquaporins, cuticular proteins, cysteine peptidases, candidate venom proteins, neuropeptides, protein hormones, biogenic amines and their G-protein-coupled receptors, as well as genes related to chemoreception, immune response, autophagy, dosage compensation, RNA interference, antioxidants, sex determination, and oxidative phosphorylation. A full description and discussion of each of these genes or gene families is given in the supplemental information (SI II.4.4; SI II.5.1–25).

6.2.2 Transposable element content and activity

Diversity and abundance of transposable elements (TEs), and consequently genome size, has been found to negatively correlate with the degree of social complexity in Apocrita (Kapheim et al. 2015). This is possibly a consequence of high recombination rates and decreased exposure to parasites and pathogens in eusocial species (Kapheim et al. 2015). We found the relative TE content in genomes of Hymenoptera, identified

with RepeatModeler (Smit and Hubley 2015) and RepeatMasker (Smit et al. 2015), to strongly correlate with genome size (Pearson's product-moment correlation $r = 0.8$, $p = 0.003$; SI II.3.5) and to range from 4.7 % (11.0 Mbp) in the honeybee (*Apis mellifera*) to 27.4 % (81.5 Mbp) in the leaf-cutting ant (*Acromyrmex echinatior*) (Fig. 6.1D; File S4). TE sequence divergence analysis, based on intra-family Kimura 2-parameter distances, indicates recent peaks in TE activity, largely caused by DNA elements, in most Hymenoptera genomes (Fig. S7). Interestingly, the *A. rosae* genome shows a TE content (5.1 %) and TE activity spectrum that is, with the exception of a very recent burst of TEs, similar to that of the honeybee (Fig. 6.1E). These results are intriguing, since they demonstrate that low TE content and overall low TE activity in Hymenoptera are not restricted to genomes of eusocial species and that consequently other ultimate factors seem to govern TE content evolution.

6.2.3 Apocrita possess more genes with reduced gene structure complexity than sawflies

The automated MAKER protein-coding gene annotation pipeline (Cantarel et al. 2008) predicted 11,894 and 10,959 genes in the draft genomes of *A. rosae* and *O. abietinus*, respectively. The numbers of genes predicted in the two sawfly draft genomes are lower than the official gene counts of most other published Hymenoptera draft genomes (Fig. 6.1D; Branstetter et al. 2018), but closely match the reported numbers of protein-coding genes in the draft genomes of *C. cinctus* (11,206; Robertson et al. 2018) and the European paper wasp, *Polistes dominula* (Fig. 6.1D; Standage et al. 2016). However, comparing features of the predicted protein-coding genes across species using COGNATE (Wilbrandt et al. 2017) revealed that the total

amount of protein-coding DNA in the two sawfly genomes (19.9 Mbp in *A. rosae* and 17.7 Mbp in *O. abietinus*) fits well into the known range of the metric in Hymenoptera (16–20 Mbp; Fig. 6.1D) and that the total amount of protein-coding DNA varies less than the number of genes across the published draft genomes of Hymenoptera. Proteins of the two sawflies are among the longest in Hymenoptera (median: 406 amino acids in *A. rosae* and 384 amino acids in *O. abietinus*; Fig. 6.1D). The protein length increase results from a larger median number of exons (5.0; note that the sizes of exons in the sawfly draft genomes do not differ markedly from the average across Hymenoptera; SI II.4.2), compared to Apocrita (4.0).

6.2.4 Gene order is constrained in Hymenoptera

Gene order is subject to change over the course of evolution due to recombination and rearrangement. Because genome-wide recombination rates vary substantially between Hymenoptera, with eusocial species likely exhibiting the highest rates (Wilfert et al. 2007), the rate of micro-synteny (gene order conservation) decay is also expected to differ between lineages. Yet, previous studies have found extensive conservation of gene order across insects (*e.g.*, Engström et al. 2007). Using protein divergence as a proxy for time, a linear decay of micro-synteny over time has been found in insect genomes (Zdobnov and Bork 2007). Capitalizing on recently published Hymenoptera divergence time estimates (Peters et al. 2017) and exploring a more extensive taxon sampling within Hymenoptera, including the two sawflies presented here, we investigated micro-synteny decay of conserved single copy orthologs (SCOs) in this insect order. Comparing the fraction of SCOs that retain the same neighboring SCO in pairwise comparisons between species in relation to the divergence times of each

species pair using a custom Perl script (included as File S39) revealed a close to linear loss of synteny over time (Fig. 6.1F). The highest degree of synteny conservation was detected between the most recently diverged lineages (*e.g.*, > 90 % between honeybee and leafcutter bee; File S38), irrespective of whether these lineages are eusocial or not. In fact, we did not observe an increase of genome shuffling in eusocial Apocrita. However, contrary to what was previously reported by Zdobnov and colleagues (Zdobnov and Bork 2007), we found a decrease in the rate of synteny loss across divergence times that span more than 240 million years (Fig. 6.1F). This retention of micro-synteny over large evolutionary distances points to the presence of functional constraints on the preservation of local genomic structures or low rates of non-homologous recombination and rearrangement. Functional annotation of genes remaining in micro-synteny, using Gene Ontology terms, revealed significant enrichment ($p < 0.05$; weighted Fisher's test and hypergeometric test) of a number of terms related to cell cycle and signaling, cellular and organelle organization, as well as development (File S2). Notably, we found consistent enrichment of Wnt and Notch signaling, both of which are vital and complex pathways in embryonic development and tissue differentiation. A specific example of a conserved gene order was also revealed by manual annotation of opsin genes (SI II.5.24): we uncovered a close linkage of the long wave sensitive (LWS) 1 and LWS 2 opsins, which was previously considered unique to the honeybee (Bao and Friedrich 2009), in the genomes of the two sawflies and of ten additional hymenopterans (interlocus distance -6–7,583 bp; File S35). The conserved LWS1/2 linkage thus represents an ancestral feature of all Hymenoptera and suggests the presence of a cis-regulatory constraint, preventing the loss of synteny between these genes.

6.2.5 Hymenoptera gene and protein domain repertoires display a reductive mode of evolution

A previous study reported that more genes were gained than lost in the evolution of protein-coding gene families in Hymenoptera (Rappoport and Linial 2015). Here, we analyzed the evolution of gene families inferred from OrthoDB (Zdobnov et al. 2017) using the CAFE software (Han et al. 2013) and exploiting recently published divergence time estimates of Hymenoptera (Peters et al. 2017). We additionally identified protein domains as well as protein domain arrangements and inferred their respective losses and gains across the Hymenoptera tree applying the Fitch parsimony optimality criterion. In contrast to the study of Rappoport and Linial (2015), we found a pronounced pattern of reduction of genes, gene families, and protein domains during the evolution of this insect order, with more losses than gains at most nodes (Fig. 6.1C; File S41; SI II.4.3). The pattern is contrasted by a large number of new protein domain arrangements uncovered at each node (Figure S11), with more new arrangements than lost arrangements (Fig. 6.1C). This result is consistent with the idea that domains can be reused and shuffled at a higher rate than new domains can emerge (Lees et al. 2016; Moore and Bornberg-Bauer 2012). Ultimately, reuse of functional units might compensate for the predominant trend of gene and domain loss as well as for gene family contractions (Lees et al. 2016).

6.2.6 Major royal jelly proteins were already synthesized by the last common ancestor of Hymenoptera

Major royal jelly proteins (MRJPs) are an important component of the honeybee's royal jelly, a gland secretion fed to developing larvae that determines the differential

development of queens and workers (Snodgrass 1925). MRJPs and MRJP-like encoding genes have only been known to occur in Apocrita, presumably having evolved from a tandem-duplication of the Yellow-family gene *y-e3* and subsequently expanded in multiple lineages (Drapeau et al. 2006, Buttstedt et al. 2014).

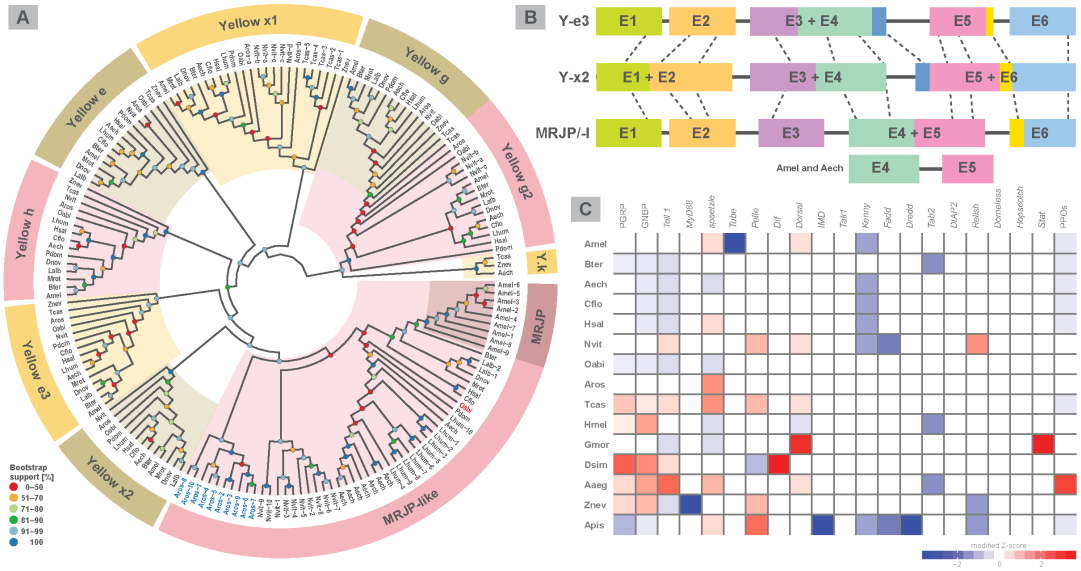

Figure 6.2. Evolution of Hymenoptera Yellow, MRJP/-like, and immune response-related genes. **(A)** Relationships of Hymenoptera Yellow, major royal jelly protein (MRJP), and MRJP-like (MRJPl) amino acid sequences, inferred under the maximum likelihood optimality criterion, modelling invariable sites, and approximating site-rate variation with a discrete gamma distribution. Branch support is estimated from 1,000 non-parametric bootstrap replicates. MRJP and MRJPl proteins of *Athalia rosae* and *Orussus abietinus* are highlighted in blue and red, respectively. **(B)** Gene structure comparison of *mrjp* and *mrjpl* genes and of two candidate sister group *yellow* genes, *y-e3* and *y-x2*. Dashed lines indicate shared amino acid motifs conserved among species within each gene and between genes (SI II.5.5). Gene and motif lengths not to scale. **(C)** Heat map visualizing copy number variation in immune response-related genes between species. Modified Z-scores indicate the deviation from the median of each gene by standard deviation units. **Abbreviations:** Aaeg = *Aedes aegypti*; Aech = *Acromyrmex echinatior*; Amel = *Apis mellifera*; Apis = *Acyrthosiphon pisum*; Aros = *Athalia rosae*; Bter = *Bombus terrestris*; Cflo = *Camponotus floridanus*; Dnov = *Dufourea novaeangliae*; Dsim = *Drosophila simulans*; Gmor = *Glossina morsitans*; Hmel = *Heliconius melpomene*; Hsal = *Harpegnathos saltator*; Lalb = *Lasioglossum albipes*; Lhum = *Linepithema humile*; Mrot = *Megachile rotundata*; Oabi = *Orussus abietinus*; Pdom = *Polistes dominula*; *Nvit* = *Nasonia vitripennis*; Tcas = *Tribolium castaneum*; Znev = *Zootermopsis nevadensis.*

Revising this scenario, manual annotation of MRJPl-encoding genes uncovered a single gene in the genome of *O. abietinus* and ten genes in the genome of *A. rosae* (Fig. 6.2A; SI II.5.5; Fig. S18). The presence of a single *mrjpl* in the genome of *O. abietinus* is consistent with the hypothesis of a single ancestral *mrjpl* in Apocrita (Drapeau et al. 2006), but with its origin already in a stem lineage of all Hymenoptera. The evolutionary origin of *mrjpls* (> 281 Ma) is thus much older than previously thought. Phylogenetic analysis recovered *mrjpls* as sister group of the Yellow-gene *y-x2* and not of the Yellow-gene *y-e3* (Fig. 6.2A; Fig. S19), despite a higher similarity of *mrjpls* in intron-exon structure with the latter (Fig. 6.2B; SI II.5.5). This phylogenetic relationship also received statistically significant support (p = 0.0048; approximately unbiased topology test). The close relationship of *mrjpls* with *y-x2* is especially surprising, given that *y-x2* is spatially distantly located from the *yellow* gene cluster containing the *mrjpls* (*e.g.*, in the genomes of *A. mellifera* and *Nasonia vitripennis*, they occur on different chromosomes; Drapeau et al. 2006, Buttstedt et al. 2014; in the genomes of *A. rosae* and *O. abietinus*, they occur on different scaffolds).

6.2.7 Hymenoptera are characterized by a small repertoire of conserved immune genes

The canonical immune response-related gene repertoire involved in recognition and signaling pathways (immune genes) of eusocial Hymenoptera was initially described as extremely reduced compared to the mostly conserved repertoire of solitary insects (Evans et al. 2006, Gadau et al. 2012). However, a more recent study suggested that a reduced immune gene repertoire might be a shared trait of Apocrita and is not strictly correlated with a eusocial lifestyle (Barribeau et al. 2015). Using profile hidden

Markov models built from reference amino acid sequences of immune genes to scan the predicted proteins of Hymenoptera and a selected set of other insects, we found the numbers of immune genes to be largely similar amongst all investigated species of Hymenoptera (28–36 genes; Fig. 6.2C; Table S25), although some lineages are characterized by the lack of specific genes (*e.g.*, the IMD pathway gene *Kenny* is absent in several Aculeata). Although the genome of *A. rosae* has the largest number of identified response-related genes among Hymenoptera, our data do not show a clear trend between immune gene repertoire reduction and eusocial lifestyle. On the contrary, we found 32 immune genes in the genome of the eusocial honeybee, but only 29 in that of the solitary *O. abietinus* (Table S25). We also found that Hymenoptera are characterized by an overall small number of immune genes (median: 30) relative to other insects (median: 38; Table S25). The reduced number of immune genes in Hymenoptera is thus likely not related to the evolution of eusociality, nor is it a characteristic of Aculeata, but rather represents the ancestral condition in Hymenoptera. However, the reduced repertoire of recognition and signaling pathway genes, which are mostly conserved across solitary insects, in Hymenoptera does not necessarily imply a reduced immune response. A study investigating *de novo* infection response genes in *N. vitripennis* identified a large repertoire of new genes involved in the immune response, many of which were taxonomically restricted and rapidly evolving (Sackton et al. 2013). It remains to be tested if and how these novel immune response-related genes compensate for the reduction of the immune gene repertoire and also whether such a compensation has evolved in other Hymenoptera.

6.2.8 Loss of a vision gene coincides with transition to a parasitoid lifestyle

Light sensing is primarily mediated by the opsin gene family of G protein-coupled transmembrane receptors. Apocrita are known to have four rhabdomeric-type opsins (r-opsins) of three wavelength-specific subfamilies: one member of the short wavelength UV-sensitive (SWS-UV) r-opsin subfamily, one member of the blue-sensitive (SWS-B) r-opsin subfamily, and two members of the long wavelength-sensitive (LWS) r-opsin subfamily, introduced above as LSW1 and LSW2 opsins (Henze and Oakley 2015, Velarde et al., 2005, Wakakuwa et al. 2005). These r-opsins are differentially expressed in the photoreceptors of the compound eye retina and the ocelli (Velarde et al. 2005). The honeybee has also been shown to possess a fifth opsin, a member of the ciliary opsin gene family (c-opsin), which is expressed in two small cell clusters of the brain, likely mediating extraretinal light sensing (Velarde et al. 2005). Using known opsin amino acid sequences as references, we identified and manually annotated all four retinal opsins that had previously been found in Hymenoptera in the genomes of the two sawflies (Fig. 6.3A and Table S27). This revealed that the molecular underpinnings underlying trichromatic compound eye vision, which has been documented by comparative physiological studies in the Hymenoptera (Peitsch et al. 1992), is highly conserved in the order. Furthermore, we found that the c-opsin is also present in the *A. rosae* genome (Fig. 6.3A) and that the *A. rosae* genome is unique among Hymenoptera in containing a sixth opsin, Rh7 (Fig. 6.3A). The Rh7 opsin is deeply conserved in arthropods (Senthilan and Helfrich-Förster 2016), but is not found in other Hymenoptera, suggesting that this opsin subfamily was lost in the stem lineage of Orussidea and Apocrita. In *Drosophila,* Rh7 opsin has been found to be expressed in the brain and is involved in the entrainment

of the circadian activity rhythm by light (Ni et al. 2017). However, Rh7 opsin is also expressed in the photoreceptor cells of a mosquito species (Hu et al. 2014). Thus, besides identifying *A. rosae* as the opsin homolog-richest hymenopteran species at this point, these findings revealed that the transition from phytophagy to a parasitoid lifestyle in Hymenoptera was accompanied by a reduction of the opsin gene repertoire. This could be related to the extreme regression of the larval visual system as ancestral parasitoid larvae are thought to have developed in wood and were thus not exposed to sunlight (Vilhelmsen and Turrisi 2011).

6.2.9 Dietary transition and specialization have not resulted in change of metabolic capabilities

Phytophagous sawfly larvae, being mobile in the environment, can utilize multiple host plants or prey. In contrast, parasitoid larvae are restricted to a single host and the finite resources contained within this host (Jervis et al. 2008). To alleviate the severely limited resources available to each parasitoid larva (Slansky 1986, Jervis et al. 2008), some highly specialized parasitoids manipulate their host to increase nutrient availability (Vinson and Iwantsch 1980). As a consequence, many of these parasitoids have in turn lost the ability to synthesize these nutrients (*e.g.*, lipids), possibly through the loss of synthesis pathway genes (Visser et al. 2010). However, the genomic changes of the metabolic gene repertoire associated with the transition from phytophagy to parasitoidism and from generalist to specialist parasitoid have not been comprehensively characterized in Hymenoptera. We functionally annotated the predicted proteins of the phytophagous sawfly (*A. rosae*), the generalist parasitoid sawfly (*O. abietinus*; host spectrum reviewed by Ahnlund and Ronquist 2002), and

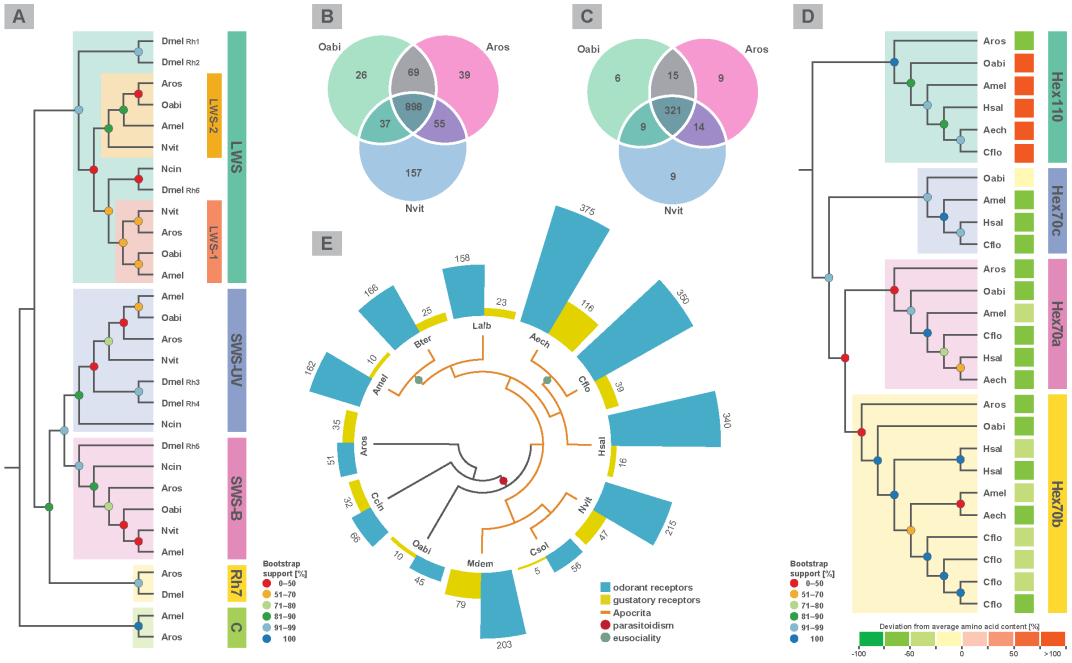

Figure 6.3. Hymenoptera vision gene, metabolic, hexamerin, and chemoreceptor repertoires.
(A) Phylogenetic relationships of Hymenoptera, *Nephotettix cincticeps* (Hemiptera), and *Drosophila* opsin genes inferred under the maximum likelihood optimality criterion. Branch support is estimated from 500 non-parametric bootstrap replicates. **(B)** Number of unique and shared enzymes (Enzyme Commission numbers) in the proteomes of *Athalia rosae*, *Orussus abietinus*, and *Nasonia vitripennis*. **(C)** Number of unique and shared metabolic pathways identified in the proteomes of *A. rosae*, *O. abietinus*, and *N. vitripennis*, inferred from enzyme and gene ontology annotations. **(D)** Phylogenetic relationships of Hymenoptera hexamerins inferred under the maximum likelihood optimality criterion. Branch support is estimated from 1,000 non-parametric bootstrap replicates. Colors indicate deviation of the amino acid glutamine (Q) from the average amino acid content in percent (%). **(E)** Copy number variation of odorant and gustatory receptor gene repertoires among Hymenoptera. Data referring to *A. rosae* and *O. abietinus* are taken from the present study, those of all remaining species from literature (Robertson et al. 2018, Zhou et al. 2015, Zhou et al. 2012, Robertson and Wanner 2006, Robertson et al. 2010, Sadd et al. 2015). Only full-length proteins comprising at least 350 amino acids were considered. Phylogenetic relationships taken from the study by Peters et al. (2017). **Abbreviations**: Aech = *Acromyrmex echinatior*; Amel = *Apis mellifera*; Aros = *Athalia rosae*; Bter = *Bombus terrestris*; Ccin = *Cephus cinctus*; Cflo = *Camponotus floridanus*; Csol = *Ceratosolen solmsi*; Dmel = *Drosophila melanogaster*; Hsal = *Harpegnathos saltator*; Lalb = *Lasioglossum albipes*; Mdem = *Microplitis demolitor*; Ncin = *Nephotettix cincticeps*; Nvit = *Nasonia vitripennis;* Oabi = *Orussus abietinus.*

the highly specialized parasitoid wasp (*N. vitripennis*; host spectrum discussed by Peters 2010 and Desjardins et al. 2010) using the CycADS pipeline (Vellozo et al. 2011). We then inferred and compared metabolic pathways in the three species through a combination of a Pathway Tools (Karp et al. 2019) analysis and manual curation of the results. We found fewer genes with predicted metabolic functions in *A. rosae* (4,090; SI II.5.10) and *O. abietinus (*3,827) than in *N. vitripennis* (4,928). Despite these differences, we found a high level of congruence in the enzyme repertoire and in the metabolic pathways between all three species (Fig. 6.3B and C). Surprisingly, the comparison of the predicted functions of the inferred enzymes and pathways did not reveal differences that can readily be attributed to dietary transitions or host specialization. The lack of any detectable reduction in the metabolic gene repertoire of the two parasitoids can possibly be explained by the propensity of adult parasitoid Hymenoptera to consume pollen, nectar, and plant tissue (Jervis et al. 1993), for which the ancestral metabolic gene repertoire is still required. The dietary transitions and specializations during the evolutionary history of Hymenoptera might consequently not have resulted in the complete loss of metabolism-related gene families, but might have instead caused a reduction in the copy-number of genes, as was shown in mammals (Kim et al. 2016), or changes in gene expression and enzyme efficiency. Consistent with this idea, the manual annotation of genes that are likely related to the ability of *A. rosae* to deal with the chemical defenses of its host plant also reflects this pattern, revealed a reduced copy number of two candidate gene families in carnivorous and secondarily phytophagous species relative to ancestrally phytophagous species (SI II.5.9). A partial repertoire reduction could thus explain how

the ability to synthesize lipids has re-evolved multiple times in parasitoid wasps (Visser et al. 2010).

6.2.10 Storage protein evolution possibly facilitated transition to parasitoidism

The efficient utilization of limited host resources by larvae and the allocation of these resources to the adult stage are essential to the reproductive success of parasitoids (Jervis et al. 2008). Most Apocrita possess four hexamerin storage proteins (Hex70a–c and Hex110; (Cristino et al. 2010, Martins et al. 2010), which provide energy and amino acids during non-feeding periods (Hagner-Holler et al. 2007). Utilizing reference amino acid sequences of known hexamerins, we identified and manually annotated all four previously known hexamerins of Hymenoptera in the genome of *O. abietinus* and all but Hex70c in the genome of *A. rosae* (Fig. 6.3D). Comparing the amino acid content of Hymenoptera hexamerins, we found a unique and substantial increase of glutamine content (> 100 % increase relative to the average) — which is important in the management of nitrogen in insects (Weihrauch et al. 2012) — in the Hex110 protein of *O. abietinus* and of all Apocrita (Fig. 6.3D). This change might have evolved in response to the increased nitrogen content in animal tissues relative to plant matter (Mattson 1980). Thus, the emergence of an additional hexamerin storage protein (Hex70c) and the increased level of glutamine in Hex110 in the stem lineage of *O. abietinus* and Apocrita possibly facilitated the transition from a herbivorous to a parasitoid lifestyle.

6.2.11 Odorant and gustatory receptors were likely key factors for the diversification of Apocrita

Chemosensation receptors are paramount for vital insect behaviors, such as host detection in parasitoid wasps (*e.g.*, Steidle and Schöller 1997). Hymenoptera detect most chemical compounds with transmembrane proteins of the odorant receptor (OR) and of the gustatory receptor (GR) multigene families. These families are very diverse in Apocrita and especially so in lineages with eusocial species (Fig. 6.3E), where they possibly facilitated the evolution of eusociality by enabling kin selection (Zhou et al. 2012). We identified and manually annotated odorant and gustatory receptors in the two sawfly genomes utilizing the antennal transcriptomes of each species and a set of reference amino acid sequences of the corresponding proteins in other Hymenoptera. In agreement with a recent study on *C. cinctus* (Robertson et al. 2018), we found considerably fewer GR- and OR-coding genes in the genomes of the two sawflies *A. rosae* and *O. abietinus* than in those of Apocrita (Fig. 6.3E; SI II.5.3). In addition, our data indicate that multiple OR and GR gene lineages present in the genomes of the herbivorous sawflies *A. rosae* and *C. cinctus* were lost during the evolution of parasitoidism in the last common ancestor of *O. abietinus* and Apocrita (Fig. 6.3E; SI II.5.3). The large OR and GR gene repertoires of Apocrita are the result of subsequent and multiple independent expansions of those OR and GR gene lineages that were retained during the evolution of parasitoidism (SI II.5.3). Most intriguingly, the 9-exon OR subfamily, which has been implicated in the detection of cuticular hydrocarbons and is particularly expanded in eusocial species (up to 139 in the red harvester ant, *Pogonomyrmex barbatus*; Zhou et al. 2012, Zhou et al. 2015, Chris R Smith et al. 2011, Pask et al. 2017) is represented by only one copy in each of the

sawfly genomes (Fig. S16). The expansion of the OR and GR gene repertoires in Apocrita likely improved the chemoreception abilities of apocritans and could thus have been a key factor in the evolutionary success of this group. Specifically, the improved chemoreception abilities may have facilitated the formation of new ecological niches by enabling efficient detection and differentiation of novel hosts in diverse habitats. Encountering new hosts is key for specialization (Schmid-Hempel 2011), which in turn enables parasitoids to evolve faster and adapt more readily to the host defense mechanisms (Kawecki 1998). Consistent with this idea, the species-poor parasitoid orussids identify potential hosts in wood via vibrational sounding (Vilhelmsen et al. 2001), which likely provides far fewer possibilities for host specialization than chemoreception. Finally, we found two of the GR genes in the genomes of *A. rosae* and *C. cinctus* to be orthologous to CO_2 receptor genes of *Drosophila* (Robertson et al. 2018, Jones et al. 2007, Kwon et al. 2007). The presence of candidate CO_2 receptor genes in the genomes of phytophagous sawflies, in contrast to their absence in the genomes of the parasitoid sawfly and Apocrita, could thus indicate the functional involvement of the encoded receptors in host plant detection.

6.3 Conclusions

The results from our comparative analyses of the *Athalia rosae* and *Orussus abietinus* genomes call several previously widely held assumptions regarding characteristics and the evolution Hymenoptera genomes in to question. It has been stated, for example, that Hymenoptera genomes are characterized by a low GC content (Branstetter et al. 2018, Standage et al. 2016). Considering the phylogenetic relationships of the investigated species, the high GC content of sawfly genomes does

not represent a simple exception from a rule, but suggests that a low GC content is a derived state of only a subordinate group of Hymenoptera, the Apocrita. Contrariwise, we uncover genomic attributes previously considered derived characteristics of highly specialized lineages (*e.g.*, bees) to actually represent Hymenoptera ground plan features (*e.g.*, presence of MRJPls and a reduced immune response gene repertoire). We also provide novel insights into genomic factors that may have facilitated the evolutionary success and the tremendous diversification of parasitoid and eusocial Apocrita (*e.g.*, changes in storage protein and chemosensory receptor repertoires). The results of our study highlight the importance of taxonomic sampling for inferring ground plan characteristics of an organismal group. They furthermore lay the foundation for a variety of future lines of research (*e.g.*, on the ancestral function of MRJPs and the possible fitness benefits of the CO_2 receptors) by providing a valuable resource for comparative studies in the mega-diverse insect order Hymenoptera, which encompasses economically (Quicke 1997, Grimaldi and Engel 2005) and medically relevant (Moreno and Giralt 2015) species as well as important model organisms (Werren et al. 2010, Weinstock et al. 2006, Branstetter et al. 2018).

6.4 Methods

6.4.1 Samples and extractions

All samples of *Athalia rosae ruficornis* Jakovlev, 1888 were derived from a strain maintained for more than 15 years, with occasional introductions of individuals from natural populations, in the laboratory of M. Hatakeyama (National Institute of Agrobiological Sciences NARO, Tsukuba, Japan). Total genomic DNA was extracted from adult haploid males originating from a single virgin female using the Gentra

Puregene Tissue kit (Qiagen, Hilden, Germany) and following the manufacturers' protocol. Total RNA was extracted from the whole body of (a) two adult males and (b) two adult females using the RNeasy mini kit (Qiagen, Hilden, Germany) as well as from the antennae of (c) 45 adult females and (d) 56 adult males using the RNeasy Micro kit (Qiagen, Hilden, Germany) and following the manufacturers' protocol. Antennae from a given sex were pooled for RNA extraction. Samples of *Orussus abietinus* (Scopoli, 1763) were derived from a natural population of the species in the vicinity of Darmstadt (Hesse, Germany). Total genomic DNA was extracted from the mesosoma and the metasoma of two adult males using the DNeasy Blood & Tissue kit (Qiagen GmbH, Hilden, Germany) and following the manufacturer's' protocol. Total RNA was extracted from (a) the mesosoma and the metasoma of an adult male using Tri-Reagent (Sigma-Aldrich, Steinheim, Germany), from (b) a whole adult female using the NucleoSpin RNA II kit (Macherey & Nagel, Düren, Germany), and from (c) the antennae of ten adult males using RNeasy Micro Kit (Qiagen, Hilden, Germany) following the manufacturers' protocols.

6.4.2 Genome and transcriptome sequencing

We applied a whole genome shotgun sequencing approach and prepared and sequenced four libraries of nominal insert sizes of 180 bp, 500 bp, 2 kbp (only *A. rosae*), 3 kbp, and 8–10 kbp. For sequencing the *A. rosae* genome, the 180-bp and 500-bp paired-end libraries and the 2-kbp mate-pair library were prepared from DNA isolated from a single male each, while the 3-kb and 8–10-kb mate-pair libraries were prepared using DNA from four and fourteen pooled males, respectively. For sequencing the *O. abietinus* genome, the 180-bp, 500-bp, and 3-kbp libraries were

prepared from DNA extracted from a single adult male wasp, while the 8–10 kbp mate-pair library was prepared using pooled DNA from two adult male wasps. To prepare the 180-bp and 500-bp libraries, we used a gel-cut paired-end library protocol. Briefly, 1 µg of the DNA was sheared using a Covaris S-2 system (Covaris, Inc. Woburn, MA, USA) using the 180-bp and 500-bp program, respectively. Sheared DNA fragments were purified with Agencourt AMPure XP beads, end-repaired, dA-tailed, and ligated to Illumina universal adapters. After adapter ligation, DNA fragments were further size selected by agarose gel separation and were subsequently PCR-amplified with 6–8 amplification cycles using the Illumina P1 and Index primer pair and the Phusion High-Fidelity PCR Master Mix (New England BioLabs, Ipswich, MA, USA). The final library was purified using Agencourt AMPure XP beads, and the library's quality was assessed with an Agilent Bioanalyzer 2100 (DNA 7500 kit) by determining the fragment size distribution. Long mate-pair libraries with 2-kb, 3-kbp and 8–10-kbp insert sizes were constructed according to the manufacturer’s protocol (Mate Pair Library v2 Sample Preparation Guide art # 15001464 Rev. A PILOT RELEASE). Briefly, 5 µg (when preparing the 2-kbp and the 3-kbp insert size libraries) or 10 µg (8–10 kb insert size library) of genomic DNA was sheared to the desired fragment size with the aid of a Hydroshear (Digilab, Marlborough, MA, USA). The obtained fragments were subsequently end-repaired and biotinylated. Fragment sizes between 1.8 kbp and 2.5 kbp (2-kbp library), between 3.0 and 3.7 kbp (3-kbp library), and between 8 and 10 kbp (8–10-kbp library) were extracted from a 1 % low melting agarose gel and then circularized by blunt-end ligation. The size-selected circular DNA fragments were then sheared to fragment sizes of 400 bp (Covaris S-2), the fragments were subsequently purified using Dynabeads M-280

Streptavidin Magnetic Beads, end-repaired, dA-tailed, and ligated to Illumina PE sequencing adapters. DNA fragments with adapter molecules on both ends were amplified for 12–15 cycles with Illumina P1 and Index primers. Amplified DNA fragments were purified with Agencourt AMPure XP beads. Quantification and size distribution of the final library were determined before sequencing as described above. All sequencing was performed on Illumina HiSeq2000 sequencers, which generated 100-bp paired-end reads. Using a genome size estimate of 170 Mbp as baseline (see section II.1.3.1), we sequenced the five *A. rosae* libraries (*i.e.*, 180-bp, 500-bp, 2-kb, 3-kbp, and 8–10-kbp) to base-coverage depths of 240 x, 62 x, 57 x, 109 x, and 57 x, respectively. Using a genome size estimate of 247 Mbp as baseline (see section II.1.3.1), we sequenced the four *O. abietinus* libraries to base-coverage depths of 77 x, 27 x, 77 x, and 44 x, respectively. The amount of DNA sequences generated from each of these libraries is given in Table S1. For RNAseq data generation, poly-A mRNA was extracted from 1 μg whole body RNA using Oligo(dT)25 Dynabeads (Life Technologies, Carlsbad, CA, USA), followed by fragmentation of the mRNA by heat at 94 °C for 3 min (for samples with a RIN value of 3 or 3.3) or 4 min (for samples with RIN value of 6.0 and above). First strand cDNA was synthesized using the Superscript III reverse transcriptase (Life Technologies, Carlsbad, CA, USA) and purified using Agencourt RNAClean XP beads (Beckman Coulter, Brea, CA, USA). During second strand cDNA synthesis, dNTP mix containing dUTP was used to introduce strand-specificity. For Illumina paired-end library construction, the resultant cDNA was processed through end-repair and A-tailing, was ligated with Illumina PE adapters, and was then digested with 10 units of Uracil-DNA Glycosylase (New England BioLabs, Ipswich, MA, USA).

Amplification of the libraries was accomplished via 13 PCR cycles using the Phusion High-Fidelity PCR Master Mix (New England BioLabs, Ipswich, MA, USA). We incorporated 6-bp molecular barcodes during this PCR amplification. The libraries were purified with Agencourt AMPure XP beads after each enzymatic reaction and were quality-assessed and quantified with the Agilent Bioanalyzer 2100 DNA Chip 7500 (Santa Clara, CA, USA). The libraries were pooled in equimolar amounts prior to their sequencing. All libraries were sequenced with 101-bp read lengths on an Illumina HiSeq2000 sequencing platform. We collected the following number of reads from the whole-body RNA extract of *A. rosae*: 24,374,007 (adult male sample 1), 23,012,651 (adult male sample 2), 17,739,404 (adult female sample 1), and 8,869,760 (adult female sample 2). We collected the following number of reads from the whole-body RNA extract of *O. abietinus*: 32,320,562 (adult male) and 30,138,682 (adult female). Library preparation of the antennal transcriptomes, including poly-A enrichment, was performed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) according to manufacturer's instructions. The libraries were sequenced on a HiSeq2500 (Illumina, San Diego, CA), which provided 100-bp long paired-end reads. In total, we sequenced 34,652,811 and 36,072,988 reads from antennal RNA extracts of *A. rosae* males and females, respectively. We collected a total of 20,906,900 reads from antennal RNA extracts of *O. abietinus* males. Antennal transcriptome reads were processed using CLC Genomics Workbench 7 (Qiagen, Hilden, Germany), removing adapters during read import. Cleaned reads were assembled using the *de novo* assembly function with its default settings, retaining only contigs of more than 200 bp in length.

6.4.3 Genome assembly

Genome sizes and individual library coverage was estimated with jellyfish (version 2.0) (Marçais and Kingsford 2011), using 17-mers of the 180-bp genome sequencing reads. Prior to assembly, we removed all adapters from the reads with the software SeqPrep (https://github.com/jstjohn/SeqPrep). The genomes were separately assembled using ALLPATHS-LG (version 35218) (Gnerre et al. 2011) and applying the program's default parameters and the haploidy option. Contigs were scaffolded and scaffold gaps were filled with the BCM tools Atlas-Link (version 1.0) and Atlas gap-fill (version 2.2). The gene space coverage of the assemblies was assessed using BUSCO (version 1.1b1, Arthropod gene set) (Simão et al. 2015) and CEGMA (version 2.4) (Parra et al. 2007).

6.4.4 Automated protein-coding gene annotation

Protein-coding genes were annotated using the Maker 2.0 annotation pipeline (Cantarel et al. 2008), tuned specifically for annotating the genomes of arthropods. Specifically, the genome assembly was first subjected to *de novo* repeat prediction and CEGMA gene space coverage analysis; the latter for generating gene models for initial training of the *ab initio* gene predictors. Three rounds of training of the gene prediction programs Augustus (version 2.5.5) (Stanke et al. 2008) and SNAP (version 1.0b6) (Korf 2004) within Maker were used to infer a high-quality training set with a bootstrap method. Input protein data included 1 million peptides from a non-redundant reduction (if proteins shared > 90 % amino acid sequence identity, only the first in the protein list was retained) of all Uniprot Ecdysozoa entries (1.25 million

peptides; accessed July 2013), supplemented with proteomes from eighteen additional species (*i.e.*, *Strigamia maritime* (Chipman et al. 2014); *Tetranychus urticae* (Grbić et al. 2011); *Caenorhabditis elegans* (The C. elegans Sequencing Consortium 1998); *Loa loa* (Desjardins et al. 2013); *Trichoplax adhaerens* (Srivastava et al. 2008); *Amphimedon queenslandica* (Srivastava et al. 2010); *Strongylocentrotus purpuratus* (Sodergren et al. 2006); *Nematostella vectensis* (Putnam et al. 2007); *Branchiostoma floridae* (Putnam et al. 2008); *Ciona intestinalis* (Dehal et al. 2002); *Ciona savignyi* (Small et al. 2007); *Homo sapiens* (International Human Genome Sequencing Consortium et al. 2001); *Mus musculus* (Mouse Genome Sequencing Consortium et al. 2002); *Capitella teleta* (Simakov et al. 2013); *Helobdella robusta* (Simakov et al. 2013); *Crassostrea gigas* (Zhang et al. 2012); *Lottia gigantean* (Simakov et al. 2013); *Schistosoma mansoni* (Berriman et al. 2009)) leading to a final non-redundant peptide evidence set of 1.03 million peptides. We additionally provided MAKER RNAseq transcription data derived from two males and two females (*A. rosae*) and a single male and single female (*O. abietinus*) to identify exon-intron boundaries. We also ran a heuristic script (included as File S42) to identify and split erroneously joined gene models.

6.4.5 Manual annotations

Gene models were manually annotated with the aid of Web Apollo (Lee et al. 2013) and the i5k interface (Poelchau et al. 2015). The manual annotation process was guided by multiple intrinsic and extrinsic evidence tracks: (a) cleaned RNAseq raw read data mapped onto the genome assembly using TopHat2 (version 2.0.12) (Kim et al. 2013) with its default settings; (b) transcripts of the respective species assembled

with Cufflinks (version 2.2.1) (Trapnell et al. 2010); (c) transcripts of the respective species assembled with Trinity (version trinityrnaseq_r20140413p1) (Grabherr et al. 2011) and mapped onto the genome assembly using the Exonerate (version 2.20) (Slater and Birney 2005) fork Exonerate-gff3 (https://github.com/hotdogee/exonerate-gff3) with the est2genome model (selected options were: --model est2genome --showtargetgff yes --gff3 yes --showalignment no --showvulgar no --geneseed 250 --bestn 2 --percent 50 --minintron 20 --maxintron 10000) and marking transcripts mapping to two locations with a custom Perl script. All manually edited gene models were submitted to an automated quality control and visual inspection before being merged with the MAKER annotations into the official gene sets (OGS). The automated QC procedure (SI II.3.2) detects ca. 50 types of formatting errors caused by manual curation. Some errors are automatically fixed, whereas other error types need to be manually reviewed by curators or administrators. Curators were provided with a list of errors to correct in Web Apollo. After a correction period, QC reports were re-generated and the procedure repeated until no errors remained. An in-depth description of the QC procedure is available on github (https://github.com/NAL-i5K/I5KNAL_OGS/wiki).

6.4.6 Taxon sampling

The genomes of *A. rosae* and *O. abietinus* were compared to those of publicly available apocritan Hymenoptera and non-Hymenoptera insects. The selected Hymenoptera comprise the honeybee, *Apis mellifera* (Weinstock et al. 2006), the bumble bee *Bombus terrestris* (Sadd et al. 2015), the alfalfa leafcutter bee, *Megachile rotundata* (Kapheim et al. 2015), the white-footed sweat bee, *Lasioglossum albipes*

(Kocher et al. 2013), the solitary bee *Dufourea novaeangliae* (Kapheim et al. 2015), the leafcutter ant *Acromyrmex echinatior* (Nygaard et al. 2011), the jumping ant *Harpegnathos saltator* (Bonasio et al. 2010), the Florida carpenter ant, *Camponotus floridanus* (Bonasio et al. 2010), the European paper wasp, *Polistes dominula* (Standage et al. 2016), and the parasitoid wasp *Nasonia vitripennis* (Werren et al. 2010). The sampling covers the most diverged lineages and a significant fraction of the ecological width of the order. A comprehensive list of all genome assemblies and gene sets analyzed, including references, version numbers, and direct links to the data are given in Table S1.

6.4.7 Transposable element annotation

Species-specific repeat libraries were generated using RepeatModeler (version open-1.0.8) (Smit and Hubley 2015) with the program's default settings. The identified TEs were classified using a reference-based similarity search against RepBase (version update 20140131) (Jurka et al. 2005). Identified TEs were verified and annotation artifacts were removed by querying the identified sequences against the NCBI nr database (downloaded 2017-02-04) with blastx of the BLAST+ (version 2.6.0) software suite (Camacho et al. 2009) using the software's default settings, discarding candidates without hits against known TE proteins and domains. The filtered library was finally combined with the TE sequences of RepBase (version 20140131) referring to Metazoa and used to annotate TEs with RepeatMasker (version open-4.0.5) (Smit et al. 2015) applying the software's default settings. Genomic TE coverage was calculated using the software 'One code to find them all' (Bailly-Bechet et al. 2014) and intra-family Kimura distances, used as a proxy for TE age distribution, were

calculated using scripts available from the RepeatMasker (version open-4.0.5) software package. The full TE annotation pipeline was implemented in a custom shell script that is available on GitHub (github.com/mptrsen/mobilome). Testing for a correlation between genome size and TE content was done by applying a linear regression, Spearman rank sum method, and Kendall's Tau within R (R Core Team 2017). We also applied the phylogenetic independent contrast (PIC) method (Felsenstein 1985) as implemented in the ape package (Paradis et al. 2004) to control for a potential phylogenetic effect.

6.4.8 Comparative analysis of gene structure

The structural properties of the MAKER-inferred protein-coding gene set of the two sawflies were compared to those of the selected apocritan Hymenoptera and the red flour beetle *Tribolium castaneum* (Richards et al. 2008) using COGNATE (version 1.01) (Wilbrandt et al. 2017) with the software's default settings. The *N. vitripennis* assembly version 2.1 was used instead of version 1.0 and the NCBI release 102 annotations of the *N. vitripennis* and *B. terrestris* genomes were used instead of the eviogene and Gnomon 1.0 annotations, respectively.

6.4.9 Orthology prediction and micro-synteny

The predicted sawfly genes were clustered along with those of other Hymenoptera in OrthoDB (version 9.1) (Zdobnov et al. 2017) and orthology assessed at the systematic level Holometabola, with *T. castaneum* as outgroup. To investigate Hymenoptera genome evolution on a micro-syntenic level, we utilized the identified single-copy

orthologs (SCOs) and the recently published Hymenoptera divergence estimates (Peters et al. 2017). SCOs represent conserved genes that likely evolve under similar constraints (*e.g.*, Ciccarelli et al. 2005) and have consequently been exploited as markers to quantify genome shuffling in insects (*e.g.*, Zdobnov and Bork 2007). Using a custom Perl script (included as File S39), the conservation of micro-synteny was inferred as the fraction of shared SCOs that retain the same neighboring SCO between two species relative to their divergences time (SI II.4.5). Positional information of the SCO was extracted from the respective official gene sets. GO terms were assigned to all groups of SCOs (SCOG) using the Argot2.5 web server (Lavezzo et al. 2016; http://www.medcomp.medicina.unipd.it/Argot2-5/) with the default settings, retaining only GO terms with a score of 200 or more, and InterPro2GO (Mitchell et al. 2019) (https://www.ebi.ac.uk/GOA/), using InterProScan with the default settings (version 5.33.72) (Mitchell et al. 2019). GO terms were assigned to each SCOG when shared by ten or more species in the group. Testing for GO term enrichment in the SCOGs which remained in synteny across all pair-wise comparisons (754) against the background of all SCOGs (3,983) was performed using topGO's weighted Fisher test (weight01) (R package version 2.30.1) (Alexa and Rahnenfuhrer 2016) and goStats hypergeometric test (R package version 2.30.1) (Falcon and Gentleman 2007).

6.4.10 Gene family and domain evolution

Gene family and domain evolution was analyzed with CAFE (version 4.1) (Han et al. 2013), with coupled birth and death rates, using the orthology predictions (see above) and an ultrametric tree derived from a recently published Hymenoptera phylogeny

(Peters et al. 2017) as input. Following the suggestions of the authors of CAFE, the birth and death rate was determined considering only gene families with fewer than 100 copies in each species before reanalyzing the full dataset with the inferred rate. Protein domains were annotated in a subset of the selected genomes with *Pfam* (version 29) (Finn et al. 2016), using the provided "pfam_scan.pl" script with the default settings. The number of unique domains and domain arrangements (the linear sequence of domains present in a protein without repeats) occurring in each species were determined. Presence and absence of domains among species were inferred using a custom python script (pyDomrates; https://github.com/sklas/pyDomrates) and the *ETE3* python module (Huerta-Cepas et al. 2016). The gain and loss of domains at nodes of the tree were inferred applying the Fitch parsimony optimality criterion. Domains are considered as gained at a node if they were inferred to not have been present at the parent node. Likewise, domains are considered as lost if they were inferred to have been present only at the parent node.

6.4.11 Major royal jelly proteins

DNA sequences of specific exons of *yellow* and *mrjp*/-like genes of *A. mellifera* and *N. vitripennis* were used as query to search them with the tblastx search algorithm with the default settings (BLAST web server hosted by the NCBI) against the reference genome assemblies of *A. rosae*, *O. abietinus*, and *L. albipes*. All found coding sequences were manually curated and aligned along those of *A. mellifera*, *B. terrestris*, *M. rotundata*, *D. novaeangliae*, *P. dominula*, *A. echinatior*, *C. floridanus*, *H. saltator*, *Linepithema humile* (Christopher D Smith et al. 2011), *N. vitripennis*, *T. castaneum*, and *Zootermopsis nevadensis* (Terrapon et al. 2014) (File S13) at the

translational level with ClustalW implemented in MEGA (version 6.0.6) (Tamura et al. 2013) with the default settings. We inferred a maximum likelihood tree from the aligned amino acid sequences, using the WAG+F+R7 amino acid substitution model. Branch support was assessed from 1,000 non-parametric bootstrap replicates. Maximum-likelihood tree reconstruction was performed in IQ-TREE (version 1.6.6) (Nguyen et al. 2015) and the best fitting model was selected using ModelFinder (Kalyaanamoorthy et al. 2017) as implemented in IQ-TREE. Topology tests were done in IQ-TREE (version 1.6.8) using (i) likelihood-mapping (Strimmer and von Haeseler 1997) with four clusters (MRJPls, Y-e3, Y-x2, and all remaining Yellow proteins) and (ii) an approximate unbiased test (Shimodaira 2002), testing the inferred ML-tree (MRJPls and Y-x2 as sister-groups) against the alternative hypothesis (MRJPls and Y-e3 as sister-groups) using 1 million RELL replicates.

6.4.12 Immune genes

A set of immune genes was selected based on the Insect Innate Immunity Database (IIID) (Brucker et al. 2012) and modified according to previous studies on Hymenoptera (Evans et al. 2006, Gadau et al. 2012, Barribeau et al. 2015). Immune genes were identified with the aid of profile hidden Markov models (HMM), utilizing reference immune response-related amino acid sequences obtained from OrthoDB (Version 9) and the NCBI protein database (including RefSeq; Pruitt et al. 2012). All amino acid sequences were aligned with MAFFT with the default settings (version 7.130) (Katoh and Standley 2013) and the HMM profiles were inferred with the software HMMER (version 3.1b1) (http://hmmer.org/) with the default settings. The HMM profiles were searched against the predicted proteins with the HMM search tool

hmmsearch with the default settings. All immunity gene candidates were evaluated with a PFAM sequence search (https://pfam.xfam.org) to exclude false positives, retaining only candidate sequences with hits against known immune genes.

6.4.13 Vision genes

Opsin-coding genes were identified using amino acid reference sequences of the corresponding proteins in *A. mellifera*, *Drosophila melanogaster*, and *T. castaneum* obtained from UniProt. Reference sequences were searched against the genome assemblies with the tblastn software of the BLAST+ software suite (version 2.6.0) with the default settings. Candidate orthologs were reciprocally searched with the aid of the BLAST+ software suite, with the default settings, against the *A. mellifera*, *D. melanogaster*, and *T. castaneum* genome assemblies to sort out false positives. Finally, all verified opsin genes were manually curated within Web Apollo. Opsin amino acid sequences of *A. rosae* and *O. abietinus* were aligned to those of *A. mellifera*, *D. melanogaster*, and *N. vitripennis* (Pultz and Leaf 2003) using ClustalW (v2.1) (Larkin et al. 2007) with the default settings. Ambiguous alignment regions were excluded using the software TrimAl (version 1.3) (Capella-Gutierrez et al. 2009), implemented on the Phylemon 2.0 server (Sanchez et al. 2011) and applying the "Automated 1" settings. A maximum likelihood tree was estimated with the MEGA software (version 6.0) and applying the JTT+G amino acid substitution model. Branch support values were estimated from 500 non-parametric bootstrap replicates.

6.4.14 Metabolism

We functionally annotated all predicted proteins of *A. rosae*, *O. abietinus*, and *N. vitripennis* with the CycADS pipeline (version 1.32) (Vellozo et al. 2011) (SI II.5.10) with the default settings. CycADS is an annotation database system that collects functional annotations predicted by multiple computational methods including Blast2Go (version 2.5) (Götz et al. 2008), InterProScan (version 5.0) (Mitchell et al. 2019), Kaas-Kegg server (version 2.0) (Moriya et al. 2007) and Priam (March 13. release) (Claudel-Renard et al. 2003). Predicted EC numbers and Gene Ontology terms (GO) collected by CycADS were then processed with the Pathway Tools software (Karp et al. 2019) to infer enzymatic reactions and metabolic pathways that were finally manually curated and compared.

6.4.15 Storage proteins

Hexamerins of selected Hymenoptera (*A. echinatior*, *A. mellifera*, *C. floridanus*, *H. saltator*) and an outgroup species, the termite *Z. nevadensis*, were downloaded from UniProt and used to identify hexamerins in the *A. rosae* and *O. abietinus* genomes using the software BLAT (Kent 2002) implemented in the i5k@NAL workspace, using an e-value cut-off of 1e-10. The reference sequences were aligned against the newly identified hexamerins of the two sawflies with MAFFT (version 7) using the E-INS-i algorithm with the default settings. The multiple amino acid sequence alignment was further processed with GBlocks (version 0.91b) (Castresana 2000) with the default settings. A maximum likelihood tree was inferred using IQ-TREE (version 1.6.6) applying the best fitting amino acid substitution model after the BIC

criterion (LG+G4) as determined by ModelFinder. Branch support values were estimated from 1,000 non-parametric bootstrap replicates.

6.4.16 Odorant and gustatory receptors

Initial candidate genes were identified by querying reference amino acid sequences from Hymenoptera (Robertson et al. 2018, Zhou et al. 2015) against the MAKER inferred gene set and the genome assemblies using tblastn (version 2.2.31) with the default settings. Candidate gene models were manually annotated or corrected in Web Apollo considering raw reads and assembled transcripts of the antennal transcriptomes which were mapped against the genomes of the respective species using the 'map to reference' function in CLC Genomics Workbench 7 (Qiagen, Hilden, Germany) with the program's default settings. Annotated gene models were queried against the assemblies along with those of other Hymenoptera to identify additional genes potentially missed by the initial annotation. Candidate nucleotide sequences were subsequently searched against the NCBI nr database with tblastx to eliminate false positives with the default settings. Predicted amino acid sequences were aligned to those of *A. echinatior*, *A. mellifera*, *N. vitripennis* (Zhou et al. 2015), and *C. cinctus* (Robertson et al. 2018) using MUSCLE (version 3.8.31) (Edgar 2004) with the default settings. All resulting alignments were visually inspected and, if necessary, manually curated. Maximum-likelihood phylogenies were built using PhyML (version 3.0) (Guindon et al. 2010) under the best-fitting substitution model as determined by SMS (Lefort et al. 2017). Branch support was estimated through an approximate likelihood-ratio test (Anisimova and Gascuel 2006). All phylogenetic trees were visualized with FigTree (version 1.4.2) (http://tree.bio.ed.ac.uk/software/figtree/).

6.4.17 Additional information

Additional information and details on all analyses, including parts not mentioned in the main text, are given in full length in the extended Supplementary Information available from: https://doi.org/10.1093/gbe/evaa106.

6.5 Data and software availability

The genome sequence assemblies of *A. rosae* and *O. abietinus* are deposited in the Ag Data Commons archives (DOI: 10.15482/USDA.ADC/1459563 and 10.15482/USDA.ADC/1459569, respectively). The raw genome sequencing reads are deposited in the NCBI Sequence Read Archives (SRA) under the accession numbers SRX276605–9 and SRX330912–15, respectively. The raw whole-body transcriptome sequencing reads of *A. rosae* and *O. abietinus* are deposited in the SRA under the accession numbers SRX903108–9 and SRX906143–4, respectively. The raw antennal transcriptome sequencing reads of *A. rosae* and *O. abietinus* are deposited in the SRA under the accession numbers ERX1404801–2 and ERX1404803, respectively. The automated genome annotation of the *A. rosae* and *O. abietinus* draft genomes are deposited in the Ag Data Commons archives (DOI: 10.15482/USDA.ADC/1459565 and 10.15482/USDA.ADC/1459561, respectively). The official gene sets (merged automated and manual annotations) of *A. rosae* and *O. abietinus* are deposited in the Ag Data Commons archives (DOI: 10.15482/USDA.ADC/1459566 and 10.15482/USDA.ADC/1459558, respectively). The custom Perl script used to infer pair-wise micro-synteny is provided in the supplementary information (File S39).

The metabolism databases of *A. rosae*, *O. abietinus*, and *N. vitripennis* have been included in the ArthropodaCyc database collection (Baa-Puyoulet et al. 2016) and are available under http://arthropodacyc.cycadsys.org/ATHRO/, http://arthropodacyc.cycadsys.org/ORUAB/, and http://arthropodacyc.cycadsys.org/NASVI/.

6.6 Acknowledgments

This genome sequencing project contributes to the i5K initiative, whose target is to sequence the genomes of 5,000 arthropods (Robinson et al. 2011; Evans et al. 2013). We thank the staff at the Baylor College of Medicine Human Genome Sequencing Center for their contributions. Genome sequencing, assembly, and annotation was funded by a grant U54 HG003273 from the National Human Genome Research Institute (NHGRI to R.A.G.). Erich Bornberg-Bauer, Alexander Donath, Steffen Klasberg, Christoph Mayer, Bernhard Misof, Oliver Niehuis, Jan Philip Oeyen, Ralph S. Peters, Malte Petersen, Lars Podsiadlowski, Jeanne Wilbrandt, and Tanja Ziesmann acknowledge the Leibniz Graduate School on Genomic Biodiversity research. Bernhard Misof, Victoria C. Moris, Oliver Niehuis, Malte Petersen, Björn Marcus von Reumont, and Jeanne Wilbrandt additionally acknowledge the German Research Foundation (DFG) for financial support (MI 649/x-1; NI 1387/2-1; NI 1387/3-1; RE 3454/2-1; RE 3454/6-1). John H. Werren thanks the US National Science Foundation (DEB 1257053 and IOS 1456233). Robert M. Waterhouse acknowledges support from the Swiss National Science Foundation (PP00P3_170664). Björn Marcus von Reumont thanks the Fraunhofer IME (Gießen) for infrastructure. Bernhard Misof and Oliver Niehuis thank Claudia Etzbauer and Sandra Kukowka (ZFMK, Bonn) for

technical support. Oliver Niehuis and Jan Philip Oeyen thank Caroline Müller (University of Bielefeld) for providing samples of *Athalia rosae* to take photographs of this species. Oliver Niehuis furthermore thanks Stefan Tischendorf (Darmstadt) for sending detailed information where and when to collect *Orussus abietinus* in the field. Victoria C. Moris and Oliver Niehuis acknowledge the support of the Freiburg Galaxy Team and Prof. Rolf Backofen, Bioinformatics, University of Freiburg, Germany, funded by Collaborative Research Centre 992 Medical Epigenetics (DFG grant SFB 992/1 2012) and German Federal Ministry of Education and Research (BMBF grant 031 A538A RBC (de.NBI)). This project was also supported by Det Frie Forskningsråd (7014-0008B to CJPG). Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

6.7 References

Abdalsamee MK, Müller C (2012) Effects of indole glucosinolates on performance and sequestration by the sawfly *Athalia rosae* and consequences of feeding on the plant defense system. Journal of Chemical Ecology. 38: 1366–1375.

Abe M (1988) A biosystematic study of the genus *Athalia* Leach of Japan (Hymenoptera: Tenthredinidae). Esakia. 26: 91–131.

Aguiar AP, Deans AR, Engel MS, Forshage M, Huber JT, et al. 2013. Order Hymenoptera. Zootaxa. 3703: 51–62.

Ahnlund H, Ronquist F (2002) Den röda parasitväxtstekelns (*Orussus abietinus*) biologi och förekomst i Norden. Entomologisk Tidskrift. 122: 1–10.

Alexa A, Rahnenfuhrer J (2016) topGO: Enrichment analysis for gene ontology. R package version 2.38.1.

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic Biology. 55: 539–552.

Ardila-Garcia AM, Umphrey GJ, Gregory TR (2010) An expansion of the genome size dataset for the insect order Hymenoptera, with a first test of parasitism and eusociality as possible constraints. Insect Molecular Biology. 19: 337–346.

Baa-Puyoulet P, Parisot N, Febvay G, Huerta-Cepas J, Vellozo AF, et al. (2016) ArthropodaCyc: a CycADS powered collection of BioCyc databases to analyse and compare metabolism of arthropods. Database 2016: baw081.

Bailly-Bechet M, Haudry A, Lerat E (2014) "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. Mobile DNA. 5: 13.

Bao R, Friedrich M (2009) Molecular evolution of the *Drosophila* retinome: exceptional gene gain in the higher Diptera. Molecular Biology and Evolution. 26: 1273–1287.

Barribeau SM, Sadd BM, du Plessis L, Brown MJ, Buechel SD, et al (2015) A depauperate immune repertoire precedes evolution of sociality in bees. Genome Biology. 16:83.

Beres BL, Dosdall LM, Weaver DK, Cárcamo HA, Spaner DM (2011) Biology and integrated management of wheat stem sawfly and the need for continuing research. Canadian Entomologist. 143: 105–125.

Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, et al. (2009) The genome of the blood fluke *Schistosoma mansoni*. Nature. 460: 352–358.

Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, et al. (2010) Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science. 329: 1068–1071.

Branstetter MG, Childers AK, Cox-Foster D, Hopper KR, Kapheim KM, et al. (2018) Genomes of the Hymenoptera. Current Opinions in Insect Science. 25: 65–75.

Brucker RM, Funkhouser LJ, Setia S, Pauly R, Bordenstein SR (2012) Insect Innate Immunity Database (IIID): An annotation tool for identifying immune genes in insect genomes. PLoS One. 7: e45125.

Buttstedt A, Moritz RFA, Erler S (2014) Origin and function of the major royal jelly proteins of the honeybee (*Apis mellifera*) as members of the *yellow* gene family. Biological Reviews. 89: 255–269.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. (2009) BLAST+: architecture and applications. BMC Bioinformatics. 10: 421.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research. 18: 188–196.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25: 1972–1973.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution. 17: 540–552.

Chipman AD, Ferrier DEK, Brena C, Qu J, Hughes DST, et al. (2014) The first myriapod genome sequence reveals conservative arthropod gene content and

genome organisation in the centipede *Strigamia maritima*. PLoS Biology. 12: e1002005.

Ciccarelli FD, von Mering C, Suyama M, Harrington ED, Izaurralde E, et al. (2005) Complex genomic rearrangements lead to novel primate gene function. Genome Research. 15: 343–351.

Claudel-Renard C, Chevalet C, Faraut T, Kahn D (2003) Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Research. 31: 6633–6639.

Cristino AS, Nunes FMF, Barchuk AR, Aguiar-Coelho VM, Simões ZLP, et al. (2010) Organization, evolution and transcriptional profile of hexamerin genes of the parasitic wasp *Nasonia vitripennis* (Hymenoptera: Pteromalidae). Insect Molecular Biology. 19: 137–146.

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science. 298: 2157– 2167.

Desjardins CA, Cerqueira GC, Goldberg JM, Hotopp JCD, Haas BJ, et al. (2013) Genomics of *Loa loa*, a *Wolbachia*-free filarial parasite of humans. Nature Genetics. 45: 495–501.

Desjardins CA, Perfectti F, Bartos JD, Enders LS, Werren JH (2010) The genetic basis of interspecies host preference differences in the model parasitoid *Nasonia*. Heredity. 104: 270–277.

Drapeau MD, Albert S, Kucharski R, Prusko C, Maleszka R (2006) Evolution of the Yellow/Major Royal Jelly protein family and the emergence of social behavior in honey bees. Genome Research. 16: 1385–1394.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32: 1792–1797.

Engström PG, Sui SJH, Drivenes Ø, Becker TS, Lenhard B (2007) Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Research. 17: 1898–1908.

Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, et al. (2006) Immune pathways and defence mechanisms in honey bees *Apis mellifera*. Insect Molecular Biology. 15: 645–656.

Evans JD, Brown SJ, Hackett KJJ, Robinson G, Richards S (2013) The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. Journal of Heredity. 104: 595–600.

Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. Bioinformatics. 23: 257–258.

Felsenstein J (1985) Phylogenies and the comparative method. American Naturalist. 125: 1–15.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, et al. (2016) The Pfam protein families database:towards a more sustainable future. Nucleic Acids Research. 44: D279–D285.

Gadau J, Helmkampf M, Nygaard S, Roux J, Simola DF, et al. (2012) The genomic impact of 100 million years of social evolution in seven ant species. Trends in Genetics. 28: 14–21.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, et al. (2011) High-quality draft assemblies of mammalian genomes from massively parallel

sequence data. Proceedings of the National Academy of Sciences of the U.S.A. 108: 1513–1518.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research. 36: 3420–3435.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 29: 644–652.

Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, et al. (2011) The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. Nature. 479: 487–492.

Gregory TR (2018) Animal genome size database. http://www.genomesize.com

Grimaldi DA, Engel MS (2005) Evolution of the Insects. Cambridge University Press, Cambridge. 772 pp.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. (2010)New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology. 59: 307–321.

Hagner-Holler S, Pick C, Girgenrath S, Marden JH, Burmester T (2007) Diversity of stonefly hexamerins and implication for the evolution of insect storage proteins. Insect Biochemistry and Molecular Biology. 37: 1064–1074.

Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Molecular Biology and Evolution. 30: 1987–1997.

Hanes J, Šimuth J (1992) Identification and partial characterization of the major royal jelly protein of the honey bee (*Apis mellifera* L.). Journal of Apicultural Research. 31: 22–26.

Hanrahan SJ, Johnston JS (2011) New genome size estimates of 134 species of arthropods. Chromosome Research. 19: 809–823.

Henze MJ, Oakley TH (2015) The dynamic evolutionary history of pancrustacean eyes and opsins. Integrative and Comparative Biology. 55: 830–842.

Hu X, Leming MT, Whaley MA, O'Tousa JE (2014) Rhodopsin coexpression in UV photoreceptors of *Aedes aegypti* and *Anopheles gambiae* mosquitoes. Journal of Experimental Biology. 217: 1003–1008.

Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Molecular Biology and Evolution. 33: 1635–1638.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature. 409: 860–921.

Jervis MA, Ellers J, Harvey JA (2008) Resource acquisition, allocation, and utilization in parasitoid reproductive strategies. Annual Review of Entomology. 53: 361–385.

Jervis MA, Kidd NAC, Fitton MG, Huddleston T, Dawah HA (1993) Flower-visiting by hymenopteran parasitoids. Journal of Natural History. 27: 67–105.

Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB (2007) Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. Nature. 445: 86–90.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, et al. (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research. 110: 462–467.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods. 14: 587–589.

Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, et al. (2015) Genomic signatures of evolutionary transitions from solitary to group living. Science. 348:1139–1143.

Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, et al. (2019) Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Briefings in Bioinformatics. Bbz 104.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution. 30: 772–780.

Kawecki TJ (1998) Red queen meets Santa Rosalia: arms races and the evolution of host specialization in organisms with parasitic lifestyles. American Naturalist. 152: 635–651.

Kent CF, Minaei S, Harpur BA, Zayed A (2012) Recombination is associated with the evolution of genome structure and worker behavior in honey bees. Proceedings of the National Academy of Sciences of the U.S.A. 109: 18012–18017.

Kent WJ (2002) BLAT —The BLAST-like alignment tool. Genome Research. 12: 656–664.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology. 14: R36.

Kim Soonok, Cho YS, Kim HM, Chung O, Kim H, et al. (2016) Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. Genome Biology. 17: 211.

Kocher SD, Li C, Yang W, Tan H, Yi SV, et al. (2013) The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. Genome Biology. 14: R142.

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics. 5: 59.

Kupke J, Spaethe J, Mueller MJ, Rössler W, Albert Š (2012) Molecular and biochemical characterization of the major royal jelly protein in bumblebees suggest a non-nutritive function. Insect Biochemistry and Molecular Biology. 42: 647–654.

Kwon JY, Dahanukar A, Weiss LA, Carlson JR (2007) The molecular basis of CO_2 reception in *Drosophila*. Proceedings of the National Academy of Sciences of the U.S.A. 104: 3574–3578.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947–2948.

Lavezzo E, Falda M, Fontana P, Bianco L, Toppo S (2016) Enhancing protein function prediction with taxonomic constraints — The Argot2.5 web server. Methods. 93: 15–23.

Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, et al. (2013) Web Apollo: a web-based genomic annotation editing platform. Genome Biology. 14: R93.

Lees JG, Dawson NL, Sillitoe I, Orengo CA (2016) Functional innovation from changes in protein domains and their combinations. Current Opinion in Structural Biology. 38: 44–52.

Lefort V, Longueville JE, Gascuel O (2017) SMS: smart model selection in PhyML. Molecular Biology and Evolution. 34: 2422–2424.

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27: 764–770.

Martins JR, Nunes FMF, Cristino AS, Simões ZLP, Bitondi MMG (2010) The four hexamerin genes in the honey bee: structure, molecular evolution and function deduced from expression patterns in queens, workers and drones. BMC Molecular Biology. 11: 23.

Mattson WJ (1980) Herbivory in relation to plant nitrogen content. Annual Review of Ecology, Evolution, and Systematics. 11: 119–161.

Mine S, Sumitani M, Aoki F, Hatakeyama M, Suzuki MG (2017) Identification and functional characterization of the sex-determining gene doublesex in the sawfly, *Athalia rosae* (Hymenoptera: Tenthredinidae). Applied Entomology and Zoology. 52: 497–509.

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Research. 47: D351–360.

Moore AD, Bornberg-Bauer E (2012) The dynamics and evolutionary potential of domain loss and emergence. Molecular Biology and Evolution. 29: 787–796.

Moreno M, Giralt E (2015) Three valuable peptides from bee and wasp venoms for therapeutic and biotechnological use: melittin, apamin and mastoparan. Toxins. 7: 1126–1150.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research. 35: W182–185.

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. Nature. 420: 520–562.

Mrinalini, Werren JH (2017) Parasitoid wasps and their venoms. In: Gopalakrishnakone P, Malhotra A (Eds.) Evolution of venomous animals and their toxins. Toxinology. Springer, Dordrecht. 187–212 pp.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution. 32: 268–274.

Ni JD, Baik LS, Holmes TC, Montell C (2017) A rhodopsin in the brain functions in circadian photoentrainment in *Drosophila*. Nature. 545: 340–344.

Niehuis O, Gibson JD, Rosenberg MS, Pannebakker BA, Koevoets T, et al. (2010) Recombination and its impact on the genome of the haplodiploid parasitoid wasp *Nasonia*. PLoS One. 5: e8597.

Nygaard S, Zhang G, Schiøtt M, Li C, Wurm Y, et al. 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. Genome Research. 21: 1339–1348.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 20: 289–290.

Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 23: 1061–1067.

Pask GM, Slone JD, Millar JG, Das P, Moreira JA, et al. (2017) Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. Nature Communications. 8: 297.

Peitsch D, Fietz A, Hertel H, de Souza J, Ventura DF, et al. (1992) The spectral input systems of hymenopteran insects and their receptor-based colour vision. Journal of Comparative Physiology A. 170: 23–40.

Peters RS (2010) Host range and offspring quantities in natural populations of *Nasonia vitripennis* (Walker, 1836) (Hymenoptera: Chalcidoidea:Pteromalidae). Journal of Hymenoptera Research. 19: 128–138.

Peters RS, Meusemann K, Petersen M, Mayer C, Wilbrandt J, et al. (2014) The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. BMC Evolutionary Biology. 14: 52.

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, et al. (2017) Evolutionary history of the Hymenoptera. *Current Biology*. 27: 1013–1018.

Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, et al. (2019) Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evolutionary Biology. 19: 11.

Poelchau M, Childers C, Moore G, Tsavatapalli V, Evans J, et al. (2015) The i5k Workspace@NAL-enabling genomic data access, visualization and curation of arthropod genomes. Nucleic Acids Research. 43: D714–719.

Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012) NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Research. 40: D130–135.

Pultz MA, Leaf DS (2003) The jewel wasp *Nasonia*: querying the genome with haplo-diploid genetics. *Genesis.* 35: 185–191.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 317: 86–94.

Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature. 453: 1064–1071.

Quicke DLJ (1997) Parasitic Wasps. Chapman & Hall: London.

Rappoport N, Linial M (2015) Trends in genome dynamics among major orders of insects revealed through variations in protein families. BMC Genomics. 16: 583.

Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, et al. (2008) The genome of the model beetle and pest *Tribolium castaneum*. Nature. 452: 949–955.

Robertson HM, Waterhouse RM, Walden KKO, Ruzzante L, Reijnders MJMF, et al. (2018) Genome sequence of the wheat stem sawfly, *Cephus cinctus*, representing an early-branching lineage of the Hymenoptera, illuminates evolution of hymenopteran chemoreceptors. Genome Biology and Evolution. 10: 2997–3011.

Robertson HM, Gadau J, Wanner KW (2010) The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. Insect Molecular Biology. 19: 121–136.

Robertson HM, Wanner KW (2006) The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. Genome Research. 16: 1395–1403.

Robinson GE, Hackett KJ, Purcell-Miramontes M, Brown SJ, Evans JD, et al. (2011) Creating a buzz about insect genomes. Science. 331: 1386.

Sackton TB, Werren JH, Clark AG (2013) Characterizing the infection-induced transcriptome of *Nasonia vitripennis* reveals a preponderance of taxonomically-restricted immune genes. PLoS One. 8: e83984.

Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, et al. (2015) The genomes of two key bumblebee species with primitive eusocial organization. Genome Biology. 16: 76.

Sanchez R, Serra F, Tarraga J, Medina I, Carbonell J, et al. (2011) Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Research. 39: W470–474.

Sáringer G (1974) Problems with *Athalia rosae* L. (Hym., Tenthredinidae) in Hungary. In: Proceedings of the 4. International Rapskongress. Giessen. 575–578 pp.

Savard J, Tautz D, Richards S, Weinstock GM, Gibbs RA, et al. (2006) Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of holometabolous insects. Genome Research. 16: 1334-1338.

Schmid-Hempel P (2011) *Evolutionary Parasitology: The Integrated Study of Infections, Immunology, Ecology, and Genetics*. Oxford University Press: Oxford.

Sekine K, Furusawa T, Hatakeyama M (2015) The *boule* gene is essential for spermatogenesis of haploid insect male. Developmental Biology. 399: 154–163.

Senthilan PR, Helfrich-Förster C (2016) Rhodopsin 7–the unusual rhodopsin in *Drosophila*. PeerJ. 4: e2427.

Sharkey MJ (2007) Phylogeny and classification of Hymenoptera. Zootaxa. 1668: 521–548.

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Systematic Biology. 51: 492–508.

Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, et al. (2013) Insights into bilaterian evolution from three spiralian genomes. Nature. 493: 526–531.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31: 3210–3212.

Slansky FJ (1986) Nutritional ecology of endoparasitic insects and their hosts: an overview. Journal of Insect Physiology. 32: 255–261.

Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6: 31.

Small KS, Brudno M, Hill MM, Sidow A (2007) A haplome alignment and reference sequence of the highly polymorphic *Ciona savignyi* genome. Genome Biology. 8: R41.

Smit A, Hubley R (2015) RepeatModeler Open-1.0. http://wwww.repeatmasker.org.

Smit A, Hubley R, Green P (2015) RepeatMasker Open-4.0. http://www.repeatmasker.org.

Smith CD, Zimin A, Holt C, Abouheif E, Benton R, et al. (2011) Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proceedings of the National Academy of Sciences of the U.S.A. 108: 5673–5678.

Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, et al. (2011) Draft genome of the red harvester ant *Pogonomyrmex barbatus*. Proceedings of the National Academy of Sciences of the U.S.A. 108: 5667–5672.

Snodgrass RE (1925) Anatomy and Physiology of the Honeybee. Mcgeaw-Hallbook Company: New York.

Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, et al. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. Science. 314: 941–952.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, et al. (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature. 466: 720–726.

Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, et al. (2008) The *Trichoplax* genome and the nature of placozoans. Nature. 454: 955.

Standage DS, Berens AJ, Glastad KM, Severin AJ, Brendel VP, et al. (2016) Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. Molecular Ecology. 25: 1769–1784.

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24: 637–644.

Steidle JLM, Schöller M (1997) Olfactory host location and learning in the granary weevil parasitoid *Lariophagus distinguendus* (Hymenoptera: Pteromalidae). Journal of Insect Behaviour. 10: 331–342.

Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proceedings of the National Academy of Sciences of the U.S.A. 94: 6815–6819.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution. 30: 2725–2729.

Tavares MG, Carvalho CR, Soares FAF (2010) Genome size variation in *Melipona* species (Hymenoptera: Apidae) and sub-grouping by their DNA content. Apidologie. 41: 636–642.

Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, et al. (2014) Molecular traces of alternative social organization in a termite genome. Nature Communication. 5: 3636.

The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: A platform for investigating biology. Science. 282: 2012–2018.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology. 28: 511–515.

Velarde RA, Sauer CD, Walden KKO, Fahrbach SE, Robertson HM (2005) Pteropsin: A vertebrate-like non-visual opsin expressed in the honey bee brain. Insect Biochemistry and Molecular Biology. 35: 1367–1377.

Vellozo AF, Véron AS, Baa-Puyoulet P, Huerta-Cepas J, Cottret L, et al. (2011) CycADS: An annotation database system to ease the development and update of BioCyc databases. Database. 2011: bar008.

Vilhelmsen L (2003) Larval anatomy of Orussidae (Hymenoptera). Journal of Hymenoptera Research. 12: 346–354.

Vilhelmsen L, Isidore N, Romani R, Basibuyuk HH, Quicke DLJ (2001) Host location and oviposition in a basal group of parasitic wasps: The subgenual organ, ovipositor apparatus and associated structures in the Orussidae (Hymenoptera, Insecta). Zoomorphology. 121: 63–84.

Vilhelmsen L, Turrisi GF (2011) Per arborem ad astra: Morphological adaptations to exploiting the woody habitat in the early evolution of Hymenoptera. Arthropod Structure and Development. 40: 2–20.

Vinson SB, Iwantsch GF (1980) Host regulation by insect parasitoids. The Quarterly Review of Biology. 55: 143–165.

Visser B, Le Lann C, den Blanken FJ, Harvey JA, van Alphen JJM, et al. (2010) Loss of lipid synthesis as an evolutionary consequence of a parasitic lifestyle. Proceedings of the National Academy of Sciences of the U.S.A. 107: 8677–8682.

Wakakuwa M, Kurasawa M, Giurfa M, Arikawa K (2005) Spectral heterogeneity of honeybee ommatidia. Naturwissenschaften. 92: 464–467.

Weihrauch D, Donini A, O'Donnell MJ (2012) Ammonia transport by terrestrial and aquatic insects. Journal of Insect Physiology. 58: 473–487.

Weinstock GM, Robinson GE, Gibbs RA, Worley KC, Evans JD, et al. (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature. 443: 931–949.

Werren JH, Richards S, Desjardins CA, Niehuis O, et al. (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science. 327: 343–348.

Whitfield JB (1998) Phylogeny and evolution of host-parasitoid interactions in Hymenoptera. Annual Review of Entomology. 43: 129–151.

Wilbrandt J, Misof B, Niehuis O (2017) COGNATE: comparative gene annotation characterizer. BMC Genomics. 18: 535.

Wilfert L, Gadau J, Schmid-Hempel P (2007) Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity. 98: 189–197.

Yamamoto DS, Sumitani M, Tojo K, Lee JM, Hatakeyama M (2004) Cloning of a decapentaplegic orthologue from the sawfly, *Athalia rosae* (Hymenoptera), and its expression in the embryonic appendages. Development Genes and Evolution. 214: 128–133.

Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, et al. (2017) OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Research. 45: D744–749.

Zdobnov EM, Bork P (2007) Quantification of insect genome divergence. Trends in Genetics. 23: 16–20.

Zhang G, Fang X, Guo X, Li L, Luo R, et al. (2012) The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 490: 49–54.

Zhou X, Rokas A, Berger SL, Liebig J, Ray A, et al. (2015) Chemoreceptor evolution in Hymenoptera and its implications for the evolution of eusociality. Genome Biology and Evolution. 7: 2407–2416.

Zhou X, Slone JD, Rokas A, Berger SL, Liebig J, et al. (2012) Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. PLoS Genetics. 8: e1002930.

Chapter 7

General Discussion

Within this thesis, I presented my research on various subjects, all related to the study of biodiversity, using a variety of tools and approaches. In the following sections, I will place my results in a larger context than what was possible in the individual research articles and also provide a discussion on the interactions and future of the respective fields.

7.1 A single species redescription in the face of the taxonomic impediment

In Chapter 2, I redescribed the Glomerida species *Eupeyerimhoffia archimedis,* that was originally described by Strasser in 1965. In the context of the global biodiversity crisis and the taxonomic impediment we are facing, this is but a small contribution that you cannot even equate to a drop in the ocean. However, despite its limitations in terms of scale and scope, the results of the study highlight the value of integrative species descriptions that employ multiple lines of evidence. This not only gives integrative descriptions a higher degree of confidence in the species hypothesis but also provides data that are useful beyond their primary goal.

Sequencing the DNA barcode of *E. archimedis* not only serves as independent evidence, but is also an important resource for future studies. In this case, the comparison to sequences from previous integrative species descriptions unveiled that the family Protoglomeridae is very likely non-monophyletic, clearly demonstrating the need for further systematic investigations within this group. However, the sequence data might also serve as an important reference for other biodiversity studies in the region (e.g. long-term monitoring of diversity) and the voucher material

deposited in the collection of the ZFMK is also conserved for long-term storage of DNA to enable future studies that might want to utilize different markers or methods.

The use of scanning electron microscopy to examine a wide range of character systems, as opposed to only documenting what is typically used to delineate species, allowed for the discovery of a previously overlooked diversity in morphological characters, such as the molar plate lacking a groove. This is of particular importance in groups that do not yet have a phylogenetic system of classification and/or have systems that depend on single, or very few characters. For example, as the numbers of species within a group increases, the characters that were initially thought informative might not suffice for delineating new species. If a majority of the species described thus far include a comprehensive documentation of their morphology using high resolution images, researchers will be able to quickly assess alternative characters systems without the need for recollecting and reexamining the whole group.

However, despite its comprehensive and fulfilling nature, the integrative approach to taxonomy does not seem like a viable option for solving the taxonomic impediment given the current state of taxonomy. The approach is inherently laborious and, at least in the case of millipedes, requires specialized skills in dissection and preparation for the morphological investigation. The adoption of integrative and detailed approaches therefore leads to a decrease in productivity, in terms of species described per taxonomist per year (Sangster and Luksenburg 2015). This decrease in productivity, coupled with the lack of training for new taxonomists (Bilton 2014), the lack of job opportunities for taxonomists (Agnarsson and Kuntner 2007), and overall decreasing

number of taxonomists (Mora et al. 2011) make it undeniably clear that there is an urgent need for innovation in taxonomy if we are to describe the diversity of life before it goes extinct.

The most commonly suggested approaches to tackle the taxonomic impediment are based exclusively on molecular data, predominantly the mitochondrial protein-coding gene *cytochrome c oxidase subunit 1* (COI barcoding or DNA taxonomy; Hebert et al. 2003, Tautz et al. 2003), often in combination with methods for automatic species delimitation (e.g. Meierotto et al. 2019). Some authors have even gone as far as arguing that the use of genetic data should be obligatory in new species descriptions (Riedel et al. 2013). These approaches' claim to superiority is the speed of species-discovery that they offer and also that they provide added value by generating DNA barcode reference libraries that enable efficient monitoring approaches (e.g. meta-barcoding of environmental DNA samples; Ji et al. 2013), which are crucial for conservation efforts (Nichols and Williams 2006). However, despite their claimed efficiency, these approaches have come under heavy criticism due to a myriad of technical problems and for being inherently uninformative for the discovery and description of biodiversity (e.g. DeSalle et al. 2005, Will et al. 2005, Meier et al. 2006, Rubinoff et al. 2006, Eberle et al. 2019, Britz et al. 2020, Dupérré 2020, Zamani et al. 2020).

The claimed efficiency of DNA barcoding based approaches has been disputed several times (Sääksjärvi et al. 2004; Hopkins et al. 2019, Zamani et al. 2020, Britz et al. 2020). For wasps, it has been shown that the field-work connected with collecting

samples is by far the most time consuming aspect of taxonomic research and providing a species description and diagnosis can be done in only a few minutes (Sääksjärvi et al. 2004; Hopkins et al. 2019, Zamani et al. 2020). This is a speed of description that cannot be matched by molecular methods, but it remains to be seen if similar approaches can be universally applied.

Apart from the claimed speed of DNA taxonomy, the technical criticism of DNA barcoding based methods of species description, delineation, and identification is mainly due to the properties of mitochondrial genome. In comparison to the nuclear genome, the mitochondrial genome has an inconsistent but also high mutation rate, experiences little to no recombination, has a small effective population size, and is also subject to evolutionary biases due to *Wolbachia* infections in some lineages (summarized in Rubinoff et al. 2006 and Eberle et al. 2019). Due to these properties, studies based exclusively on COI data often fail to correctly delineate species (e.g. Meier et al. 2006, Eberle et al. 2016) and overestimate the number of species as well as the observed genetic structure (Eberle et al. 2016). To overcome these issues, authors have argued for the addition of multiple nuclear loci for the automatic delineation of species (e.g. Yang and Rannala 2010, Dupuis et al. 2012). More recently, Eberle et al. (2019) argued for the application of next-generation sequencing to produce a genomic-scale dataset comprising of hundreds of standardized nuclear markers. This approach will not only overcome the limitations of mitochondrial data, but also drastically expands the number of characters available, potentially to a level that allows phylogenetic analyses, and that could also provide detailed insights into the genetic structure and history of populations. However, the application of novel

sequencing strategies and the resulting increase in data for analysis do not address the most central part of the criticism, that the use of molecular markers alone is uninformative for the discovery of species and the description of biodiversity.

Species described exclusively on the grounds of molecular data are essentially numbered anonymous units in a space occupied by other numbered anonymous units that together produce an illusion of a complete description of biodiversity. The information included in integrative descriptions (e.g. morphology, life-history, ecology, behavior etc) are what allows us to see them not only as a numbered unit, but as a species in the context of their ecosystem. Furthermore, describing new species based solely on DNA characters provides no way to associate these species with previously described species and museum specimens that lack molecular data (Dupérré 2020, Zamani et al. 2020).

Another problem with the application of DNA taxonomy to solve the global taxonomic impediment is that most species occur in highly threatened tropical forests (Wright 2005, Phillips et al. 2017) and most undiscovered species occur in Asia and the southern hemisphere (Costello et al. 2012). Countries in these regions are often still developing and do not have the financial resources, technology, or knowledge required to apply the DNA taxonomy approach, especially at a genomic scale. Furthermore, classical taxonomy currently represents an accessible field of research to developing countries and these countries are increasingly important for the employment of taxonomists (e.g. Pearson et al. 2011). Enforcing DNA taxonomy, as some authors have suggested (Riedel et al. 2013), would therefore only further reduce

the number of active taxonomists and also hinder the progress of science in developing countries (Dupérré 2020). The lack of resources and knowledge to establish and run a DNA taxonomy initiative in developing countries can likely be overcome via cooperation and the exchange of technology as well as the knowledge required to apply it. However, the inherent uninformative nature of DNA based species descriptions remains unresolved.

Next to or along with DNA taxonomy, it is commonly suggested that increasing the participation of amateurs (i.e. citizen science) in taxonomy can compensate for the ever declining funding for professionals (Pearson et al. 2011, Hill et al. 2012, Konrat et al. 2018). Citizen scientists have a long history of making significant contributions to other fields, such as astronomy (e.g. Kyba et al. 2013), and amateurs are already making substantial contributions to taxonomy, having described almost half of the recently discovered European species (Fontaine et al. 2012). However, there are at least two notable problems that have to be overcome in order for this to be a (partial) solution to the taxonomic impediment. Firstly, there are strong societal preferences for certain taxa, while others are ignored, actively disliked or even remain entirely unknown to the public (Cardoso et al. 2011, Troudet et al. 2017). Although this taxonomic bias is present also among professionals, or at least in the groups that professionals are hired to study, it is much stronger in the wider society (Troudet et al. 2017). Therefore, it will be necessary to devise a strategy to increase the public interest in neglected groups. This can, for example, be done by increasing the public awareness of the biodiversity, ecological importance, and even exciting peculiarities of "less charismatic" groups (Cardoso et al. 2011). Secondly, amateur taxonomists

have caused a substantial amount of problems by disregarding the scientific standards that are set by the taxonomic community (Kaiser et al. 2013, Troncoso-Palacios et al. 2019). These scientific errors are not only time-consuming to correct but also negatively impact every field that rely on a stable taxonomy, especially conservation biology (Kaiser et al. 2013, Garnett and Christidis 2017, Troncoso-Palacios et al. 2019). As an example for the extent of damage that can be done: a single amateur herpetologist renamed and reclassified two super-families, one family, three sub-families, 89 tribes, 113 genera, 64 sub-genera, 25 species and 53 subspecies of snakes, geckos, skinks and crocodiles, all within the span of 12 years, without providing scientific evidence for his changes or submitting his studies to peer-review (Kaiser et al. 2013). These problems are to a large extent made possible by the absence of requirements for the quality of the taxonomic work in International Code of Zoological Nomenclature that is governed by the International Commission on Zoological Nomenclature (ICZN). There are no requirements for peer-review nor for the quality of the evidence supporting the proposed taxonomic changes. Although these issues are a large in both scale and consequence, they can be overcome by the application of stricter rules in the ICZN Code and also by strategically coupling amateurs with professionals that can guide them in their pursuits (Pearson et al. 2011), as long as the professionals still exist.

In summary, it is clear that there is no simplistic and one-dimensional solution to the taxonomic impediment. If we are to discover and describe the biodiversity of our planet before it goes extinct, a multi-pronged approach, uniting traditions with technological development, must be employed in unison with a campaign to change

perception of the field among researchers, officials, and society in generall. Only by raising awareness about the importance of taxonomy that is done with no other goal but to explore and describe the basic units of the biosphere can we hope to increase public funding of taxonomy and ensure recruitment of the next generation of taxonomists as well as educated citizen scientists. Funding basic research must be made the popular and only acceptable political stance in the eyes of the public. One way to illustrate the vast resources made accessible by taxonomy to a wider audience is to focus on biomimicry (or bionics) – the study of nature to inform the invention of designs, materials, processes, and products (Wheeler 2020). Framing the biomimicry approach with human societies' desperate need for sustainable solutions and the millions of years of natural selection that have produced solutions for virtually every problem that life has faced shows the importance of the field in a manner that is easily communicated and understood by both professionals and the wider public. Finding a sustainable future for humanity must become synonymous with sustainable taxonomy.

Next to producing a focused campaign to increase the appreciation of, funding for, and recruitment to taxonomy, taxonomy as a field needs to provide realizable solutions to increase the efficiency of species discovery and description. A wide range of technological solutions are being continuously developed (e.g. Earl et al. 2020, Ärje et al. 2020) and steps have already been taken in developing strategies to tackle the problem (e.g. Wheeler et al. 2012), but they need to be formalized and employed. Taxonomists also need to realize the great opportunities that lie within their discoveries and capitalize on them. For example, if the application of genome scale DNA taxonomy becomes feasible, it will enable taxonomists to use the generated

population genomic data to address questions at the core of evolutionary biology and thereby substantially increase the reach and impact of their work. Although chasing impact is a perversion brought on by the structuring of modern academia, it is a reality that we must face and take advantage of when possible.

Lastly, it is up to the taxonomic community to show agency in paving the way for its own future by widely testing and then adopting the new approaches. Only by doing so, taxonomy can show that it will live up to the promises made and secure its rightful position as one of the most important sciences.

7.2 Phylogenetic systematics: Morphology

In Chapter 3, I presented the results of our attempts to reconstruct the evolutionary history of the Glomerida based on morphological characters. The presented analysis for the first time enabled us to infer the apomorphic characters for the Glomerida and also provided the first volvation-independent characters in support of a Glomerida+Glomeridesmida relationship. However, it provided disappointingly little insight into the relationships among the genera and families of the Glomerida, only confirming that the currently known characters are not phylogenetically informative at this level. Although this is might be seen as a failure, it was not expected that we fully resolve the relationships based on such a limited character matrix. Rather, the compilation of the character matrix, including all characters known from literature, represents an important first step towards a well-founded understanding of the group and as such is a major result of the study.

The classical literature on Diplopoda is on one hand a treasure trove of knowledge, but on the other hand a significant barrier for current researchers (Blanke and Wesener 2014). A large amount of knowledge is buried in large and complex German and Italian textbooks published in the early 20th century that are seldom available as digital copies (e.g. Silvestri 1903, Attems 1926–1930, Verhoeff 1928). Summarizing all the relevant knowledge from this literature therefore increases the efficiency of future studies and significantly lowers the barrier of entry for new researchers. Furthermore, the character matrix, once it has been sufficiently expanded to resolve the intraordinal relationships, will allow taxonomists and, most importantly, paleontologists to confidently place taxa within the group.

The value of these resources will only increase in the future. Given the rapid development of efficient scanning and reconstruction methods (see Wipfler et al. 2016), pushing towards increased automation (e.g. O'Mahoney et al. 2020), it is conceivable that in the not too distant future species descriptions will regularly be accompanied by scanning data (e.g. Akkari et al. 2015). With the increased application range of deep learning algorithms and other similar technologies (Vasques et al. 2016, Glaser et al. 2019), one can further envision that the interpretation of the scans through the identification of structures, characters, and even character states that can be seamlessly integrated in the work-flow and existing matrixes. When these technological developments are realized, the rapid and efficient generation of data in combination with recent advancements made in the application of probabilistic methods for the analysis of morphological data (e.g. Puttick et al. 2019, Vernygora et

al. 2020) could bring about a future in which morphology again becomes a central approach in both taxonomy and phylogenetic systematics.

7.3 Phylogenetic systematics: Molecules

In Chapter 5, I present the results from tackling the same questions as in Chapter 3, the phylogenetic relationships of the millipede order Glomerida, utilizing a comprehensive set of analyses on a modern phylogenomic-scale molecular datasets derived from transcriptome sequencing. The results of these analyses clearly substantiate that the current system of the order consists of predominantly non-monophyletic groups. The analyses further uncovered multiple independent origins of dwarfism and aberrant morphology, as well as the biogeographic affinities of some taxa. However, despite the tremendous amount of data, our approach failed to fully resolve the relationships among the sampled taxa. In particular, three species were recovered at different positions across the tree depending on the dataset and reconstruction method.

The failure to produce a robust phylogeny of the Glomerida was unexpected as as similar approaches have been successfully applied in other groups on a comparable evolutionary time scale (e.g. Darwell et al. 2018). As discussed in section 5.4, it is likely that increasing the density of the taxon sampling in combination with a more specific set of orthologous genes would eliminate some of the problems observed in the here presented analysis. However, this might not be enough to produce a robust phylogenetic hypothesis. A multitude of unsolved methodological problems, that cannot be addressed by further increasing the amount of data, still remain at virtually

all levels of data processing in phylogenomics and the progress towards solving them is slow (Young and Gillung 2019, Jermiin et al. 2020, Kapli et al. 2020). Furthermore, the application of a wider set of tree reconstruction methods and data filtering methods increasingly results in unresolved phylogenetic relationships due to factors that we still do not sufficiently understand. For example, in a recent study, Zhang et al. (2020) found that there was a significant functional bias between genes supporting two alternative hypotheses. As the authors note, it will likely be necessary to compare whole genomes to uncover what processes produce this bias and how it can be overcome to produce a well-resolved phylogeny.

The persisting issues in phylogenomic methods are highlighted particularly well by the problems associated with the fundamental procedure of generating multiple sequence alignments. A plethora of methods exist for the alignment of both protein and nucleotide data. These methods differ in their approaches, perform differently based on the nature of the input data, and produce diverging results that have a strong influence on the reconstruction of phylogenetic relationships (Ogden and Rosenberg 2006, Wong et al. 2008). This becomes a substantial problem through the lack of methods to determine and compare the performance of the different alignment algorithms on our data. Currently, there are no real options other than subjectively picking what produces results that are consistent with our, often unfounded, expectations. Once this step has been completed, the next issue arises with the masking of the newly generated alignments. It is necessary to mask alignments in order to remove wrongly aligned regions and to reduce bias (e.g. introduced through hyper-variable regions) that violates the assumptions of the models available for tree

reconstruction (Young and Gillung 2019, Jermiin et al. 2020). Again, we are lacking methods to inform the choice of masking software and also methods to inform the stringency with which we filter the data. We are thus rendered victims to a strongly subjective balancing act between removing too little and too much, with both options having a very realistic potential to skew the inference of phylogenetic relationships (Jermiin et al. 2020). Due to these problems, it is hard to find confidence in phylogenies that are proclaimed as resolved on the basis of molecular data alone, especially so if only a narrow set of methods have been applied. One immediately wonders how different it would look if another method had been applied at any of the numerous steps of the analyses.

In order to circumvent the issues outlined above, but also to provide an independent control for the results of other analyses, some authors have now recruited genomic meta-characters, such as near intron pairs (Niehuis et al. 2012, Lehmann et al. 2013) and transposable element insertions (Suh et al. 2015), for the inference of phylogenetic relationships. Similar approaches, especially those using gene-level synteny comparisons between genomes, have seen an increase in development (e.g. Bourque and Pevzner 2002, Liu et al. 2005, Luo et al. 2012, Feng et al. 2017, Drillon et al. 2020, Malik et al. 2020). It remains to be seen if these approaches will be establish as core tools in phylogenetics. However, given the decreasing costs of whole genome sequencing of non-model organisms, they might, in time, become a viable independent control for phylogenetic hypotheses and provide insights into the biological processes that currently hamper phylogenomic reconstructions (e.g. Zhang et al. 2020).

It is somewhat ironic that molecular phylogenetics, failing to fully deliver on its promises, is turning to morphological characters of the genome, essentially bringing the field to a full circle. However, great advances have been made on the back of molecular phylogenetics. Molecular methods provide a tremendous amount of characters that allow insights into groups that have a highly simplified morphology and thus few or no morphological characters available for analyses (e.g. Morse and Normark 2006). Molecular approaches have also provided well-founded evolutionary hypotheses for relationships that were previously considered highly problematic based on morphology alone and also provide insights into the timing of their divergence (e.g. Misof et al. 2014, Upham et al. 2019). Ultimately, the molecular approach to phylogenetic systematics remains a central and important tool, but will not always provide the easy out-of-the-box solution that many researchers were hoping for.

7.4 Comparative genomics, species, and speciation

In Chapter 6, I presented the results of the sequencing and comparative analysis of the genomes of the phytophagous sawfly *Athalia rosae* and the parasitoid sawfly *Orussus abietinus*. By comparing the two sawfly genomes to those of ten other Hymenoptera, spanning divergence times of over 250 myr, the study provided new insights into the genome composition of the ancestral hymenopteran, the dynamics of hymenopteran genome evolution as well as the genomic traits related to the transitions from phytophagy to parasitoidism and the disparate success of parasitoid lineages.

Next to its primary results, our study exemplifies the power of comparative genomic studies in non-model organisms on the back of detailed knowledge of the species' natural history in combination with a resolved and dated phylogeny. Together, these resources enable us to observe trends and patterns (e.g. in gene-content or genome structure) that we can interpret in light of the evolutionary relationships and biology of the sampled species to formulate hypotheses about the processes and mechanisms that could have shaped the observed patterns. However, in order to interpret these patterns, we often rely heavily on extrapolating from knowledge accrued by studying model organisms (e.g. functions of a specific domain or gene), which can be problematic.

In our study on the sawfly genomes, the analysis of genes related to light sensing and vision (opsins) exemplify the problematic nature of interpreting results based on functional characterization from other species (Chapter 6.2.8). Analyzing the opsin repertoire of the sawflies, an ortholog of the Rh7 opsin was identified in *A. rosa*, but not in any other of the Hymenoptera. The Rh7 opsin has been shown to be expressed in the photoreceptor cells of two mosquito species (Hu et al. 2014) and is highly conserved across arthropods (Senthilan and Helfrich-Förster 2016). Given that the larvae of all other investigated Hymenoptera have a partially or completely reduced visual system, likely related to the transition to develop in wood (Vilhelmsen and Turrisi 2011), it seems logical that Rh7 might have been lost along with this. However, Ni et al. (2017) showed that Rh7 is not expressed in the photo receptors, but in the brain of *Drosophila* species, where it is likely involved in controlling the circadian rhythm (Ni et al. 2017). The latter finding renders our hypothesis of vision-

gene loss along with the reduction of the larval visual system a lot weaker than what was initially believed. If one wishes to test this hypothesis, it will be necessary to functionally characterize Rh7 in both the imago and larva of *A. rosae*.

The interpretation of differences in gene-family copy numbers between species in respect to potential functional diversity is another difficulty. The copy number, especial of highly dynamic gene families, is commonly thought to correlate to the functional requirements and versatility of a species (e.g. Xu et al. 2009, Zhou et al. 2012). Although this might hold true in some or possible a majority of cases, results from other studies have shown that this interpretation is not always as straight forward: In *Drosophila,* the gene encoding for DSCAM (Down syndrome cell adhesion molecule), which is an important part of the immune system and nervous system development, encodes for ~150,000 protein isoforms that are generated through alternative splicing (Watson et al. 2005). In comparison, both the tick *Ixodes scapularis* and the centipede *Strigamia maritima* have multiple copies of the same gene (Brites et al. 2013), with the genome of *S. maritima* containing ~80 canonical and ~20 non-canonical genes encoding for DSCAM proteins (Chipman et al. 2014). Although the exact functional differences between these two strategies to generate protein diversity are not known in detail, it shows that copy numbers, in the absence of supporting data, need to be interpreted carefully.

Despite the problems inherent with the interpretation of comparative genomic analyses of non-model organisms, it is clear that these analyses represent an incredibly powerful approach to discover patterns and even evolutionary novelties.

However, in order to fully understand these observations, a more focused approach is required. This is becoming increasingly feasible, even for non-model organisms, through the development of a wide range of methods for the functional validation, such as genetically modifying non-model organisms using CRISPR/Cas (reviewed in Ford et al. 2019), using RNA interference to silence target genes (reviewed by Gu and Knipple 2013), studying chromatin biology via ATAC-seq (reviewed in Yan et al. 2020), conducting expression studies with single-cell resolution (reviewed in Kulkarni et al. 2019), and analyzing epigenetic effects using methylome sequencing (reviewed in Non and Thayer 2015). For some hypotheses, in particular hypotheses on more general evolutionary mechanisms, it will however be necessary to expand the sampling to the level where evolution acts, the population level (Ehrlich and Raven 1969). With the history of research on transposable elements in mind, it quickly becomes clear what tremendous potential the omics approaches have when applied in concert.

Transposable elements (TEs), or jumping genes (McClintock 1950), were first considered a part of the non-functional 'Junk DNA' (Ohno 1972). Later, TEs were described as genomic parasites (parasitic genes) that invade the host genome (Dawkins 1976, Doolittle and Sapienza 1980, Orgel and Crick 1980) under great costs to the host by inducing deleterious effects such as the disruption of endogenous coding genes (e.g. Kazazian 1988). Since then, through the rapidly increasing number of sequenced genomes, we are beginning to get a more complete understanding of TEs and their effects on the host genome (Guerreiro 2012, Maumus et al. 2015). TEs are now recognized as key drivers of genome evolution (e.g. Jangam et al. 2017,

Peccoud et al. 2017, Petersen et al. 2019), that through their mutational properties can increase the evolutionary potential of species (Guerreiro 2012) and even act as a mechanism of speciation (Serrato-Capuchina and Matute 2018). This potential is not only realized by transpositions within the genome, but also through the horizontal transfer of DNA between genomes of closely (e.g. Hill and Betancourt 2018) and distantly related species (e.g. Peccoud et al. 2017). Through the study of TEs within populations and closely related species, our understanding of their potential as drivers of evolution has been further refined (e.g. *Drosophila,* reviewed by Mérel et al. 2020). For example, it is known that TEs can be activated by a range of factors, including abiotic and biotic stress (Guerreiro 2012), and that they can affect both the development and behavior of the host organism (e.g. Rech et al. 2019). Recently it was shown that, in the thale cress (*Arabidopsis thaliana*), some heat-stress-activated TEs are preferentially integrated into regions that contain environmental response genes and away from essential genes (Quadrana et al. 2019). Although this is highly speculative and likely over-simplified, these results could indicate that that the genome has a powerful source of directed mutagenesis that can be activated only when needed and targeted as required depending on the source of the stress, all while protecting the essential regions of the genome from potentially damaging mutational loads. This might, almost instantaneously, increase the genetic diversity within the population and thereby providing a substrate for selection to allow rapid adaptation to various sources of stress, like changes in the environment. It might therefore be evolutionarily beneficial to keep a set of active TEs in the genome. If this is confirmed and found to be a more general mechanism of genome evolution, it might require us to rethink how we view evolution and thus redefine evolutionary biology.

Another field that is heavily influenced by the increased ability to sequence and compare genomes on the population level is the study of speciation (i.e. population genetics/genomics; Wolf et al. 2010, Seehausen et al. 2014, Wolf and Ellegren 2016, Campbell et al. 2018). The introduction of the biological species concept (BSC) by Mayr (1942) for the first time provided a framework for the study of the speciation process. Mayr postulated that species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups, and that the build-up of reproductive barriers could arise through geographic isolation. Since then, a plethora of species concepts have been suggested (e.g. Coyne 1994, Harrison 1998, Jiggins & Mallet 2000). However, the majority of alternative species concepts are centered around ways to delineate species rather than the process of speciation (Mayr 2001) and the BSC has remained the most central framework for studying speciation (Gao and Reiseberg 2020) through the study of reproductive barriers and reproductive isolation.

It is now known that these reproductive barriers and isolation can be initiated either by intrinsic (i.e. evolution of genetic incompatibility through genetic drift, indirectly via selection, or through genomic conflict) or extrinsic (i.e. divergent selection through ecological or sexual selection) factors (see Box 2 in Seehausen et al. 2014). Through the use of model organisms, the intrinsic mechanisms of reproductive isolation were historically the most thoroughly investigated (especially genetic incompatibility; Wu 2001, Wolf et al. 2010). However, with the advent of molecular methods, such as the amplification of anonymous loci (e.g. microsatellites; Goldstein and Schlotterer 1999), researchers were increasingly able to gain insights into the

genetic structure and history of natural populations, including gene-flow (Campbell et al. 2018). This revealed that gene flow among species or diverging species could be more common than what was predicted by the BSC (Wu 2001).

The combination of observed reproductive barriers as well as introgression and hybridization, in particular in *Drosophila,* led to the introduction of an alternative species concept, the 'genic view of speciation', by Wu (2001). Wu states that under the BCS the whole genome acts as a cohesive co-adapted unit and that it would loose its integrity from exchanging genetic material with another species (hybridization and introgression). In the genic view, species are not defined by their whole genome, but by a set of speciation genes that only make up a fraction of the genome. These are genes that influence the species' morphological, reproductive, or ecological characters (Wu 2001) and are suggested to be the only part of the genome that would result in a fitness-reduction when affected by introgression, while the rest of the genome can be more freely (ex)changed. The speciation genes, or non-introgressable regions of the genome, are expected to be more divergent than introgressable regions (Wu 2001) and have been referred to as genomic islands (reviewed in Wolf and Ellegren 2016).

Although the methods employed up to that point enabled the study of non-model organisms, they did not readily provide insights into the functional or structural factors that influence the observed genomic patterns and inferred processes, such as the presence of genomic islands in natural populations (Wu 2001). It was not until more recently, with the arrival of the next-generation sequencing revolution, that researchers were able to fully investigate the onset of reproductive isolation, the

selective forces affecting it and the roles of ecology, geography, and demography (Seehausen et al. 2014, Campbell et al. 2018).

The genomic era of speciation research has already, in less than a decade, produced a number of astonishing results, including the documentation of isolation by adaption (e.g. Nosil et al. 2008, Shafer et al. 2013), genome-wide responses to selection (e.g. Burke 2012, Soria-Carrasco et al. 2014, Riesch et al. 2017, Moest et al. 2020), and that the tendency of lineages to diverge or remain static is dependent on genomic architecture (Campbell et al. 2018). Mechanisms, such as cryptic genetic variation (Zheng et al. 2019) and epigenetic modifications (e.g. Flatscher et al. 2012, Burggren 2016, Smith et al. 2016). that were previously not considered have now also been shown to have effects on adaptive evolution and therefore potentially on speciation.

A growing number of studies have documented the prevalence of hybridization among diverging lineages (Payseur Rieseberg 2016, Wang et al. 2020) and across exceedingly strong reproductive barriers (Sambatti et al. 2012, Roux et al. 2013, Brandvain et al. 2014, Zhang et al. 2016), indicating that hybridization is a common feature of speciation and evolution in general. In the butterfly genus *Heliconius* it was shown that adaptive introgression of color-pattern loci across several lineages was important for the early divergence of species (Edelman 2019, Moest et al. 2020), but that later stages of speciation might be shaped more by reproductive barriers such as habitat partitioning, hybrid sterility, and chemically mediated mate choice (Mérot et al. 2017). Hybridization has also been shown to be a driving force in the adaptive radiation of cichlid fish by enabling the formation of novel phenotypes (Stelkens et al.

2009). Moreover, in accordance with the genic view of speciation, the presence of genomic islands of genetic divergence has been confirmed in several taxa, including crows and butterflies, where color-pattern loci are located in such regions (Poelstra et al. 2014, Nadeau et al. 2012, Nadeau et al. 2014). However, it remains unknown whether this is indeed a general feature of the early phases of speciation, as suggested by Wu (2001). A study on diverging lineages of the stick insect genus *Timema*, that are in the early stages of speciation, did not find evidence of growing genomic islands, despite genome-wide differentiation in loci associated with adaptive color-pattern loci and mate choice in the presence of gene-flow (Riesch et al. 2017). Furthermore, both empirical and simulation based studies have shown that genomic islands of divergence can develop through a multitude of other factors including structure of the genome, variation in recombination rates, initial genetic diversity, time since divergence, genetic drift, background selection, selective sweeps, and adaption to local environments (Noor and Bennett 2009, Cruickshank and Hahn 2014, Campagna et al. 2015, Quilodrán et al. 2019).

It is clear that great strides have been made towards understanding speciation through the comparison of genomes. However, we still do not know how general the observed patterns, forces, and mechanisms are. This can be attributed to the fact that taxon sampling in population genomic studies still remains biased towards a few model systems of speciation research (Wolf and Ellegren 2016), most studies address lineages that are in the early stages of speciation (Wang et al. 2020), and in less studied systems it can be difficult to study genomic divergence as there might be more than two gene-pools that are diverging (e.g. Gladieux et al. 2015, Leroy et al. 2017).

Because of these limitations, the progress that has been made thus far has more than anything accentuated the complexities and diversity of speciation processes (Galtier 2018) and evolution therefore remains difficult to predict (Nosil et al. 2018). In order to progress towards a more general understanding and universal hypothesis of speciation, investigations across all stages of the speciation continuum with a much wider taxon sampling are essential (Seehause et al. 2014, Wolf and Ellegren 2016, Wang et al. 2020).

Our incomplete understanding of speciation and the forces governing it also directly affects species delimitation approaches that only consider DNA evidence (Galtier 2018; Cadena and Zapata 2020). The problems arise, at least in part, through the uncertainties surrounding species concepts (Galtier 2018) and the attempted verification of phenotypic species hypotheses using molecular data (i.e. when 'good' morphological species do not represent distinct genetic lineages; Cadena and Zapata 2020). The former is a problem that arises through the geneic-view of speciation, and other alternatives to the BSC, that do not have a strict definition as to when speciation is complete and generally accept that gene-flow is possible after speciation is complete (Mallet 2020). This is especially problematic in birds, where post-zygotic reproductive isolation is often poorly developed (Cadena and Zapata 2020). For example, in the redpoll genus *Acanthis*, three distinct forms are distinguishable based on plumage and bill morphology but cannot be separated by molecular data (Mason and Tyler 2015). However, comparisons of gene expression data revealed patterns that correlate with with the three phenotypes, indicating that the speciation-loci under divergent selection might not be found in protein-coding genes but in regulatory

regions or networks. This is by some authors interpreted as evidence of species boundaries (Cadena and Zapata 2020), but could just as easily be argued to be evidence of a nacent speciation process or as a descrete polymorphism. Due to the large diversity in empirical speciation data and the difficulties involved in placing the observation in a specific phase of the speciation continuum (e.g. Wang et al. 2020), it is currently not possible to determine an objective threshold for when speciation is complete if one does not apply Mayrs criterion of absolute reproductive isolation. Although these issues might be considered an 'end-all' argument against genomic species delimitation, I argue that it is far too early to pass judgment. As stated in the previous paragraph, our current understanding of the genomics of speciation is severely limited by a restricted taxonomic sampling and it is not clear where on the speciation continuum examples such as the case of *Acanthis* should be placed. Rather than abandoning the genomic evidence, these cases should motivate us to pursue an improved understanding of what species are and how they arise.

A limiting factor in the expansion of the taxon sampling is the amount of resources necessary to screen large numbers of natural populations from a wide range of taxa. Although the costs of whole genome sequencing is rapidly decreasing, it still represents a barrier for studies that sequence hundreds of individuals across multiple populations and species. If the genome-scale DNA taxonomy based on universal single-copy orthologs, as suggested by Eberle et al. (2019), is widely adopted as a part of integrative species description, it could provide a solution for this problem. Single-copy orthologs have already successfully been employed to study patterns of allopatric divergence in plants (Dong et al. 2019) and could prove an efficient

heuristic to discover genomic and phenotypic patterns in a wide range of speciation processes across the tree of life as well as to identify lineages in different stages of speciation for further study. Similar to our changed perception of transposable elements, going from being seen as parasitic junk to an integral part of the dynamics of genome evolution as a powerful source of variation induction and phenotypic novelty, an increased density and resolution of data on speciation processes, together with detailed study of model organisms, might revolutionize our view of speciation and species concepts. Even if a revolution is not realized, it might reveal objective diagnostic characters to distinguish different stages of the speciation continuum and bring us closer to widely and empirically supported species concept.

7.5 Concluding remarks

If we hope to truly understand how the diversity and complexities of nature originated and, especially so, if we hope to make predictions on how it will develop and adapt to the continuation of the Anthropocene, one-dimensional fast-track solutions are unlikely to provide meaningful progress. Thus, in conclusion, it is clear that no matter the level of biodiversity research, a holistic or integrative approach should be viewed as a requirement.

7.6 References

Agnarsson I, Kuntner M (2007) Taxonomy in a changing world: Seeking solutions for a science in crisis. Systematic Biology. 56: 531–539.

Akkari N, Enghoff H, Metscher BD (2015) A new dimension in documenting new species: High-detail imaging for myriapod taxonomy and first 3D cybertype of

a new millipede species (Diplopoda, Julida, Julidae). PLoS One. 10: e0135243.

Ärje J, Melvad C, Jeppesen MR, Madsen SA, Raitoharju J, et al. (2020) Automatic image-based identification and biomass estimation of invertebrates. Methods in Ecology and Evolution. 11: 922–931.

Attems, C. 1926-1930. Myriapoda. In Handbuch der Zoologie Walter de Gruyter, Berlin.

Bilton DT (2014) Whats in a name? What have taxonomy and systematics ever done for us? Journal of Biological Education. 48: 116–118.

Blanke A, Wesener T (2014) Revival of forgotten characters and modern imaging techniques help to produce a robust phylogeny of the Diplopoda (Arthropoda, Myriapoda). Arthropod Structure and Development. 43: 63–75.

Bourque G, Pevzner PA (2002) Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Research. 12: 26–36.

Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL (2014) Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. PLoS Genetics. 10: e1004410.

Brites D, Brena C, Ebert D, Du Pasquier L (2013) More than one way to produce protein diversity: duplication and limited alternative splicing of an adhesion molecule gene in basal arthropods. Evolution. 67: 2999–3011.

Britz R, Hundsdörfer AK, Fritz U (2020) Funding, training, permits – the three big challenges of taxonomy. Megataxa. 1: 59–52.

Burggren W (2016) Epigenetic inheritance and its role in evolutionary biology: Re-evaluation and new perspectives. Biology. 5: 1–22.

Burke MK (2012) How does adaptation sweep through the genome? Insights from long-term selection experiments. Proceedings of the Royal Society B: Biological Sciences. 279: 5029–5038.

Cadena CD, Zapata F (2020) The genomic revolution and species delimitation in birds (and other organisms): gene flow matters, but what about other evolutionary forces. EcoEvoRxiv Preprints. Accessed on 24.09.20: 10.32942/osf.io/s3dhk.

Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ (2015) Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Molecular Ecology. 24: 4238–4251.

Campbell CR, Poelstra JW, Yoder AD (2018) What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Biological Journal of the Linnean Society. 124: 561–583.

Cardoso P, Erwin TL, Borges PAV, New TR (2011) The seven impediments in invertebrate conservation and how to overcome them. Biological Conservation. 144: 2647–2655.

Chipman AD, Ferrier DEK, Brena C, Qu J, Hughes DST, et al. (2014) The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. PLOS Biology. 12: e1002005.

Costello MJ, Wilson SP, Houlding B (2012) Predicting total global species richness using rates of species description and estimates of taxonomic effort. Systematic Biology. 61: 871–88.

Coyne JA (1994) Ernst Mayr on the origin of species. Evolution 48: 19–30.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology. 23: 3133–3157.

Darwell CT, Ayyampalayam S, Leebens-Mack J, Smith CI, Segraves KA, Althoff DM (2018) Phylogenomic reconstruction of transcriptome data confirms the basal position of Prodoxidae moths within the order Lepidoptera. Arthropod Systematics and Phylogeny. 76: 59–64.

Dawkins R (1976) The Selfish Gene. Oxford University Press, Oxford. 224 pp.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: Taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences. 360: 1905–1916.

Dong Y, Chen S, Cheng S, Zhou W, Ma Q, et al. (2019) Natural selection and repeated patterns of molecular evolution following allopatric divergence. Elife. 8: 1–35.

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature. 284: 601–603.

Drillon G, Champeimont R, Oteri F, Fischer G, Carbone A (2020) Phylogenetic Reconstruction Based on Synteny Block and Gene Adjacencies. Molecular Biology and Evolution. 37: 2747–2762.

Dupérré N (2020) Old and new challenges in taxonomy: what are taxonomists up against? Megataxa. 1: 59–62.

Dupuis JR, Roe AD, Sperling FAH (2012) Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. Molecular Ecology. 21: 4422–4436.

Earl C, White A, Trizna M, Frandsen P, Kawahara A, et al. (2019) Discovering Patterns of Biodiversity in Insects Using Deep Machine Learning. Biodiversity Information Science and Standards. 3.

Eberle J, Warnock RCM, Ahrens D (2016) Bayesian species delimitation in *Pleophylla* chafers (Coleoptera) – the importance of prior choice and morphology. BMC Evolutionary Biology. 16: 94.

Eberle J, Ahrens D, Mayer C, Niehuis O, Misof B (2020) A Plea for Standardized Nuclear Markers in Metazoan DNA Taxonomy. Trends in Ecology and Evolution. 35: 336–345.

Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, et al. (2019) Genomic architecture and introgression shape a butterfly radiation. Science. 366: 594–599.

Ehrlich PR, Raven PH (1969) Differentiation of Populations. Science. 165: 1228–1232.

Feng B, Lin Y, Zhou L, Guo Y, Friedman R, Xia R, Hu F, Liu C, Tang J (2017) Reconstructing yeasts phylogenies and ancestors from whole genome data. Scientific Reports. 7: 15209.

Flatscher R, Frajman B, Schönswetter P, Paun O (2012) Environmental heterogeneity and phenotypic divergence: Can heritable epigenetic variation aid speciation? Genetics Research International. 2012: 1–9.

Fontaine B, van Achterberg K, Alonso-Zarazaga MA, Araujo R, Asche M, et al. (2012) New Species in the Old World: Europe as a Frontier in Biodiversity Exploration, a Test Bed for 21st Century Taxonomy. Plos ONE. 7: e36881.

Ford K, McDonald D, Mali P (2019) Functional Genomics via CRISPR–Cas. Journal of Molecular Biology. 431: 48–65.

Galtier N (2018) Delineating species in the speciation continuum: A proposal. Evolutionary Applications. 12: 657–663.

Gao L, Rieseberg LH (2020) While neither universally applicable nor practical operationally, the biological species concept continues to offer a compelling framework for studying species and speciation. National Science Review. 7: 1398–1400.

Garnett ST, Christidis L (2017) Taxonomy anarchy hampers conservation. Nature. 546: 25–27.

Gladieux P, Wilson BA, Perraudeau F, Montoya LA, Kowbel D, et al. (2015) Genomic sequencing reveals historical, demographic and selective factors associated with the diversification of the fire-associated fungus *Neurospora discreta*. Molecular Ecology. 24: 5657–5675.

Glaser JI, Benjamin AS, Farhoodi R, Kording KP (2019) The roles of supervised machine learning in systems neuroscience. Progress in Neurobiology. 175: 126–137.

Goldstein DB, Schlotterer C (1999) Microsatellites: Evolution and Applications. Oxford University Press, Oxford. 368 pp.

Gu L, Knipple DC (2013) Recent advances in RNA interference research in insects: Implications for future insect pest management strategies. Crop Protection. 45: 36–40.

Guerreiro M (2012) What makes transposable elements move in the *Drosophila* genome? Heredity. 108: 461–468.

Harrison RG (1998) Linking evolutionary pattern and process: The relevance of species concepts for the study of speciation. In: Howard D and Berlochers S (eds.) Endless Forms: Species and Speciation. Oxford University Press, Oxford. 19–31 pp.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences. 270: 313–321.

Hill T, Betancourt AJ (2018) Extensive exchange of transposable elements in the *Drosophila pseudoobscura* group. Mobile DNA. 9: 22–24.

Hill A, Guralnick R, Smith A, Sallans A, Gillespie R, et al. (2012) The notes from nature tool for unlocking biodiversity records from museum records through citizen science. Zookeys. 209: 219–233.

Hopkins T, Roininen H, van Noort S, Broad GR, Kaunisto K, et al. (2019) Extensive sampling and thorough taxonomic assessment of Afrotropical Rhyssinae (Hymenoptera, Ichneumonidae) reveals two new species and demonstrates the limitations of previous sampling efforts. ZooKeys. 878: 33–71.

Hu X, Leming MT, Whaley MA, O'Tousa JE (2014) Rhodopsin coexpression in UV photoreceptors of *Aedes aegypti* and *Anopheles gambiae* mosquitoes. Journal of Experimental Biology. 217: 1003–1008.

Jangam D, Feschotte C, Betrán E (2017) Transposable element domestication as an adaptation to evolutionary conflicts. Trends in Genetics. 33: 817–831.

Jermiin LS, Catullo RA, Holland BR (2020) A new phylogenetic protocol: dealing with model misspecification and confirmation bias in molecular phylogenetics. NAR Genomics and Bioinformatics. 2: 1–14.

Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, et al. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters. 16: 1245–1257.

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. Trends in Ecology and Evolution. 15: 250–255.

Kaiser H, Crother BI, Kelly CM, Luiselli L, O'Shea M, et al. (2013) Best Practices: In the 21st Century, Taxonomic Decisions in Herpetology are Acceptable Only When Supported by a Body of Evidence and Published via Peer-Review. Herpetological Review. 44: 8–23.

Kapli P, Yang Z, Telford MJ (2020) Phylogenetic tree building in the genomic age. Nature Reviews Genetics. 21: 428–444.

Kazazian HH, Wong C, Youssoufian H, Scott AE, Phillips DG (1988) Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism of mutation in man. Nature. 332: 164–166

Konrat M von, Campbell T, Carter B, Greif M, Bryson M, et al. (2018) Using citizen science to bridge taxonomic discovery with education and outreach. Applications in Plant Sciences. 6: 1–12.

Kulkarni A, Anderson AG, Merullo DP, Konopka G (2019) Beyond bulk: a review of single cell transcriptomics methodologies and applications. Current Opinion in Biotechnology. 58: 129–136.

Kyba CCM, Wagner JM, Kuechly HU, Walker CE, Elvidge CD, et al. (2013) Citizen science provides valuable data for monitoring global night sky luminance. Scientific Reports. 3:1–6.

Lehmann J, Stadler PF, Krauss V (2013) Near intron pairs and the metazoan tree. Molecular Phylogenetics and Evolution. 66: 811–823.

Leroy T, Roux C, Villate L, Bodénès C, Romiguier J, et al. (2017) Extensive recent secondary contacts between four European white oak species. New Phytologist. 214: 865–878.

Liu T, Tang J, Moret B (2005) Quartet-based phylogeny reconstruction from gene orders. Proceedings of the 11th International Computing and Combinatorics Conference (COCOON’ 05), LNCS. Vol. 3595. Springer, Berlin. 63–73 pp.

Luo H, Arndt W, Zhang Y, Shi G, Alekseyev MA, et al. (2012) Phylogenetic analysis of genome rearrangements among five mammalian orders. Molecular Phylogenetics and Evolution. 65: 871–882.

Malik AJ, Poole AM, Allison JR (2020) Structural Phylogenetics with Confidence. Molecular Biology and Evolution. 37: 2711–2726.

Mallet J (2020) Alternative views of biological species: reproductively isolated units or genotypic clusters? National Science Review. 7: 1401–1407.

Mason NA, Taylor SA (2015) Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. Molecular Ecology. 24: 3009–3025.

Maumus F, Fiston-Lavier AS, Quesneville H (2015) Impact of transposable elements on insect genomes and biology. Current Opinion in Insect Science. 7: 30–36.

Mayr E (1942) Systematics and the Origin of Species from the Viewpoint of a Zoologist. Harvard University Press, Cambridge. 334 pp.

Mayr E (2001) Wu's genic view of speciation. Journal of Evolutionary Biology. 14: 866–867.

McClintock B (1950) The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences of the U.S.A. 36: 344–355.

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. Systematic Biology. 55: 715–728.

Meierotto S, Sharkey MJ, Janzen DH, Hallwachs W, Hebert PDN, et al. (2019) A revolutionary protocol to describe understudied hyperdiverse taxa and overcome the taxonomic impediment. Deutsche Entomologische Zeitschrift. 66: 119–145.

Mérel V, Boulesteix M, Fablet M, Vieira C (2020) Transposable elements in *Drosophila*. Mobile DNA. 11: 1–20.

Mérot C, Salazar C, Merrill RM, Jiggins CD, Joron M (2017) What shapes the continuum of reproductive isolation? Lessons from *Heliconius* butterflies. Proceedings of the Royal Society B: Biological Sciences. 284: 20170335.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. Science. 346: 763–767.

Moest M, Van Belleghem SM, James JE, Salazar C, Martin SH, et al. (2020) Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. Plos Biology. 18: e3000597.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? PLoS Biology. 9: 1–8.

Morse GE, Normark BB (2006) A molecular phylogenetic study of armoured scale insects (Hemiptera: Diaspididae). Systematic Entomology. 31: 338–349.

Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, et al. (2012) Genome wide patterns of divergence and gene flow across a butterfly radiation. Molecular Ecology. 22: 814–826.

Nadeau NJ, Ruiz M, Salazar P, Counterman B, Medina JA, et al. (2014) Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. Genome Research. 24: 1316–1333.

Ni JD, Baik LS, Holmes TC, Montell C (2017) A rhodopsin in the brain functions in circadian photoentrainment in *Drosophila*. Nature. 545: 340–344.

Nichols JD, Williams BK (2006) Monitoring for conservation. Trends in Ecology and Evolution. 21: 668–673.

Niehuis O, Hartig G, Grath S, Pohl H, Lehmann J, et al. (2012) Genomic and morphological evidence converge to resolve the enigma of Strepsiptera. Current Biology. 22: 1309–1313.

Non AL, Thayer ZM (2015) Epigenetics for anthropologists: An introduction to methods. American Journal of Human Biology. 27: 295–303.

Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert: Examining the role of restricted recombination in maintaining species. Heredity. 103: 439–444.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: “isolation by adaptation” and multiple roles for divergent selection. Evolution. 62: 316–336.

Nosil P, Villoutreix R, De Carvalho CF, Farkas TE, Soria-Carrasco V, et al. (2018) Natural selection and the predictability of evolution in timema stick insects. Science. 359: 765–770.

Ogden TH, Rosenberg MS (2006) Multiple sequence alignment accuracy and phylogenetic inference. Systematic Biology. 55: 314-328.

Ohno S (1972) So much junk DNA in our genome. In Smith HH (ed.) Evolution of genetic systems, Brookhaven Symposium on Biology. 366-370 pp.

O’Mahoney T, Mcknight L, Lowe T, Mednikova M, Dunn J (2020) A machine learning based approach to the segmentation of micro CT data in archaeological and evolutionary sciences. bioRxiv. Accessed on 24.09.20: https://doi.org/10.1101/859983.

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature. 284: 604–607.

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. Molecular Ecology. 25: 2337–2360.

Pearson DL, Hamilton AL, Erwin TL (2011) Recovery plan for the endangered taxonomy profession. BioScience. 61: 58–63.

Peccoud J, Loiseau V, Cordaux R, Gilbert C (2017) Massive horizontal transfer of transposable elements in insects. Proceedings of the National Academy of Sciences of the U.S.A. 114: 4721–4726.

Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, et al. (2019) Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evolutionary Biology. 19: 11.

Phillips HRP, Newbold T, Purvis A (2017) Land-use effects on local biodiversity in tropical forests vary between continents. Biodiversity Conservation. 26: 2251–2270.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, et al. (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science. 344: 1410–1414.

Puttick MN, O'Reilly JE, Pisani D, Donoghue PCJ (2019) Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. Palaeontology. 62: 1–17.

Quadrana L, Etcheverry M, Gilly A, Caillieux E, Madoui MA, et al. (2019) Transposition favors the generation of large effect mutations that may facilitate rapid adaption. Nature Communcations. 10: 1–10.

Quilodrán CS, Ruegg K, Sendell-Price AT, Anderson EC, Coulson T, et al. (2019) The multiple population genetic and demographic routes to islands of genomic divergence. Methods in Ecology and Evolution. 11: 6–21.

Rech GE, Bogaerts-Márquez M, Barrón MG, Merenciano M, et al. (2019) Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. PLOS Genetics. 15: e1007900.

Riedel A, Sagata K, Suhardjono YK, Tänzler R, Balke M (2013) Integrative taxonomy on fast track-towards more sustainability in biodiversity research. Frontiers in Zoology. 10: 10–15.

Riesch R, Muschick M, Lindtke D, Villoutreix R, Comeault AA, et al. (2017) Transitions between phases of genomic differentiation during stick-insect speciation. Nature Ecology and Evolution. 1: 1–13.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: Genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. Molecular Biology and Evolution. 30: 1574–1587.

Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. Journal of Heredity. 97: 581–594.

Sääksjärvi IE, Haataja S, Neuvonen S, Gauld ID, Jussila R, et al. (2004) High local species richness of parasitic wasps (Hymenoptera: Ichneumonidae; Pimplinae and Rhyssinae) from the lowland rainforests of Peruvian Amazonia. Ecological Entomology. 29: 735–743.

Sambatti JBM, Strasburg JL, Ortiz-Barrientos D, Baack EJ, Rieseberg LH (2012) Reconciling extremely strong barriers with high levels of gene exchange in annual sunflowers. Evolution. 66: 1459–1473.

Sangster G, Luksenburg JA (2015) Declining rates of species described per taxonomist: Slowdown of progress or a side-effect of improved quality in taxonomy? Systematic Biology. 64: 144–151.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, et al. (2014) Genomics and the origin of species. Nature Reviews Genetics. 15: 176–192.

Senthilan PR, Helfrich-Förster C (2016) Rhodopsin 7–the unusual rhodopsin in *Drosophila*. PeerJ. 4: e2427.

Serrato-Capuchina A, Matute D (2018) The role of transposable elements in speciation. Genes. 9: 254.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. Ecology Letters. 16: 940–950.

Silvestri F (1903) Classis Diplopoda. Vol. 1a. Anatome: Pars I, Segmenta, Tegumentum, Musculi. Acari, Myriapoda et Scorpiones huscque in Italia reperta. 346: 1–272.

Smith TA, Martin MD, Nguyen M, Mendelson TC (2016) Epigenetic divergence as a potential first step in darter speciation. Molecular Ecology. 25: 1883–1894.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, et al. (2014) Stick insect genomes reveal natural selection’s role in parallel speciation. Science. 344: 738–742.

Strasser K (1965) I diplopodi della fauna siciliana. Atti della Accademia gioenia di Scienze naturali in Catania. 7: 1–35.

Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. PLoS Biology. 13: 1–18.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends in Ecology and Evolution. 18: 70–74.

Troncoso-Palacios J, Gamboa MR De, Langstroth R, Rtiz JC, Labra A (2019) Without a body of evidence and peer review, taxonomic changes in Liolaemidae and Tropiduridae (Squamata) must be rejected. Zookeys. 813: 39–54.

Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017) Taxonomic bias in biodiversity data and societal preferences. Scientific Reports. 7: 1–14.

Upham NS, Esselstyn JA, Jetz W (2019) Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. Plos Biology. 17: e3000494.

Vasques X, Vanel L, Villette G, Cif L (2016) Morphological neuron classification using machine learning. Frontiers in Neuroanatomy. 10: 1–12.

Verhoeff KW (1928) Diplopoda I. In: Bronn's Klassen und Ordnungen des Tierreiches 5 (2.II). Akademischer Verlag, Leipzig. 1–1071 pp.

Vernygora OV, Simões TR, Campbell EO (2020) Evaluating the performance of probabilistic algorithms for phylogenetic analysis of big morphological datasets: A simulation study. Systematic Biology. syaa020: 1–49.

Vilhelmsen L, Turrisi GF (2011) Per arborem ad astra: morphological adaptations to exploiting the woody habitat in the early evolution of Hymenoptera. Arthropod Structure and Development. 40: 2–20.

Wang X, He Z, Shi S, Wu CI (2020) Genes and speciation: is it time to abandon the biological species concept? National Science Review. 7: 1387–1397.

Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, et al. (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science. 309: 1874–1878.

Wheeler QD (2020) A taxonomic renaissance in three acts. Megataxa. 1: 4–8.

Wheeler QD, Knapp S, Stevenson DW, Stevenson J, Blum SD, et al. (2012) Mapping the biosphere: Exploring species to understand the origin, organization and sustainability of biodiversity. Systematics and Biodiversity. 10: 1–20.

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology. 54: 844–851.

Wipfler B, Pohl H, Yavorskaya MI, Beutel RG (2016) A review of methods for analysing insect structures — the role of morphology in the age of phylogenomics. Current Opinion in Insect Science. 18: 60–68.

Wolf JBW, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics. 18: 87–100.

Wolf JBW, Lindell J, Backström N (2010) Speciation genetics: Current status and evolving approaches. Philosophical Transactions of the Royal Society B: Biological Sciences. 365: 1717–1733.

Wong KM, Suchard MA, Huelsenbeck JP (2008) Alignment uncertainty and genomic analysis. Science. 319: 473–476.

Wright SJ (2005) Tropical forests in a changing environment. Trends in Ecology and Evolution. 20: 553–560.

Wu CI (2001) The genic view of the process of speciation. Journal of Evolutionary Biology. 14: 851–865.

Xu G, Ma H, Nei M, Kong H (2009) Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. Proceedings of the National Academy of Sciences of the U.S.A. 106: 835–840.

Yan F, Powell DR, Curtis DJ, Wong NC (2020) From reads to insight: A hitchhiker's guide to ATAC-seq data analysis. Genome Biology. 21: 1–16.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences of the U.S.A. 107: 9264–9269.

Young AD, Gillung JP (2019) Phylogenomics — principles, opportunities and pitfalls of big data phylogenetics. Systematic Entomology. 45: 225–247.

Zamani A, Vahtera V, Sääksjärvi IE, Scherz MD (2020) The omission of critical data in the pursuit of 'revolutionary' methods to accelerate the description of species. Systematic Entomology.

Zhang J, Lindsey ARI, Peters RS, Heraty JM, Hopper KR, et al. (2020) Conflicting signal in transcriptomic markers leads to a poorly resolved backbone phylogeny of chalcidoid wasps. Systematic Entomology. 45: 783–802.

Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR (2016) Genome-wide introgression among distantly related *Heliconius* butterfly species. Genome Biology. 17: 25.

Zheng J, Payne JL, Wagner A (2019) Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. Science. 365: 347–353.

Zhou X, Slone JD, Rokas A, Berger SL, Liebig J, et al. (2012) Phylogenetic and Transcriptomic Analysis of Chemosensory Receptors in a Pair of Divergent Ant Species Reveals Sex-Specific Signatures of Odor Coding. Plos Genetics 8: e1002930.

List of Figures

List of Tables

List of Supplementary Material

Chapter 3

All supplementary files are available under: doi.org/10.1016/j.asd.2018.02.005.

Chapter 5

All supplementary files and figures can be found on the enclosed CD under: ./Chapter_5-Supplementary_Material.

Supplementary Files

S1 Methods

S2 Summaries of FcML Permutations.

S3 Table S1–S3: Sample data, orthology reference data, and PhyloBayes summary statistics.

Supplementary Figures

S1 SymTest results for all datasets.

S2 AliStat results for all datasets.

S3 All trees inferred from the amino acid matrix.

S4 All trees inferred from the BMGE filtered amino acid matrix.

S5 All trees inferred from the nucleotide matrix.

S6 All trees inferred from the BMGE filtered nucleotide matrix.

S7 FcLM of hypotheses derived from ML and MSC analyses of the amino acid matrix.

S8 FcLM permutations of hypotehses derived from ML and MSC analyses of the amino acid matrix.

Chapter 6

The supplementary material is available online under: doi.org/10.1093/gbe/evaa106.

Acknowledgments

Firstly, I want to express my gratitude to my main supervisors, Bernhard Misof and Oliver Niehuis, for giving me the opportunity to do research under their guidance. I am grateful for the exciting and intellectually challenging discussions, the freedom to pursue what I found intriguing, and especially so for their continued support as well as belief in me and my work in times where I had lost confidence.

I also thank my supervisor Thomas Wesener for his guidance and support for anything related to, or not related to, millipedes. Your door was always open, you always had time for me, and without the opportunities you granted me, this thesis would certainly not have been completed.

I thank my office mates and good friends Jeanne Wilbrandt, Panagiotis Provataris, and Alexandros Vasilikopoulos for friendship and for sharing the good and the bad moments. I also thank my co-students and the staff at the ZFMK for making me feel very at home among you. I particularly want to thank Hajo Krammer, Ricarda Wistuba, Trevor Burt, Ameli Kirse, Leon Hilgers, Neha Singh, Ximo Mengual, Malte Petersen, and Tanja Ziesmann.

For their support in the molecular lab, I thank Claudia Etzbauer and Sandra Kukowka and Karin Ulmen for her support in anything related to the SEM. I also thank, in no particular order, Leif Moritz, Thorsten Klug, Philipp Erkeling, Peter Rühr, Helen Read, Paul Richards, Robert Mesibov, Stelios Simaiakis, Axel Schönhofer, Jason Dunlop, Hans Reip, Paul Lee, Karin Voigtländer, Nattarin Wongthamwanich, Peter Kautt, and Christodoulos Makris for their support in collecting specimens for my projects. I also thank Guenther Pass, Daniela Bartel, Nikolaus Szucsich, and Karen Mausemann for granting me access to their unpublished data. Your contributions are invaluable.

I thank Prof. Dr. Lukas Schreiber and Prof. Dr. Thomas Döring for agreeing to participate as members of my examination commission.

Lastly, I want thank my family, especially my parents, Dan Åge and Susanne Øyen, my siblings, and my grandparents, Ruth and Werner Franke, for their unwavering love and support throughout my endeavours, irrespective of direction or reason.