Understanding the evolution of Boraginales: insights from morphology, phylogeny, biogeography and taxonomy

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Maria-Anna Vasile

aus

Maroussi, Griechenland

Bonn, Oktober 2023

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

Gutachter / Betreuer: **Prof. Dr. Maximilian Weigend**

Gutachter: **Prof. Dr. Alexander Suh**

Tag der Promotion: **11. März 2024**
Erscheinungsjahr: **2024**

Contents

Acknowledgements **7**

1. General Introduction **9**

1.1 Foreword 11
1.2 Study group 12
1.2.1 Geographical distribution and systematics of Boraginales 12
1.2.2 Morphology and reproductive biology of Boraginales 13
1.3 Evolutionary diversification of flowers and fruits: a morphological perspective 17
1.4 Towards reconstructing the tree of angiosperms: a molecular perspective 18
1.5 Hypotheses 19
1.6 Research questions 20
1.7 Overview of the dissertation 21
1.8 Contributions to chapters 23
1.9 Literature cited 24

2. Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia* **29**

2.1 Introduction 31
2.2 Materials and methods 33
2.3 Results 35
2.4 Discussion 40
2.5 Acknowledgements 46
2.6 References 47

3. Is it all *Phacelia pinnatifida*? Molecular delimitations and taxonomic revision of *Phacelia* sect. *Glandulosae* in South America based on ddRADseq **51**

3.1 Introduction 53
3.2 Material and Methods 58
3.3 Results 61
3.4 Discussion 65
3.5 Conclusions 70
3.6 Taxonomic Treatment 70
3.7 Acknowledgments 76
3.8 Literature Cited 77

4. Flower and fruit development of three species from Hydrophylleae sheds new light on flower evolution in Hydrophyllaceae **83**

4.1 Introduction 85
4.2 Material and Methods 86
4.3 Results 87
4.4 Discussion 98
4.5. Acknowledgments 100
4.6 Literature Cited 100

5. The odd one out: a comparison of flower and fruit development in holoparasitic *Pholisma arenarium* (Lennoaceae, Boraginales) to that in closely related Ehretiaceae. **103**

5.1 Introduction 105
5.2 Material and Methods 107
5.3 Results 107
5.4 Discussion 119
5.5 Conclusions 121
5.6 Acknowledgments 121
5.7. Literature Cited 121

6. Unravelling Boraginales: an updated phylogeny based on the Angiosperms353 probe set.....123
6.1 Introduction..... 125
6.2 Material and Methods 130
6.3 Results..... 133
6.4 Discussion 139
6.5 Conclusions..... 144
6.6 Acknowledgments..... 144
6.7 Literature cited 144

7. General conclusions 151
7.1. Insights into the evolutionary history of the basal grade of Boraginales II 153
7.1.1. Phylogenetic relationships and historical biogeography 154
7.1.2. Species delimitation in absence of distinct morphological characters..... 155
7.2. Underlying processes of fruit evolution in Boraginales II..... 156
7.3. Evolution of Boraginales based on an updated phylogeny. 158
7.3.1. Phylogenetic relationships and systematics..... 158
7.3.2. Fruit evolution 159
7.4. Conclusion and possible directions for further studies based on this dissertation..... 161
7.5 Literature cited 163

Appendix A 168
Appendix B 186
Appendix C 252
Appendix D 254

Acknowledgements

First and foremost, I would to like to express my sincere gratitude to my supervisor and mentor Maximilian Weigend for his faith in me; for all the opportunities that he has given me since my Master's studies and until this moment; for introducing me to such an exciting study group, the Boraginales and for being supportive and patient throughout my doctoral journey. His calm and generous guidance all these years helped me grow as a scientist.

I would also like to thank Alexander Suh for agreeing to be my second supervisor. I am grateful for his help and contribution.

Julius Jeiter and Tim Böhnert deserve more thanks than I can ever give them. They have been wonderful friends and colleagues and I deeply appreciate all the time that they devoted to me. They taught me the ropes in lab, helped me to run analyses, discussed with me for hours about my projects and science in general, read and corrected my manuscripts, advised me whenever I needed it and helped me with my teaching obligations. Julius transmitted to me his passion about morphology and showed me its importance and Tim his passion for phylogenetics and bioinformatics. I feel grateful for the powerful tools and knowledge you have provided me with.

Federico Luebert may have left Nees-Institute at the beginning of my PhD, however he had a tremendous influence on me. I would like to thank him for his positivity and encouragement, for helping me understand phylogenetic and biogeographic analyses, but above all for introducing me to the scientific way of thinking.

This dissertation would not be possible without the help of our technical assistants and therefore, I would like to thank Nicole Schmandt, Yaron Malkowsky, Thomas Joßberger and Claudia Schütte for always being there for me. Each one of them contributed in her/his way to the success of this work.

I would also like to thank all the present and former members of Nees-Institute: the always supportive Elke Friesecke, Jens Mutke, Florian Losch, Jonathan Ruhm, Julia Gravendyck, Stefan Abrahamczyk, Hans Jürgen Einsikat, Dietmar Quandt, Julia Bechteler, Khaoula Ferchichi and her dog Diva who were the best office-neighbors I could have for the last 3 years, Karola Maul, Jovani Pereira, Paloma Morales, Lara Iaboli, Rosa Mbulu, Annette Scheersoi and all the members of her working group. We shared lunch and coffee breaks, had fruitful discussions, enjoyed barbeques and the other festivities of our institute, went to conferences together, and helped each other when necessary. Thank you all for making Nees-Institute a lively and warm place to work.

Plant material for this study was partially collected from Bonn University Botanic Gardens and I would like to express my thanks to the staff and the scientific curator Cornelia Löhne for taking care of the collections and for always giving me a helping hand.

I am deeply indebted to all the Herbaria and collectors of specimens cited in this research, but I am particularly grateful to Nataly O'Leary, Bridget Lee and Dena Grossenbacher for their prompt response and willingness to provide us with hard-to-find plant material. Furthermore, I deeply appreciate all the long-term efforts for digitizing collections and generally making metadata accessible for everyone. This research would not be possible without databases such as those of Royal Botanic Gardens Kew, JSTOR, GBIF or Github.

I am especially grateful to Thomas Bartolomaeus and Alexander Blanke for permitting the conduction of the micro-computed tomography at the Institute of Evolutionary Biology and Ecology at the

University of Bonn and Alexander Ziegler and Dagmar Wenzel for their technical support. Funding for the Skyscan 1272 μCT was provided by the Deutsche Forschungsgemeinschaft (INST 217/849-1 FUGG).

I wish to express my gratitude also to Benjamin Wipfler for allowing the conduction of the micro-computed tomography at the Museum König and Juliane Vehof for her technical support.

I would also like to thank the students Emma Lewin, Laura Eckert, Andrew Marcil (who is now a good friend), Alexander Menidzabal, Marlon Martin and Esma Colak for all the data they produced for me. Even though most of them could not be included in this thesis, they will be valuable to future publications. I hope that you enjoyed the time that we spent together in the SEM lab as much as I did.

All the co-authors of the individual chapters have already been listed above but I would like to thank them once again for their valuable contributions to this dissertation and our fruitful collaborations. My sincere thanks also go to all the editors and reviewers for their meticulous efforts in critically reviewing and enhancing my manuscripts during the peer-review process that my published chapters underwent.

I often think how much my life has changed for the better since I came to Bonn and I know that this might not have been possible without the Bodossaki Foundation granting me a scholarship for the Plant Science Master's program in Bonn. I will always be grateful for this lifechanging opportunity.

There are so many people not related to University of Bonn, or at least not anymore, who I am deeply grateful to for their infinite support and positivity, but also for all the beautiful moments that we had together the past years. I thank my “Plant Sciences Friends” Tyler Hopewell, Jordy Perez, Gulmairam Sultanova and Srijana Raj, especially for our relaxing plant science summer meetings in Berlin. I would also like to thank my “academic sisters” Elena Vasilopoulou, Elizabeth Ioannou, Anastasia Apodiakou, Andressa Cabral and Berivan Mansurolgou for sharing this very long journey from the Bachelor’s or the Master’s until the PhD.

There are many friends outside academia, currently in Greece and in the Netherlands, who I would like to thank for always being there for me and for making the past few years lively. I cannot name them all but I explicitly want to thank Christina Goudroupi for her constant and unconditional support.

It has been already six years since I left Greece and came to Germany to continue my studies and I know that the time I can devote to my family is never enough anymore. Therefore, I would like to thank my extended family, for always encouraging me to follow my dreams and my wills. They all mean a lot to me, but most of all I would to thank my mother Despina for her unwavering faith in me and for planting inside me the seed of loving plants since I was a child.

I am also grateful to Foula and Yorgos who were always like parents to me. Thank you for your endless support all these years.

My stay in Bonn would not be the same without my lovely dog Gigi. She was always welcoming me happily no matter how tired or moody I was returning back home. She is also the main reason I kept my sanity during the Corona lockdowns. Thank you Gigi!

The importance of the last person I would like to thank cannot be really described with words. Indisputably, this journey would not be the same without the love, encouragement and selflessness of my partner in life. He believed in me when I doubted myself, he gave me energy when I had none, he made me laugh even on the most difficult days, he always listened to me and always advised me. Thank you a million Vasilis for being by my side at every step all these years that we have been together.

General Introduction



Chapter 1

Alkanna tinctoria (Boraginaceae) photographed at mountain Ymittos, Athens, Greece, long ago, when I could have never imagined that one day I will study these beautiful plants.

"We must consider the distinctive characters and the general nature of plants from the point of view of their morphology, their behavior under external conditions, their mode of generation, and the whole course of their life."

Theophrastus, περὶ φυτῶν ἱστορία
(De historia plantarum)

1. General introduction

1.1 Foreword

My interest in plant morphology and phylogenetics arose with a study regarding gynoecium and fruit development in Hydrophyllaceae and their ovule/seed number evolution, while I was still a Master's student. I have soon become enthusiastic about morphology and wanted to learn and discover more about fruit evolution of Boraginales, a plant order with such divergent fruit morphology.

While I was working on Hydrophyllaceae seed number evolution, I needed a modern phylogeny for the family in order to reconstruct the ancestral character state for this trait. This is how the idea of chapter 2 came up. Chapter 2 is tightly connected with my research regarding seed number evolution in Hydrophyllaceae (Vasile & al., 2021; not part of this thesis) and was even published beforehand. On the one hand, chapter 2 aims to provide more insights into the phylogenetic relationships of the basal grade of Boraginales II (Hydrophyllaceae and Namaceae) by using plastid and nuclear markers and open the way for morphological studies under an up-to-date phylogenetic context. On the other hand, it specifically addresses for the first time the phylogenetic position of the South American representatives of these two families and subsequently their historical biogeography. The focus was set on *Phacelia* Juss., the only genus from Hydrophyllaceae with S. American representatives, and *Wigandia* Kunth. from Namaceae. Although most of the questions were successfully answered, the interspecific relationships of the South American species of *Phacelia* sect. *Glandulosae* remained unresolved. This topological uncertainty was the trigger for more research and created the idea of Chapter 3, which aims at resolving the phylogenetic relationships of *P.* sect. *Glandulosae* in South America, this time with a next generation sequencing approach (ddRADseq). This clade presents a lot of phenotypic variation but lacks distinct morphological traits. Therefore, we paired a genomic and a classical taxonomic approach for species delimitation.

Since the beginning of my investigations of fruit and seed number evolution in Hydrophyllaceae, I realized that understanding internal ovary architecture and organization is a prerequisite for understanding fruit type transitions and the underlying processes which resulted in structural innovations, not only within this family, but across the order. Modifications of the internal ovary architecture as a driving force for fruit evolution in Boraginales II are consequently the central point in two of my chapters (4 and 5). Chapter 4 deals with the clade Hydrophylleae from the Hydrophyllaceae which presents interesting modifications in placentation, while chapter 5 addresses the aberrant fruit morphology of the parasitic Lennoaceae compared to that of two closely related species of Ehretiaceae. In both chapters, the complete floral development is studied in order to explore the synorganization of the gynoecium and the other organs during development and their integration into the floral architecture.

I might have started small, by looking at the phylogenetic relationships and the fruit evolution in some taxa of Boraginales II clade, but I was interested since the beginning in the greater picture and the history of fruit evolution across the whole order. There are many morphological data sets available across Boraginales, collected either before my times in the Nees Institute, or during my PhD with the help of many bachelor and master's students. However, we cannot make the most of these data without an updated and modern phylogeny. Even though recent studies have set the fundamentals for familial (Weigend & al., 2014; Luebert & al., 2016) and in some cases infrafamilial classifications (Luebert & al., 2011; Chacón & al., 2016), many relationships remain ambiguous or unresolved. The last research

chapter of my thesis, chapter 6, is the first big step towards a comprehensive understanding of the evolution of Boraginales. For the first time in Boraginales, a phylogenomic approach, namely target capture sequencing with the Angiosperms353 bait kit, was used in order to infer the phylogenetic relationships of the order.

The study of morphological characters in a phylogenetic context is of vital importance to better understand angiosperms evolution. However, there are two main parameters which often make this type of research difficult: (i) the lack of well resolved phylogenies holds us back from defining species limits and therefore from understanding trait evolution in these species; (ii) the lack of suitable, high quality developmental and morphological data in many species across angiosperms obstructs our understanding of processes involved in character evolution and prevents consistent use of data in scoring floral traits and reconstructing evolution. These two tightly correlated parameters were the main drivers for formulating hypotheses for my study group and put the focus of this thesis on two basic directions: molecular phylogenetics and evolutionary morphology of Boraginales.

1.2 Study group

1.2.1 Geographical distribution and systematics of Boraginales

Boraginales is a cosmopolitan medium-size order nested within Lamiidae and comprising 11 families with about 125 genera and 2700 species (Luebert & al., 2016). They consist of herbs, shrubs, trees and rarely lianas, often covered with hirsute or hispid indumentum as their name implies (*Borago* L. possibly comes from the Latin word burra, meaning hairy garment; Simpson, 2010). They are predominantly distributed in seasonally arid habitats across temperate and tropical regions (Weigend & al., 2016). The monogeneric Codonaceae and Wellstediaceae are plants from deserts and semi-deserts of southwestern Africa (Codonaceae) or of both southwestern and northeastern Africa (Wellstediaceae; Kadereit & Bittrich, 2016). Boraginaceae as the largest family can be mostly found in (warm-) temperate and Mediterranean climates, while in the tropics they are restricted to higher elevations (Weigend & al., 2016). Hydrophyllaceae and Namaceae, are mainly found in western North America, with few species in Central America and western South America and a handful of *Nama* L. species in the Caribbean and Hawaiian Islands (Luebert & al., 2016). Heliotropiaceae, Ehretiaceae and Cordiaceae are of worldwide distribution, but mainly found in tropics and subtropics (Kadereit & Bittrich, 2016). Lennoaceae is a small parasitic family, comprising only four species distributed in southwestern North America, Mexico, and northwestern South America (Yatskievych & Mason, 1986). The monogeneric Hoplestigmataceae has two species endemic to central and western Africa and the monotypic Coldeniaceae are widespread in the Old World tropics and introduced elsewhere (Luebert & al., 2016). Boraginales may have originated in West Gondwana during the Early Cretaceous ca. 109 Mya, and an initial West Gondwana vicariance event is a likely explanation for the origin of the two major clades (Boraginales I and II) which diverged during the Late Cretaceous (Luebert & al., 2017). However, alternative explanations for the origin of Boraginales cannot be ruled out and these results should be cautiously interpreted due to uncertainties associated with the sister relationship of Boraginales (Refulio-Rodriguez & Olmstead, 2014; Weigend & al., 2014; Stull & al., 2015) and with ancestral area reconstruction (Luebert & al., 2017).

Boraginales has been only recently recognized as an order by APG IV (2016), but taxonomists have not yet agreed on a stable finer subdivision of this plant group. Their classification has undergone dramatic changes, especially since the establishment of molecular phylogenetic analyses (Luebert & al., 2016).

They have been traditionally treated as one family (Boraginaceae s.l.) containing four subfamilies (Boraginoideae, Cordioideae, Ehretioideae, and Heliotropioideae) by various authors (De Candolle, 1845, 1846; Bentham & Hooker, 1876; Gürke, 1893), a subdivision that was followed in many subsequent studies. In the late 80s, Takhtajan (1987) subdivided Boraginales in seven families (Boraginaceae, Cordiaceae, Ehretiaceae, Hoplestigmataceae, Hydrophyllaceae, Lennoaceae, and Wellstediaceae), but he later revised this classification and included Boraginaceae with five subfamilies (Boraginoideae, Cordioideae, Ehretioideae, Heliotropioideae, and Wellstedioideae), Hoplestigmataceae, Hydrophyllaceae, Lennoaceae, and Tetrachondraceae (Takhtajan, 1997). Gradually, with the use of molecular data, all the unrelated taxa formerly included in Boraginales (e.g. Tetrachondraceae, *Hydrolea* L., *Pteleocarpa* Oliv.) found their place on the tree of life (Soltis & al., 2000; Wagstaff & al., 2000; Mabberley, 2008). Since then, many authors, who were consistently working with Boraginales, elevated the subfamilies to family rank (i.e. Boraginaceae s.str., Cordiaceae, Ehretiaceae, and Heliotropiaceae; Gottschling & al., 2004; Luebert & Wen, 2008; Mansion & al., 2009; Gottschling & al., 2014a). The use of chloroplast molecular data by Weigend & al. (2014), revealed that Boraginales consist of two major clades, namely Boraginales I and II. According to that study, Boraginales I consists of Codonaceae, Wellstediaceae and Boraginaceae s. str., and Boraginales II includes Hydrophyllaceae I (Hydrophyllaceae s.str.), Hydrophyllaceae II (Nameae Choisy), Heliotropiaceae, Ehretiaceae, Lennoaceae (an exclusive relationship of Lennoaceae with core-Ehretiaceae was supported) and Cordiaceae including *Coldenia* L. and *Hoplestigma* Pierre. Following this study, Luebert & al. (2016) provided the most recent familial classification of the order, which is compatible with phylogenetic data, but also adequately reflects the morphological heterogeneity of the group (see figure 1.1). As a result, they further recognized Hydrophyllaceae II as Namaceae, *Coldenia* L. as Coldeniaceae and *Hoplestigma* Pierre as Hoplestigmataceae. Overall, despite the disagreements on the classification of the order, the morphological — mostly fruit-centric—heterogeneity of the group, was always a central point of research in Boraginales (Chance & Bacon, 1984; Chuang & Constance, 1992; Hilger, 1992; Di Fulvio & al., 1997; Diane & al., 2002; Gottschling & al., 2005, 2014b; Hilger, 2014; Luebert & al., 2016; Jeiter & al., 2018; Otero & al., 2019; Heigl & al., 2020; Ebadi & Nikzat, 2021) and often provides strong evidence for its finer subdivision into several families. However, when it comes to genera and species delimitation, the lack of distinct traits is sometimes a prominent problem and significant obstacle for the conduction of systematic, evolutionary and conservation studies. This thesis tries to address some of the ambiguous inter- and infrafamilial (chapter 2) or interspecific (chapter 3) relationships, but also tries to provide an up-to-date phylogenetic framework for the whole order with a fresh approach (chapter 6).

1.2.2 Morphology and reproductive biology of Boraginales

Boraginales are a morphologically very diverse group of angiosperms in terms of flower architecture and fruit morphology (Fig. 1.1; Weigend & al., 2014; Jeiter, 2020), in spite of their generally conserved floral organization (i.e. the number and insertion of floral organs; Jeiter, 2020). Numerous morphological characters of taxonomic importance are widespread in Boraginales and have been used to characterize individual subgroups, but characters uniting the entire order or its two main subgroups Boraginales I and Boraginales II have not been identified so far (Kadereit & Bittrich, 2016; Luebert & al., 2016).

Flower organization of Boraginales exhibits the general trends of the asterids, namely mostly tetracyclic, sympetalous, and pentamerous flowers, predominantly with radial symmetry (Weigend & al., 2014),

usually arranged in scorpioid inflorescences (Buys & Hilger, 2003). What is really variable in Boraginales is the floral architecture (i.e. the result of differential growth rates, synorganisation, organ fusion and modification; Jeiter, 2020) largely due to the morphologically diversity in their gynoecial nectaries and stamen-corolla tubes. The modifications of the stamen-corolla tube in Hydrophyllaceae (Jeiter & Weigend, 2018) as well as the modifications and septa of the remotely related *Codon* L. (Codonaceae; Jeiter & al., 2016) compartmentalize the internal floral space and therefore each compartment encloses one nectary gland, leading in both cases to a revolver architecture (Jeiter, 2020). In Boraginaceae, the stamen-corolla tube modifications (faucal and basal scales), greatly impact the floral architecture, but they have a totally different function. Faucal scales sometimes contribute to the formation of anther cones, or they simply narrow the entrance to the floral reward; conversely, the function of the basal scales remains unclear, but most probably they act as a protective layer for the ovary and the nectary (Jeiter, 2020; Jeiter & al., 2020). Presence/absence of stamen-corolla tube modifications (scales) and their pilosity are important taxonomic characters (Walden & Patterson, 2012; Weigend & al., 2014), but the functional integration of these structures into the floral architecture has not yet been studied in depth, despite its potentially high significance for understanding evolution of flower architecture in correlation to pollination syndromes of Boraginales. The dominant breeding system across the order is zoophilous allogamy, with insect pollination being the most common syndrome and nectar being the primary floral reward (Kadereit & Bittrich, 2016). Bird pollination has been suggested for the genera *Wigandia* (Namaceae; Hofmann & al., 2016) and *Cordia* L. (Cordiaceae; Gottschling & al., 2016). Selfing is also common for most of the families, and many other phenomena such as cleistogamy, protandry, transitions between homostyly and distyly, dioecy and buzz-pollination have also been recorded (Kadereit & Bittrich, 2016). We are far from understanding the evolutionary consequences of plant-pollinator interactions on the flower morphology of Boraginales, but Jeiter and colleagues, recently revisited the flowers of this group with a fresh look and new approaches, emphasizing how different perspectives can provide new insights into evolution (Jeiter & al., 2016, 2018, 2020; Jeiter & Weigend, 2018; Jeiter, 2020), opening the way for more research into the same direction.

The evidently diverse fruit morphology of the order on the other hand, has drawn a lot more attention (Hilger, 1984, 1987, 1992, 2014; Di Fulvio & al., 1997, 1999; Hofmann, 1999; Gottschling & Hilger, 2004; Gottschling & al., 2014b; Jeiter & al., 2018; Heigl & al., 2020; Vasile & al., 2021), in a try to explain the multiple fruit type transitions at all different taxonomic ranks. Both major clades, Boraginales I and II have a basal grade of two small to medium-sized families with capsules (Boraginales I: Codonaceae, Wellstediaceae; Boraginales II: Hydrophyllaceae, Namaceae), while in the remaining families of both clades can be found drupes and mericarpids (nutlets) of either eremocarpic (Boraginaceae) or schizocarpic (Boraginales II) origin (Luebert & al., 2016; Jeiter & al., 2018). Especially for the two largest families, Boraginaceae (~1700 species in ~97 genera; Chacón & al., 2016) and Heliotropiaceae (~450 species in four genera; Diane & al., 2016), gynoecium and fruit morphology, and in the case of Boraginaceae also nutlet ornamentation, have played a significant role for infrafamilial classification. The complexity of fruit architecture arises from the development of various secondary septa subdividing the two carpels of the gynoecium (Hilger, 1992, 2014; Gottschling, 2004), resulting in variable types of diaspores and dozens of terms trying to describe this diversity. In Box 1, a glossary summarizes and, when necessary, redefines the most important terms, many of which will be found across this thesis. The different levels of seed and pericarp integration (see Box 1) heavily impact diaspore function and consequently result in various dispersal modes. In families with capsular fruits (Codonaceae, Wellstediaceae, Hydrophyllaceae, Namaceae and partially Lennoaceae) the infructescences remain attached to the plants and the seeds, or sometimes part of the plants that break off, are

gradually dispersed by wind and gravity (Kadereit & Bittrich, 2016). Zoochory is the main mechanism in Boraginaceae (Weigend & al., 2016), but also anemochory, hydrochory and myrmecochory are important in both Boraginaceae and Heliotropiaceae (Diane & al., 2016; Weigend & al., 2016). For the drupaceous fruits of Ehretiaceae, Cordiaceae and Heliotropiaceae, endozoochory has been suggested (Diane & al., 2016; Gottschling & al., 2016).

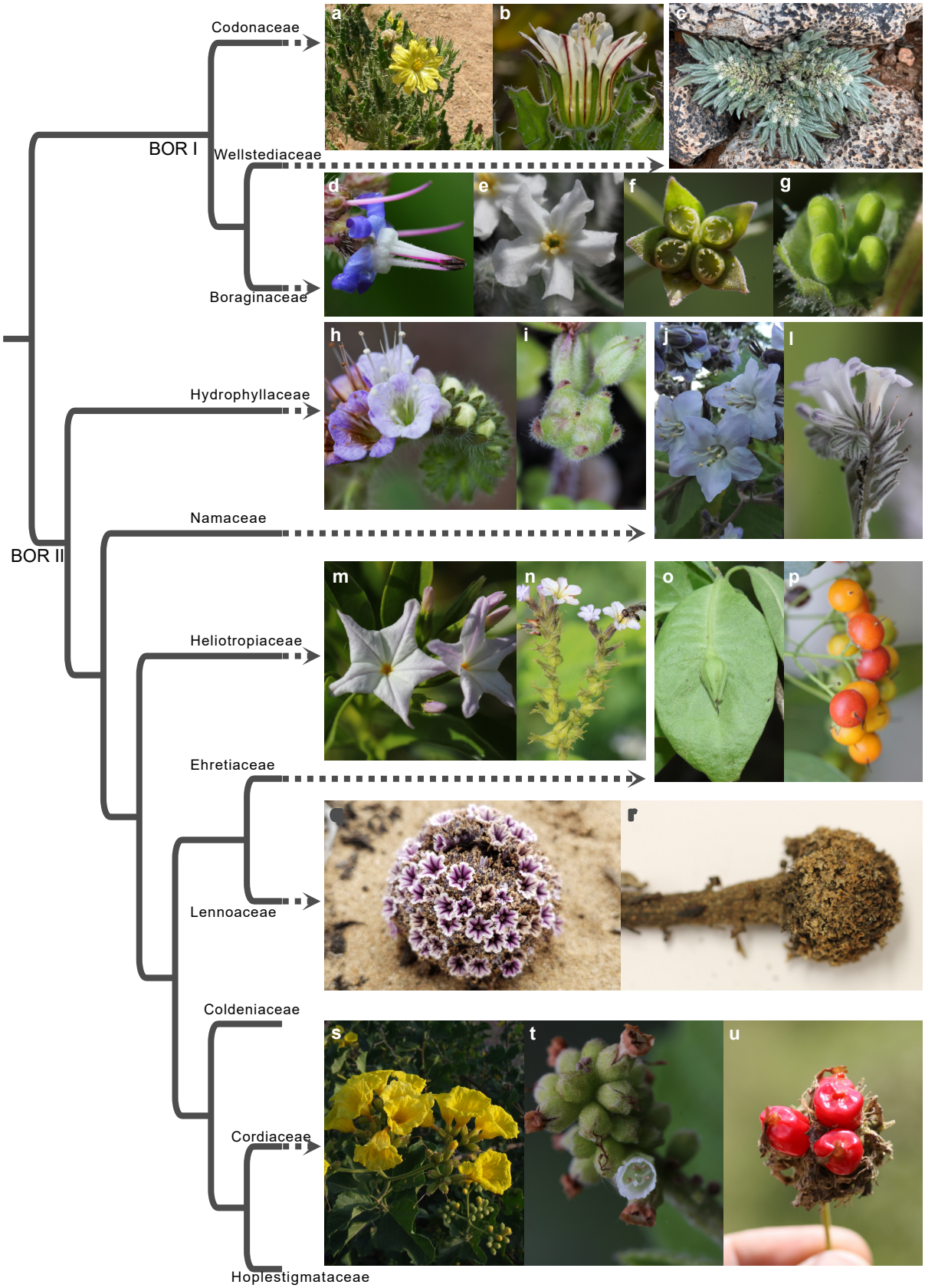


Figure 1.1 Phylogenetic relationships and morphological diversity of Boraginales. On the left, phylogenetic tree showing familial classification of Boraginales according to Luebert et al. 2016. On the right, flowers and fruits showing part of diversity across the order: a, *Codon schenckii* Schinz (photo: T. Camenzind, Namibia); b, *Codon royenii* L., photo: J. Jeiter); c, *Wellstedia cf. dinteri* Pilg (photo and collection: Q. Dietmar & J. Bechteler, SFB963 Namibia); d, *Trachystemon orientalis* (L.) D.Don (GE-0-BONN-33857); e, *Oreocarya spiculifera* Piper (xx-0-BONN-15133); f, *Omphalodes nitida* (Willd.) Hoffmanns. & Link (xx-0-BONN- 46154); g, *Nonea lutea* (Desr.) DC. (UA-0-BONN-23732); h, *Phacelia secunda* J.F.Gmel. (CL-0-BONN-38409); i, *Romanzoffia californica* Greene (xx-0-BONN-14661); j, *Wigandia cf. caracasana* (photo: E. Jiménez, Costa Rica); l, *Eriodictyon crassifolium* Benth. (US-0-BONN-36707); m, *Ixorhea tschudiana* Fenzl (AR-0-BONN-39837); n, *Heliotropium corymbosum* Ruiz & Pav. (PE-0-BONN-33458); o, *Keraunea brasiliensis* Cheek & Simão-Bianchini (photo and collection: D. Cardoso, 4902 Brazil); p, *Bourreria ovata* Miers (xx-0-BONN- 35343); q, inflorescence of *Pholisma arenarium* Nutt. (photo and collection: B. Lee & D. Grossenbacher 1, California USA); r, infructenscence of *Pholisma arenarium* (collection: B. Lee & D. Grossenbacher 1, California USA); s, *Cordia lutea* Lam (photo: M. Weigend, Peru); t, *Varronia cylindristachya* Ruiz & Pav. (PE-0-BONN- 36749); u, *Varronia macrocephala* Desv. (photo and collection: M. Weigend, MW9643, Peru).

Box 1

Glossary

Endomericarp: part of a fruit (usually drupe, schizocarp, fleshy capsule) where the lignified endocarp encloses and protects the seeds of a complete carpel of a syncarpous gynoecium.

Endomericarpid: part of a fruit (usually drupe, schizocarp, fleshy capsule) where the lignified endocarp encloses and protects the seeds of part of carpel of a syncarpous gynoecium.

Eremocarp: one-seeded mericarpid 'nutlet' derived from a syncarpous ovary via ab initio separate development (Hilger, 2014; Luebert & al., 2016).

Gynobase: style and basal part of the gynoecium where the mericarpids are attached (columella) and that remains on the mother plant at fruit dispersal (Hilger 2014, Jeiter & al., 2018).

Mericarp: dispersal unit derived from a complete carpel of a syncarpous gynoecium.

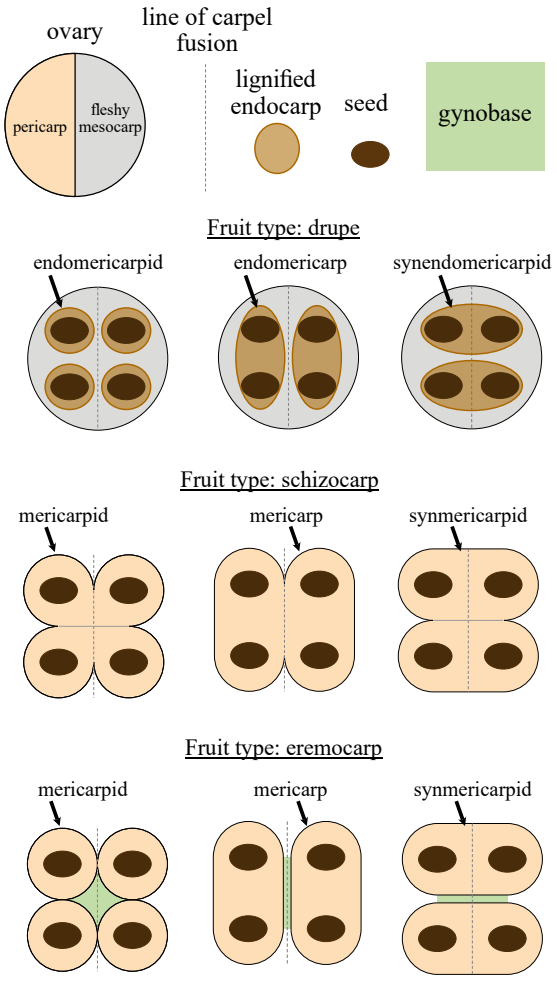
Mericarpid: dispersal unit derived from part of a carpel of a syncarpous gynoecium.

Schizocarp: derived from a two or more loculed syncarpous ovary in which the locules separate at maturity (fruit type; Simpson 2010).

Synendomericarpid: part of a fruit (usually drupe, schizocarp, fleshy capsule) where the lignified endocarp encloses and protects the seeds of part of two (or more) different carpels of a syncarpous gynoecium.

Synmericarpid: dispersal unit derived from parts of two (or more) different carpels of a syncarpous gynoecium.

Schematic drawings & terminology



Hitherto, the external morphology has drawn a lot of attention and the focus was placed on the transition from dehiscent capsules to indehiscent mericarps and drupes. Several authors highlighted that septation

is a key feature involved in the complex fruit architecture (Hilger, 1985, 1987, 1992; Hofmann, 1999; Gottschling & al., 2014b) and that one of the underlying mechanisms for the transition from dehiscent capsules to indehiscent mericarps and drupes is the reduction of ovule number from many to four (Weigend & al., 2014; Luebert & al., 2016). However, none of these or other studies were explicitly addressed the internal ovary architecture. Lately, Vasile & al. (2021) investigated these overlooked aspects using Hydrophyllaceae as a model family, due to the fact that it is the only family among Boraginales which presents a great variation in ovule/seed number, namely from four to ~220. This study revealed independently evolved developmental mechanisms (i.e. multiplication of ovule layers, multiplication of ovule rows per placenta, locule expansion) crucial for fruit evolution of Hydrophyllaceae. At the same time, the most valuable insight was that reduction and multiplication of seed number is reversible as long as fruits remain dehiscent, and that the most probable process involved to seed number reduction is heterochrony (paedomorphosis). The morphological studies of this thesis (chapters 4 and 5) are a continuation of this work, with carefully selected sampling, in order to address the role of different structures and processes in fruit evolution. Investigations of Hydrophylleae from the Hydrophyllaceae, a clade with very diverse placentation and transitions between pluri- and four-ovulate gynoecia (chapter 4) as well as of the aberrant fruit morphology of Lennoceae, likely driven from the evolution of parasitism, in comparison to that of the closely related Ehretiaceae (chapter 5), are expected to reveal crucial insights into fruit evolution of Boraginales.

1.3 Evolutionary diversification of flowers and fruits: a morphological perspective

Morphology, one of the oldest biological disciplines, has been traditionally used to describe and analyze organismal forms, in order to understand how the overwhelming phenotypic diversity has evolved (Wanninger, 2015). Morphology was the main, if not the only tool of scientists and naturalists for describing, classifying and understanding diversity before the advent of molecular systematics and developmental and molecular genetics. From the 1980s onwards, the integration of molecular techniques into systematics, phylogenetics and developmental biology led to a neglect of morphology (Wanninger, 2015), but not for long. A renaissance of morphology cropped up under the term MorphoEvoDevo, a discipline investigating how structures or process combinations change during plant development and evolution, ontogeny and phylogeny (Sattler & Rutishauser, 2023).

This thesis implements aspects of classical morphology and MorphoEvoDevo by looking at flowers and fruits not purely as structures consisting of strictly defined organs, but also as processes and as complex three-dimensional functional units. This approach becomes more comprehensible by contemplating the complimentary relationship of classical, continuum and process morphology. Classical morphology states that there are only three mutually exclusive organ categories in vascular plants, i.e. root, stem (caulome), and leaf (phyllome); continuum morphology recognizes a continuum between these organ categories, meaning that membership in a category and hence homology is a matter of degree; process morphology supersedes structure/process dualism and therefore treats structures as processes or process combinations and not as structures which have processes (Sattler & Rutishauser, 2023). Chapters 4 and 5 are dedicated to ontogeny and morphology of specific Boraginales lineages and look into structures as processes of growth, decay, differentiation and dedifferentiation both externally and superficially by using scanning electron microscopy (SEM), but also internally and as three-dimensional functional units by using micro-computed tomography (μCT).

The key to an understanding of flower and fruit diversification, is therefore not only to detect structural innovations, but also to identify the processes involved. One of the most important processes for attaining new structural levels is synorganization between floral organs (Endress, 2011). Numerous structural innovations arose in this way, many of which were key innovations for angiosperm diversification (e.g. the advent of syncarpy and an intragynoecial compitum; Endress, 2011). First and foremost, ovules and a leaflike structure were successfully and probably rapidly (as we still lack intermediate fossils) synorganized in an earlier phase of evolutionary history (Endress, 2019). This synorganization, the generation of carpel or gynoecium, came to stay in angiosperm evolution. In a second step, other specializations or modifications of ovary structure such as carpel bulging (Endress, 2011), loss of dehiscence (Beaulieu & Donoghue, 2013) or changes in seed number (Vasile & al., 2021) are crucial for transitions between fruit types. Furthermore, it has been shown that the basic structure of the pericarp is retained in cases of transformation of an apocarpous gynoecium/fruit into syncarpous one and between the different fruit types (Bobrov & Romanov, 2019), however, the functional integration of seed and pericarp can be complex, resulting in a range of different diaspores. The directionality of evolutionary transformation sometimes varies, but there are also transformations which likely lead to evolutionary dead ends.

For the interpretation of the phenotype and in order to gain insights into transitional states and the homology of structures, except for the synorganization of floral organs, there are other fundamental processes which we need to observe and comprehend. According to Zimmerman (1930) these are: heteromorphy (i.e. exhibiting diversity of forms), heterotopy (i.e. positional shifting of structures), and heterochrony (i.e. evolutionary change in the rate or timing of development). By documenting the development of the structures of interest in two of the chapters of this thesis, the focus is primarily on the detection of these processes. Secondarily, understanding the developmental trajectories of the morphologies here investigated prior to defining characters for ancestral state reconstructions of their evolution in future studies, is a crucial, if not mandatory, step. As highlighted by Jeiter & Smets (2023), the resemblance or sameness of characters at maturity can be misleading, and very often similar, but homoplasious characters, initially thought to be the same, turn out to follow parallel and/or independent developmental and evolutionary trajectories.

1.4 Towards reconstructing the tree of angiosperms: a molecular perspective

Reconstructing and comprehending the tree of life remains one of the ultimate challenges in science (Hinchliff & al., 2015). This challenge is simultaneously a necessity if we want to navigate the evolution as well as the properties of life on Earth on a central biological roadmap (Baker & al., 2022). Data gaps need to be filled and blur points need to be pulled into focus regarding the phylogenetic relationships of angiosperms, as they are potentially of enormous importance for dealing with global challenges such as climate change, biodiversity crisis and human well-being (Antonelli & al., 2020).

DNA, despite its simplicity (i.e. consists of only four nucleotides/character states), is undoubtedly a powerful tool and has been proven to be superior to morphological datasets for phylogenetic reconstructions (Doyle & Endress, 2000; Jeiter & Smets, 2023). DNA cannot lie and encloses all the evolutionary history of organisms on earth. However, this does not necessarily mean that our phylogenetic reconstructions cannot lie either, as they are prone to numerous errors. Beginning from the laboratory work, and moving to data processing, the chances of introducing errors at each step are high. Alignment uncertainty can affect phylogeny estimation in various ways (Wong & al., 2008; Redeling & Suchard,

2009; Vialle & al., 2018) or in case of working with methods identifying single nucleotide polymorphisms (SNPs), artifacts during locus reconstruction can result in erroneous SNP identification (O'Leary & al., 2018). Selecting the correct genes is another crucial step. During filtering of the aligned putative orthologues, contaminants, paralogues or xenologues may be erroneously identified as orthologues, leading to errors of the phylogenetics inference accuracy, for example by yielding longer branches, biased model parameters or topological differences (Kapli & al., 2020). Another essential factor is the genealogical heterogeneity across genes. Different genes might have different histories, and thus, different gene trees might be due to divergent biological processes, such as gene duplication and loss, horizontal gene transfer or polymorphism in ancestral species which leads to the phenomenon of incomplete lineage sorting (ILS; Maddison, 1997; Nichols, 2001). Therefore, by concatenating all genes into a supermatrix we introduce the assumption that a single gene tree underlies all genes and corresponds to the species tree (Kapli & al., 2020). Last but not least, even researcher's bias (presenting/interpreting part of carefully selected data) or ignorance can result to working with inappropriate data for answering a given research question and to misinterpretations.

Despite all these probabilities of miscalculating a phylogeny, molecular data are revolutionizing our perception of systematics. Phylogenetic reconstructions based on molecular data have gradually provided a quite stable backbone for the angiosperm tree of life and a relatively stable classification system up to the family level (Chase & Reveal, 2009; APG IV, 2016; Li & al., 2019). Everything regarding molecular systematics, from sequencing methods, to evolutionary models, alignment tools and phylogenetic reconstruction methods, is being constantly advanced. Nevertheless, all show that we still have a long way to go and a lot to discover until we reveal the tree of life. Scientists should remind themselves that DNA is not a panacea and that even reliable phylogenies will be pointless if we are ignorant of the phenotype and function of the taxa. Interdisciplinary research, despite its complexity, is urgently needed because interest in evolutionary biology derives from the morphology and the function of the taxa combined with knowledge of their relationships. The present thesis, in a try to understand some aspects of the evolution of Boraginales and, at a later stage to contribute to reconstructing the tree of life and its evolution, pairs the disciplines of plant molecular phylogenetics, phylogenomics, systematics and morphology and formulates the following hypotheses.

1.5 Hypotheses

Boraginales is clearly a very diverse group, lacking synapomorphies which can justify the accommodation of such diversity into one family, i.e., Boraginaceae sensu lato. A finer subdivision into eleven families has been already suggested based on molecular phylogenetic evidence and morphological, mainly fruit characters (Weigend & al., 2014; Luebert & al., 2016). However, the exact phylogenetic relationships across the order as well as the underlying processes of fruit evolution remain largely unknown. Due to its medium size, but yet its high flower and fruit diversity, Boraginales consists an excellent study group to investigate evolutionary trends across asterids and angiosperms in general. Consequently, the following hypotheses were formed aiming to understand different evolutionary aspects at various taxonomic levels.

The subdivision of the traditional Hydrophyllaceae into two families, Hydrophyllaceae and Namaceae, is not in conflict with morphology, since the capsular fruit is a plesiomorphic key morphological character of the two clades, however, neither the available molecular phylogenetic and historical biogeographic data, nor morphology do provide strong support for their separation (Luebert & al., 2016, 2017).

Considering these premises, **(H1)** increased phylogenetic resolution of interfamilial, infrafamilial and intergeneric relationships is expected by elaborating the taxonomic and geographic representation of the sampling across Hydrophyllaceae and Namaceae when using plastid and/or nuclear markers. This will subsequently permit revising historical biogeography which is expected to provide additional support for the separation of the two clades, i.e. Hydrophyllaceae and Namaceae. However, interspecific relationships of recently diversified clades within these two families are not expected to be clarified by this phylogenetic approach.

The basal grade of Boraginales II (Hydrophyllaceae and Namaceae) is characterized by four- to pluriovulate ovaries developing into one- to many-seeded capsules (Luebert & al., 2016). The remaining clades of Boraginales II have drupes or nutlets, except for the parasitic Lennoaceae, which present an apomorphic fruit type due to secondary multiplication and subdivision of the carpels (Yatskievych & Mason, 1986; Weigend & al., 2014). Vasile & al. (2021) with a case study in Hydrophyllaceae showed that reduction and multiplication of seed numbers in capsular fruits is readily reversible as long as fruits remain dehiscent, while the reduction in ovule number to four likely occurs due to truncation of ovary development after the initiation of the first layer of four ovules. **(H2)** Transitions in seed number alone do not explain the evolution of mericarpic and endomericarpic fruits from capsular fruits. Therefore, ontogenetic studies of internal ovary architecture of various apomorphic lineages within Hydrophyllaceae, as well as within Boraginales II (i.e., Lennoaceae) are expected to reveal the contribution of other structures, such as placentation, but also of processes, such as heterochrony, to fruit type transitions.

Despite the rapid and constant advances in molecular phylogenetics and systematics, there is still a lack of robust phylogenies in many plant groups within angiosperms, including Boraginales, limiting our understanding of evolution. Although, chloroplast and nuclear markers have been for long widely used and significantly contributed to largely resolve phylogenetic relationships, they are inadequate compare to NGS techniques. **(H3)** By using hundreds of loci retrieved with target capture sequencing and a comprehensive sampling across Boraginales, a highly resolved phylogeny at interfamilial and infrafamilial level is expected, yet with some topological incongruences compare to previously published hypotheses.

1.6 Research questions

Based on the aforementioned hypotheses, the following questions arose and constitute the general guidelines of this thesis:

(Q1) Can standard plastid and nuclear markers provide adequate evidence for the separation of the basally branching clades of Boraginales II (Hydrophyllaceae and Namaceae) and for clarification of the infrafamilial relationships?

(Q2) Can we elucidate the phylogenetic position of South American representatives of *Phacelia* and *Wigandia* as well as their origin and diversification history by only using standard plastid and nuclear markers?

(Q3) Can we resolve the interspecific relationships of the recently diversified clade *Phacelia* sect. *Glandulosae* in South America by implementing a genomic approach?

(Q4) Are the morphological characters traditionally used to classify South American species of *Phacelia* sect. *Glandulosae* useful for differentiating clades inferred using ddRAD data?

(Q5) In presence of an up-to-date phylogenetic framework, are ontogenetic and morphological data able to reveal which modifications of the internal ovary morphology have the greatest influence on internal ovary/fruit architecture and how are these modifications and the evolutionary processes involved connected to transitions in fruit types?

(Q6) Are floral and fruit ontogenetic and morphological data able to provide additional insights into the exact relationships of Lennoaceae and Ehretiaceae, which remain doubtful due to conflicting phylogenetic signals?

(Q7) Is the evolution of the aberrant fruit morphology of Lennoaceae somehow connected to the evolution of parasitism?

(Q8) Is Angiosperms353 bait kit a powerful tool to obtain genomic data and subsequently to infer relationships at family, subfamily and tribe level across Boraginales? If not, how can we explain any topological uncertainties?

(Q9) In presence of an up-to-date Boraginales phylogeny and considering all published data about boraginalean fruits, is it possible to provide a hypothetical yet comprehensive evolutionary series for the fruit diversity of the order?

1.7 Overview of the dissertation

This section is dedicated to a brief description of the here presented dissertation, which is a cumulative work comprising five research articles. Chapters 2, 4 and 5 have been published in peer-reviewed journals and are included here in their published form. Chapter 3 is submitted and included here as manuscript, and chapter 6 is included as manuscript which will be later submitted for publication. All chapters have separate sections of introduction, materials and methods, results, discussion and literature cited, with minor variations in formatting and structure depending on the journal style. Supplementary material of chapters 2, 3, 5 and 6 can be found in Appendices A, B, C and D respectively. Sequences generated for this thesis have either already been submitted to GenBank (chapter 2) or will be submitted (chapters 3 and 6). More details for the individual research chapters are described below:

Chapter 2 is a phylogenetic study of two new world families, Hydrophyllaceae and Namaceae (Boraginales), mainly distributed in western North and South America, with a few species in Central America. A plastid and a nuclear marker were used in order to reconstruct the phylogenetic relationships of almost all genera of the two families. The aim is to examine the systematic position of South American species of *Phacelia* (Hydrophyllaceae) and *Wigandia* (Namaceae) and the historical biogeography of Hydrophyllaceae and Namaceae using molecular dating and ancestral area reconstruction. The South American representatives of *Phacelia* are nested in three different clades and two colonizations of South America by North American species took place in the Miocene and at least one in the late Pliocene. *Wigandia* forms a well-supported monophylum with partially unresolved interspecific relationships. Within Namaceae a colonization of South America by North American species occurred during the Oligocene–Miocene transition.

Chapter 3 is a phylogenetic study of the South American representatives of *Phacelia* sect. *Glandulosae* (Hydrophyllaceae). The lack of resolution achieved through previous molecular phylogenetic analyses, as well as the taxonomic chaos—lack of typifications, dozens of synonyms, different names used for the same taxon between different countries—have long confounded the species delimitation and taxonomy of this plant group in South America. This chapter adopts a ddRADseq strategy to investigate phylogenetic relationships of all five South American representatives of *P.* sect. *Glandulosae* and a classical taxonomic approach in order to re-evaluate morphological characters traditionally used to define species in this section. Our results show that the currently recognized morphospecies are retrieved as poly- and/or paraphyletic and that most morphological characters are of little diagnostic value. We suggest the recognition of a single, polymorphic species *P. pinnatifida* Griseb. ex Wedd. and its sister species *P. artemisioides* Griseb. and we provide the necessary taxonomic revision.

Chapter 4 is a morphological study with a particular focus on the gynoecium and modifications of the stamen–corolla tube evolution of tribe Hydrophylleae (Hydrophyllaceae). Despite insights from recent studies of the flower and fruit evolution of the family, the lack of ontogenetic studies obstructs a complete understanding of the underlying processes. By implementing scanning electron microscopy and micro–computed tomography we investigated the flower and fruit ontogeny of two genera of Hydrophylleae, the monotypic *Emmenanthe penduliflora* Benth. and two out of three *Pholistoma* Lilja species. Our results indicate that changes in developmental rate or timing (heterochrony) can cause profound changes in otherwise very conserved morphologies. A hypothetical evolutionary series for the internal ovary architectural diversity of Hydrophyllaceae is proposed.

Chapter 5 deals with the small root-parasitic plant family Lennoaceae. Their exact position within the Boraginales has not been yet clearly defined and their aberrant morphology, probably correlated to their holoparasitic lifestyle, leaves many open questions regarding their morphological evolution. Investigations of the flower and fruit ontogeny and morphology of *Pholisma arenarium* Nutt. from the Lennoaceae, in comparison to that in the closely related *Bourreria succulenta* Jacq. and *Ehretia dicksonii* Hance from the Ehretiaceae using scanning electron microscopy and micro–computed tomography, support a position of Lennoaceae within Ehretiaceae. The results also confirm the hypothesis that holoparasitism can be a major driver of phenotypic diversity, leading to changes in merosity and fruit type.

Chapter 6 studies Boraginales, a subcosmopolitan and highly diverse order incl. 11 families and about 2700 species, by utilizing the Angiosperms353 probe set to reconstruct the phylogenetic relationships. Phylogenomic data were obtained from 208 samples representing 10 Boraginales families and 79 % of ca. 133 accepted genera. The study confirms and expands upon previous findings, revealing well-resolved species trees with consistently high support. The topology of previously studies is largely recovered, but there are several novelties, among them the paraphyly of Ehretiaceae, suggesting the inclusion of Lennoaceae. A few uncertainties remain regarding for example the intergeneric relationships in Cynoglosseae. These can probably be resolved by strategically expanding sampling. Overall, the results highlight the effectiveness of the target enrichment probe-set for understanding the evolutionary history of Boraginales and set the fundamentals for detailed studies on the evolution of group-specific morphological traits.

1.8 Contributions to chapters

Chapter 2: **Vasile MA**, Weigend M, Jeiter J., Weigend M., Luebert F. 2020. Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia*. *Systematics and Biodiversity* 18 (8): 757–770. https://doi.org/10.1080/14772000.2020.1771471

> **Own contributions: S**tudy design (together with F. Luebert and M. Weigend), methodology (together with F. Luebert), lab-work, data analyses (together with F. Luebert), writing the first draft, manuscript preparation (together with F. Luebert), revision (together with all co-authors).

Chapter 3: **Vasile MA**, Böhnert T, O'Leary N, Weigend M. *In review*. Is it all *Phacelia pinnatifida*? Molecular delimitations and taxonomic revision of *Phacelia* Sect. *Glandulosae* in South America based on ddRADseq. *Submitted to Taxon*.

> **Own contributions:** Conceptualization and study design (together with T. Böhnert and M. Weigend), methodology (together with T. Böhnert), lab-work, data analyses (together with T. Böhnert), examination and georeferencing of the herbarium specimens, researching virtual herbaria and taxonomic literature, taxonomic revision (together with T. Böhnert and M. Weigend), writing the first draft, manuscript preparation (together with T. Böhnert and M. Weigend), revision (together with all co-authors).

Chapter 4: **Vasile MA**, Weigend M, Ziegler A., Jeiter J. 2022. Flower and fruit development of three species of Hydrophylleae sheds new light on flower evolution in Hydrophyllaceae. *International Journal of Plant Sciences* 183 (5): 331–347. https://doi.org/10.1086/719907

> **Own contributions:** Conceptualization and study design (together with J. Jeiter), plant material collection, samples preparation, SEM, processing of µCT-data (together with A. Ziegler), visualisation of µCT-data, data analyses and images editing, writing the first draft, manuscript preparation and revision (together with all co-authors).

Chapter 5: Jeiter J, **Vasile MA**, Lewin EM, Weigend M. 2023. The odd one out: a comparison of flower and fruit development in holoparasitic *Pholisma arenarium* (Lennoaceae, Boraginales) to that in closely related Ehretiaceae. *International Journal of Plant Sciences* 184 (1): 1–18. https://doi.org/10.1086/722593

> **Own contributions:** Conceptualization and study design (together with J. Jeiter), plant material collection (together with J. Jeiter), samples preparation (together with J. Jeiter and EM. Lewin), SEM (together with J. Jeiter and EM. Lewin), processing and visualisation of µCT-data (together with J. Jeiter), data analyses and images editing (together with J. Jeiter), manuscript preparation and revision (together with J. Jeiter and M. Weigend).

Chapter 6: **Vasile MA**, Böhnert T, Jeiter J., Weigend M. *In prep.* Unravelling Boraginales: an updated phylogeny based on the Angiosperms353 probe set. To be submitted to *Molecular Phylogenetics and Evolution.*

Own contributions: Conceptualization and study design (together with T. Böhnert and M. Weigend), taxon sampling (together with all co-authors), methodology (together with T. Böhnert), lab-work, data analyses (together with T. Böhnert), writing the first draft, manuscript preparation and revision (together with all co-authors).

1.9 Literature cited

Antonelli, A., Smith, R.J., Fry, C., Simmonds, M.S.J., Kersey, P.J., Pritchard, H.W., Abbo, M.S., Acedo, C., et al. 2020. *State of the World's Plants and Fungi*. Research Report. Royal Botanic Gardens (Kew); Sfumato Foundation.

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20. https://doi.org/10.1111/boj.12385

Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers, T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B., Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., Mcginnie, C., Pokorny, L., Roy, S., Stone, M., Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J., & Forest, F. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* 71: 301–319.

Beaulieu, J.M., & Donoghue, M.J. 2013. Fruit evolution and diversification in campanulid angiosperms. *Evolution* 67: 3132–3144.

Bentham, G., & Hooker, J.D. 1876. Borragineae. Pp. 832–865 in: *Genera Plantarum (Ad Exemplaria Imprimis in Herbariis Kewensibus Servata Definita)*, vol. 2. London: Lovell Reeve, William & Norgate. Reprint 1965, Weinheim: Cramer.

Bobrov, A.V.F.Ch., & Romanov, M.S. 2019. Morphogenesis of fruits and types of fruit of angiosperms. *Botany Letters* 166: 366–399.

Buys, M.H., & Hilger, H.H. 2003. Boraginaceae cymes are exclusively scorpioid and not helicoid. *TAXON* 52: 719–724.

Chacón, J., Luebert, F., Hilger, H.H., Ovchinnikova, S., Selvi, F., Cecchi, L., Guilliams, C.M., Hasenstab-Lehman, K., Sutorý, K., Simpson, M.G., & Weigend, M. 2016. The borage family (Boraginaceae s.str.): A revised infrafamilial classification based on new phylogenetic evidence, with emphasis on the placement of some enigmatic genera. *TAXON* 65: 523–546.

Chance, G.D., & Bacon, J.D. 1984. Systematic implications of seed coat morphology in *Nama* (Hydrophyllaceae). *American Journal of Botany* 71: 829–842.

Chase, M.W., & Reveal, J.L. 2009. A phylogenetic classification of the land plants to accompany APG III. *Botanical Journal of the Linnean Society* 161: 122–127.

Chuang, T.I., & Constance, L. 1992. Seeds and Systematics in Hydrophyllaceae: Tribe Hydrophylleae. *American Journal of Botany* 79: 257–264.

De Candolle, A.P. 1845. Borragineae. Pp. 466–566 in: De Candolle, A.L.P.P., (ed.), *Prodromus Systematis Naturalis Regni Vegetabilis*, vol. 9. Paris: Treuttel & Würtz.

De Candolle, A.P. 1846. Borragineae. Pp. 1–178 in: De Candolle, A.L.P.P., (ed.), *Prodromus Systematis Naturalis Regni Vegetabilis*, vol. 10. Paris: Treuttel & Würtz.

Di Fulvio, T.E., Cosa, M.T., & Dottori, N. 1997. Morfología y vascularización floral en *Turricula*, *Eriodictyon* y *Wigandia* (Phacelieae; Hydrophyllaceae) con relación a la taxonomía. *Kurtziana* 25: 47–66.

Di Fulvio, T.E., Cosa, M.T., & Dottori, N. 1999. Morfología y vascularización floral de *Draperia*, *Emmenanthe*, *Hesperochiron*, *Romanzoffia* y *Tricardia* (Phacelieae, Hydrophyllaceae). *Kurtziana* 27: 187–209.

Diane, N., Förther, H., & Hilger, H.H. 2002. A systematic analysis of *Heliotropium*, *Tournefortia*, and allied taxa of the Heliotropiaceae (Boraginales) based on ITS1 sequences and morphological data. *American Journal of Botany* 89: 287–295.

Diane, N., Hilger, H.H., Förther, M., Weigend, M., & Luebert, F. 2016. Heliotropiaceae. Pp. 203–211 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*, vol. XIV. Heidelberg: Springer

Doyle, J.A., & Endress, P.K. 2000. Morphological phylogenetic analysis of basal angiosperms: comparison and combination with molecular data. *International Journal of Plant Sciences* 161: S121–S153.

Ebadi, M., & Nikzat, S. 2021. Nutlet micromorphology and character evolution of *Lappula* species (Boraginaceae) and its systematic implications. *Botanical Studies* 62: 17.

Endress, P.K. 2011. Evolutionary diversification of the flowers in angiosperms. *American Journal of Botany* 98: 370–396.

Endress, P.K. 2019. The morphological relationship between carpels and ovules in angiosperms: pitfalls of morphological interpretation. *Bot J Linn Soc* 189: 201–227.

Gottschling, M. 2004. Floral ontogeny in *Bourreria* (Ehretiaceae, Boraginales). *Flora - Morphology, Distribution, Functional Ecology of Plants* 199: 409–423.

Gottschling, M., Diane, N., Hilger, H.H., & Weigend, M. 2004. Testing hypotheses on disjunctions present in the primarily woody Boraginales: Ehretiaceae, Cordiaceae, and Heliotropiaceae, inferred from ITS1 sequence data. *International Journal of Plant Sciences* 165: S123–S135.

Gottschling, M., & Hilger, H.H. 2004. Characterisation of a novel fruit type found in *Ehretia* (Ehretiaceae, Boraginales). *Blumea - Biodiversity, Evolution and Biogeography of Plants* 49: 145–153.

Gottschling, M., Luebert, F., Hilger, H.H., & Miller, J.S. 2014a. Molecular delimitations in the Ehretiaceae (Boraginales). *Molecular Phylogenetics and Evolution* 72: 1–6.

Gottschling, M., Miller, J.S., Weigend, M., & Hilger, H.H. 2005. Congruence of a phylogeny of Cordiaceae (Boraginales) inferred from ITS1 sequence data with morphology ecology, and biogeography. *Annals of the Missouri Botanical Garden* 92: 425–437.

Gottschling, M., Nagelmüller, S., & Hilger, H.H. 2014b. Generative ontogeny in *Tiquilia* (Ehretiaceae: Boraginales) and phylogenetic implications. *Biol J Linn Soc* 112: 520–534.

Gottschling, M., Weigend, M., & Hilger, H.H. 2016. Ehretiaceae. Pp. 165–178 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants,* vol. XIV. Heidelberg: Springer.

Gürke, M. 1893. Borraginaceae (Asperifoliaceae). Pp. 71–131 in: Engler, A., Prantl, K. (eds.), *Die Natürlichen Pflanzenfamilien,* vol. IV, 3a. Engelmann, Leipzig.

Heigl, H.M.L., Kretschmann, J., Hilger, H.H., & Gottschling, M. 2020. Flower and fruit anatomy of *Cordia nodosa* Lam. and *Varronia bonplandii* Desv. (Cordiaceae, Boraginales) with phylogenetic implications. *Org Divers Evol* 20: 9–24.

Hilger, H.H. 1984. Wachstum und Ausbildungsformen des Gynoeceums von *Rochelia* (Boraginaceae). *Pl Syst Evol* 146: 123–139.

Hilger, H.H. 1985. Ontogenie, Morphologie und systematische Bedeutung geflügelter und glochidientragender Cynoglosseae- und Eritricheae-Fruchte (Boraginaceae). *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie*.

Hilger, H.H. 1987. Flower and fruit development in *Wigandia caracasana* (Hydrophyllaceae). *American Journal of Botany* 74: 250–259.

Hilger, H.H. 1992. Morphology of *Heliotropium* (Boraginaceae) dispersal units. *Botanica Acta* 105: 387–393.

Hilger, H.H. 2014. Ontogeny, morphology, and systematic significance of glochidiate and winged fruits of *Cynoglosseae* and *Eritrichieae* (Boraginaceae). *Plant Diversity and Evolution* 131: 167–214.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., Crandall, K.A., Deng, J., Drew, B.T., Gazis, R., Gude, K., Hibbett, D.S., Katz, L.A., Laughinghouse, H.D., McTavish, E.J., Midford, P.E., Owen, C.L., Ree, R.H., Rees, J.A., Soltis, D.E., Williams, T., & Cranston, K.A. 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences* 112: 12764–12769. https://doi.org/10.1073/pnas.1423041112

Hofmann, M. 1999. Flower and fruit development in the genus *Phacelia* (Phacelieae, Hydrophyllaceae): characters of systematic value. *Systematics and Geography of Plants* 68: 203–212.

Hofmann, M., Walden, G.K., Hilger, H.H., & Weigend, M. 2016. Hydrophyllaceae. Pp. 221–238 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants,* vol. XIV. Heidelberg: Springer

Jeiter, J. 2020. *Evolution of floral organisation and architecture in Boraginales and Geraniales.* Dissertation, University of Bonn, Bonn, Germany.

Jeiter, J., Danisch, F., & Hilger, H.H. 2016. Polymery and nectary chambers in *Codon* (Codonaceae): Flower and fruit development in a small, capsule-bearing family of Boraginales. *Flora - Morphology, Distribution, Functional Ecology of Plants* 220: 94–102.

Jeiter, J., Langecker, S., & Weigend, M. 2020. Towards an integrative understanding of stamen–corolla tube modifications and floral architecture in Boraginaceae s.s. (Boraginales). *Bot J Linn Soc* 193: 100–124.

Jeiter, J., & Smets, E. 2023. Integrating comparative morphology and development into evolutionary research. *TAXON* 72: 724–732. https://doi.org/10.1002/tax.12983

Jeiter, J., Staedler, Y.M., Schönenberger, J., Weigend, M., & Luebert, F. 2018. Gynoecium and fruit development in *Heliotropium* sect. *Heliothamnus* (Heliotropiaceae). *International Journal of Plant Sciences*

Jeiter, J., & Weigend, M. 2018. Simple scales make complex compartments: ontogeny and morphology of stamen–corolla tube modifications in Hydrophyllaceae (Boraginales). *Biol J Linn Soc* 125: 802–820.

Kadereit, J.W., & Bittrich, V. (eds.) 2016. *The families and genera of vascular plants, vol. 14, Flowering Plants. Eudicots: Aquifoliales, Boraginales, Bruniales, Dipsacales, Escalloniales, Garryales, Paracryphiales, Solanales (except Convolvulaceae), Icacinaceae, Metteniusaceae, Vahliaceae*. Heidelberg: Springer

Kapli, P., Yang, Z., & Telford, M.J. 2020. Phylogenetic tree building in the genomic age. *Nat Rev Genet* 21: 428–444.

Li, H.-T., Yi, T.-S., Gao, L.-M., Ma, P.-F., Zhang, T., Yang, J.-B., Gitzendanner, M.A., Fritsch, P.W., Cai, J., Luo, Y., Wang, H., van der Bank, M., Zhang, S.-D., Wang, Q.-F., Wang, J., Zhang, Z.-R., Fu, C.-N., Yang, J., Hollingsworth, P.M., Chase, M.W., Soltis, D.E., Soltis, P.S., & Li, D.-Z. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* 5: 461–470.

Luebert, F., Brokamp, G., Wen, J., Weigend, M., & Hilger, H.H. 2011. Phylogenetic relationships and morphological diversity in Neotropical *Heliotropium* (Heliotropiaceae). *TAXON* 60: 663–680.

Luebert, F., Cecchi, L., Frohlich, M.W., Gottschling, M., Guilliams, C.M., Hasenstab-Lehman, K.E., Hilger, H.H., Miller, J.S., Mittelbach, M., Nazaire, M., Nepi, M., Nocentini, D., Ober, D., Olmstead, R.G., Selvi, F., Simpson, M.G., Sutorý, K., Valdés, B., Walden, G.K., & Weigend, M. 2016. Familial classification of the Boraginales. *TAXON* 65: 502–522.

Luebert, F., Couvreur, T.L.P., Gottschling, M., Hilger, H.H., Miller, J.S., & Weigend, M. 2017. Historical biogeography of Boraginales: West Gondwanan vicariance followed by long-distance dispersal? *Journal of Biogeography* 44: 158–169.

Luebert, F., & Wen, J. 2008. Phylogenetic analysis and evolutionary diversification of *Heliotropium* sect. *Cochranea* (Heliotropiaceae) in the Atacama Desert. *Systematic Botany*. 32: 390-–402

Mabberley, D.J. 2008. *Mabberley's Plant Book*, 3rd ed. Cambridge: Cambridge University Press.

Maddison, W.P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

Mansion, G., Selvi, F., Guggisberg, A., & Conti, E. 2009. Origin of Mediterranean insular endemics in the Boraginales: integrative evidence from molecular dating and ancestral area reconstruction. *Journal of Biogeography* 36: 1282–1296.

Nichols, R. 2001. Gene trees and species trees are not the same. *Trends in Ecology & Evolution* 16: 358–364.

O'Leary, S.J., Puritz, J.B., Willis, S.C., Hollenbeck, C.M., & Portnoy, D.S. 2018. These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology* 27: 3193–3206.

Otero, A., Jiménez-Mejías, P., Valcárcel, V., & Vargas, P. 2019. Being in the right place at the right time? Parallel diversification bursts favored by the persistence of ancient epizoochorous traits and hidden factors in Cynoglossoideae. *American Journal of Botany* 106: 438–452.

Redeling, B.D., & Suchard, M.A. 2009. Robust inferences from ambiguous alignments. In: M. Rosenberg (ed.), *Sequence Alignment: Methods, Models, Concepts, and Strategies*. University of California Press.

Refulio-Rodriguez, N.F., & Olmstead, R.G. 2014. Phylogeny of Lamiidae. *American Journal of Botany* 101: 287–299.

Sattler, R., & Rutishauser, R. 2023. Fundamentals of Plant Morphology and Plant Evo-Devo (Evolutionary Developmental Morphology). *Plants* 12: 118.

Simpson, M.G. 2010. *Plant Systematics*, second edition. Elsevier.

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Prince, L.M., Kress, W.J., Nixon, K.C., & Farris, J.S. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

Stull, G.W., Stefano, R.D. de, Soltis, D.E., & Soltis, P.S. 2015. Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. *American Journal of Botany* 102: 1794–1813.

Takhtajan, A. 1987. *Systema Magnoliophytorum.* Leningrad: Soviet Sciences Press.

Takhtajan, A. 1997. *Diversity and Classification of Flowering Plants.* New York: Columbia University Press.

Vasile, M.-A., Luebert, F., Jeiter, J., & Weigend, M. 2021. Fruit evolution in Hydrophyllaceae. *American Journal of Botany* 108: 925–945.

Vialle, R.A., Tamuri, A.U., & Goldman, N. 2018. Alignment modulates ancestral sequence reconstruction accuracy. *Molecular Biology and Evolution* 35: 1783–1797

Wagstaff, S.J., Martinsson, K., & Swenson, U. 2000. Divergence estimates of *Tetrachondra hamiltonii* and *T. patagonica* (Tetrachondraceae) and their implications for austral biogeography. *New Zealand Journal of Botany* 38: 587–596.

Walden, G.K., & Patterson, R. 2012. Nomenclature of subdivisions within *Phacelia* (Boraginaceae: Hydrophylloideae). *Madr* 59: 211–222.

Wanninger, A. 2015. Morphology is dead – long live morphology! Integrating MorphoEvoDevo into molecular EvoDevo and phylogenomics. *Frontiers in Ecology and Evolution* 3.

Weigend, M., Luebert, F., Gottschling, M., Couvreur, T.L.P., Hilger, H.H., & Miller, J.S. 2014. From capsules to nutlets—phylogenetic relationships in the Boraginales. *Cladistics* 30: 508–518.

Weigend, M., Selvi, F., Thomas, D.C., & Hilger, H.H. 2016. Boraginaceae. Pp. 41–102 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*. Heidelberg: Springer.

Wong, K.M., Suchard, M.A., & Huelsenbeck, J.P. 2008. Alignment uncertainty and genomic analysis. *Science* 319: 473–476.

Yatskievych, G., & Mason, C.T. 1986. A revision of the Lennoaceae. *Systematic Botany* 11: 531–548.

Zimmermann, W. 1930. *Die Phylogenie der Pflanzen: ein Überblick über Tatsachen und Probleme*. G. Fischer.

Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special referenceto *Phacelia* and *Wigandia*



Chapter 2

Phacelia secunda (Hydrophyllaceae) in cultivation in Bonn Botanical Gardens. It is one of the few South American species of *Phacelia*

Vasile MA, Weigend M, Jeiter J., Weigend M., Luebert F. 2020. Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia*. *Systematics and Biodiversity* 18 (8): 757–770.
https://doi.org/10.1080/14772000.2020.1771471

This is an Accepted Manuscript of an article published by Taylor & Francis in *Journal of Systematics and Biodiversity* on *18 Jun 2020*, available online: *https://dx.doi.org/10.1080/14772000.2020.1771471*

Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia*

MARIA-ANNA VASILE[1], JULIUS JEITER[1], MAXIMILIAN WEIGEND[1] & FEDERICO LUEBERT[1, 2]

[1] *Nees-Institut für Biodiversität der Pflanzen, Universität Bonn, Meckenheimer Allee 170, 53115, Germany Bonn*

[2] *Departamento de Silvicultura y Conservación de la Naturaleza, Universidad de Chile, Santiago, Chile*

Correspondence to: Maria-Anna Vasile, E-mail: m.vasile@uni-bonn.de

Abstract

This study aimed to examine the systematic position of South American species of *Phacelia* (Hydrophyllaceae) and *Wigandia* (Namaceae) and the historical biogeography of Hydrophyllaceae and Namaceae using molecular dating and ancestral area reconstruction. To this end, we constructed two data sets, one with a plastid (*ndhF*) and one with a nuclear marker (ITS), using previously published and newly generated sequences. We inferred the phylogeny of Hydrophyllaceae and Namaceae implementing both likelihood and Bayesian methods. We also estimated divergence times and ancestral areas for all major clades using a relaxed Bayesian uncorrelated molecular clock and the dispersal-extinction-cladogenesis (DEC) approach, respectively. The South American representatives of *Phacelia* are placed in three different clades of the genus and two colonizations of South America by North American species took place in the Miocene and at least one in the late Pliocene. *Wigandia* forms a well-supported monophylum with interspecific relationships partly unresolved. Within Namaceae a colonization of South America by North American species occurred during the Oligocene-Miocene transition. The MRCA of *Wigandia* was distributed in North and Central America in late Oligocene. Long-distance dispersal may have been necessary for the colonization of South America by Namaceae during Oligocene-Miocene transition, when North and South America were not connected.

Key words: Amphitropical disjunction, ancestral area reconstruction, Boraginales II, molecular dating, South America

2.1 Introduction

Hydrophyllaceae and Namaceae are two new world families mainly distributed in western North and South America and a few species in Central America. Mostly herbaceous Hydrophyllaceae comprise twelve genera and ca. 240-260 species. *Phacelia* Juss. is the largest and most diverse member of the family (ca. 207 species; Hofmann et al., 2016). Namaceae are a smaller family of

herbs, shrubs, or small, soft-wooded trees. They comprise 4 genera with approximately 75 species in total and most species in *Nama* L. (ca. 50 spp.).

Hydrophyllaceae and Namaceae are successively sister families to the remainder of the Boraginales II clade (including also Coldeniaceae, Cordiaceae, Ehretiaceae, Heliotropiaceae, Hoplestigmataceae, and Lennoaceae) based on plastid data (Luebert et al., 2016; Stull et al., 2015; Weigend et al., 2014). Phylogenetic studies over the last two decades have resolved the relationships in these two to families: Hydrophyllaceae falls into three major clades (Ferguson, 1998; Walden, 2010). The first clade includes *Phacelia* and *Romanzoffia* Cham. (equivalent to Romanzoffieae, Walden et al., 2014), the second clade comprises *Hydrophyllum* L., *Pholistoma* Lilja, *Nemophila* Nutt., *Emmenanthe* Benth., *Ellisia* L. and *Eucrypta* Nutt. (equivalent to Hydrophylleae; Walden, 2015), and the third clade includes *Draperia* Torr., *Tricardia* Torr. ex S. Watson, *Howellanthus* (Constance) Walden & R. Patt. and *Hesperochiron* S. Watson (unnamed; Luebert et al., 2016; Walden et al., 2014). Namaceae falls into two clades: a *Nama* clade (about 50 spp.; Hofmann et al., 2016), and a clade comprising woody *Wigandia* Kunth (6 spp.; Hofmann et al., 2016), *Eriodictyon* Benth., and *Turricula* J.F. Macbr. plus two species of *Nama* (rendering *Nama* polyphyletic Ferguson, 1998; Luebert et al., 2016; Taylor, 2012).

Most species of Hydrophyllaceae and Namaceae are found in western North America. Only one genus of Hydrophyllaceae (*Phacelia*) and two of Namaceae (*Nama*, *Wigandia*) have representatives in South America. *Phacelia* and *Nama* thus represent amphitropical disjunctions, the latter also with one species in Hawaii (Hofmann et al., 2016). *Wigandia* is present in southwestern North America, Central America and its South American distribution is restricted to the northern part of that continent.

Phacelia has been subject to several molecular phylogenetic studies with plastid and nuclear sequence data (Gilbert et al., 2005; Hansen et al., 2009; Walden et al., 2014). These studies aimed at evaluating existing classifications of the genus based on a broad sampling. However, none of them included South American species.

In *Nama*, one dissertation (Taylor, 2012) addressed the phylogenetic relationships of its species based on plastid and nuclear sequence data and a sampling that included one out of three South American species. However, this study and its sequence data have not been published so far.

In *Wigandia*, there is no specific phylogenetic study available and only few large-scale phylogenies (e.g., Ferguson, 1998; Walden et al., 2014) have included one or two species of the genus from North America.

Therefore, the systematic placement of the South American species of Hydrophyllaceae and Namaceae remains to be assessed with molecular data. As a consequence, the biogeographic history of these genera, across their ranges, has not been investigated.

Late Eocene divergence times of the crown node of Hydrophyllaceae and Namaceae were recovered in the first historical biogeography study of the order by Luebert et al. (2017). The study further retrieves a North American ancestral area for both families and subsequent disper-

sal into Central and South America. Due to low sampling density in the families specific questions about the biogeographic history were not addressed.

The present study uses previously published sequence data and supplements them with a broad sampling of South American representatives of *Phacelia* and *Wigandia* to address the following questions: (1) what is the phylogenetic position of South American representatives of *Phacelia* and *Wigandia*? (2) When and where did these genera originate and diversify? (3) How many dispersal events between North and South America took place and in which directions? To address these questions, we inferred the phylogenetic relationships of Hydrophyllaceae and Namaceae with an expanded taxon sampling for *Phacelia* and *Wigandia* and estimated their divergence times and ancestral ranges based on molecular clock analyses and the dispersal-extinction-cladogenesis (DEC) approach to historical biogeography.

2.2 Materials and methods

Phylogenetic analyses

Taxon sampling and outgroup selection. We built two datasets, one with plastid NADH dehydrogenase F (*ndhF*) and one with nuclear Internal transcribed spacer (ITS). In both cases we combined newly generated and published sequences. Newly generated sequences include seven out of ten South American species of *Phacelia* and four species of *Wigandia.* The *ndhF* dataset consisted of 241 accessions, 117 Hydrophyllaceae (110 taxa), 29 Namaceae (23 taxa), and includes representatives from all major clades of each family of Boraginales (see Luebert et al., 2017; Weigend et al., 2014 for details). All *ndh* genes are absent in the chloroplast of the holoparasitic Lennoaceae (Schneider et al., 2018), so they could not be included in the analyses. Four taxa from the Gentianales were chosen as an outgroup, which is likely the sister order to Boraginales (Leebens-Mack et al., 2019; Stull et al., 2015). The ITS dataset consisted of 131 accesions, 110 Hydrophyllaceae (107 taxa), and 19 Namaceae, (19 taxa). One species each of Ehretiaceae and Heliotropiaceae were used as outgroups, because a further expansion of the outgroup sampling made the alignment equivocal. Sampling within *Phacelia* for both markers included representatives of all subgenera and sections recognized by Walden and Patterson (2012). New sequences were deposited in GenBank. Voucher information and GenBank accession numbers is given in Appendix S1, online supplemental material, which is available from the article's Taylor & Francis Online page at http://dx.doi.org/10.1080/14772000.2020.1771471.

DNA extraction, amplification and sequencing. DNA was extracted from samples of silica-gel-dried leaves or herbarium material with a modified CTAB method (Doyle & Doyle, 1987). PCR amplifications were performed in a Trio-Thermoblock thermal cycler (Biometra, Göttingen, Germany) and a Mastercycler ep (Eppendorf AG, Hamburg) in a 25 µl volume containing 0.5 U of Taq Polymerase, 5 µl 5 x Taq Buffer, 25 mM $MgCl_2$, 1.25 mM of each dNTP, 0.2 mM of each primer and about 50 ng of genomic DNA. Amplification primers and cycling conditions followed Moore and Jansen (2006). PCR products were purified according to Werle et al. (1994). Sanger sequencing was performed by GATC Biotech (Eurofins Genomics Germany

GmbH). The resulting sequences were assembled and aligned automatically using Geneious v.8.1.9 (https://www.geneious.com). Manual adjustments followed using PhyDE 0.9971 (Müller et al., 2010).

Phylogenetic inference. Previous studies showed that nuclear and plastid markers are significantly heterogeneous in Hydrophyllaceae and Namaceae and should therefore be analyzed separately (Ferguson, 1998; Hansen et al., 2009; Moore & Jansen, 2006; Nazaire & Hufford, 2012; Taylor, 2012; Walden et al., 2014; Weeks et al., 2010). Maximum likelihood (ML) analysis for each marker was conducted through CIPRES Science Gateway (Miller et al., 2010) using RAxML-HPC BlackBox (8.2.12) tool (Stamatakis, 2014) on XSEDE (Extreme Science and Engineering Discovery Environment). The substitution model GTR+Γ was used in both analyses (*ndhF* and ITS), with 1,000 bootstrap replicates and best tree search. Bayesian inferences (BI) were conducted in MrBayes 3.2.6 using XSEDE through the CIPRES portal using default parameters, random starting trees and four independent runs. The number of generations was set to 1,000,000 using four chains and sampling every 1,000 generations. The GTR+Γ model was chosen and the number of substitution types was set to 6.

Historical biogeography

Fossil record and node calibration. The phylogeny was calibrated by fixing the age of the Boraginales stem node (Boraginales & Gentianales) with a secondary calibration from Magallón et al. (2015), namely a normal calibration with a mean of 88.25 Ma and SD of 11 Ma. Following Luebert et al. (2017), four fossil calibrations for internal nodes of Ehretiaceae, Heliotropiaceae, and Boraginaceae s.str. were set as lognormal distributions with an offset of the minimum age of the fossil and with a lognormal mean so that the median age falls in the middle of the time period assigned to the fossil. Fossils used for calibration are listed in Table 1.

Estimating divergence times. To estimate divergence times in Boraginales the *ndhF* dataset was used. The alignment of ITS across Boraginales and Gentianales is unreliable due to extensive sequence variation. We used Gentianales as a root of the tree and a sister group of Boraginales based on previous phylogenetic studies (Stull et al., 2015). Uncorrelated Bayesian dating was conducted with the relaxed molecular clock model implemented in BEAST v1.8.4. (Drummond et al., 2012). Following previous analysis (Luebert et al., 2017), an uncorrelated lognormal relaxed-clock and the GTR+Γ+I model were selected and a Yule prior (Gernhard, 2008) was assigned to the branching process. One run of 100,000,000 chains with sampling every 10,000 generations was performed and convergence was assessed by analyzing the results in Tracer v1.7.1 (Rambaut et al., 2018). After removing 10,000,000 burn-in samples, median ages and 95% High Posterior Density (HPD) intervals with a posterior probability limit of 0.95 were summarized with TreeAnnotator v1.8.4 (https://beast.community/treeannotator), prior to visualization in FigTree v1.4.4 (https://beast.community/figtree).

Biogeographical analysis. Biogeographical analyses were performed separately for Hydrophyllaceae and Namaceae. Three biogeographical areas were defined according to the extant distribution patterns of the two families: A: North America, B: Central America, C: South America.

Hydrophyllaceae are found in North or South America only, whereas for Namaceae all three biogeographical areas are represented. The chronogram obtained from the BEAST analysis was first edited using the R-package 'ape' (Paradis & Schliep, 2019), in order to extract the clades corresponding to Hydrophyllaceae and Namaceae. Using the R-package 'BioGeoBEARS' 1.1.2 (Matzke, 2013), the Dispersal-Extinction-Cladogenesis approach (DEC) described by Ree and Smith (2008), was implemented. Maxareas was set to the number of areas in the respective analyses (i.e., two for Hydrophyllaceae and three for Namaceae). The coded areas correspond to the distributions at species level, except for *Wigandia,* for which we coded the distribution of the specimens due to uncertainties in the species delimitations and distributions and a lack of up-to-date taxonomic revision.

Table 2.1 List of fossils used for calibration.

Family	Taxon	Fossil type	Collectors	Epoche	Estimated Age (Ma)	Region	Offset (Ma)	Mean (Ma)	Median age (Ma)
1. Boraginaceae s.str.	*Cryptantha* Lehm ex. Fisch & C.A. Mey	Nutlets	Thomasson (1987); Gabel et al. (1998)	Upper Miocene	13.6 – 10.3	USA	10.3	0.5	11.95
2. Boraginaceae s.str.	*Ogastemma* Brummitt	Nutlets	Hammouda et al. (2015)	Eocene	56– 41.2	SW Algeria	41.2	2	48.6
3. Heliotropiaceae	*Tournefortia* L. sect. *Tournefortia*	Pollen grains	Graham & Jarzen (1969)	Lower Oligocene	33.9 – 28.1	Puerto Rico	28.6	1.1	31.1
4. Ehretiaceae	*Ehretia* P. Browne	Fruit parts	Reid & Chandler (1933); Chandler (1961)	Lower Eocene	56– 47.8	London Clay	47.8	1.1	51.8

2.3 Results

Phylogenetic analysis and age estimates

The aligned *ndhF* data set consisted of 2,228 sites, with 1,347 distinct alignment patterns and the aligned ITS data set was 889 sites in length with 517 distinct alignment patterns. Separate analyses for the plastid and nuclear datasets resulted in similar tree topologies, with the same well-supported major clades in terms of Bayesian posterior probabilities (BPP) and ML bootstrap values (BS; Figs 1, 2).

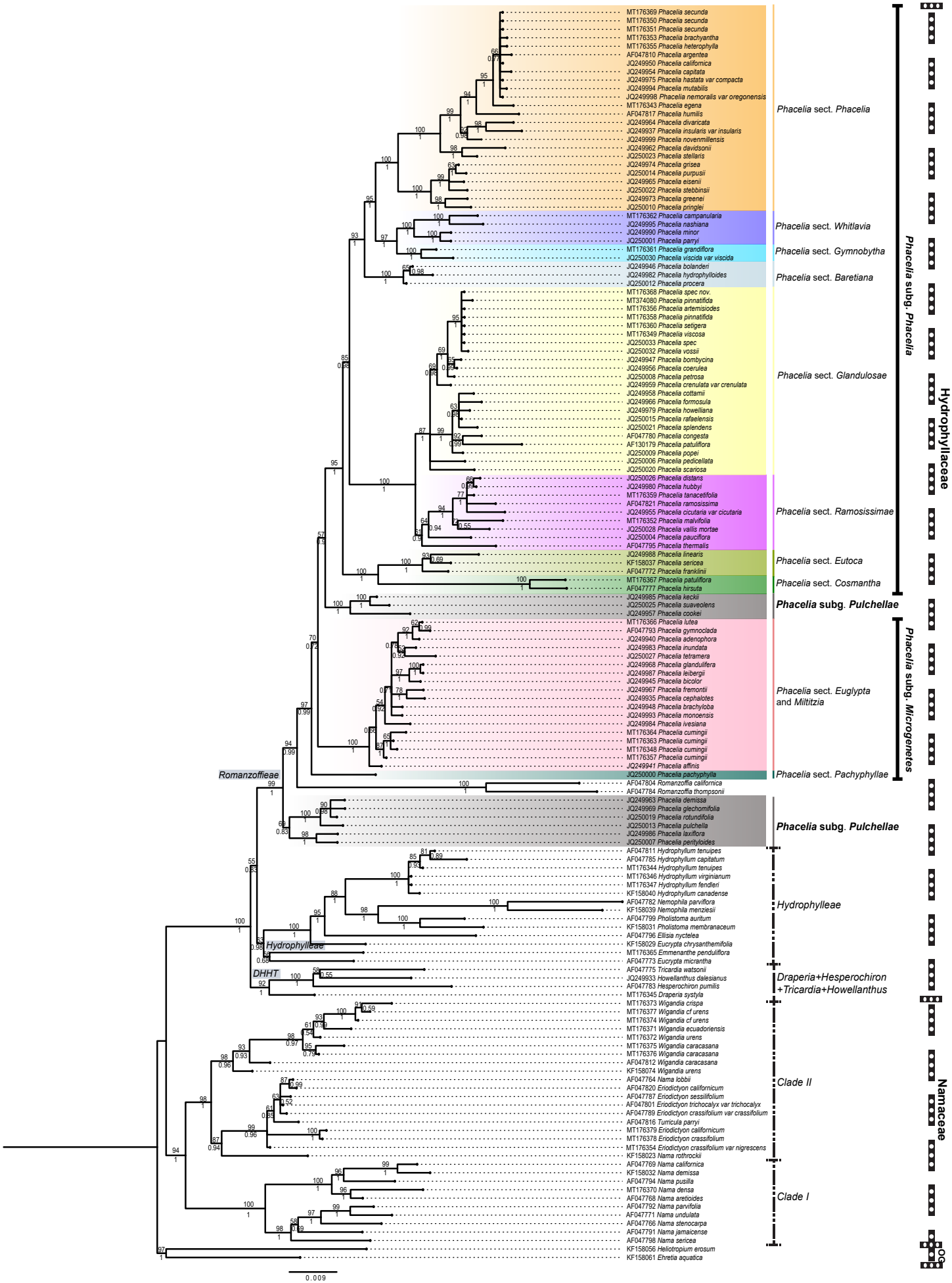

MT176369 Phacelia secunda
MT176350 Phacelia secunda
MT176351 Phacelia secunda
MT176353 Phacelia brachyantha
MT176355 Phacelia heterophylla
AF047810 Phacelia argentea
JQ249950 Phacelia californica
JQ249954 Phacelia capitata
JQ249975 Phacelia hastata var compacta
JQ249994 Phacelia mutabilis
JQ249998 Phacelia nemoralis var oregonensis
MT176343 Phacelia egena
AF047817 Phacelia humilis
JQ249964 Phacelia divaricata
JQ249937 Phacelia insularis var insularis
JQ249999 Phacelia novenmillensis
JQ249962 Phacelia davidsonii
JQ250023 Phacelia stellaris
JQ249974 Phacelia grisea
JQ250014 Phacelia purpusii
JQ249965 Phacelia eisenii
JQ250022 Phacelia stebbinsii
JQ249973 Phacelia greenei
JQ250010 Phacelia pringlei
Phacelia sect. Phacelia
MT176362 Phacelia campanularia
JQ249995 Phacelia nashiana
JQ249990 Phacelia minor
JQ250001 Phacelia parryi
Phacelia sect. Whitlavia
MT176361 Phacelia grandiflora
JQ250030 Phacelia viscida var viscida
Phacelia sect. Gymnobytha
JQ249946 Phacelia bolanderi
JQ249982 Phacelia hydrophylloides
JQ250012 Phacelia procera
Phacelia sect. Baretiana
MT176368 Phacelia spec nov.
MT374080 Phacelia pinnatifida
MT176356 Phacelia artemisiodes
MT176358 Phacelia pinnatifida
MT176360 Phacelia setigera
MT176349 Phacelia viscosa
JQ250033 Phacelia spec
JQ250032 Phacelia vossii
JQ249947 Phacelia bombycina
JQ249956 Phacelia coerulea
JQ250008 Phacelia petrosa
JQ249959 Phacelia crenulata var crenulata
JQ249958 Phacelia cottamii
JQ249966 Phacelia formosula
JQ249979 Phacelia howelliana
JQ250015 Phacelia rafaelensis
JQ250021 Phacelia splendens
AF047780 Phacelia congesta
AF130179 Phacelia patuliflora
JQ250009 Phacelia popei
JQ250006 Phacelia pedicellata
JQ250020 Phacelia scariosa
Phacelia sect. Glandulosae
JQ250026 Phacelia distans
JQ249980 Phacelia hubbyi
MT176359 Phacelia tanacetifolia
AF047821 Phacelia ramosissima
JQ249955 Phacelia cicutaria var cicutaria
MT176352 Phacelia malvifolia
JQ250028 Phacelia vallis mortae
JQ250004 Phacelia pauciflora
AF047795 Phacelia thermalis
Phacelia sect. Ramosissimae
JQ249988 Phacelia linearis
KF158037 Phacelia sericea
AF047772 Phacelia franklinii
Phacelia sect. Eutoca
MT176367 Phacelia patuliflora
AF047777 Phacelia hirsuta
Phacelia sect. Cosmantha
JQ249985 Phacelia keckii
JQ250025 Phacelia suaveolens
JQ249957 Phacelia cookei
Phacelia subg. Pulchellae
MT176366 Phacelia lutea
AF047793 Phacelia gymnoclada
JQ249940 Phacelia adenophora
JQ249983 Phacelia inundata
JQ250027 Phacelia tetramera
JQ249968 Phacelia glandulifera
JQ249987 Phacelia leibergii
JQ249945 Phacelia bicolor
JQ249967 Phacelia fremontii
JQ249935 Phacelia cephalotes
JQ249948 Phacelia brachyloba
JQ249993 Phacelia monoensis
JQ249984 Phacelia ivesiana
MT176364 Phacelia cumingii
MT176363 Phacelia cumingii
MT176348 Phacelia cumingii
MT176357 Phacelia cumingii
JQ249941 Phacelia affinis
Phacelia sect. Euglypta and Miltitzia
JQ250000 Phacelia pachyphylla
Phacelia sect. Pachyphyllae
Romanzoffieae
AF047804 Romanzoffia californica
AF047784 Romanzoffia thompsonii
JQ249963 Phacelia demissa
JQ249969 Phacelia glechomifolia
JQ250019 Phacelia rotundifolia
JQ250013 Phacelia pulchella
JQ249986 Phacelia laxiflora
JQ250007 Phacelia perityloides
Phacelia subg. Pulchellae
AF047811 Hydrophyllum tenuipes
AF047785 Hydrophyllum capitatum
MT176344 Hydrophyllum tenuipes
MT176346 Hydrophyllum virginianum
MT176347 Hydrophyllum fendleri
KF158040 Hydrophyllum canadense
AF047782 Nemophila parviflora
KF158039 Nemophila menziesii
AF047799 Pholistoma auritum
KF158031 Pholistoma membranaceum
AF047796 Ellisia nyctelea
KF158029 Eucrypta chrysanthemifolia
MT176365 Emmenanthe pendulifloraa
AF047773 Eucrypta micrantha
Hydrophylleae
Hydrophylleae
DHHT
AF047775 Tricardia watsonii
JQ249933 Howellanthus dalesianus
AF047783 Hesperochiron pumilis
MT176345 Draperia systyla
Draperia+Hesperochiron +Tricardia+Howellanthus
MT176373 Wigandia crispa
MT176377 Wigandia cf urens
MT176374 Wigandia cf urens
MT176371 Wigandia ecuadoriensis
MT176372 Wigandia urens
MT176375 Wigandia caracasana
MT176376 Wigandia caracasana
AF047812 Wigandia caracasana
KF158074 Wigandia urens
AF047764 Nama lobbii
AF047820 Eriodictyon californicum
AF047787 Eriodictyon sessilifolium
AF047801 Eriodictyon trichocalyx var trichocalyx
AF047789 Eriodictyon crassifolium var crassifolium
AF047816 Turricula parryi
MT176379 Eriodictyon californicum
MT176378 Eriodictyon crassifolium
MT176354 Eriodictyon crassifolium var nigrescens
KF158023 Nama rothrockii
Clade II
AF047769 Nama californica
KF158032 Nama demissa
AF047794 Nama pusilla
MT176370 Nama densa
AF047768 Nama aretioides
AF047792 Nama parvifolia
AF047771 Nama undulata
AF047766 Nama stenocarpa
AF047791 Nama jamaicense
AF047798 Nama sericea
KF158056 Heliotropium erosum
KF158061 Ehretia aquatica
Clade I
0.009
Phacelia subg. Phacelia
Phacelia subg. Microgenetes
Hydrophyllaceae
Namaceae
og

Fig. 2.1 Phylogram obtained from the Bayesian analysis based on *ndhF*. Only the Hydrophyllaceae and Namaceae clades and their sister genera *Heliotropium* and *Ehretia* (OG = outgroup) are shown. The remaining clades were pruned. Numbers above branches are ML bootstrap values (≥50) and below branches Bayesian posterior probabilities (BPP). Scale bar = mean number of nucleotide substitutions per site. Major clades of *Phacelia* are indicated with different colours.

Hydrophyllaceae and Namaceae were retrieved as monophyletic (1 BPP, 100% BS and 1 BPP, 94% BS respectively for *ndhF*; 0.98 BPP, 75% BS and 0.99 BPP, 90% BS respectively for ITS).

In Hydrophyllaceae, *Draperia* forms a clade with *Tricardia*, *Howellanthus* and *Hesperochiron* (1 BPP, 92% BS for *ndhF*; 0.96 BPP, 60% BS for ITS; hereafter DHHT clade). The *ndhF* dataset supports that DHHT clade is sister to Hydrophylleae + Romanzoffieae (1 BPP, 100% BS). The ITS dataset on the other hand shows the DHHT clade as sister to Hydrophylleae with low support (0.53 BPP). The genera *Emmenanthe*, *Eucrypta*, *Ellisia*, Hy*drophyllum*, *Pholistoma,* and *Nemophila* form a clade (Hydrophylleae; 0.98 BPP, 63% BS for *ndhF*; 1 BPP, 91% BS for ITS) sister to *Phacelia* + *Romanzoffia* for the *ndhF* dataset (Romanzoffieae; 0.83 BPP, 55% BS). The placement of *Romanzoffia* was incongruent between the plastid and the nuclear topologies: the ITS tree (Fig. 2) recovers *Phacelia* as monophyletic (0.97 BPP, 72% BS) and sister to *Romanzoffia* (0.99 BPP, 83% BS), while the *ndhF* tree (Fig. 1) retrieves *Romanzoffia* nested in *Phacelia* (0.99 BPP, 94% BS).

Most of the currently recognized taxonomic groups of *Phacelia* were retrieved as monophyletic with high support values in both plastid and nuclear analyses. The only exceptions in the majority-rule consensus tree (BI) derived from *ndhF*, were *Phacelia* subg. *Pulchellae*, *Phacelia* sect. *Euglypta* and *Phacelia* sect. *Miltitzia* (Fig. 1). In the majority-rule consensus tree (BI) derived from ITS, *Phacelia* sect. *Whitlavia* was retrieved as paraphyletic (Fig. 2), resulting in a non-mophyletic *Phacelia* subg. *Phacelia*.

The South American species of *Phacelia* were not retrieved as monophyletic, but are rather placed in three different clades within *Phacelia* in both analyses. Andean *P. secunda* J.F. Gmel. and south-temperate *P. brachyantha* Benth. belong to *Phacelia* sect. *Phacelia* clade and the ITS analysis suggests that they are sister taxa (0.78 BPP, 80% BS) and deeply nested in a North American clade. *P. artemisioides* Griseb., *P. pinnatifida* Griseb. Ex. Wedd., *P. setigera* Phil., and *P. viscosa* Phil. belong to *Phacelia* sect. *Glandulosae.* They form a well-supported polytomy along with *P. vossii* N.D. Atwood for *ndhF* (1 BPP, 95% BS) and a well-supported clade for ITS (1 BPP, 100% BS) deeply nested in a large clade of western North American species. In the *ndhF* tree *P. affinis* A. Gray is sister to all other species of the *Phacelia* sect. *Euglypta* + *Miltitzia* clade, including South American *P. cumingii* (Benth.) A. Gray (1 BPP, 100 % BS). The ITS analysis suggests that *P. affinis* and *P. cumingii* are direct sister taxa in the *Phacelia* sect. *Euglypta* + *Miltitzia* clade (1 BPP, 90% BS).

In Namaceae, most *Nama* species are grouped together in one clade (1 BPP, 100% BS for *ndhF*; 1 BPP, 100% BS for ITS), except for *Nama rothrockii* A. Gray and *Nama lobbii* A. Gray. These two species are retrieved in the second clade with *Eriodictyon*, *Turricula* and *Wigandia*

(1 BPP, 98% BS for *ndhF*; 0.99 BPP, 93% BS for ITS). *Wigandia* is recovered as monophyletic with high support in both analyses (0.96 BPP, 98% BS for *ndhF*; 1 BPP, 100% BS for ITS).

Age estimates for Hydrophyllaceae and Namaceae are shown in Table 2. The median ages estimated for the nodes are placed within the range of the 95% HPD intervals. Our analysis suggests that Hydrophyllaceae and Namaceae started their diversification almost simultaneously, between late Cretaceous and middle Paleogene. Hydrophylleae is estimated as the oldest clade of Hydrophyllaceae, with a median age in the Paleocene. The crown node of Romanzoffieae was dated in early Eocene and the crown node of the DHHT clade in middle Eocene, albeit with large 95% HPD intervals. Among the South American *Phacelia* species, the crown node of south temperate to Mediterranean *P. cumingii* is retrieved with the highest age (4.3 Ma). The essentially Andean species of *Phacelia* sect. *Phacelia* follow with a crown node median age estimate of 3.3 Ma, whereas the deserticolous clade in *Phacelia* sect. *Glandulosae* is dated to a median age of 3.2 Ma for the crown node. Slightly younger ages are suggested for most well-supported clades of Namaceae (see Table 2). The crown node age of *Wigandia* is 25.4 Ma.

Ancestral area reconstructions of Hydrophyllaceae

The ancestral area reconstructions of Hydrophyllaceae resulted from the DEC analysis are presented in Fig. 3. The most recent common ancestor (MRCA) of Hydrophyllaceae probably occurred in North America. North America is suggested as the ancestral area of all three major clades of the family and virtually all major subclades (see Table 3). At least three dispersal events from North to South America were inferred. South America was first colonized by representatives of the basal grade of *Phacelia* sect. *Euglypta* and *Miltitzia. Phacelia affinis* as a lineage (current distribution in North America) had split from *P. cumingii* (current distribution in South America) ca. 8.6 Ma. Our ancestral area reconstruction indicates that the dispersal from North to South America occurred before the inferred split. South American representatives of *P.* sect. *Glandulosae* (*P. artemisioides*, *P. pinnatifida*, *P. setigera* and *P. viscosa*) form a well-supported clade which split from its North American sister group in the late Miocene (ca. 6.8 Ma). The most recent colonization of South America occurred in *P.* sect. *Phacelia* during late Pliocene (ca. 3.6 Ma) and could be the result of a single dispersal event. *Phacelia brachyantha* occurs only in South America while *P. secunda* in both North and South America. Due to lack of resolution in the *ndhF* topology and low sample size in the ITS dataset, the exact number of dispersal events remains to be confirmed.

Ancestral area reconstructions of Namaceae

The ancestral area reconstructions of Namaceae resulting from the DEC analysis are presented in Fig. 4. The MRCA of Namaceae probably occurred in North America. Both the MRCA of clade I and clade II likely occured in North America (see Table 4). The MRCA of *Wigandia* occured in North and Central America in the Oligocene (ca. 25 Ma) and the colonization of South America took place during the Oligocene – Miocene transition.

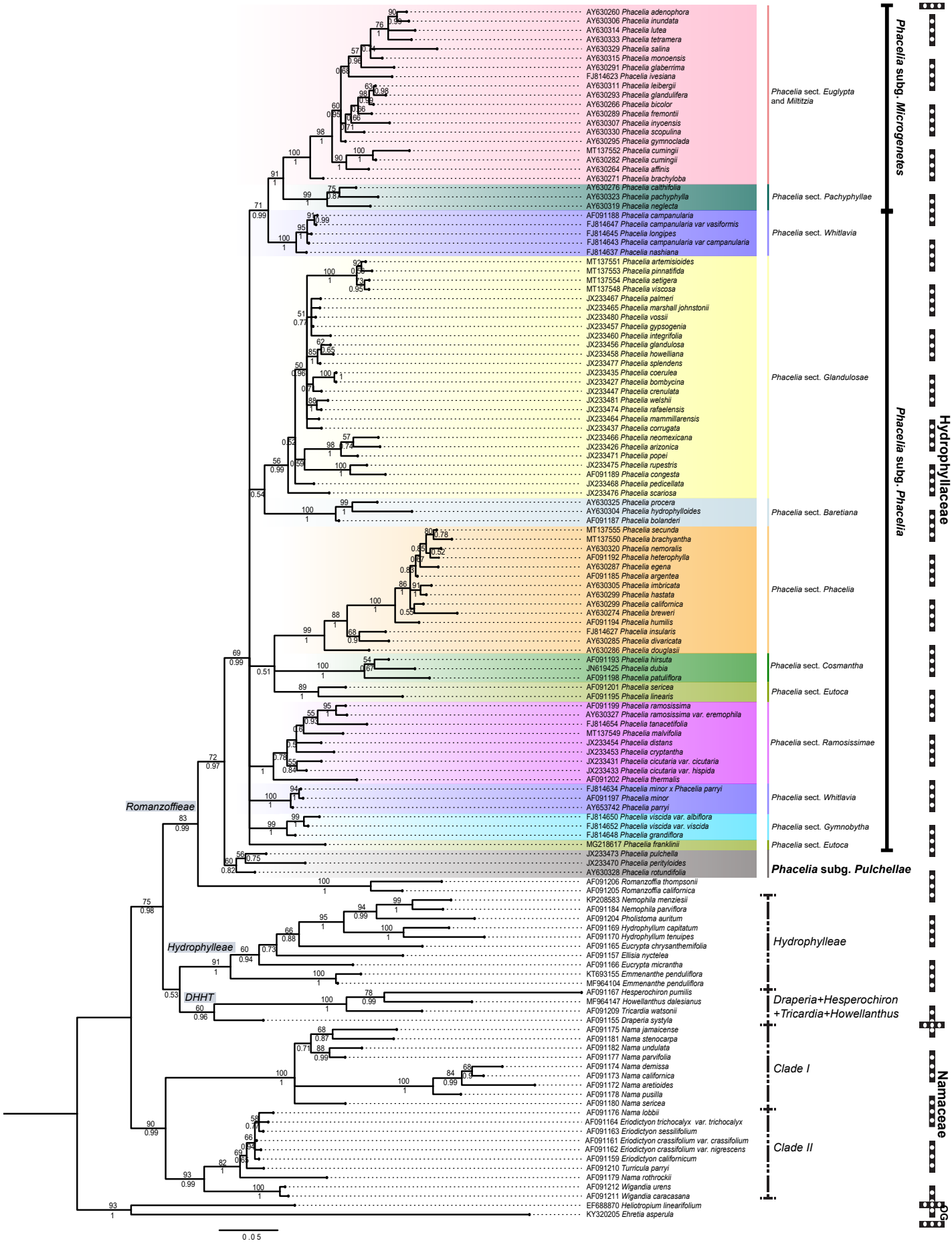

AY630260 Phacelia adenophora
AY630306 Phacelia inundata
AY630314 Phacelia lutea
AY630333 Phacelia tetramera
AY630329 Phacelia salina
AY630315 Phacelia monoensis
AY630291 Phacelia glaberrima
FJ814623 Phacelia ivesiana
AY630311 Phacelia leibergii
AY630293 Phacelia glandulifera
AY630266 Phacelia bicolor
AY630289 Phacelia fremontii
AY630307 Phacelia inyoensis
AY630330 Phacelia scopulina
AY630295 Phacelia gymnoclada
MT137552 Phacelia cumingii
AY630282 Phacelia cumingii
AY630264 Phacelia affinis
AY630271 Phacelia brachyloba
Phacelia sect. Euglypta and Miltitzia
AY630276 Phacelia calthifolia
AY630323 Phacelia pachyphylla
AY630319 Phacelia neglecta
Phacelia sect. Pachyphyllae
AF091188 Phacelia campanularia
FJ814647 Phacelia campanularia var vasiformis
FJ814645 Phacelia longipes
FJ814643 Phacelia campanularia var campanularia
FJ814637 Phacelia nashiana
Phacelia sect. Whitlavia
MT137551 Phacelia artemisioides
MT137553 Phacelia pinnatifida
MT137554 Phacelia setigera
MT137548 Phacelia viscosa
JX233467 Phacelia palmeri
JX233465 Phacelia marshall johnstonii
JX233480 Phacelia vossii
JX233457 Phacelia gypsogenia
JX233460 Phacelia integrifolia
JX233456 Phacelia glandulosa
JX233458 Phacelia howelliana
JX233477 Phacelia splendens
JX233435 Phacelia coerulea
JX233427 Phacelia bombycina
JX233447 Phacelia crenulata
JX233481 Phacelia welshii
JX233474 Phacelia rafaelensis
JX233464 Phacelia mammillarensis
JX233437 Phacelia corrugata
JX233466 Phacelia neomexicana
JX233426 Phacelia arizonica
JX233471 Phacelia popei
JX233475 Phacelia rupestris
AF091189 Phacelia congesta
JX233468 Phacelia pedicellata
JX233476 Phacelia scariosa
Phacelia sect. Glandulosae
AY630325 Phacelia procera
AY630304 Phacelia hydrophylloides
AF091187 Phacelia bolanderi
Phacelia sect. Baretiana
MT137555 Phacelia secunda
MT137550 Phacelia brachyantha
AY630320 Phacelia nemoralis
AF091192 Phacelia heterophylla
AY630287 Phacelia egena
AF091185 Phacelia argentea
AY630305 Phacelia imbricata
AY630299 Phacelia hastata
AY630299 Phacelia californica
AY630274 Phacelia breweri
AF091194 Phacelia humilis
FJ814627 Phacelia insularis
AY630285 Phacelia divaricata
AY630286 Phacelia douglasii
Phacelia sect. Phacelia
AF091193 Phacelia hirsuta
JN619425 Phacelia dubia
AF091198 Phacelia patuliflora
Phacelia sect. Cosmantha
AF091201 Phacelia sericea
AF091195 Phacelia linearis
Phacelia sect. Eutoca
AF091199 Phacelia ramosissima
AY630327 Phacelia ramosissima var. eremophila
FJ814654 Phacelia tanacetifolia
MT137549 Phacelia malvifolia
JX233454 Phacelia distans
JX233453 Phacelia cryptantha
JX233431 Phacelia cicutaria var. cicutaria
JX233433 Phacelia cicutaria var. hispida
AF091202 Phacelia thermalis
Phacelia sect. Ramosissimae
FJ814634 Phacelia minor x Phacelia parryi
AF091197 Phacelia minor
AY653742 Phacelia parryi
Phacelia sect. Whitlavia
FJ814650 Phacelia viscida var. albiflora
FJ814652 Phacelia viscida var. viscida
FJ814648 Phacelia grandiflora
Phacelia sect. Gymnobytha
MG218617 Phacelia franklinii
Phacelia sect. Eutoca
JX233473 Phacelia pulchella
JX233470 Phacelia perityloides
AY630328 Phacelia rotundifolia
Phacelia subg. Pulchellae
Phacelia subg. Microgenetes
Phacelia subg. Phacelia
Hydrophyllaceae
Romanzoffieae
AF091206 Romanzoffia thompsonii
AF091205 Romanzoffia californica
KP208583 Nemophila menziesii
AF091184 Nemophila parviflora
AF091204 Pholistoma auritum
Hydrophylleae
AF091169 Hydrophyllum capitatum
AF091170 Hydrophyllum tenuipes
AF091165 Eucrypta chrysanthemifolia
AF091157 Ellisia nyctelea
AF091166 Eucrypta micrantha
KT693155 Emmenanthe penduliflora
MF964104 Emmenanthe penduliflora
Hydrophylleae
DHHT
AF091167 Hesperochiron pumilis
MF964147 Howellanthus dalesianus
AF091209 Tricardia watsonii
AF091155 Draperia systyla
Draperia+Hesperochiron +Tricardia+Howellanthus
AF091175 Nama jamaicense
AF091181 Nama stenocarpa
AF091182 Nama undulata
AF091177 Nama parvifolia
AF091174 Nama demissa
AF091173 Nama californica
AF091172 Nama aretioides
AF091178 Nama pusilla
AF091180 Nama sericea
AF091176 Nama lobbii
Clade I
AF091164 Eriodictyon trichocalyx var. trichocalyx
AF091163 Eriodictyon sessilifolium
AF091161 Eriodictyon crassifolium var. crassifolium
AF091162 Eriodictyon crassifolium var. nigrescens
AF091159 Eriodictyon californicum
AF091210 Turricula parryi
AF091179 Nama rothrockii
AF091212 Wigandia urens
AF091211 Wigandia caracasana
Clade II
EF688870 Heliotropium linearifolium
KY320205 Ehretia asperula
Namaceae
OG
0.05

Fig. 2.2 Phylogram obtained from the Bayesian analysis based on ITS. Hydrophyllaceae and Namaceae clades and their sister genera *Heliotropium* and *Ehretia* (OG = outgroup) are shown. Numbers above branches are ML bootstrap values (≥50) and below branches Bayesian posterior probabilities (BPP). Scale bar = mean number of nucleotide substitutions per site. Major clades of *Phacelia* are indicated with different colours.

Table 2.2 Crown (C) and stem (S) node age estimates in major groups of Boraginales phylogeny. Time is given in millions of years (Ma) before present and in parentheses are noted the 95% highest posterior density intervals.

Node	Age estimates (Ma)
(C) H*ydrphyllum*+*Pholistoma*+*Nemophila*+*Ellisia*+*Eucrypta*+ *Emmenanthe* (Hydrophylleae)	58.2 (43.5-73.9)
(C) *Phacelia*+*Romanzoffia* (Romanzoffieae)	54.9 (41.6-68.5)
(C) *Draperia*+*Howellanthus*+*Hesperochiron*+*Tricardia* (DHHT)	47.9 (26.3-68.8)
(C) Hydrophyllaceae	66.1 (51.7-83.1)
(C) *Nama* (Clade I)	42.2 (27.5-59.8)
(C) *Wigandia* clade (Clade II)	25.4 (14.3-39.4)
(S) *Wigandia* clade (Clade II)	38.6 (23.9-58.3)
(C) *Nama*+*Eriodictyon*+*Turricula* (Clade II)	30.3 (15.6-47.4)
(C) Namaceae	66.8 (46-88)

2.4 Discussion

Phylogenetic analysis

Our results largely confirm previous studies on Hydrophyllaceae and Namaceae. We recovered similar *ndhF* topologies as Ferguson (1998) and Walden et al. (2014), and similar ITS topologies and support values as Gilbert et al. (2005), Hansen et al. (2009) and Walden et al. (2014). The plastid analysis resulted in better resolution regarding the sister relationships between major clades, whereas the nuclear analysis resulted in better resolution within the individual clades. Weaker support for deeper nodes in the ITS trees was also shown by Hansen et al. (2009) and Walden et al. (2014). However, divergent taxon sampling between the plastid and the nuclear analyses in this study may also explain some of the topological differences.

Hydrophyllaceae is divided into three major clades as shown before (Ferguson, 1998; Luebert et al., 2016; Walden, 2010). However, some of our results differ from previous studies. *Draperia* is well-supported sister to the DHHT clade in both analyses, which is probably a result of the increased sample size compared to Walden et al. (2014). The placement of *Emmenanthe* is incongruent between the ITS and *ndhF* phylogenies. In the ITS analysis *Emmenanthe*

is sister to the remaining Hydrophylleae clade, whereas in the *ndhF* analysis it forms a monophylum with *Eucryptha micrantha* (Torr.) A. Heller with low support (0.68 BPP, 56% BS), which is then sister to the remaining Hydrophylleae (Walden et al., 2014; see also Walden, 2015). *Eucrypta* is retrieved as paraphyletic (Ferguson, 1998; Walden, 2015). The relationship between *Phacelia* and *Romanzoffia* remains incongruent between the two markers. The monophyly of the genus *Phacelia* with *Romanzoffia* sister to it, is supported only by the ITS analysis (Gilbert et al., 2005; Hansen et al., 2009; Walden et al., 2014), whereas in the *ndhF* phylogeny *Romanzoffia* is nested in *Phacelia*.

Phacelia is currently subdivided in three subgenera, 11 sections and 12 subsections (Figs 1, 2; Walden & Patterson, 2012). The majority of these units were retrieved as monophyletic in both analyses. The paraphyly of *Phacelia* subg. *Pulchellae* in the *ndhF* analysis could not be confirmed in the ITS tree: no ITS-sequences were available for the relevant species *P. keckii* Munz & I.M. Johnst, *P. suaveolens* Greene and *P. cookei* Constance & Heckard. Sections *Euglypta* and *Miltitzia* were not monophyletic (see also Ferguson, 1998; Gilbert et al., 2005; Hansen et al., 2009; Walden et al., 2014), but they were retrieved in one clade exactly as in Walden et al. (2014). Section *Whitlavia* was paraphyletic in the nuclear analysis and consequently *Phacelia* subg. *Phacelia* was not retrieved as monophyletic (see also Hansen et al., 2009; Walden et al., 2014). The monophyly of section *Whitlavia* is supported in the *ndhF* analysis (also in Walden et al., 2014) and additionally in a combined *rpl16* intron + ITS dataset in Hansen et al. (2009). The placement of section *Baretiana* was incongruent in plastid and nuclear analyses, as in Walden et al. (2014). Section *Eutoca* was monophyletic only in the *ndhF* tree. Walden et al. (2014) retrieved a monophyletic section *Eutoca* for ITS without sampling *P. franklinii,* which in our case did not group together with the other two representatives of the section.

The South American *Phacelia* species are not monophyletic, but are placed in three different clades in both phylogenies. *P. viscosa, P. artemisioides, P. pinnatifida*, and *P. setigera* are placed in *Phacelia*. sect. *Glandulosae*. *Phacelia brachyantha* belongs to *Phacelia* subsect. *Phacelia* (not *Phacelia* subsect. *Humiles* as suggested by Walden and Patterson (2012)). The placement of *P. cumingii* in *Phacelia* sect. *Euglypta* and its sister relationship to the North American *P. affinis* is in congruence with previous studies (Gilbert et al., 2005; Hansen et al., 2009; Walden et al., 2014).

The topology of Namaceae retrieved in the present study is congruent with previous phylogenetic analyses (Ferguson, 1998; Luebert et al., 2016; Taylor, 2012). Our extensive sampling of *Wigandia* confirmed its monophyly.

Wigandia urens (Ruiz & Pav.) Kunth and *Wigandia caracasana* Kunth were retrieved as non-monophyletic in the *ndhF* tree. Cornejo (2006, 2007) recognizes six species in *Wigandia*, treating *W. caracasana* as a synonym of *W. urens* s.l., but this would still not render *W. urens* monophyletic. A critical revision of this widespread genus is urgently required.

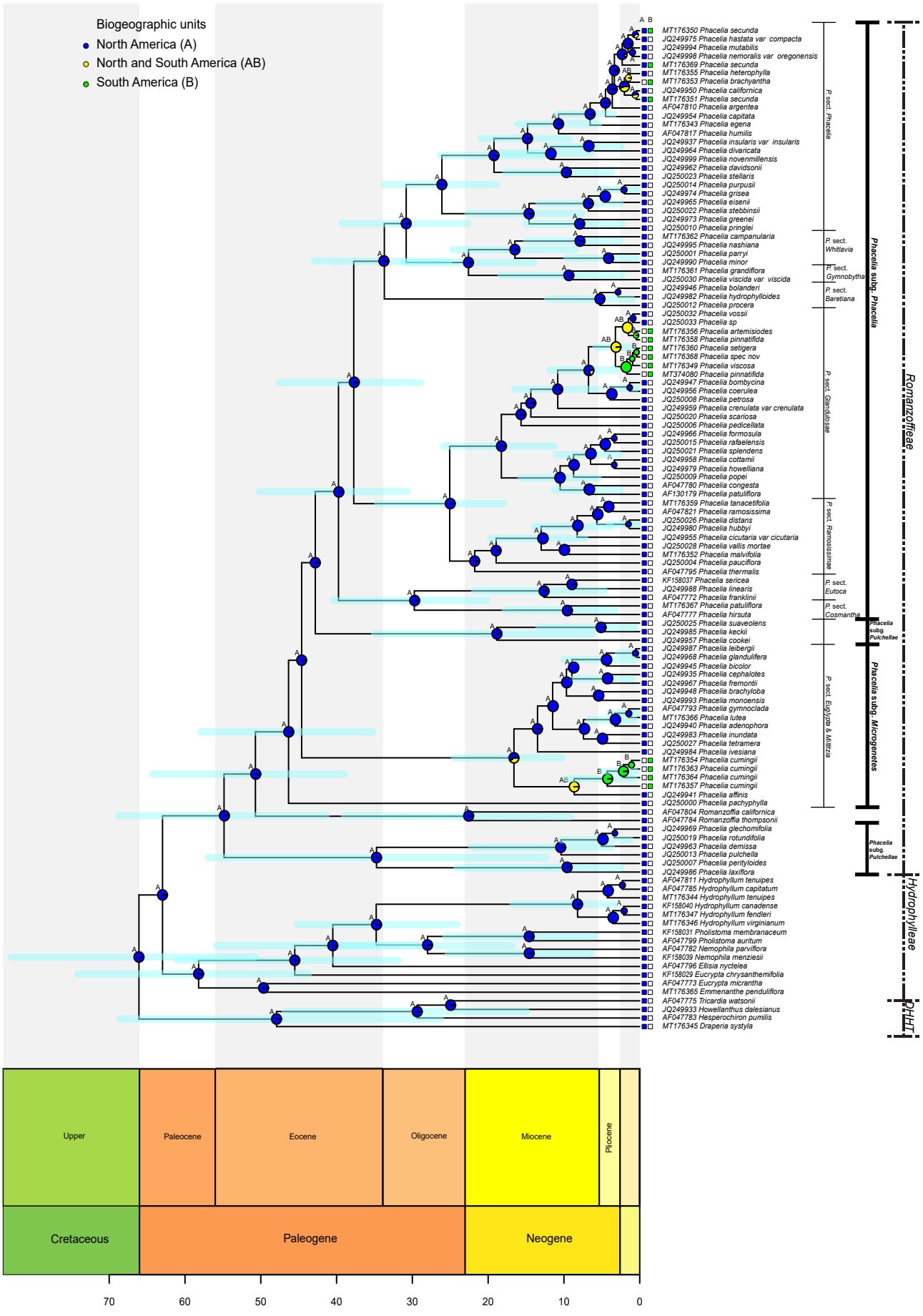

Biogeographic units
North America (A)
North and South America (AB)
South America (B)
MT176350 Phacelia secunda
JQ249975 Phacelia hastata var compacta
JQ249994 Phacelia mutabilis
JQ249998 Phacelia nemoralis var oregonensis
MT176369 Phacelia secunda
MT176355 Phacelia heterophylla
MT176353 Phacelia brachyantha
JQ249950 Phacelia californica
MT176351 Phacelia secunda
AF047810 Phacelia argentea
JQ249954 Phacelia capitata
MT176343 Phacelia egena
AF047817 Phacelia humilis
JQ249937 Phacelia insularis var insularis
JQ249964 Phacelia divaricata
JQ249999 Phacelia novenmillensis
JQ249962 Phacelia davidsonii
JQ250023 Phacelia stellaris
JQ250014 Phacelia purpusii
JQ249974 Phacelia grisea
JQ249965 Phacelia eisenii
JQ250022 Phacelia stebbinsii
JQ249973 Phacelia greenei
JQ250010 Phacelia pringlei
MT176362 Phacelia campanularia
JQ249995 Phacelia nashiana
JQ250001 Phacelia parryi
JQ249990 Phacelia minor
MT176361 Phacelia grandiflora
JQ250030 Phacelia viscida var viscida
JQ249946 Phacelia bolanderi
JQ249982 Phacelia hydrophylloides
JQ250012 Phacelia procera
JQ250032 Phacelia vossii
JQ250033 Phacelia sp
MT176356 Phacelia artemisiodes
MT176358 Phacelia pinnatifida
MT176360 Phacelia setigera
MT176368 Phacelia spec nov
MT176349 Phacelia viscosa
MT374080 Phacelia pinnatifida
JQ249947 Phacelia bombycina
JQ249956 Phacelia coerulea
JQ250008 Phacelia petrosa
JQ249959 Phacelia crenulata var crenulata
JQ250020 Phacelia scariosa
JQ250006 Phacelia pedicellata
JQ249966 Phacelia formosula
JQ250015 Phacelia rafaelensis
JQ250021 Phacelia splendens
JQ249958 Phacelia cottamii
JQ249979 Phacelia howelliana
JQ250009 Phacelia popei
AF047780 Phacelia congesta
AF130179 Phacelia patuliflora
MT176359 Phacelia tanacetifolia
AF047821 Phacelia ramosissima
JQ250026 Phacelia distans
JQ249980 Phacelia hubbyi
JQ249955 Phacelia cicutaria var cicutaria
JQ250028 Phacelia vallis mortae
MT176352 Phacelia malvifolia
JQ250004 Phacelia pauciflora
AF047795 Phacelia thermalis
KF158037 Phacelia sericea
JQ249988 Phacelia linearis
AF047772 Phacelia franklinii
MT176367 Phacelia patuliflora
AF047777 Phacelia hirsuta
JQ250025 Phacelia suaveolens
JQ249985 Phacelia keckii
JQ249957 Phacelia cookei
JQ249987 Phacelia leibergii
JQ249968 Phacelia glandulifera
JQ249945 Phacelia bicolor
JQ249935 Phacelia cephalotes
JQ249967 Phacelia fremontii
JQ249948 Phacelia brachyloba
JQ249993 Phacelia monoensis
AF047793 Phacelia gymnoclada
MT176366 Phacelia lutea
JQ249940 Phacelia adenophora
JQ249983 Phacelia inundata
JQ250027 Phacelia tetramera
JQ249984 Phacelia ivesiana
MT176354 Phacelia cumingii
MT176363 Phacelia cumingii
MT176364 Phacelia cumingii
MT176357 Phacelia cumingii
JQ249941 Phacelia affinis
JQ250000 Phacelia pachyphylla
AF047804 Romanzoffia californica
AF047784 Romanzoffia thompsonii
JQ249969 Phacelia glechomifolia
JQ250019 Phacelia rotundifolia
JQ249963 Phacelia demissa
JQ250013 Phacelia pulchella
JQ250007 Phacelia perityloides
JQ249986 Phacelia laxiflora
AF047811 Hydrophyllum tenuipes
AF047785 Hydrophyllum capitatum
MT176344 Hydrophyllum tenuipes
KF158040 Hydrophyllum canadense
MT176347 Hydrophyllum fendleri
MT176346 Hydrophyllum virginianum
KF158031 Pholistoma membranaceum
AF047799 Pholistoma auritum
AF047782 Nemophila parviflora
KF158039 Nemophila menziesii
AF047796 Ellisia nyctelea
KF158029 Eucrypta chrysanthemifolia
AF047773 Eucrypta micrantha
MT176365 Emmenanthe penduliflora
AF047775 Tricardia watsonii
JQ249933 Howellanthus dalesianus
AF047783 Hesperochiron pumilis
MT176345 Draperia systyla
P. sect. Phacelia
P. sect. Whitlavia
P. sect. Gymnobytha
P. sect. Baretiana
P. sect. Glandulosae
P. sect. Ramosissimae
P. sect. Eutoca
P. sect. Cosmantha
P. sect. Euglypta & Miltitzia
Phacelia subg. Phacelia
Phacelia subg. Pulchellae
Phacelia subg. Microgenetes
Phacelia subg. Pulchellae
Romanzoffieae
Hydrophylleae
DHHT
Upper
Paleocene
Eocene
Oligocene
Miocene
Pliocene
Cretaceous
Paleogene
Neogene
70
60
50
40
30
20
10
0

Fig. 2.3 Divergence time estimates for the phylogeny of Hydrophyllaceae. Major clades are indicated. The distribution corresponding to each species is indicated by the coloured squares next to the tips. Pie charts at the nodes depict relative probabilities of areas as estimated from the Dispersal-Extinction-Cladogenesis (DEC) analysis with BioGeoBEARS (unconstrained, single-area analysis). Significant areas (relative probability > 0.05) are also indicated next to the pie charts (for details see Table 3). Bars around nodes are 95% highest posterior density intervals.

Table 2.3 Results of the Dispersal-Extinction-Cladogenesis (DEC) analysis for the major crown (C) or stem (S) nodes of the phylogeny of Hydrophyllaceae. The log-likelihood of the analysis is lnL = -43. Letters indicate the ancestral areas at the node and the numbers in parenthesis are their relative probabilities. Displayed are only reconstructions with relative probabilities (RP) ≥ 0.05. A: North America, B: South America.

Node	**Unconstrained, total area lnL = -43**
(C) Hydrophyllaceae	A (0.99)
(C) *Draperia+Howellanthus+Hesperochiron+Tricardia* (DHHT)	A (1)
(C) Hydrophylleae	A (0.99)
(C) Romanzoffieae	A (0.99)
(S) Romanzoffieae	A (0.99)
(C) *Phacelia* sect. *Miltitzia* & *Euglypta*	A (0.6); AB (0.4)
(C) *P. affinis* & *P. cumingii* clade	AB (1)
(C) *P. cumingii* clade	B (1)
(C) *Phacelia* sect. *Glandulosae*	A (1)
(C) *P. vossii*, *P. artemisiodes*, *P. pinnatifida*, *P. setigera*, *P. viscosa* clade	AB (0.99)
(C) *P. viscosa* & *P. setigera* clade	B (1)
(C) *P. vossii*, *P. artemisiodes*, *P. pinnatifida* clade	AB (1)
(C) *Phacelia* sect. *Phacelia*	A (1)
(C) *P. brachyantha* & *P. heterophylla* clade	AB (1)

The odd placement of two species of *Nama* (*Nama rothrockii* and *N. lobbi*) further underlines the need for generic alignments in Namaceae. Taylor (2012) already indicated the isolation of the latter two species, with *N. lobbii* (= *Eriodictyon lobbii* Greene) nested in *Eriodictyon* and *Nama rothrockii* as sister to *Eriodictyon*. An expansion of the genus *Eriodictyon* to include both species, as well as *Tirricula parryi* which appears nested within *Eriodictyon* clade (Figs 1, 2), would thus be the simplest solution to resolve the taxonomy.

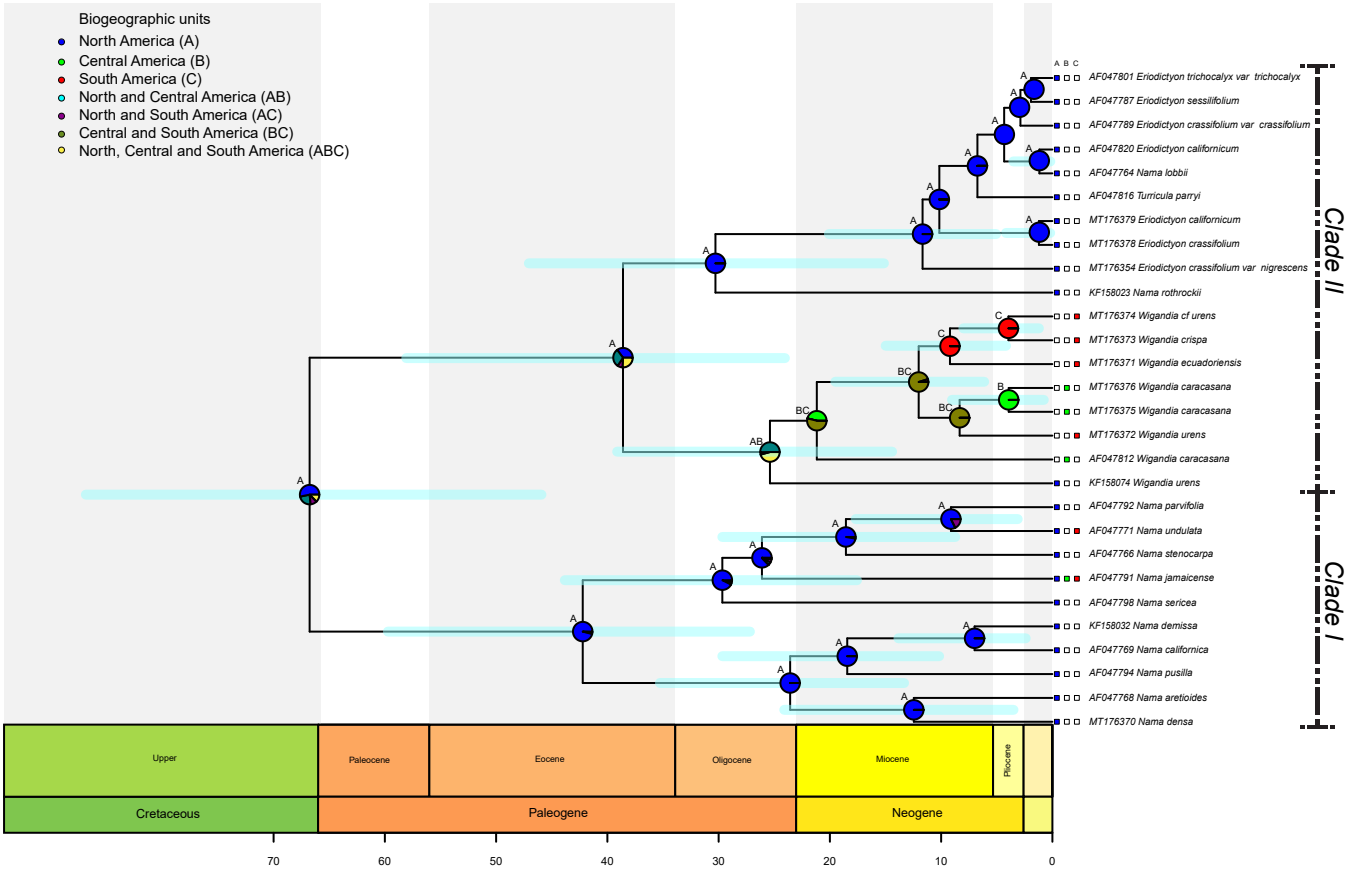


Fig. 2.4 Divergence time estimates for the phylogeny of Namaceae. Major clades are indicated. The distribution corresponding to each species is indicated by the coloured squares next to the tips. Pie charts at the nodes depict relative probabilities of areas as estimated from the Dispersal-Extinction-Cladogenesis (DEC) analysis with BioGeoBEARS (unconstrained, single-area analysis). Significant areas (relative probability > 0.05) are also indicated next to the pie charts (for details see Table 4). Bars around nodes are 95% highest posterior density intervals.

Table 2.4 Results of the Dispersal-Extinction-Cladogenesis (DEC) analysis for the major crown nodes (C) of the phylogeny of Namaceae. The log-likelihood of the analysis is lnL = -20.93. Letters indicate the ancestral areas at the node and the numbers in parenthesis are their relative probabilities. Displayed are only reconstructions with relative probabilities (RP) ≥ 0.05. A: North America, B: Central America, C: South America.

Node	**Unconstrained, total area lnL = -20.93**
(C) Namaceae	A (0.53); AB (0.23); AC (0.11); ABC (0.13)
(C) Clade I (*Nama* species)	A (0.96)
(C) Clade II (Shrubby genera)	A (0.35); AB (0.30); AC (0.08); ABC (0.26)
(C) *Wigandia*	AB (0.50); AC (0.04); ABC (0.46)
(C) *Eriodictyon*, *Nama*, *Tirricula*	A (1)

Divergence times and historical biogeography

We obtained similar age estimates to previous studies (Bremer et al., 2004; Luebert et al., 2011, 2017) for the stem node of Boraginales (112.7 (95% HPD 96.4-129 Ma)). The age estimates for the crown nodes of Hydrophyllaceae (66.1 (95% HPD 51.7-83.1 Ma)) and Namaceae (66.8 (95% HPD 46-88 Ma)) are dated to the early Paleocene whereas Luebert et al. (2017) suggested that both families may have diversified in parallel during the Middle Eocene. This difference is probably due to the expanded taxon sampling for these families in our study, or due to the use of a single marker (*ndhF*). Reduced ingroup sampling has been shown to lead to lower age estimates (Linder et al., 2005; Luebert et al., 2017; Soares & Schrago, 2015).

Our expanded sampling of Hydrophyllaceae and Namaceae provides the first detailed insights into their historical biogeography. They both appear to have originated in North America. Our results are consistent with the general trend that amphitropical disjunctions are mainly the outcome of long distance dispersal (LDD) during the Miocene to Holocene and that the most common directionality is from North to South America (Raven, 1963; Simpson et al., 2017; Wen and Ickert-Bond, 2009). Direct dispersal by migratory birds may have occurred by epizoochory or endozoochory.

Hydrophyllaceae first colonized South America during the early to middle Miocene (*Phacelia* sect. *Euglypta*) and at least two more dispersal events occurred later, one in the late Miocene (*Phacelia* sect. *Glandulosae*) and one in the Pliocene (*Phacelia* sect. *Phacelia*). *P. vossii,* which occurs in North America, groups together with the South American species of *P.* sect. *Glandulosae* with low support and no resolution, indicating the possibility of a back-dispersal to North America. The phylogenetic relationships of *P. secunda* and *P. brachyantha* in the clade of *P.* sect. *Phacelia* are not well-resolved. and South American divergence and subsequent back-dispersal can not be excluded. *P. secunda* could be an ancestral paraphyletic species or a species complex with cryptic taxa embedded.

Independent of the individual timing and the taxa involved, long-distance dispersal must be assumed for the exchange between North and South America. Although some species of *Phacelia* show no obvious dispersal mode (Heckard, 1963), a range of dispersal mechanisms are realized in Hydrophyllaceae and Namaceae, including myrmecochory (Hofmann et al., 2016) and epizoochory (Heckard, 1963; Schenk & Saunders, 2017). However, very little is known about the real dispersal mechanisms in Hydrophyllaceae and Namaceae in general, although many species, especially in Namaceae (*Wigandia, Eriodictyon*) have dust-fine, wind-dispersed seeds (Hofmann et al., 2016). Especially *Wigandia* and many species of *Nama* and *Phacelia* are early colonizing species with naturally small, disjunct populations, depending on regular dispersal for their preservation. *Wigandia* is thus common on road-cuttings in the Andes and *Phacelia viscosa* is commonly found on recent landslides and scree-slopes. *Nama dichotomum* is an extremely widespread weed on roadsides and fallows. All these species evidently have high dispersability. Due to its broad ecological range and tropical nature, stepping stone dispersal across much of its current range is thus likely for *Wigandia* – which is fully tropical and present in tropical Central America – ever since the development of the Isthmus of Panama (15–13 Ma; Coates & Stallard, 2013; Hoorn & Flantua, 2015; or even earlier: Bacon et al., 2013) and possibly also for *Nama dichomtomum.* Based on ecological preferences of extant species, this can be ruled out for *Pha-*

celia, which today has a large distribution gap in Central America. For this genus, only epizoochory can be invoked (Heckard, 1963; Schenk & Saunders, 2017).

Phacelia shows an amphitropical disjunct distribution and has been listed as an example of temperate disjunctions (Raven, 1963; Simpson et al., 2017; Wen and Ickert-Bond, 2009), although desert and temperate American amphitropical disjunctions (AAD) are occasionally difficult to differentiate (Simpson et al., 2017). We suggest that the case of *P. affinis* and *P.cumingii* represents a desert disjunction (dating to 8.64 Ma bp) based on current distribution ranges (see details about AAD bioregions in Simpson et al., 2017) and divergence times. The mean divergence time of desert AAD has been found to be 5.96 Ma, whereas that of temperate AAD is 2.76 Ma (Simpson et al., 2017), which correlates to our dates for the respective clades in *Phacelia* (*P.* sect. *Euglypta* and *Miltitzia* ca. 8.64, *P.* sect. *Glandulosae* ca. 6.8 Ma, *P.* sect. *Phacelia* ca. 3.6 Ma).

Colonization of South America by Namaceae occurred earlier, during the Oligocene-Miocene transition or later. Species of the genus *Nama* have been also listed as examples of temperate disjunctions for which LDD has been suggested (Raven, 1963; Simpson et al., 2017; Wen & Ickert-Bond, 2009). For species in Nama with continuous distributions in the Neotropics, stepwise migration is a possible explanation.

Wigandia also shows a continuous distribution range from northern Mexico to northwestern South America and it is found in subtropical and tropical areas and in semiarid situations (Hofmann et al., 2016). The timing of the diversification in *Wigandia* coincides with the timing of clononization of South America during the Oligocene-Miocene transition, indicating that LDD may have to be invoked.

Luebert et al. (2017) suggested a minimum of one dispersal event in each Hydrophyllaceae and Namaceae. The present study demonstrates several dispersal events in the two families, at least three in Hydrophyllaceae and at least three in Namaceae, which took place across a wide chronological range and must be explained in the context of different biogeographical processes, with LDD as the overruling mechanism. Future research should focus on filling the sampling gaps that still exist in the phylogenies of both families and increasing phylogenetic resolution by adding more markers or employing next-generation sequencing. Most importantly, the taxonomic uncertainties, especially in South American *Phacelia* and *Wigandia,* need to be addressed in order to design an adequate sampling strategy.

2.5 Acknowledgments

The authors wish to thank the Botanic Gardens of the Rheinische Friedrich Wilhelms Universität Bonn for providing plant material. We greatfully acknowledge the plant collectors R. Acuña, S. Boyd, B. Erten, R. Halse, J. Miller, G. Sancho, A. Thiem, F. Trabert and L. Woodruff, as well as the curators of the herbaria B, BSB, MO and GH for permitting access to their collections. We also thank T. Joßberger and N. Schmandt for their assistance during this study and A. Cabral for her valuable comments. We are greatful to the editors and to Matt Guilliams and an anonymous reviewer for their valuable feedback on an earlier version of this manuscript. The

author M. A. Vasile further wish to thank the Bodossaki Foundation for funding her studies at Plant Sciences Master's program at Bonn University.

2.6 References

Bacon, C. D., Mora, A., Wagner, W. L., & Jaramillo, C. A. (2013). Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. *Botanical Journal of the Linnean Society*, *171*(1), 287–300. https://doi.org/10.1111/j.1095-8339.2012.01281.x

Bremer, K., Friis, E., & Bremer, B. (2004). Molecular phylogenetic dating of Asterid flowering plants shows early Cretaceous diversification. *Systematic Biology*, *53*(3), 496–505. https://doi.org/10.1080/10635150490445913

Chandler, M. E. J. (1961). *The Lower Tertiary floras of Southern England. I. Paleocene floras.* London Clay flora (supplement), The British Museum.

Coates, A. G., & Stallard, R. F. (2013). How old is the Isthmus of Panama? *Bulletin of Marine Science, 89*(4), 801–813. https://doi.org/info:doi/10.5343/bms.2012.1076

Cornejo, X. (2006). *Wigandia ecuadorensis* (Hydrophyllaceae), una Nueva Especie del Bosque Muy Seco Tropical al Occidente de Ecuador. *Novon: A Journal for Botanical Nomenclature, 16*(3), 324–327.

Cornejo, X. (2007). *Wigandia brevistyla*: una nueva Hydrophyllaceae de la Repüblica Dominicana. *Anales del Jardín Botánico de Madrid, 64*(1), 63–67.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*(1), 11-15.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Ferguson, D. M. (1998). Phylogenetic analysis and relationships in Hydrophyllaceae based on *ndhF* sequence data. *Systematic Botany*, *23*(3), 253–268. https://doi.org/10.2307/2419504

Gabel, M. L., Backlund, D. C., & Haffner, J. (1998). The Miocene macroflora of the northern Ogallala Group, northern Nebraska and southern South Dakota. *Journal of Paleontology*, *72*(2), 388–397. https://doi.org/10.1017/S0022336000036362

Gernhard, T. (2008). The conditioned reconstructed process. *Journal of Theoretical Biology*, *253*(4), 769–778. https://doi.org/10.1016/j.jtbi.2008.04.005

Gilbert, C., Dempcy, J., Ganong, C., Patterson, R., & Spicer, G. S. (2005). Phylogenetic relationships within *Phacelia* subgenus *Phacelia* (Hydrophyllaceae) inferred from nuclear rDNA ITS sequence data. *Systematic Botany*, *30*(3), 627–634. https://doi.org/info:doi/10.1600/0363644054782251

Graham, A., & Jarzen, D. M. (1969). Studies in Neotropical paleobotany. I. The Oligocene communities of Puerto Rico. *Annals of the Missouri Botanical Garden*, *56*(3), 308–357. doi:10.2307/2394849

Hansen, D. R., Spicer, G. S., & Patterson, R. (2009). Phylogenetic relationships between and within *Phacelia* sections *Whitlavia* and *Gymnobythus* (Boraginaceae). *Systematic Botany*, *34*(4), 737–746. https://doi.org/info:doi/10.1600/036364409790139637

Hammouda, S. A., Weigend, M., Mebrouk, F., Chacón, J., Bensalah, M., Ensikat, H. J., & Adaci, M. (2015). Fossil nutlets of Boraginaceae from the continental Eocene of Hamada of Meridja (southwestern Algeria): the first fossil of the borage family in Africa. *American Journal of Botany*, *102*(12), 2108–2115. https://doi.org/10.3732/ajb.1500350

Heckard, L. R. 1963. The Hydrophyllaceae. The Quarterly Review of Biology 38: 117–123. https://doi.org/10.1086/403794

Hofmann, M., Walden, G. K., Hilger, H. H., & Weigend, M. (2016). Hydrophyllaceae. In: Kadereit JW, Bittrich V, eds. *Families and genera of vascular plants XIV*. Heidelberg: Springer, 221–238. https://doi.org/10.1007/978-3-319-28534-4_20

Hoorn, C., & Flantua, S. (2015). An early start for the Panama land bridge. *Science*, *348*, 186–187. https://doi.org/10.1126/science.aab0099

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J. et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, *574*, 679–685. https://doi.org/10.1038/s41586-019-1693-2

Linder, H. P., Hardy, C. R., & Rutschmann, F. (2005). Taxon sampling effects in molecular clock dating: An example from the African Restionaceae. *Molecular Phylogenetics and Evolution*, *35*(3), 569–582. https://doi.org/10.1016/j.ympev.2004.12.006

Luebert, F., Cecchi, L., Frohlich, M. W., Gottschling, M., Guilliams, C. M., Hasenstab-Lehman, K. E., Hilger, H. H., Miller, J. S., Mittelbach, M., Nazaire, M., Nepi, M., Nocentini, D., Ober, D., Olmstead, R. G., Selvi, F., Simpson, M. G., Sutorý, K., Valdés, B., Walden, G. K., & Weigend, M. (2016). Familial classification of the Boraginales. *TAXON*, *65*(3), 502–522. https://doi.org/10.12705/653.5

Luebert, F., Couvreur, T. L. P., Gottschling, M., Hilger, H. H., Miller, J. S., & Weigend, M. (2017). Historical biogeography of Boraginales: West Gondwanan vicariance followed by long-distance dispersal?. *Journal of Biogeography*, *44*(1), 158–169. https://doi.org/10.1111/jbi.12841

Luebert, F., Hilger, H. H., & Weigend, M. (2011). Diversification in the Andes: Age and origins of South American *Heliotropium* lineages (Heliotropiaceae, Boraginales). *Molecular Phylogenetics and Evolution*, *61*(1), 90–102. https://doi.org/10.1016/j.ympev.2011.06.001

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, *207*(2), 437–453. https://doi.org/10.1111/nph.13264

Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, *5*(4), 242-248. https://doi.org/10.21425/F5FBG19694

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. https://doi.org/10.1109/GCE.2010.5676129

Moore, M. J., & Jansen, R. K. (2006). Molecular evidence for the age, origin, and evolutionary history of the American desert plant genus *Tiquilia* (Boraginaceae). *Molecular Phylogenetics and Evolution*, *39*(3), 668–687. https://doi.org/10.1016/j.ympev.2006.01.020

Müller, K., Müller, J., Quandt, D., & Neinhuis, C. (2010). *PhyDe®— Phylo-genetic data editor.* (Version v0.9971). Retrieved from http://www.phyde.de/index.html

Nazaire, M., & Hufford, L. (2012). A broad phylogenetic analysis of Boraginaceae: Implications for the relationships of *Mertensia*. *Systematic Botany*, *37*(3), 758-783. https://doi.org/info:doi/10.1600/036364412X648715

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics (Oxford, England)*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Raven, P. H. (1963). Amphitropical relationships in the floras of North and South America. *The Quarterly Review of Biology*, *38*(2), 151–177. https://doi.org/10.1086/403797

Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, *57*(1), 4–14. https://doi.org/10.1080/10635150701883881

Reid, E. M., & Chandler, M. E. J. (1933). *The London Clay flora*. British Museum.

Schenk, J. J., and K. Saunders. 2017. Inferring long-distance dispersal modes in American amphitropically disjunct species through adaptive dispersal structures. American Journal of Botany 104: 1756–1764. https://doi.org/10.3732/ajb.1700178

Schneider, A. C., Braukmann, T., Banerjee, A., & Stefanović, S. (2018). Convergent plastome evolution and gene loss in holoparasitic Lennoaceae. *Genome Biology and Evolution*, *10*(10), 2663–2670. https://doi.org/10.1093/gbe/evy190

Simpson, M. G., Johnson, L. A., Villaverde, T., & Guilliams, C. M. (2017). American amphitropical disjuncts: Perspectives from vascular plant analyses and prospects for future research. *American Journal of Botany*, *104*(11), 1600–1650. https://doi.org/10.3732/ajb.1700308

Soares, A. E. R., & Schrago, C. G. (2015). The influence of taxon sampling on Bayesian divergence time inference under scenarios of rate heterogeneity among lineages. *Journal of Theoretical Biology*, *364*, 31–39. https://doi.org/10.1016/j.jtbi.2014.09.004

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stull, G. W., Stefano, R. D. de, Soltis, D. E., & Soltis, P. S. (2015). Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. *American Journal of Botany*, *102*(11), 1794–1813. https://doi.org/10.3732/ajb.1500298

Taylor, S. E. (2012). Molecular systematics and the origins of gypsophily in *Nama* L. (Boraginaceae). [Dissertation]. University of Texas.

Thomasson, J. R. (1987). Late Miocene plants from northeastern Nebraska. *Journal of Paleontology*, *61*(5), 1065–1079. https://doi.org/10.1017/S0022336000029413

Walden, G. K., Garrison, L. M., Spicer, G. S., Cipriano, F. W., & Patterson, R. (2014). Phylogenies and chromosome evolution of *Phacelia* (Boraginaceae: Hydrophylloideae) inferred from nuclear ribosomal and chloroplast sequence data. *Madroño*, *61*(1), 16–47. https://doi.org/10.3120/0024-9637-61.1.16

Walden, G. K., & Patterson, R. (2012). Nomenclature of subdivisions within *Phacelia* (Boraginaceae: Hydrophylloideae). *Madroño*, *59*(4), 211–222. https://doi.org/10.3120/0024-9637-59.4.211

Walden, G. K. (2010). Phylogeny of infrageneric relationships within *Phacelia* (Boraginaceae) inferred from chloroplast sequence data [M.Sc. Thesis]. San Francisco State University.

Walden, G. K. (2015). Systematics of *Emmenanthe*, *Eucrypta*, and *Phacelia* sect. *Ramosissimae* (Hydrophyllaceae; Boraginales). [Dissertation]. University of California, Berkeley.

Weeks, A., Baird, K. E., & McMullen, C. K. (2010). Origin and evolution of endemic Galápagos *Varronia* species (Cordiaceae). *Molecular Phylogenetics and Evolution*, *57*(2), 948–954. https://doi.org/10.1016/j.ympev.2010.08.014

Weigend, M., Luebert, F., Gottschling, M., Couvreur, T. L. P., Hilger, H. H., & Miller, J. S. (2014). From capsules to nutlets—phylogenetic relationships in the Boraginales. *Cladistics*, *30*(5), 508–518. https://doi.org/10.1111/cla.12061

Wen, J., & Ickert-Bond, S. M. (2009). Evolution of the Madrean–Tethyan disjunctions and the North and South American amphitropical disjunctions in plants. *Journal of Systematics and Evolution*, *47*(5), 331–348. https://doi.org/10.1111/j.1759-6831.2009.00054.x

Werle, E., Schneider, C., Renner, M., Völker, M., & Fiehn, W. (1994). Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Research*, *22*(20), 4354–4355. https://doi.org/10.1093/nar/22.20.4

Is it all *Phacelia pinnatifida*? Molecular delimitations and taxonomic revision of *Phacelia* sect. *Glandulosae* in South America based on ddRADseq



Chapter 3

Phacelia pinnatifida (Hydrophyllaceae) in cultivation in Bonn Botanical Gardens. Photographed by Maximilian Weigend. Seeds collected at Moquegua, Peru by M. Weigend.

Vasile MA, Böhnert T., O'Leary N., Weigend M. (In review).Is it all *Phacelia pinnatifida*?Molecular delimitations and taxonomic revision of *Phacelia* sect. *Glandulosae*in South America based on ddRADseq. Submitted to *Taxon*

Is it all *Phacelia pinnatifida*? Molecular delimitations and taxonomic revision of *Phacelia* Sect. *Glandulosae* in South America based on ddRADseq.

Maria-Anna Vasile[1], Tim Böhnert[1], Nataly O'Leary[2], Maximilian Weigend[1]

[1] *Nees Institute for Biodiversity of Plants, University of Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany*

[2] *Instituto de Botánica Darwinion, Labardén 200, CC 22, B1642HYD San Isidro, Buenos Aires, Argentina*

Address for correspondence: *Maria-Anna Vasile, m.vasile@uni-bonn.de*

Abstract

The *Phacelia* sect. *Glandulosae,* with ca. 62 taxa, is the most diverse of the 11 sections of the genus *Phacelia* (Hydrophyllaceae). Although the family is mainly restricted to North and Central America, *Phacelia* has an amphitropical disjunct group with nine species reported from western South America, five of them are assigned to belong to *P*. sect. *Glandulosae*. These species are found in arid to semi-arid Andean habitats between 1000 and 5000 m.a.s.l. The lack of resolution achieved through the analyses of previous molecular datasets, as well as the taxonomic chaos—lack of typifications, dozens of synonyms, different names used for the same taxon between different countries—have long confounded the species delimitation and taxonomy of this plant group in South America. Here, we adopted a ddRADseq strategy to investigate phylogenetic relationships of all South American representatives of *P*. sect. *Glandulosae*. A total of 33 specimens were sampled, and single nucleotide polymorphism (SNP) data were identified. Both ML and species tree inference analyses of the ddRADseq data resulted in a well resolved topology. All five species currently recognized in *P*. sect. *Glandulosae* in South America form a highly supported monophylum including *P. nana*, previously considered as belonging to *Phacelia* sect. *Euglypta*. The currently recognized morphospecies are, however, retrieved as poly- and/or paraphyletic. The different morphospecies and ecotypes do not represent defined monophyla and we propose recognizing only two species in South American Sect. *Glandulosae*: a single, polymorphic species *P. pinnatifida* and its sister species *P. artemisioides*. The morphological characters traditionally used to define species in sect. *Glandulosae* can be shown to be of little diagnostic value.

Key words american amphitropical disjunction; Andes; Boraginales; Hydrophyllaceae; species delimitation

3.1 INTRODUCTION

Andes, the longest continental mountain range in the world and one of the most fascinating mountain systems, is found in South America. These incredibly species-rich mountains (Mutke & Barthlott, 2005; Cai & al., 2022) are home to many rapid radiations (Lagomarsino & al., 2016) including the páramo as the fastest evolving biodiversity hotspot worldwide (Madriñán & al., 2013). It is increasingly recognized that geologically dynamic areas, like those undergoing active mountain uplift, sustain higher biodiversity than geologically stable areas, due to emergence of ecological gradients and

landscape and climatic shifts (Linder, 2008; Fjeldså & al., 2012; Hoorn & al., 2013). Biodiversity in mountain belts is determined by the interplay between migration, speciation and extinction (Hoorn & al., 2013). The tendency for organisms to maintain their environmental preferences over evolutionary time (e.g. niche conservatism; Antonelli, 2015) crucially affects these processes.

American amphitropical disjunctions (AAD) have contributed to the Andean biodiversity and diachronically stoked species richness. Species diversification rate can increase following an AAD event, likely driven by new niche availability and reduced competition (Hughes & Eastwood, 2006; Scherson & al., 2008; Simpson & al., 2017). Desert AAD plants, representing up almost 22% of the total AAD lineages, occur in the Atacama, Monte and Patagonian deserts, and probably the tropical dry forest (Chaco) of Argentina (Simpson & al., 2017). The origins and main migratory routes of Andean plants can be very complex to understand, but Pérez-Escobar et al. (2022) tried to address by what routes did the Andean floras assemble. This and a previous study (Antonelli & al., 2018) revealed that species from Cerrado and Chaco and the Patagonia Steppes – among other regions – colonized the Central Andes. This could indicate that there are cases where diversification of a plant group in Central Andes initiated after a desert AAD. It is obvious that dispersal events, adaptive radiation and gene flow in the American continent can make the evolutionary history of some plant groups extremely complicated, and even though these phenomena have always intrigued botanists, these topics have only recently started to be resolved with any detail by using modern molecular phylogenetic methods (Pennington & al., 2010; Särkinen & al., 2012; Böhnert & al., 2020, 2022; Lörch & al., 2021; Murillo-A. & al., 2022.) or by combining phylogenetic and Earth science models (Lagomarsino & al., 2016).

Phacelia Juss. is the largest and most diverse genus of Hydrophyllaceae (Luebert & al., 2016), currently subdivided in three subgenera, 11 sections and 12 subsections (Walden & Patterson, 2012), and comprising about 207 species (Hofmann & al., 2016). Hydrophyllaceae are largely restricted to North and Central America, however, *Phacelia* has an amphitropical disjunct group of nine species in western South America, occurring in arid to semi-arid Andean habitats between 1000 and 5000 m.a.s.l. (Deginani, 1982; Walden & al., 2014; Vasile & al., 2020). There is a single taxonomic revision available for the genus *Phacelia* in Argentina (Deginani, 1982), which covers all nine south American species. *Phacelia* sect. *Euglypta* S. Watson is represented by *P. cumingii* (Benth.) A.Gray, distributed across the Andean region of Chile and Argentina, and *P. nana* Wedd., in the Andes of Bolivia, Chile and Argentina. *Phacelia* sect. *Phacelia* is represented by *Phacelia secunda* J. F. Gmel. as the most widespread species ranging from the southwestern United States to Argentina along the American Cordillera, and *P. brachyantha* Benth. in Andean region of Chile and Argentina and Patagonia. In *Phacelia* sect. *Glandulosae* (Rydb.) Walden & R. Patt., a total of about 62 taxa are reported and a total of five species are recognized in South America (Figs. 3.1 & 3.2). *Phacelia artemisioides* Griseb. is reported from Argentina, Uruguay and southern Brazil, *P. pinnatifida* Griseb. ex Wedd. is reported from the Andes of Peru, Bolivia, Chile and Argentina, *P. sanzinii* Hicken is considered as endemic to Argentina, *P. setigera* Phil. is reported from the Andes of Chile and Argentina and *P. sinuata* Phil. is reported from Argentina and central Chile.

There has been a constant shift in species delimitation within *P.* sect. *Glandulosae* since the early 20th century, when Brand (1913) provided a family-wide revision of Hydrophyllaceae (incl. at that point Namaceae; Luebert & al., 2016). Names have been accepted or rejected within *P.* sect. *Glandulosae* based on important taxonomic characters such as leaf morphology, stamen-corolla tube modifications, sepal, gynoecium, fruit and seed morphology. Geographical distribution was often reflected in nomenclature. Apparently, very often different names were given for the same species among the different

countries e.g. *P. viscosa* Phil. from Chile and *P. boliviana* Brand from Bolivia, both currently considered as synonyms of *P. pinnatifida*. The problems were complicated by the lack of consistent typifications for many names until recently (Mirra & al., 2021). The concept of infraspecific names was often made the taxonomy even more complicated, as former varieties (e.g. *P. pinnatifida* var. *setigera* (Phil.) Reiche), were subsequently recognized as separate species, namely *P. setigera*. Today, a handful of formerly accepted species and/or varieties are placed under the oldest name *P. pinnatifida* (Weddell, 1857), but a critical, modern revision of the complicated group has not been provided.



Figure 3.1 A–N, Morphological diversity within *P. pinnatifida*; **A–H**, Morphotype I; **A**, B 100641749, Sama, Bolivia; **B**, MW9739, DNA Nr. W5599, Oyon, Peru; **C**, B 100270296, DNA Nr. W6080, Cuesta del Obispo, Argentina; **D–F**, *P. pinnatifida* in Canon de Colca, Arequipa, Peru; **E**, young rosette; **F**, flowers; **G–H**, *P. pinnatifida* in Chuquibamba, Peru; **H**, flowers; **I-L**, Morphotype II; **I**, MW9974, DNA Nr. W5636; **J**, habitat of MW9974 in Moquegua, Peru; **K**, flowers of MW9974; **L**, habitat of MW9991, DNA Nr. W5637, Moquegua, Peru; **M–N**, Morphotype III; **M,** B 100763983, DNA Nr. W6079, Lake Titicaca, Peru; **N**, Ferreyra 405 BONN, Azangaro Peru. All three morphotypes fall under the name *P. pinnatifida* following the new classification here proposed. Images D-H, J-L by M. Weigend.

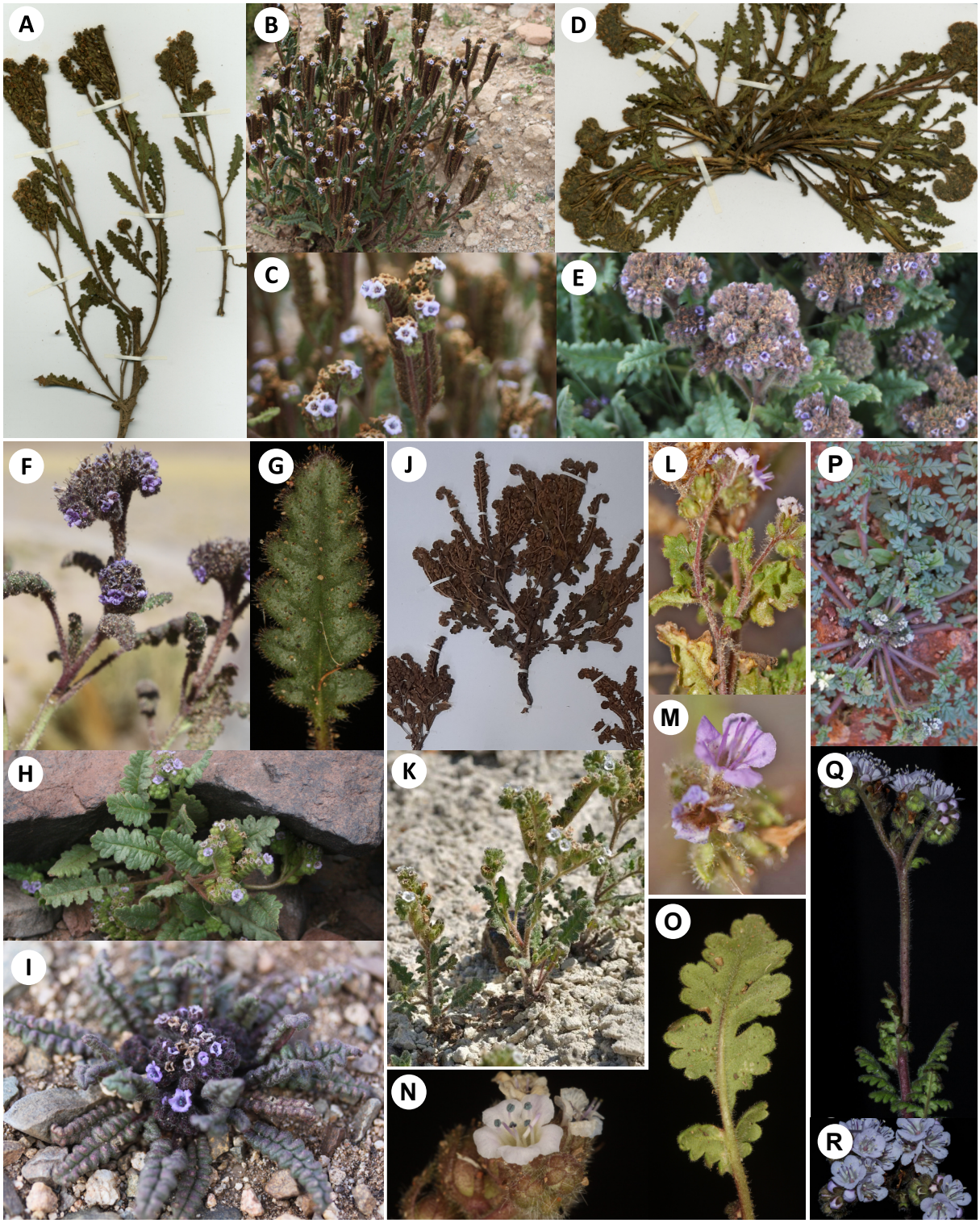


Figure 3.2 A–R, Morphological diversity of *P.* sect. *Glandulosae*; **A–O,** taxa synonymized in this study with *P. pinnatifida* (for more information regarding the nomenclature see the section of taxonomic treatment and regarding the vouchers see Table S1); **P–R,** *Phacelia artemisioides*, sister species of *P. pinnatifida s.l*; **A**–**E,** morphological diversity of *P. setigera* in Chile; **A–C**, *P. viscosa* Moreira & Luebert AM2410, Parinacota, Chile; **A**, AM2410 BONN; **B**, habitat; **C**, inflorescences; **D**–**E**, *P. setigera* Moreira & Luebert AM2440, Parinacota, Chile; **D,** AM2440 BONN; **E**, inflorescences and leaf morphology; **F-I**, morphological diversity of *P. setigera* in Argentina; **F**, Acosta 1046, DNA Nr. W6375, Catamarca, Argentina; **G**, Acosta 1046, leaf; **H**, Aagesen 12082, DNA Nr. W6370, habitat, La Rioja, Argentina; **I**, *P. setigera* var. *humahuaquense* Nicola 13528, DNA Nr. W6371, Jujuy, Argentina; **J**–**K**, *P. sinuata;* **J**, Brigone 31 SI, Mendoza, Argentina; **K**, habitat, Chubut, Argentina; **L–M**, *P. sanzinii* Deginani 2301, Mendoza, Argentina; **N–O**, *P.* nana Acosta 1159, DNA Nr. W6373, Jujuy, Argentina; **N**, flowers; **O**, leaf; **P–R**, *P. artemisioides*; **P**, habitat, Zuloaga 14955, La Rioja, Argentina; **Q**, Sanguinetti 177, Maldonado, Uruguay leaves and inflorescences; **R**, Sanguinetti 177, flowers. Images B, C, E by F. Luebert; F, G, N, O by C.A. Zanotti; H, I, P by F. Zuloaga; L, M by N. Deginani, K by D. Quintana; Q, R by A. Sanguinetti. Images A–E were obtained by Nees Institute for Biodiversity of plants, while F–R were obtained by Instituto de Botánica Darwinion.

The traditional taxonomic framework cannot be replaced (Ebach & al., 2011; Crisci & al., 2020; Baldini & al., 2021). Molecular taxonomy has its limits, e.g. due to the low genetic divergence in recently diverged species (Rouhan & Gaudeul, 2021), but pairing morphology-based taxonomy with molecular phylogenetics is highly efficient in inferring complex evolutionary processes more precisely (Pennington, 1996; Beier & al., 2003; Scotland & al., 2003; Martynov, 2012). Modern tools, such as Next Generation Sequencing techniques, for example Whole Genome Sequencing, can facilitate the process of identifying and describing new species (Hajibabaei & al., 2007; Rouhan & Gaudeul, 2021; Smith & Carstens, 2022). High-throughput techniques based on restriction site-associated DNA sequencing (RADseq) makes it possible to resolve a wide range of ecological, evolutionary and conservation-related questions (Lemmon & Lemmon, 2013; Andrews & al., 2016; Liu & al., 2020; Vaux & al., 2023). Furthermore, double-digest restriction site-associated DNA sequencing (ddRADseq) is applicable specifically for shallow-level phylogenetic and phylogeographic questions, population structure and introgression (Vaux & al., 2023) and has been successfully applied using silica-dried and herbarium tissue (Jordon-Thaden & al., 2020; Lang & al., 2020), making this method potentially suitable for taxonomic revisions that require a phylogenetic framework where morphological characters are difficult to assess (Prata & al. 2018; Böhnert et al. 2023). The phylogenetic relationships of the genus *Phacelia* have been addressed by few studies with a limited sampling, using chloroplast and nuclear markers (Ferguson, 1998; Gilbert & al., 2005; Hansen & al., 2009). Walden et al. (2014) significantly expanded the sampling of these studies, including approximately 43% of the species of the genus, but only two South American species. Vasile et al. (2020), in an attempt to address the historical biogeography of Hydrophyllaceae and Namaceae and the phylogenetic position of the South American representatives of *Phacelia* and *Wigandia* Kunth, supplemented previously published datasets including four more, but not all South American species. Vasile & al. (2020) found at least three independent dispersal events to South America, one of them in *P.* sect. *Glandulosae*, but the phylogenetic relationships of the South American species within *P.* sect. *Glandulosae* remained unresolved.

We believe that morphology alone is likely not enough for understanding speciation and species delimitation in this recently diverged clade (ca. 3 Mya according to Vasile & al., 2020). Based on numerous field studies by one of us (MW) we assumed that current taxonomy does not accurately reflect the diversity of *Phacelia* sect. *Glandulosae* in South America. Several collections from southern Peru appeared to represent morphotypes not covered by the named species in this group and the present study aimed at clarifying their phylogenetic position relative to the described species. Based on the field

studies and a comparison to herbarium material, we expected the morphotypes to fall into distinct clades, representing the annual, mid-elevation morphotypes and the perennial, high-elevation and temperate morphotypes, with further subdivision reflecting, e.g. leaf dissection as a classically used diagnostic character. To test our hypothesis, we used a ddRADseq strategy to investigate the phylogenetic relationships of all South American representatives of *P.* sect. *Glandulosae* and recognize the monophyla in this group. In a second step, we expected to revise species limits and provide the necessary taxonomic changes.

3.2 MATERIALS AND METHODS

Taxon Sampling. —Based on geographical distribution as well as availability of herbarium collections of the south American taxa of *P.* sect. *Glandulosae*, 48 herbarium specimens from BONN, B, HUSA, LP and SI were used for the evaluation of morphological characters and the taxonomic revision. The specimens include two representatives of *P. artemisioides,* one of *P. sinuata*, one of *P. sanzinii*, eight of *P. setigera* including one *P. setigera var. humahuaquense* Deginani and 36 of the *P. pinnatifida* complex including multiple specimens which could not be clearly assigned to any of the formally described taxa (determined as "*P. spec. nov.*" by F. Luebert and M.-A. Vasile in 2019 and "*P. cf. viscosa*" by M. Weigend). Depending on collection year, locality and state of the specimens, 28 samples were selected for DNA extraction, representing all the currently accepted taxa. As outgroup, we selected 10 species, representing five additional sections, including all remaining South American species (*P. cumingii* and *P. nana* from *P.* sect. *Euglypta*, *P. campanularia* A. Gray from *P.* sect. *Whitlavia* (Harv.) Benth. & Hook.f., *P. egena* (Brand) Greene ex C.F.Baker*, P. secunda* and *P. brachyantha* from *P.* sect. *Phacelia*, *P. tanacetifolia* Benth. and *P. malvifolia* Cham. & Schltdl. from *P.* sect*. Ramosissimae* (Rydb.) Walden & R. Patt., *P. grandiflora* (Benth.) A.Gray from *P.* sect*. Gymnobytha* (A. de Candolle) Benth. & Hook.f.) and *Hydrophyllum virginianum* L. as a representative of the sister clade to the genus *Phacelia* and outgroup. All information about vouchers and DNA numbers is provided in Table S1. In regard to the taxonomic revision, type specimens from all 6 currently accepted taxa (5 species and 1 variety), as well as those from previously accepted and infraspecific species as well as those of *P. nana* were examined through JSTOR Global Plants (https://plants.jstor.org). Additionally, virtual herbarium specimens were also examined through JSTOR Global Plants and the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/).

DNA extraction and ddRADseq Data Acquisition. —Plant tissue from herbarium collections and whenever available silica dried leaves was used for the extraction of genomic DNA (see Table S1), applying a modified CTAB method (Doyle & Doyle, 1987) with increased incubation time of up to 2 hours. High quality genomic DNA is needed for the ddRADseq method, therefore initial quality control and a rough quantification of DNA concentration was carried out using electrophoresis on 1 % agarose gels using Lonza GelStar Nucleic Acid Gel Stain (100x; Lonza Bioscience, Basel, Switzerland) and lambda DNA (250µg, Promega, Madison, Wisconsin, U.S.A.). For some samples we repeated DNA extractions and pooled them together, aiming to increase the DNA concentration, except for *P. sinuata* for which two samples for the same accession were retained separately due to the use of different type of material (leaf and stem) and differences in concentration, respectively. Five samples failed the initial quality control. The remaining samples were quantified using Quantus Fluorometer (Promega, Madison, Wisconsin, U.S.A.). Thirty-four samples (ca. 600 ng genomic DNA per sample), ten of which belong to the outgroup, were sent to a sequencing service (LGC Genomics GmbH, Biosearch Technologies, Berlin, Germany) for library preparation and sequencing. Following Böhnert & al. (2023), the enzyme

combination *Pst*I and *Msp*I were used. An Illumina NextSeq 500/550 v2 (Illumina, San Diego, California, U.S.A.) was used for sequencing. Short single-end reads of 75 bp were generated and subsequently deposited at the European Nucleotide Archive (ENA) under the accession xxxxx. Demultiplexing of the sequencing libraries was done with the Illumina bcl2fastq v2.20 conversion software (Illumina, San Diego, California, U.S.A.). Mismatches of Ns were not allowed for the barcodes except for the restriction site. Reads with final lengths <20 bases and reads with 5′ ends not matching the restriction enzyme were discarded. Finally, FastQC v.0.11.9 (Andrews & al., 2012) was used for evaluating raw reads quality and checking the presence of remaining adapter sequences.

Assembling ddRAD Sequences. —De novo assembly was performed in ipyrad v.0.9.84 (Eaton & Overcast, 2020), an interactive assembly and analysis toolkit that includes most of the analytical tools used in this study. The default parameters were used, except for the datatype, which was set to ddRAD, and the restriction overhangs, which were specified for the combination of restriction enzymes used here. Moreover, during the 7 steps of the ipyrad assembly workflow, we branched our data after step 2 in order to test three different clustering thresholds [85 % (default value), 90 %, 95 %], as this parameter specifies the level of sequence similarity when constructing loci. This resulted in three parallel data sets (data_clust85, data_clust90, data_clust95). After step 6 for each one of the three data sets, one sample (*P. setigera* var. *humahuaquense*, W6371) was excluded due to low consensus read number (< 1000). The amount of missing data allowed in the final alignments significantly influences the tree topology and the support values, thus allowing 60 to 90 percent in the final assembly step 7 can lead to more stable results (Huang & Knowles, 2016; Eaton & al., 2017; Tripp & al., 2017). Hence, each of the three datasets was branched again prior to step 7 in order to produce six final assemblies with 60, 65, 70, 75, 80 and 85 percent of missing data for the ingroup taxa, but not for the outgroup taxon (*Hydrophyllum virginianum*). This resulted in a total of 18 assemblies subsequently used for downstream analyses. Due to problematic topology (see phylogenetic inference results section below), we redefined the outgroup taxa for step 7 of the assembly and produced another two sets of 18 assemblies. For the second set, not only *H. virginianum*, but also all the *Phacelia* species which do not belong in *P.* sect. *Glandulosae*, plus the distantly related *P. artemisioides* were defined as outgroup (10 outgroup taxa, 23 ingroup taxa). *Phacelia nana* was considered as ingroup hereafter, as the first analysis demonstrated that it is nested within *P.* sect. *Glandulosae* with high support. For the third and last set, all the outgroup taxa were excluded from the assembly, except for *P. artemisioides* (1 outgroup taxon, 23 ingroup taxa). The downstream analyses of the third set are presented here, while those of the two first sets can be found in the Supplementary material S1–S3 (in Appendix B). Documentation of the assembly workflow can be found in a Jupyter-Notebook available in Appendix B.

Phylogenetic Inference. —To estimate phylogenetic relationships among *P.* sect. *Glandulosae* as well as to counter the effects of different data assembly protocols on phylogenetic inference, maximum likelihood (ML) tree searches were utilized on all three sets of 18 assemblies described in "Assembling ddRAD Sequences" section, while a quartet-based species tree approach under the multispecies coalescent model was utilized only on the third set of the assembly. The ML analysis was conducted using RAxML v.8.2.12 (Stamatakis, 2014) with 200 rapid bootstrap replicates for assessing the relationships' support and the GTRGAMMA substitution model. The coalescent analysis was conducted using TetRAD v.0.9.13 implemented in ipyrad v. 0.9.84. TetRAD employs the inference method originally developed in SVDquartets (Chifman & Kubatko, 2014) for inferring quartet trees using phylogenetic invariants, and joining those into a super tree. TetRAD, in comparison to the original program, allows bootstrap resampling and thus provides node support. Depending on the input data, either *H.*

virginianum or *P. artemisioides* was used to root the produced trees. Phylogenetic trees were visualized using the Python library Toytree v.2.0.5 (Eaton, 2020).

Species-level Analyses. —To evaluate the cases of polyphyly and paraphyly detected in the topologies recovered from both the concatenated maximum likelihood and the coalescence analyses, we first compared all the inferred phylogenies for the three sets of 18 assemblies and selected the best tree based on the stability of the topology and the support values. Among the trees inferred for the third set of 18 assemblies (23 taxa ingroup, *P. artemisioides* outgroup), we selected for both ML and TetRAD those datasets with 60% missing data and 90% clustering threshold. To further check the impact of missing data on species-level analysis, we produced an assembly with clustering threshold 90% and without specifying missing data and one including 50% missing data. STRUCTURE analysis performs more accurately if there is fewer missing data and when base calls are more accurate (Pritchard & al. 2010). Due to the higher number of unlinked SNPs recovered for the assembly with 50 % missing data included, the following analyses were performed with this data set. Despite the fact that *P. artemisioides* belongs to *P.* sect. *Glandulosae*, it was excluded from all species level-analyses due to its clear taxonomic identity (morphologically and phylogenetically).

Genetic structure

For examining the genetic structure among and within our ingroup taxa, we used STRUCTURE v.2.3.4 (Pritchard & al. 2010) implemented in ipyrad v.0.9.84. STRUCTURE analysis aims at delineating clusters of individuals on the basis of their genotypes at multiple loci using a Bayesian approach. (Evanno & al., 2005; Pritchard & al. 2000). Predefined populations are not a prerequisite, but do allow the examination of concordance between STRUCTURE results and expected groupings of individuals (Evanno & al., 2005). Definition of populations is usually based on geographical distribution of samples or phenotypes, even though none of these criteria necessarily reflects genetically differentiated populations. For example, not discretely distributed populations can nevertheless be genetically structured, due to unidentified barriers to gene flow (see more in Evanno & al., 2005). Here, we defined the populations according to the topology and the geographical distribution which is reflected on the well supported clades. Due to the wide range of phenotypic variation of the *P.* sect. *Gladulosae* in South America, phenotype is not an appropriate criterion for defining the populations. We created a python dictionary including seven well defined groups of populations, which are not incorporated in the analysis, but are only used for data filtering and results plotting. Randomly sampled SNPs shared among 50 % of the samples were allowed in the analysis in order to reduce linkage among SNPs ('mincov' and 'minmap' were set to 0.5). The burn-in was set to 100000, the number of replicates to 500000 and the number of runs per K to 10. We tested seven K-values (i.e. K= 2 to 8).

PCA

Principal component analysis (PCA) was performed in ipyrad v.0.9.84, aiming to identify and describe clusters of genetically related individuals. For filtering data and plotting results, we first used the same general grouping as for the STRUCTURE analysis with 50 % read coverage in each population. Subsequently, a second general grouping was used separating the samples in three groups considering only the geographical distribution (Peru, Chile, Argentina) and not the topology. For each grouping we generated a matrix of SNPs containing the genotype calls and the linkage information used for subsampling of unlinked SNPs and bootstrap resampling. Twenty-five replicates of the PCA analysis were performed, with a randomly sampled set of unlinked SNPs to reduce the effect of linkage from the data. The results of the first three dimensions (PC0, PC1, PC2) were visualized and shown in lower opacity each replicate and only the centroids of those replicates per sample are printed as solid points.

TreeMix

To estimate migration rate and detect hybridization we used TreeMix v.1.13 (Pickrell & Pritchard, 2012) implemented in ipyrad v.0.9.84, a tool which utilizes the allele frequencies in the modern populations. The individuals were grouped into populations in the same way as for the STRUCTURE analysis in order to filter the data and plot the results. The minimum percentage of samples that must be present in each SNP from each group was set to 50%. In order to find the best value for admixture edges (*m*), we evaluated the log likelihood score of each model fit for *m* = 0–9. The admixture edge for which the likelihood of trees reached an initial plateau and the magnitude of migration rates decreased (i.e. m = 3) was chosen. The basis for the resulting topology is a composite maximum likelihood of the local optimum tree as suggested by Felsenstein (1981), while the branch lengths are proportional to the amount of genetic drift per branch. According to Peter (2016), inference is based on "shared genetic drift" between sets of populations, under the premise that shared drift implies a shared evolutionary history. Due to the fact that TreeMix can be biased by its starting tree, especially when numerous populations are involved in admixture events (Lipson & al., 2013; Lipson, 2020), we further checked over nine different subsampled iterations by restarting the TreeMix analysis with different seed values. The constructed maximum likelihood trees which accommodate possible admixture events were rooted in *P. sinuata*, which is sister to all remaining samples according the results of the RAxML analysis.

Taxonomic revision. —All available specimens were examined with the use of stereomicroscope and in some cases a light microscope (e.g. for indument). The material is deposited in BONN, B, HUSA, LP and SI (see Table S1). Type specimens from all 6 currently accepted taxa of *P.* sect. *Glandulosae* (5 species and 1 variety), as well as those from previously accepted species and infraspecifics and of *P. nana* which belongs until now to *P.* sect. *Euglypta* were examined through JSTOR Global Plants. The collection locality of the specimens was georeferenced according to the available information on the labels. Previously published morphological data by Vasile et al. (2021) for *P. viscosa* (Luebert & Böhnert 3640, BONN) and *P. setigera* (Moreira & Luebert 2440, BONN) were re-examined and additionally seed micromorphology was studied for *P. artemisioides* (B100763933; B100763932), *P. pinnatifida* (Cordova et al. 025. HUSA, B100844687, B100267427), *P. viscosa* (Luebert & Böhnert 3640, BONN), *P. setigera* (M. Belov CL-0-BONN-34622), *P. nana* (B100763992) and *P. sinuata* (B100763870) following standard scanning electron microscopy preparations (Vasile & al., 2021).

3.3 RESULTS

De novo assembly and phylogenetic inference. —The sequencing of the ddRADseq library including 34 samples (24 ingroup, 10 outgroup) yielded a total of 46,740,966 raw reads, corresponding to an average of 1,374,734 reads per sample, while the median consensus read number per sample was 14,157. One sample was excluded due to low number of consensus reads (<1000). Different de novo assembly strategies (variation in the amounts of missing data and the clustering threshold parameter) had a clear impact on the phylogenetic reconstruction. The first two sets of assemblies, with either *H. virginianum* only, or *H. virginianum* and nine *Phacelia* species defined as outgroups, resulted in highly incongruent topologies recovered by the concatenated maximum likelihood analyses (see Supplementary Material S1 in Appendix B). However, it could be shown that *P. nana*, defined as part of the outgroup, actually belongs to the ingroup, i.e., *P.* sect. *Glandulosae* and not *P.* sect. *Euglypta*. Therefore, the ingroup dataset, including *P. nana,* was analyzed independently (third set of assemblies) with *P. artemisioides* as outgroup, which also belongs to *P.* sect. *Glandulosae,* but is both phylogenetically and morphologically clearly separate from the remainder of the ingroup-taxa. The 18 possible combinations

of parameter settings of the third set of assemblies yielded largely congruent topologies and support values for both the concatenated and the coalescent trees. Considering the stability of the topology, especially that of the backbone, as well as the high support values, we present here the most robust RAxML tree (Fig. 3.3) and the best TetRAD tree (Fig. 3.4), both reconstructed for the datasets with 60% missing data and 90% clustering threshold. The remaining combinations of the third set of assemblies can be found in Supplementary Material S2 (Appendix B) for RAxML and S3 for TetRAD.

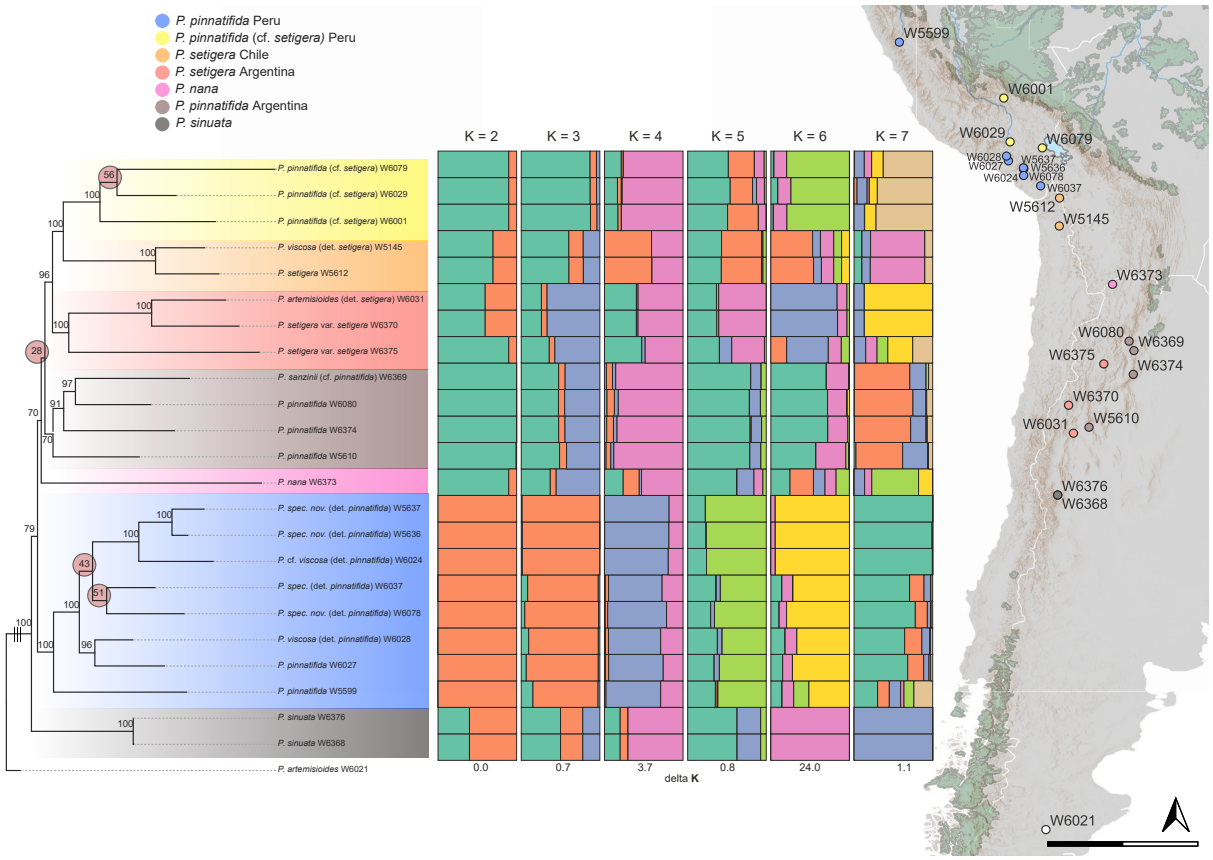


Figure 3.3 Phylogeny and population structure of the south American *P.* sect. *Glandulosae*. The phylogenetic reconstruction is based on a concatenated Maximum Likelihood analysis using RAxML. Support values are shown above branches and low bootstrap values leading to polytomies are highlighted in red. The three vertical lines indicate that the branch length was modified for visualization purposes. In parentheses, taxa names subsequent to determination are indicated. Different clusters of taxa used for the structure analysis are also displayed in different colours on the phylogram. Individuals in the structure plot are ordered according to their phylogenetic proximity. The higher ΔK value is 24.0 for K =6. K = 6 indicates that there is a lot of shared ancestry between the currently accepted taxa and shows that the "*P. pinnatifida* Peru" is the most genetically distinct cluster. Occurrences of specimens are displayed on a map of the Peruvian, Chilean and Argentinian Andes at the right side of the figure and the DNA number of each specimens is noted. The colour of occurrence points corresponds to the colour of the clades in the phylogram. *Phacelia sanzinii* is included in "*P. pinnatifida* Argentina" clade. *P. artemisioides* (W6021) is shown in white. All the taxa mentioned in the figure have been synonymized in this study with *P. pinnatifida* except for *P. artemisioides*. Scale bar: 250 km.

The concatenated tree retrieves six well supported clades largely reflecting geographical distribution (Fig. 3.3). The Peruvian *P. pinnatifida* clade (100% BS) includes both annuals and perennials (W5599, W6027, W6028) without well supported separation. The morphospecies *P. setigera* is retrieved in three distinct clades from Peru, Chile and Argentina respectively, each with maximum support and all of them forming a monophylum (96% BS). This morphotype is currently classified as *P.*

pinnatifida in Peru, *P. setigera* or *P. viscosa* in Chile and *P. setigera* var. *setigera* in Argentina. The remaining taxa are from Argentina and form another two clades, one including *P. sinuata* (100% BS) and one including representatives of *P. pinnatifida* and *P. sanzinii* (70% BS). The placement of *P. nana* is not clear and this leads to a polytomy including all clades of the *P. setigera* morphotype and additionally the Argentinian *P. pinnatifida* clade. *Phacelia sinuata* is recovered as sister to the remaining taxa (79% BS).

The backbone of the coalescent tree is more unstable and less resolved (Fig. 3.4; see also Supplementary Material S3 in Appendix B) compared to the concatenated tree. This is further highlighted by the recovered cloud tree (Fig. 3.4). The topology of the Peruvian *P. pinnatifida* clade (100% BS) is almost identical to the concatenated tree and the same goes for the morphotype of *P. setigera*. However, the three clades of the latter (Peruvian, Chilean and Argentinian), form a monophylum with moderate support in this case (60% BS). In contrast to the concatenated tree, the Peruvian *P. pinnatifida* clade and the *P. setigera* morphotype are here sister clades with moderate support (60% BS). The backbone of all remaining Argentinian lineages, i.e., *P. nana,* the Argentinian *P. pinnatifida*, *P. sanzinii,* and *P. sinuata*, remains unresolved and a polytomy is formed.

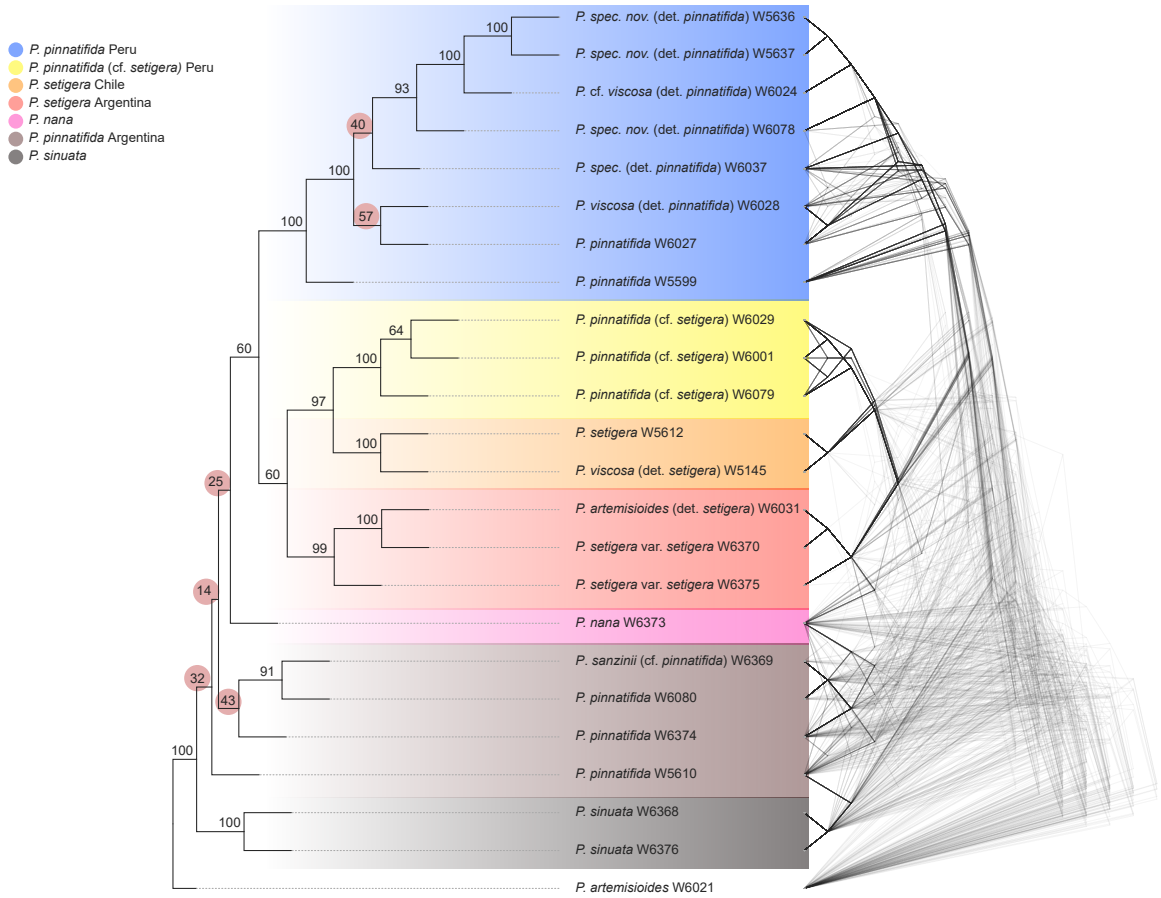


Figure 3.4 Phylogeny of the south American *P.* sect. *Glandulosae* using tetRAD. On the left a multispecies coalescent consensus tree is shown (90% clustering threshold, 60% missing data), while on the right side the cloud tree shows the variation over bootstrap replicates. Support values are shown above branches and low bootstrap values leading to polytomies are highlighted in red. In parentheses, taxa names subsequent to determination are indicated. Different clusters of taxa used for the structure analysis are displayed in different colours on the phylogram as in Figure 3.3. *Phacelia sanzinii* is included in "*P. pinnatifida* Argentina" clade. All the taxa mentioned in the figure have been synonymized in this study with *P. pinnatifida* except for *P. artemisioides*.

Species-level analyses. —The clades retrieved by the RAxML and TetRAD analyses were treated as seven different populations for the species-level analyses. The degree of genetic admixture was estimated with the STRUCTURE analysis performed for 4200 unlinked SNPs. Based on Ln P(K) and delta K, we determined that the optimal value of K was 6 (Fig. 3.3), which reveals extensive admixture among the different populations. All populations comprise 5% to maximum 38% (in *P. pinnatifida* W5610) of the *P. sinuata* cluster. *Phacelia nana* and the Chilean *P. setigera* clade are both highly admixed with almost all other clusters. Populations of *P. pinnatifida* in Peru and Argentina (clades marked as "*P. pinnatifida* Peru" and "*P. pinnatifida* Argentina" in Fig. 3.3) have different and distinct genetic structure but some individuals present slight admixture.

The first PCA performed for the same general grouping as that of STRUCTURE considered 8587 SNPs, of which 4200 were unlinked. It clustered the samples in three (PC2, PC0) to four (PC1, PC0) groups. The first eigenvector (PC0), which explains 12.8% of the data variation, separates samples of "*P. pinnatifida* Peru" from all the rest which are clustered close to each other (Fig. 3.5A). The second eigenvector (PC1), which explains 9.2% of the data variation, clusters "*P. setigera* Chile" and "*P. pinnatifida* (cf. *setigera)* Peru" together, close to the Argentinian populations (including "*P. nana*", "*P. sinuata*", "*P. pinnatifida* Argentina") and the "*P. pinnatifida* Peru", while the "*P. setigera* Argentina" form another two groups because one individual (*P. setigera* var. *setigera* W6375) is further apart from the rest. The third eigenvector (PC2) explains 8.3% of data variation and groups the samples in three clusters separating "*P. pinnatifida* Peru" in one cluster, all individuals of *P. setigera* morphotype in a second cluster (except for *P. setigera* var. *setigera* W6375, which is again placed further apart) and the remaining Argentinian samples in a third cluster. The second PCA analysis performed using a general grouping based only on geographical distribution (Argentina, Chile, Peru) considered a total of 10977 SNPs, of which 5189 were unlinked. The retrieved clusters are similar with PC0 explaining 12.8%, PC1 9% and PC2 7.7% of data variation (Fig. 3.5B). In this case, PC0 separates the Peruvian samples from the rest, but two Peruvian *P.* cf. *setigera* (W6001, W6029) are clustered closer to Argentinian samples. PC1 places the two Chilean *P. setigera* and two of the Argentinian *P. setigera* at opposite edges, while the PC2 is much more similar to PC1 of the first analysis (Fig. 3.5A). The signal of genetic admixture is strong for both PCA analyses performed here and only the clade "*P. pinnatifida* Peru" forms a distinct cluster. Overall, both PCAs provide additional evidence that a clear separation of the examined taxa into distinct clades is not possible.

To estimate the rate of geneflow and possible migration events, TreeMix was applied to the ingroup only, which was grouped in seven populations as done for STRUCTURE. TreeMix considered 8587 SNPs from which 4200 were unlinked and indicated that three migration events have the highest likelihood (Fig. 3.6). It shows that the occurrence of gene flow from the "*P. pinnatifida* Peru" to "*P. sinuata*" has the highest migration weight, while gene flow from "*P. pinnatifida* Argentina" to the crown node of "*P. pinnatifida* Peru", "*P. pinnatifida* (cf. *setigera*) Peru" and "*P. setigera* Chile" follows, and gene flow from "*P. setigera* Chile" to "*P. pinnatifida* Peru" has minor migration weight (Fig. 3.6). The results of TreeMix over nine different subsampled iterations, include the aforementioned migration events, but indicate different trends with higher frequency. We present two sets of plotted results of 9 iterations for m=3 in the Supplementary Material S4 (Appendix B), to show that even though this analysis gives slightly different results every time that it is repeated with a different seed value, the overall trends of potential gene flow are retained. Gene flow with either direction between "*P. pinnatifida* Argentina" and "*P. sinuata*" appears most often and with high migration weight. Furthermore, "*P. nana*" is very often involved in migration events. As in all species-level analyses, strong signal of genetic admixture is present in the TreeMix results.

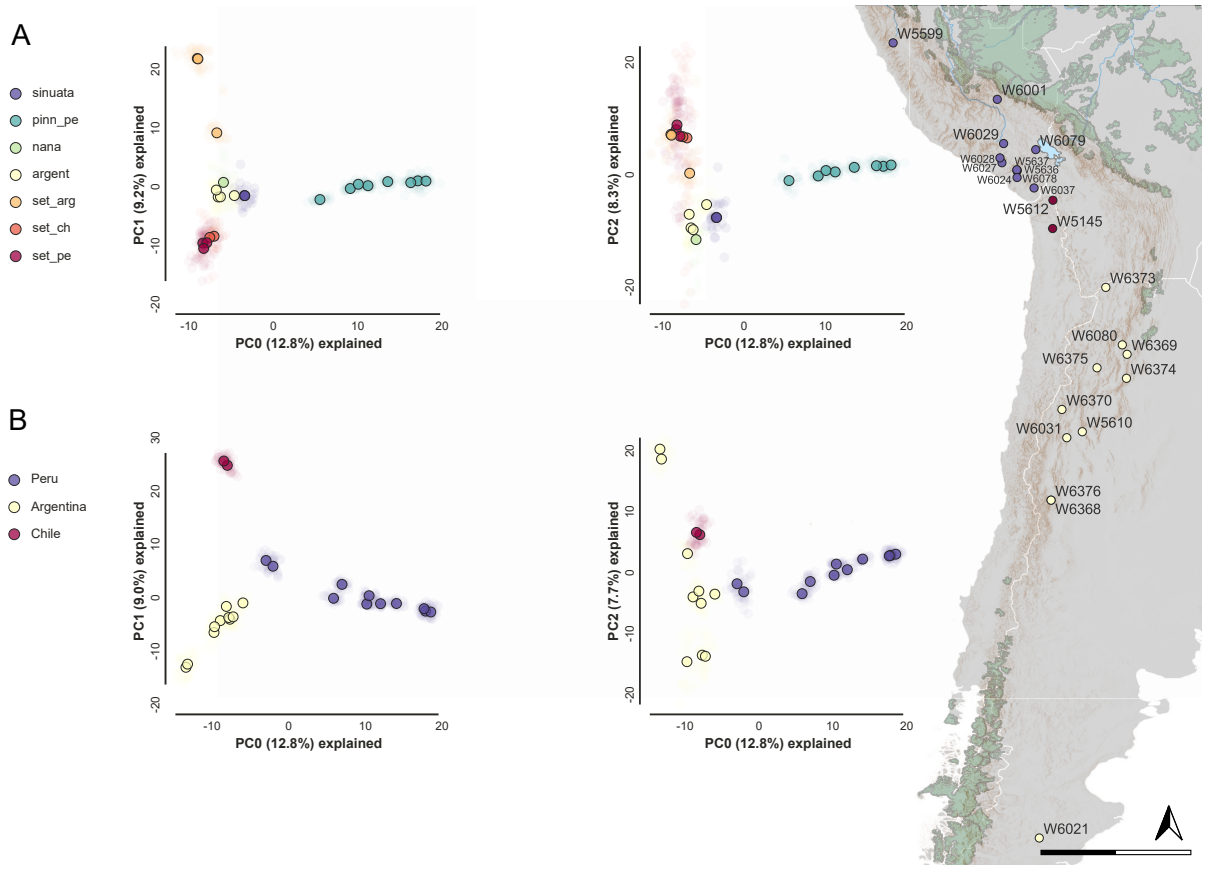


Figure 3.5 Results of the Principal Component Analyses (PCA) with randomly sampled, unlinked SNPs and 25 replicates of 23 ingroup samples. In this figure, genetic diversity of the ingroup taxa is represented in two ways: A. according to their phylogenetic relationships, B. according to their geographical distribution. For both A and B, PCA1 and PCA2 are displayed. Points with lower opacity are the results from the replicates, while the solid points represent the centroid of each sample. Occurrences of specimens are displayed on a map of the Peruvian, Chilean and Argentinian Andes at the right side of the figure and the DNA number of each specimens is noted. The colour of occurrence points corresponds to the colour of the three geographical areas defined for B. Scale bar 250 km.

3.4 DISCUSSION

The taxonomic history of *Phacelia* is puzzling (Brand, 1913; Voss, 1935; Howell, 1946; Constance, 1949, 1952, 1963; Atwood, 1975; Deginani, 1982; Walden & Patterson, 2011, 2012), partially due to extensive interspecific and intraspecific phenotypic variation as well as the recognition of polyploidy and hybridization (Constance, 1950; Gillett, 1960; Heckard, 1960; Levy, 2021; Levy & al., 2021). Furthermore, the few molecular phylogenetic studies available often suffer from a limited taxon sampling and low phylogenetic resolution, making species delimitation based on those data impossible (Ferguson, 1998; Gilbert & al., 2005; Hansen & al., 2009; Glass & Levy, 2011; Walden & al., 2014; Vasile & al., 2020). To overcome this, we present here the first study using a ddRADseq approach within Hydrophyllaceae in a small fraction of its most diverse genus *Phacelia*. The approach is able to resolve much of the complex evolutionary processes of these New World herbaceous plants.

Phylogenetic analysis. — In the current study, SNP data from ddRADseq shed light on the evolutionary relationships within *Phacelia* at different taxonomic levels, like in many other studies using RADseq-like methods for non-model plants (Helmstetter & al., 2020; Liu & al., 2020; Pearman & al., 2021; Böhnert & al., 2022; Otero & al., 2022). The different percentages of allowed missing data seem to heavily impact the tree topologies and lead to resolved and highly supported relationships within the

outgroup only when they were set to high levels (see RAxML trees with 85% missing data for the 1st set of assemblies in Supplementary Material S1 in Appendix B). Interestingly, *P.* sect. *Ramosissimae* is sister to *P.* sect. *Glandulosae* and *P.* sect. *Whitlavia* sister to *P.* sect. *Gymnobatha* as in the *ndhF* phylogenies of Walden & al. (2014) and Vasile & al. (2020), but the placement of *P.* sect. *Phacelia* and *P.* sect. *Euglypta* are incongruent with previous studies using *ndhF*, while those using ITS provided no resolution for the backbone (Gilbert & al., 2005; Hansen & al., 2009; Walden & al., 2014; Vasile & al., 2020). However, the inclusion of all outgroup taxa in the downstream analysis did not allow to resolve the phylogenetic relationships of the ingroup and reduced missing data resulted in highly incongruent tree topology overall. Therefore, the ingroup was analyzed independently. These findings highlight the importance of exploration of different parameters on the one hand—especially clustering threshold and different amounts of missing data—and of different specimens combinations on the other hand, as done already in various previous studies (Pante & al., 2015; Tripp & al., 2017; Pearman & al., 2021). As suggested by Vaux & al. (2023), a practical solution would be to tolerate a higher proportion of missing data in an overall dataset to identify divergent lineages, while independently analyzing subsets of the dataset with reduced missing data can better resolve intraspecific variation.

In order to test our hypothesis, we were aiming to resolve the interspecific relationships of the recently diversified South American *Phacelia* species of *P.* sect. *Glandulosae* (ca. 3 Mya according to Vasile & al., 2020). However, the phylogenetic analyses—both RAxML and TetRAD—provide evidence that the ingroup samples do not fall into independent clades (species) with separate evolutionary histories, but are best understood as different populations of one species, namely *P. pinnatifida* sensu lato (Figs 3.3, 3.4). The ingroup includes *P. nana*—previously placed in *P.* sect. *Euglypta*, resp. sect. *Phacelia* by Brand (1913). Within the ingroup, the different morphotypes are not retrieved as different lineages and molecular data contradict the recognition of these morphotypes as separate taxa. The only taxon clearly retrieved outside *P. pinnatifida* s.l. is the morphologically distinct *P. artemisioides*.

Species-level analyses. — Although *P. pinnatifida* and *P. setigera* are poly- and/or paraphyletic, we retrieved highly supported geographical clades for Peru, Chile and Argentina (Fig. 3.3). This fact underscores that we have some degree of ecogeographical differentiation and we performed some analyses for population genetics. Sampling is admittedly limited, but Vaux & al. (2023) demonstrated that genotyping-by-sequencing or RADseq-like methods permit smaller sample sizes compared to traditional sequencing methods (e.g., microsatellites) due to the much larger number of sequenced loci. We are able to present only a first estimate of population structure and admixture (Figs. 3.3, 3.5, 3.6). A more extensive taxon sampling would be necessary to document geneflow in more detail and possibly retrieve evidence for a clear split into one or more subspecies as done by Schilling et al. (2018) for the genus *Boechera* Á.Löve & D.Löve. The same goes for understanding the exact effect of the Andes as a geographical barrier and of the dispersal ability for the diversification of this group. Analysis of RADseq data on species or population level can generally reveal the influence of geographical barriers, such as rivers (Nazareno & al., 2017), or mountains (Vargas & al., 2017; Sassone & al., 2021; Böhnert & al., 2022) to the genetic differentiation, but also ancient or recent introgression despite the geographic proximity of the species (Kim & al., 2018).

Our current SNP dataset clarifies the monophyly of the *P. pinnatifida* s. l. and demonstrates some degree of admixture, which varies across different clades (Figs. 3.3, 3.5), as well as multiple recent or ancient dispersal events (Fig. 3.6). As in other plant groups (e.g., in Eriogoneae; Pearman & al., 2021), discerning detailed evolutionary relationships in the South-American *P.* sect. *Glandulosae* can be

challenging without more extensive genetic data, due to the fine morphological differences separating the current accepted taxa, but also due to the varying geographical origins and distributions.

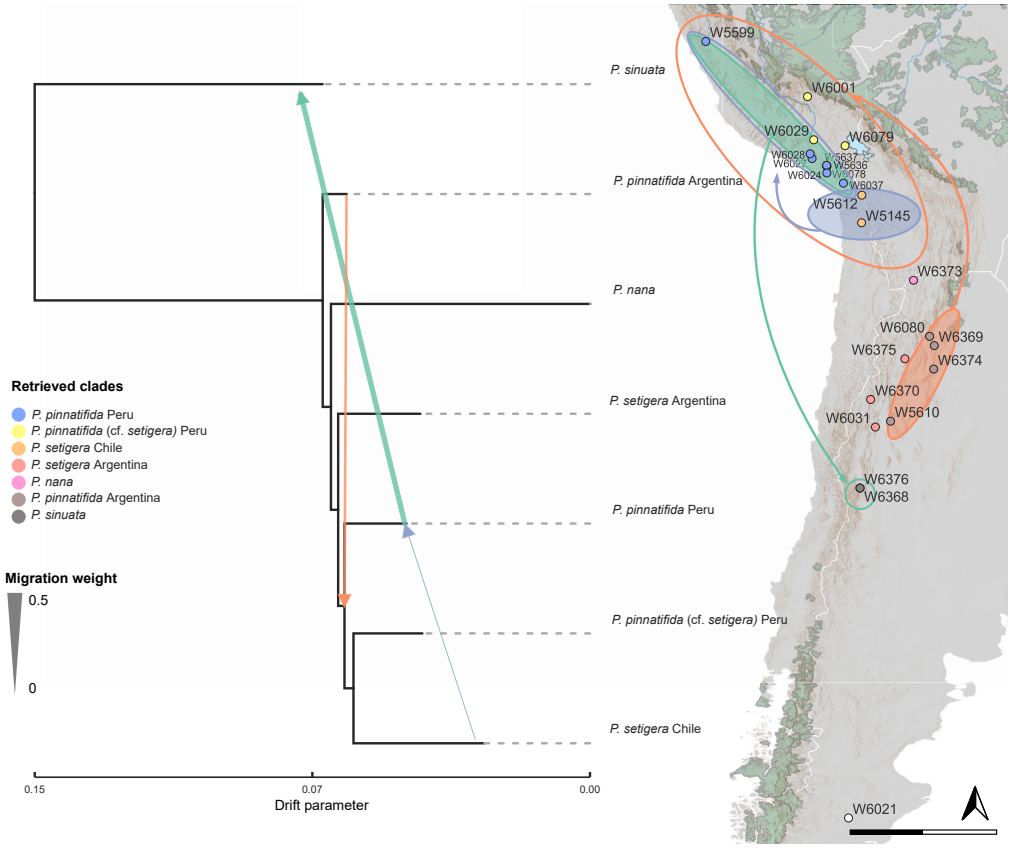


Figure 3.6 Treemix analysis showing inferred gene flow among the 7 populations detected with the structure analysis at K=6. The best number of migration events was 3. Arrows of different colours on the ML tree represent migration events. The same likely migration events are illustrated on the map, using the respective colours. Occurrences of the individual taxa on the map are coloured according to the colour of the retrieved clades. *Phacelia sanzinii* is included in "*P. pinnatifida* Argentina" clade. *Phacelia sinuata* was used to root the tree concerning the retrieved topology. All the taxa mentioned in the figure have been synonymized in this study with *P. pinnatifida.* Scale bar 250 km.

Evaluation of diagnostic characters and taxonomic implications. —Taxonomy is a phenotypic science and molecular systematics in our case is just a supplementary tool to the taxonomy and not the other way around. Already in 1950, Constance highlighted that the study of "Evolution" has in recent years been so monopolized mainly by geneticists, that it is often overlooked that taxonomists have been concerned with phylogenetic considerations for at least a century or longer. In our initial attempts to identify distinct morphotypes within South American *P.* sect. *Glandulosae*, we encountered considerable difficulties: traditionally recognized character complexes in indument, calyx, corolla (including scales and coloration), gynoecium, androecium, fruit and seed morphology (Brand, 1913; Heckard, 1960; Deginani, 1982; Hofmann, 1999; Walden & Patterson, 2012), showed considerable and largely uncorrelated variation between specimens and populations. No single suite of characters corresponded uniquely to any of the six currently recognized taxa. Based on leaf morphology and the overall plant habit we could very roughly delimit three morphotypes within *P. pinnatifida* (Fig. 3.1). Morphotype I (Fig. 3.1A-H) includes perennials, usually erect and strongly branched plants, longer than 30 cm (usually up to 60 cm), with leaves oblong-lanceolate, pinnatifid (or bipinnatifid) to pinnatisect (or bipinnatisect), petiolate to long petiolate with 4 to 9 pairs of toothed, sometimes trilobate, pinnae. Morphotype I usually occurrs at elevations of (1500–) 2000–4000 m.a.s.l.

and closely resembles the "typical" material of *P. pinnatifida*, i.e., the specimens the name was originally based on. Morphotype II (Fig. 3.1I–L) includes annuals, usually short plants (5 to 15 cm), lanceolate, pinnatifid, lobate, petiolate leaves, with 3 to 4 pairs of pinnae (lobate, trilobate or toothed). Morphotype II is found at lower elevations (ca. 1400–2600 m.a.s.l) and significantly resembles *P. crenulata* Torr. ex S.Watson, a polymorphic North American taxon that has been recently revised (Voss, 1937; Atwood, 1972, 1975; Walden & al., 2016) and that has been considered as the genetic source of the amphitropical disjunct South American *P.* sect. *Glandulosae* in the past (Heckard, 1963). Morphotype II includes what has been called *P. viscosa* in the past. Morphotype III (Fig. 3.1M–N) is found mostly at high elevations (ca. 3000–4200 m.a.s.l) and comprises plants with oblong to linear lanceolate leaves, pinnatifid to pinnatisect, petiolate to long petiolate, with 9 to 12 pairs of pinnae (toothed, sometimes trilobate). Morphotype III includes what has been called *P. setigera* in the past. However, many individual specimens are impossible to assign to any of these morphotypes since leaf shape and size are extremely plastic. This variability has not only been found in *P. pinnatifida*, but also across all currently accepted taxa of *P.* sect. *Glandulosae* in South America (Fig. 3.2). There are few diagnostic characters. The putative absence of corolla scales in *P. sanzinii*, *P. sinuata* and *P. nana* is one of them, but minute scales, which are barely visible even under the stereomicroscope, have been observed also in specimens of *P. setigera*, leading to misidentifications of *P. pinnatifida* or *P. setigera* as *P. sanzinii*, *P. sinuata* or *P. nana.* Therefore, absence of scales is neither reliable nor a consistent character for *P.* sect. *Glandulosae*, in contrast to - for example - South American *P. cumingii* from *P.* sect. *Euglypta*, which always lacks scales. The lack of reliable differentiating characters and the phenotypic plasticity is reflected in the numerous misidentifications (including even characteristic *P. artemisioides*) in herbarium collections, the lack of determinations (samples named as "*P. spec."*) and repeated renaming of various specimens in collections. Characters such as position of style cleft, capitate stigmatic heads and exserted stamens are simply non-useful, especially for herbarium specimens' observations, because ontogenetic phenomena like protandry are being ignored. Seed micromorphology also reveals phenotypic overlap among the currently accepted species, except for *P. artemisioides* which has different type of reticulation (larger pits and sometimes slightly echinate on the dorsal side) and an unlobed hilum (Fig. 3.7).

Deginani (1982) already realized that her species concepts did not entirely reflect the distribution of characters states across the group, explicitly mentioning specimens and/or characters deviating from her taxon concept. In the absence of a well-resolved phylogeny, recent authors have tried to make sense of the complex phenotypic variation and to square it with extant material (Deginani, 1982; O'Leary & al., 2022). The data here presented demonstrate that the misgivings and problems with regards to the delimitation of the currently recognized species simply go back to the complex mosaic of character states in this phylogenetically recent and incompletely diverged complex. Comparing material across the distribution range of the *P. pinnatifida* complex and evaluating character states with the help of a phylogeny it becomes apparent that there is no correlation between these individual characters states and exclusive lineages, i.e., taxa worth recognizing.

Morphological plasticity in plants due to environmental factors is a well-known phenomenon (Grenier & al., 2016; Mizutani & Kanaoka, 2018). Depending on the environmental conditions the development, physiology and life cycle of organisms can alter with usually intense modifications in regards to photosynthetic leaf area and branching (Sultan, 2000; Nicotra & al., 2010). Based on field observations by one of us (MW), the different morphotypes within *P. pinnatifida s.l.* are usually found in entirely homogenous populations at any given locality, but may be distinctly different from other populations in a separate catchment area or at a different elevation. This may be due to adaptations to specific conditions resulting in divergent life histories and morphology (Damasco & al., 2019) or simply

a result of genetic drift. However, our data not only document the non-monophyly of *P. pinnatifida* morphotypes resp. previously recognized segregate taxa, but also demonstrate considerable geneflow both along and across the Andes (Fig. 3.3), strongly arguing for the recognition of a single, more widely defined, polymorphic species, which then takes the name *P. pinnatifida*. It can be readily distinguished from its sister species, *P. artemisioides*.

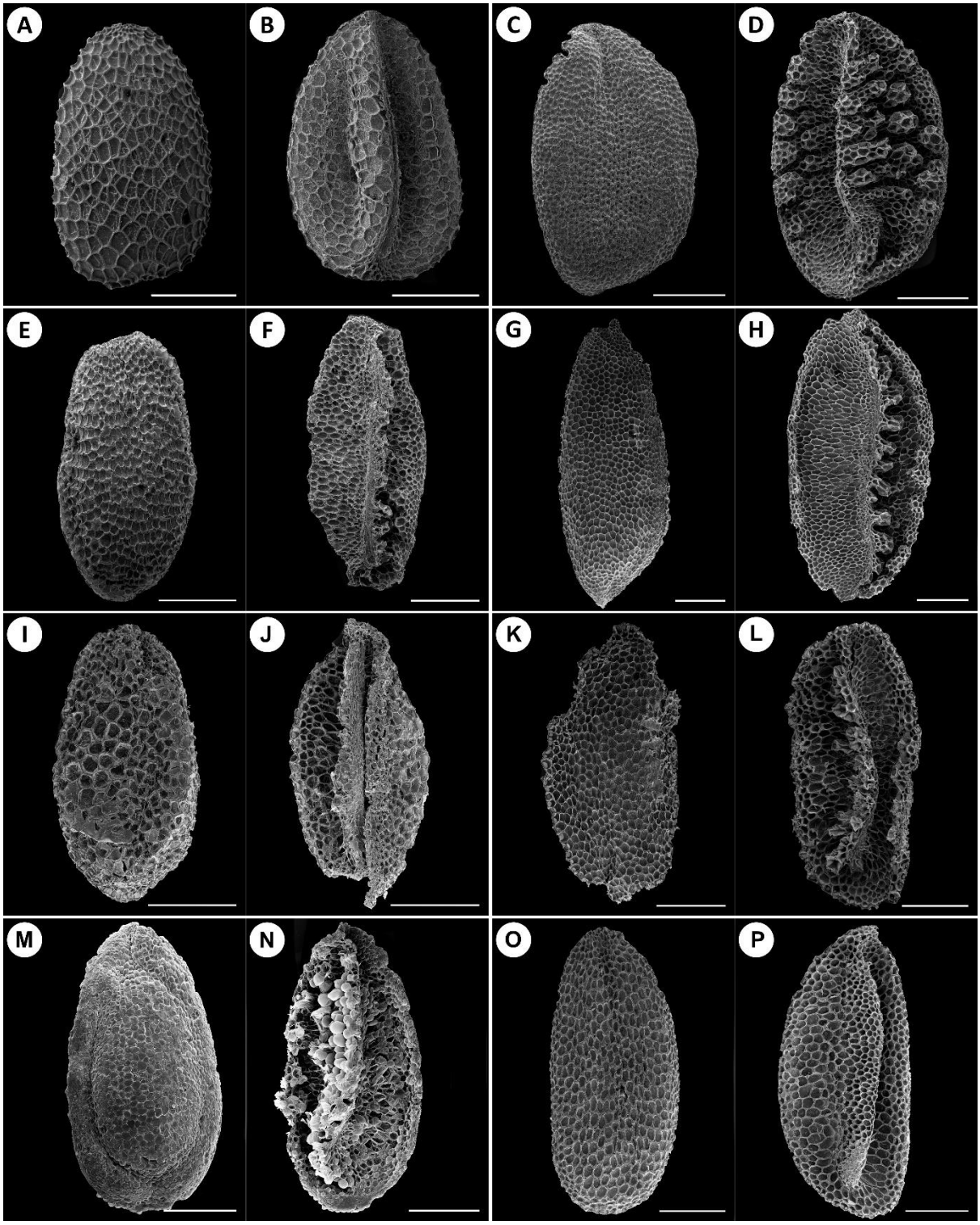


Figure 3.7 Seed micromorphology of *P.* sect. *Glandulosae* in South America; all the taxa mentioned in C–P have been synonymized in this study with *P. pinnatifida*; **A–B**, *P. artemisioides*, B 100763932, DNA Nr. W6021; **C–D**, *P. nana*, B 100763992; **E–F**, *P. pinnatifida* (Morphotype I), Cordova et al. 025. HUSA, DNA Nr. W6027; **G–**

H, *P. pinnatifida* (Morphotype III), B 100267427, DNA Nr. W6001; **I–J**, *P. pinnatifida* (Morphotype II), B 100844687, DNA Nr. W6024; **K–L**, *P. setigera*, M. Belov CL-0-BONN-34622, DNA Nr. W5612; **M–N**, *P. viscosa*, Luebert & Böhnert 3640 BONN, DNA Nr. W5145; **O–P**, *P. sinuata,* B 100763870. All seeds are from herbarium specimens, except for *P. viscosa*, Luebert & Böhnert 3640 BONN which is form material stored in FAA and then critical point dried. All photos are by Y. Malkowsky, except for M, N by M.-A.Vasile. Scale bars: 700 µm.

3.5 CONCLUSIONS

Information derived from coalescent-based approaches can be valuable to species delimitation and hence the status and history of species taxa, and this information can be employed as the basis for integrative taxonomic efforts, whereas a phylogenetic species concept that uses monophyly as a criterion would not be able to delimit species, even those with modest effective population size (Wilkins & al., 2022). The intuition of the taxonomist in order to answer the question "one species or more?" is an integral part of species delimitation (Wilkins & al., 2022), given that naming all populations as species is simply pointless (Zachos & al., 2020), but also given that gene flow among populations can be complex and can challenge our perception of what constitutes a species (see Jackson & al., 2017 and literature cited in there). Characters used in the past for species delimitation within *P.* sect. *Glandulosae* in South America neither are representative for the wide range of intermediate character states which are found between the supposed extremes, nor define monophyletic groups, as our phylogenetic results show. Therefore, we suggest that only the names *P. artemisioides* and *P. pinnatifida* – including all the remaining South American taxa – should be retained in South American sect. *Glandulosae*.

3.6 TAXONOMIC TREATMENT

Phacelia Juss. in Gen. Pl.: 129. 1789 – Lectotype (designated by Gmelin, in Syst. Nat. ed. 13, 2:330. 1791): *Phacelia secunda* J. F. Gmel.

Phacelia Juss. subg. ***Phacelia*** sect. ***Glandulosae*** (Rydberg) Walden & R. Patt. in Madroño 59(4): 216. 2012 – Type: *Phacelia glandulosa* Nutt.

Phacelia artemisioides Grisebach in Abh. Königl. Ges. Wiss. Göttingen 19: 230. 1874 – Holotype: Argentina, Capital, Córdoba, "Im Camp südlich von Córdoba", 10 Sep 1870, *P. G. Lorentz 97* (GOET barcode GOET004432 [photo!]; isotype CORD barcode CORD00006097 [photo!]).

= *Phacelia glandulosa* Nutt. var. *patagonica* Ball, J. Linn. Soc. Bot. 21: 227. 1886 – Type: Argentina, Buenos Aires. "South of province of Buenos Ayres (No. 211, G. C.)", *G. Claraz 211* (not located).

Description. — Herbaceous plant, annuals (4–)10–20 (–30) cm long, pilose, hispid–glandular, with a branched base and tendent or ascending stems. Leaves bipinnatisect with 3–5 pairs of obovate pinnules, 2–4 lobed and a larger, blunt terminal one. Occasionally intercalated monolobular pinnules between the pinnules. Petiole 5-30 mm long, lamina 10–40 mm long, 5–20 mm wide, hispid (glandular) on both sides. Flowers nearly sessile flowers, in dense, terminal, scorpioid compound cymes. Calyx hispid(–glandular), 2–3 mm long, sepals oblong, obtuse. Corolla tubular (rarely campanulate), white (rarely lilac), 2–5 mm long, sometimes hairy at the dorsal side of the corolla lobes, with a pair of oblong scales next to the insertion point of the stamens. Stamens inserted 0.5 mm from base of corolla. Stamens and style even with the corolla (rarely slightly exserted). Style glabrous (rarely pilose), bifid from its midpoint (rarely at 2/3). Ovary subspherical, densely pilose, hispid–glandular, with 4 ovules and with nectary disc at the base. Capsule spherical, hispid–glandular, 2–3 mm long with calyx not exceeding its height. Seeds 1–4, ovate, cymbiform, 2 mm long, 0.5–1 mm wide, reticulate on both sides, dorsal side

slightly echinate, ventral side corrugated, dark brown to black. It inhabits the south of Brazil, Uruguay and Argentina. It grows in dunes and dry rivers, and can be found from sea level to heights ca. 4500 m a.s.l. It flowers from July to March.

Notes. —The leaves of this species show a great variation in size, with the lower leaves being generally larger with longer petiole. The number and the size of the pinnules also varies drastically, with the upper leaves having usually more lobed pinnules than the lower. Specimens G. Scott-Elliot 462 & 473 (E), are probably incorrectly determined as *P. artemisioides* and therefore the presence of the species in Chile, close to the Argentinian, borders remains dubious. Based on photos, G. Scott-Elliot 462 (E) is *P. cumingii* and G. Scott-Elliot 473 (E) is *P. pinnatifida* s.l.

Additional material examined. — **ARGENTINA**. CATAMARCA: **Andalgalá:** 6 Oct 1916, *P. Jörgensen 1050* (US). **Capital:** Las Carreras, 700 m, 22 Apr 1928, *S. Venturi 6566* (US). **Santa Maria:** Campo Colorado, 4400 m, 20 Jan 1925, *S. Venturi 6558* (US). CÓRDOBA: **Capital:** Córdoba, 1908, *W.Gourlay s.n.* (E); in the vicinity of the city, 12 Oct 1876, *Hieronymus s.n.* (B); in the vicinity of the city, 12 Oct 1876, *Hieronymus s.n.* (US); Dec 1891, *C.E.O. Kuntze s.n.* (US); No precise locality, Córdoba (illegible), *s.coll. s.n.* (K). JUJUY: **Tilcara:** Tilcara 3000 m, 8 Feb 1927, *S. Venturi 4975* (US). LA PAMPA: **Rancul:** *P.G. Lorentz 598* (F). MENDOZA: **Las Heras:** Uspallata, Ostfuss d. Anden, 1950 m, 18 Oct 1898, *Prinzessin Therese von Bayern 261* (M). NEUQUÉN: **Zapala:** 20 km. south of Zapala, -70.32°W, -39.2°S, 15 Nov 1948, *H.A. Senn 4310* (US). RIO NEGRO: sine loc., 10 Sep 1874, *C. Berg 22* (LP). **General Roca:** vicinity of Río Negro valley, 250–360m, 25 Sep 1914, *W. Fischer 39* (US). SALTA: **Cachi:** Ruta Provincial 33 de Cachi a Escoipe, 25º13'53"S, 65º55'57"W, 2150 m, 03 Feb 2005, *F.O. Zuloaga et al. 8457* (NY). SANTA CRUZ: **Rio Chico:** Ruta Provincial 521, a 37 km al E de Gobernador Gregores, Gran Altiplanicie Central, estepa, 17 Jan 1987, *Leuenberger & Arroyo Nr. 3708* (B). SANTA FE: **Ceres:** Oct 1892, *C.E.O. Kuntze s.n.* (US). TUCUMÁN: **Burruyacú:** Rio Medina, playa del río 700 m, 12 Oct 1929, *S. Venturi 9575* (US). **Chicligasta:** Estancia Las Pavas, 3200 m, 21 Mar 1925, *S. Venturi 6568* (US). **Leales:** Sueldos, 300 m, 19 Nov 1919, *S. Venturi 571* (US). **Tafi:** Sierra del Cajon, 3700 m, 16 Feb 1924, *S. Venturi 10095* (US). **Trancas:** Leocadio Paz., 1000 m, 14 Mar 1927, *S. Venturi 7369* (US). **CHILE**. VALPARAISO: **Los Andes:** "Chile & Argentina, above Las Cuevas, 3200–3400 m, in shade of rocks on slopes", 23 Jan 1904, G. *Scott-Elliot 462* (E); "Chile & Argentina, above Las Cuevas, 3200–3400 m, in shade of rocks on slopes", 23 Jan 1904, *G. Scott-Elliot 473* (E). **URUGUAY**. MONTEVIDEO: Barra de Sta. Lucia, 1 Dec 1876, *M. Fruchard s.n.* (US). SAN JOSE: San José, Barra Sta. Lucia, 10 m, Nov 1927, *G. Herter 973* (US); Barra del río Sta Lucia, 10 Nov 1935, *B. Rosengurtt 2040* (US).

Phacelia pinnatifida Griseb. ex Wedd. in Chlor. Andina 2: 85. 1857 – Holotype (erroneous lectotype designated by Mirra & al. in Darwiniana, n.s. 9(1): 212. 2021): Peru, "In rupibus Cordiller. pr. San Antonio", Jan 1854, *W. Lechler 1801* (GOET barcode GOET006991 [photo!]).

= *Phacelia nana* Wedd. in Chlor. Andina 2: 86. 1857 – Lectotype (or perhaps holotype, designated by Mirra & al. in Darwiniana, n.s. 9(1): 211. 2021): Bolivia, Potosí, "Voisinage de la laguna de Potosí", s.d., *A.D. d'Orbigny 1449* (P barcode P00648953 [photo!]), **syn. nov.**

= *Phacelia pinnatifida* var. *robusta* Wedd. in Chlor. Andina 2: 86. 1857 – **Lectotype (designated here):** [Bolivia], "Amériq. méridionale, Perou Austral [Péninsule de Copa-Cabana, Lac Titicaca]", Mar 1847, *H.A. Weddell 4388* (P barcode P00648959 [photo!]; isolectotype: P barcode P00648960 [photo!].

= *Phacelia viscosa* Phil. in Fl. Atacam. 37. 1860 – Lectotype (designated by Mirra & al. in Darwiniana n.s. 9(1): 212. 2021): Chile, Atacama, "In valle Sandoncum priore legi", Feb 1854, *R.A. Philippi s.n.*

(SGO barcode SGO000003993 [photo!]; isolectotypes: SGO barcode SGO000003994 [photo!], SGO barcode SGO000003995 [photo!]).

= *Phacelia sinuata* Phil. in Linnaea 33: 185. 1864 – Lectotype (designated by Mirra & al. in Darwiniana n.s. 9(1): 212. 2021): Chile, Coquimbo, Cordillera de DoñaAna, 1860, *H. Volckmann s.n.* (SGO barcode SGO000003992 [photo!]; isolectotypes SGO barcode SGO000003991 [photo!], SI barcode SI001769 [photo!], SI barcode SI002110 [photo!]), **syn. nov.**

= *Phacelia pinnatifida* var. *elatior* Griseb. in Abh.Königl. Ges. Wiss. Göttingen 24: 267. 1879 – Lectotype (designated by Mirra & al. in Darwiniana n.s. 9(1): 212. 2021): Argentina, Salta, Los Potreros, Nevadodel Castillo, 25 Mar 1873, *P. G. Lorentz & G. Hieronymus 192* (GOET barcode GOET004433 [photo!]; isolectotypes: CORD barcode CORD00006100 [photo!], US barcode US00110552 [photo!]).

= *Phacelia foliosa* Phil. in Anales Mus. Nac. Santiago de Chile 8: 53. 1891 ≡ *Phacelia pinnatifida* var. *foliosa* (Phil.) Reiche in Anales Univ. Chile 120:810. 1907 – Lectotype (designated by Mirra & al. in Darwiniana n.s. 9(1): 212. 2021): Chile, Antofagasta, Socaire, Feb 1885, *F. Philippi s.n.* (SGO barcode SGO000003986 [photo!]; isolectotype: SGO barcode SGO000003987 [photo!]).

= *Phacelia setigera* Phil. in Anales Mus. Nac., Santiago de Chile 8: 54. 1891 ≡ *Phacelia pinnatifida* var. *setigera* Reiche in Anales Univ. Chile 120: 810. 1907 – Holotype: Chile, Colana, *F. Philippi s.n.* (SGO barcode SGO000003990 [photo!]), **syn. nov.**

= *Eutoca pedunculosa* Phil. in Anales Mus. Nac., Santiago de Chile 1891: 54. 1891 ≡ *Phacelia pinnatifida* var. *pedunculosa* in Reiche, Anales Univ. Chile 120: 810. 1907 – Holotype: Chile, Tarapacá, Vegas del Diablo, 3700 m, 10 Jan 1885, *F. Philippi s.n.* (SGO barcode SGO000003976 [photo!])

= *Eutoca lomarifolia* Phil. in Anales Mus. Nac., Santiago de Chile 8: 54. 1891 – Lectotype (designated by Mirra & al. in Darwiniana Nueva Serie 9(1): 211. 2021): Chile, Atacama, Breas, Jan 1885, *F. Philippi s.n.* (SGO barcode SGO000003972 [photo!]; isolectotype: SGO barcode SGO000003973 [photo!]).

= *Phacelia pinnatifida* var. *decumbens* Brand in Pflanzenr. (Engler) IV, 251 [Heft 59]: 77. 1913 – **Lectotype (designated here):** Argentina, Jujuy, in reg. Chaco-Andinis, 28 Dec 1901, *K.R.E. Fries 958* (S barcode S12-4437 [photo!]; syntypes: Argentina, Jujuy, Yavi, "Exped. Suec. in reg. Chaco-Andinis", 3300 m, 02 Jan 1901, *G.H.E. von Hofsten 1701* (S barcode S12-4423 [photo!]; Argentina, Chubut, Biedma, Puerto San José, 11 Dec 1904, *Per K.H. Dusén 5266* S barcode S12-4435 [photo!]).

= *Phacelia boliviana* Brand in Pflanzenr. (Engler) IV, 251[Heft 59]: 77. 1913 – Lectotype (designated by Mirra & al. in Darwiniana n.s. 9(1): 212. 2021): Bolivia, Tarija, 30 Jan 1904, *K. Fiebrig 3012* (S barcode S12-4420 [photo!]; isolectotypes E barcode E00433725 [photo!], GH barcode GH00093561 [photo!], HBG barcode HGB517917 [photo!], K barcode K000612632 [photo!], U barcode U0002502 [photo!], US barcode US00110481 [photo!]).

= *Phacelia sanzinii* Hicken in Physis (Buenos Aires) 1: 385. 1914. – Lectotype (designated by Mirra & al. in Darwiniana n.s. 9(1): 212. 2021): Argentina, Mendoza, Potrerillos, 25 Aug 1913, *R. Sanzin 33* (SI barcode SI001767 [photo!]), **syn. nov.**

= *Phacelia pinnatifida* var. *lobulata* Hicken in Darwiniana 1: 134. 1924 – Holotype: Argentina, Salta. Rosario de Lerma. Quebradas del Río Toro y del Río Blanco. 24º53'50"S, 65º40'15"W, Jan 1923, *I.C. Vattuone 99* (SI barcode SI001761[photo!]).

= *Phacelia setigera* var. *humahuaquense* Deginani in Darwiniana 24(1–4): 420. 1982 – Holotype: Argentina, Jujuy, Mina Aguilar, nivel 1, 19 Feb 1970, *A.L. Cabrera, J. Frangi 20678* (LP barcode LP004233 [photo!]), **syn. nov.**

= *Phacelia pinnatifida* f. *minor* Hicken in Darwiniana 1: 33. 1922 – Holotype: Argentina, La Rioja, Famatina, Sierra de Famatina, Quebrada Encrucijada, 3500–5000 m, Mar 1913, *A. Flossdorf, 45* (SI barcode SI001762 [photo!]).

Description. — Herbaceous plant, perennial or annual, erect to decumbent, with stems usually bushy and branched, sometimes unbranched, (5–)20–40(–70) cm tall, hispid–glandular (hirsute). Leaves (oblong) lanceolate, pinnatifid to pinnatisect (less often bipinnatisect, or lobate with either sinuate or obtuse lobes, or blunt at apex and base), (1.5–)5–8(–11) cm long, 0.9–4 cm wide, (sessile) petiolate to long petiolate, with (3)4–9(12) pairs of pinnae (toothed sometimes lobate or trilobate) of variable sizes and interpolated segments (1–)4–15(–20) mm long, with irregular margins, usually crenulate or dentate. Flowers short petiolate to sessile, in dense compound scorpioid cymes. Calyx 3-6 mm long, increasing in size to 5–7 mm as the capsule matures; sepals (linear) ovate to oblong-spatulate, obtuse, hispid-glandular, distinct from each other or sometimes overlapping. Corolla campanulate, lilac to purple (bluish), (3–)4–7(–8) mm long, glabrous or slightly hairy on dorsal side; lobes obtuse, imbricate; scales semi-oval to longitudinal, two next to the insertion of the stamens, sometimes minute or absent. Stamens inserted 0.5–1(–1.5) mm from base of corolla. Stamens and style inclusive, even or exserted. Style pilose–glandular at the base to midpoint, bifid at the lower 1/3 to its midpoint (sometimes at the base). Ovary ovoid, hispid–glandular, with 4 ovules and nectary disc at the base. Capsule subglobose to ovoid, 3–4(–6) mm long, hispid–glandular. Seeds 1–4, ovate, cymbiform, compressed, marginate, reticulate on both sides with pits of very variable size, ventral side carinate, with more or less corrugated margins, (2–)3–4 mm long, 1–2 mm wide. It grows in mountainous and highland areas of Peru, Bolivia, Chile and Argentina from (700–) 1400–4800 (–5500). It flowers from August to March.

Nomenclatural Notes. — **(1)** The typification by Mirra & al. (2021) designated the collection *Lechler 1801*, with several duplicates in various herbaria, as lecto-/isolectotypes of the species, choosing one Paris-specimen as lectotype. They overlooked that Weddell (1857) specifically referred to a single specimen annotated by Griesebach as the sole specimen of the typical variety ("var. α"), whereas citing both the other specimens of *Lechler 1801* and collections by *Weddell* himself and *d'Orbigny* as original material of var. *robusta*. The *Lechler 1801* specimen in GOET has the original handwritten note by Griesebach and is therefore clearly the holotype of var. *pinnatifida*, whereas the other specimens are available as types for var. *robusta*. We propose choosing a *Weddell*-collection as lectotype for that name, in order to avoid further confusion with the type of the typical variety. **(2)** In the protologue of *P. pinnatifida* var. *robusta* ("var. β"), Weddell mentions the collection number for the specimens of *Lechler 1801*, whereas for the specimens of himself and *d'Orbigny,* he only mentions the locality. Based on the localities, it is clear that he is referring to *H.A. Weddell 4388* and *A. D. d'Orbigny 1342*. However, among all these specimens and their duplicates, two of the *A. D. d'Orbigny 1342* specimens (P00648956 and P00648957) have a label annotated probably by Weddell with an "α", indicating that although he does not mention these under the typical variety, he apparently considered these particular sheets as belonging to var. α. Only *H.A. Weddell 4388* (P00648959) is annotated as *P. pinnatifida* var. *robusta* and therefore it is designated here as a lectotype. **(3)** Brand (1913) mentions three syntypes in the protologue of *P. pinnatifida* var. *decumbens*, *Hofsten* 1701, Fries *958* and *Dusén 5266*, all deposited in the herbarium Stockholm. Among these three specimens *Fries 958* better shows the characters mentioned in the protologue and therefore is here designated as lectotype, while the other two specimens are

mentioned as syntypes. These specimens are all types of *P. pinnatifida* var. *decumbens*, a synonym of *P. pinnatifida* since long ago, *Fries 958* and *Hofsten* 1701 were determined as *P. setigera* respectively *Dusén 5266* as *P. sinuata* by O.H. Borsini in 1966. **(4)** All the following names are synonymized in this study with *P. pinnatifida*: *Phacelia nana* Wedd., *Phacelia sanzinii* Hicken, *Phacelia setigera* Phil., *Phacelia setigera* var. *humahuaquense* Deginani, *Phacelia sinuata* Phil.

Morphological Notes. — Our phylogenetic analyses show the non-monophyly of morphotypes within *P. pinnatifida* as well as of previously accepted taxa (i.e. *P. nana*, *P. sanzinii*, *P. setigera*, *P. sinuata*). According to this new information all the morphological characters previously used for species delimitation, do not define monophyletic groups and are now included in species description of *P. pinnatifida s.l.* Important diagnostic characters formerly used for species delimitation are summarized here: The inclusion of stamens and stigma in the corolla, has been used to differentiate *P. nana*, *P. setigera* and *P. sinuata* from *P. pinnatifida* (usually long exerted stigma and stamens). *Phacelia sinuata* (Fig. 3.2J–K) has leaves distinctively blunt at apex and base; *P. sanzinii* has characteristic sinuate lobes (Fig. 3.2L–M); *P. nana* (Fig. 3.2N–O) has notably obtuse, smooth lobes, with the lower ones much smaller and distant (Fig. 3.2O). Scales are usually absent in *P. sinuata*, *P. sanzinii* and *P. nana*, while in *P. setigera* they can be minute or well developed. All characters, i.e., included or exserted style and stamens, leaf morphology, absence or presence of scales are very plastic and there is a wide range of intermediate states between the supposed extremes. For all the aforementioned reasons, specimens traditionally named as *P. nana*, *P. sanzinii*, *P. setigera*, *P. setigera* var. *humahuaquense* and *P. sinuata* should be from now on be referred to as *P. pinnatifida.*

Additional material examined. —**ARGENTINA.** Sine loc. *Pearce s.n* (K000612631); *s. Coll. s.n.* (K000612630). CATAMARCA: **Antofagasta de la Sierra:** -26.276667°S, -67.063611°W, 4305m, 19 Feb 2022, *J.M. Acosta 1046* (SI). **Santa Maria:** Laguna del Tesoro, 4600 m, 22 Mar 1924, *S. Venturi 6557* (US); Laguna del Tesoro, 4600 m, 21 Jan 1925, *S. Venturi 6560* (US). CHUBUT: **Biedma:** Puerto San José, 11 Dec 1904, *P. Dusén 5266* (SI); Puerto Madryn, 06 Apr 1914, *P. Madryn 9* (SI). CÓRDOBA: Cerca del puesto de San José, arriba de los Allendes, Sierra Chica, 26 Jan 1876, *Hieronymus 474* (P); 10 km north of Alta Gracia 755 m, 19 Jan 1996, *R.R. Brooks et al. MS136* (E). JUJUY: **Cochinoca:** Abra de Rachaite, 4140 m, 16 Mar 1994, *M.E. Múlgura et al. 1316* (SI). **Humahuaca**: Mina Aguilar, 4500m, 13 Jan 1969, *J. Frangi & R. Kiesling 52* (LP); Mina Aguilar, 4350m, 03 Mar 1983, *J.H. Hunziker et al. 10539* (SI). **Rinconada:** Mina Pirquitas 4300m, 01 Mar 1964, *H. Schwabe et al. 912* (LP); RP 85, de Cusi-Cusi a Lagunillas del Farallón hacia Vilama, en cascada, 4351m, 24 Feb 2022, *J.M. Acosta 1159* (SI). **Susques:** Lomas al pié del Cerro Tuzgle, *A.L. Cabrera 8402* (LP); A 65 km. Al W de Susques, entre el Salar de Olaroz y Archibaca, 3850m, 16 Mar 1994, *M.E. Múlgura 1276* (SI). **Tilcara:** Tilcara, 3000m, 21 Jan 1926, *S. Venturi 6562* (US); Chorru Valley, 3810m, 13 Feb 1939, *E. K. Balls 6033* (E). **Tumbaya:** Nevado del Chani, 4500m, 5 Feb 1925, *S. Venturi 10089* (US). **Yavi:** Abra de Cajas, 18 Jan 1966, *A.L. Cabrera et al. 17571* (LP). LA RIOJA: **Chilecéto:** Nonogasta, 900m, 27 Dec 1928, *S. Venturi 7819* (US). MENDOZA: Sine loc., 20 Sep 1913, *R. Sanzin*, *s.n.* (CAS); Puntilla, 05 Oct 1913, *R. Sanzin 43* (SI), Parque del Oeste, 26 Oct 1913, s.c. s.n. (SI); Sine loc., Sep 1915, *J.N. Rose & P.G. Russell 20983* (US); Sine loc., 1931, *Negrete s.n.* (SI); Sine loc., 05 Dec 1942, *H.H. Bartlett 19187* (SI). **Las Heras:** in aridus collinis subandinis prope La Puntilla, 05 Oct 1913, *R. Sanzin s.n.* (GH); Potrerillos, 10 Jan 1943, *H.H. Bartlett 19302* (SI); El Chalco, 16 Jan 1943, *H.H. Bartlett 19187* (SI); Uspallata, 12 Mar 1944*, A. Soriano 1087* (SI); Cruz de Paramillo, 2900m, 24 Jan 1950, *M.A. Palacies & F.A. Barkley 20Mz366* (L); 14 km NE Uspallata camino a Villavicencio, 14 Jan 1963, *Boelcke et al. 9894* (SI); Sierra de Uspallata, El Balcon, 2750m, 05 Jan 1988, *Leuenberger & Arroyo 3783* (B); 32º37’08’’S, 69º 16’ 00’’W, 2296m. 05 Apr 2017, *N. Brignone 31* (SI). **Malargüe:** Río Malargüe,

Dec 1911, *Girth 38* (SI). RÍO NEGRO: **General Roca:** Sargento Vidal, Sep 2005*, Simon & Simon 898* (US); A ca. 7 km. Antes de llegar a los Castillos de Pincheiras, viniendo desde Malargüe, 1700m, 17 Feb 2007, *J. Chiapella et al. 1920* (SI). **San Rafael:** Los Molles arriba del Cuechillo, 2850m, 31 Dec 1949, *H. Sleumer 648* (B); Sa. Del Nevado, arroyo Chacay-co, puesto Barroso, 1300m, 03 Dec 1973, *O. Boelcke 15493* (SI). NEUQUÉN: **Confluencia:** Hills north of Neuquen, 17 Aug 1925, *H.F. Comber 15* (E). SALTA: Cercanías del Nevado del Castillo, Mar 1873, *Lorentz & Hieronymus 24* (CORD). **Capital:** Astillero, 1400-1500m, 06 Feb 1988, *L.J. Novara 7643* (B). **La Poma:** Mina Esperanza, 12 Feb 1960, *C.H. Hernandez 16* (LP). **Los Andes:** Hombre Muerto, 3800m, 24 May 1927, *Catalano 11* (SI); San Antonio de los Cobres, Quebrada de Polvorillos, 02 Feb 1944, *A.L. Cabrera 8394* (LP); Ruta 59, subida al Salar del Diablo, 17 Feb 1980, *A.L. Cabrera et al. 31799* (SI); Socompa, 29 Jan 1981, *S. Chalukian 1357* (SI). **Rosario de Lerma:** Campo Quijano-Dique las Lomitas, 1500m, 18 Nov 1985, *L.J. Novara 4681* (B). **San Carlos:** Cerro de Cachí, 4000m, 14 Mar 1927, *S. Venturi 6743* (US). SAN JUAN: Río Colorado (zona Ramada), 21 Jan 1951, *J. Semper 13971* (SI). **Calingasta:** Castaño Viejo, 23 Jan 1971, *Volponi & Zardini 130* (SI); Entre Ojo de Agua y El Molle, 2300-2850m, 15 Dec 1976, *Luti 5742* (SI); Entre Arroyo Los Hornillos y Colovado, 2250m, 31 Jan 1979, *Kopta 5782* (SI); Quebrada de la Alumbrera, 2050-2300 meters, 16 Feb 1988, *R. Kiesling* (SI); Quebrada de Las Burras, 3000-3300m, 13 Feb 1990, *R Kiesling et al. 7454* (SI); Quebrada del Arroyo del Medio, 11 Jan 1997, *Haene & Izquierdo 1622* (SI). **Iglesia:** Near Junta del Rio Jagua y Rio Cura, 2900m, 12 Jan 1926, *I.M. Johnston 6158* (US); Quebrada del Agua Negra, 3800m, 10 Jan 1976, *A.L. Cabrera et al. 27069* (SI); Ojo de Agua, 3500m, 12 Dec 1979, *A.L. Cabrera et al. 31176* (US); San Guillermo. Los Caserones laderas pedregosas, 3450m, 22 Feb 1981, *E.G. Nicora 8302* (SI); San Guillermo, Cerro Imán, 4200m, 23 Feb 1981*, J.L. Cajal 108* (SI); Reserva de San Guillermo, Baños de San Crispín cono de deyección sobre las vegas termales, 3200m, 24 Feb 1983, *J.*C. Pujalte 215 (SI); Reserva de San Guillermo, Río de la Sal, 3600m, 04 Jan 1984, *S. Meglioli 13* (SI); Quebrada de la Pancha, 10 Apr 1989, *E.R. Guaglianone 2465* (SI); Quebrada de Agua Negra. Ojos de Agua, 3250m, 22 Feb 1992, *R. Kiesling 7924* (SI); Quebrada de la Vicuñita, 2600m, 28 Feb 1992, *R. Kiesling 8036* (SI); San Guillermo, Agua del Godo, 3200m, 09 Jan 1997, *R Kiesling et al. 8771* (SI); Quebrada de Agua Negra, 3570m, 21 Jan 1997, *R. Kiesling et al. 8856* (SI); -30.095 S, -69.538056 W, 3240 m, 18 Jan 2009, *J. Chiapella 2368 & E. Vitek 09-0383* (NY). **Sarmiento**: Quebrada de la Flecha, 14 Nov 1982, *R. Kiesling 4043* (SI). TUCUMAN: **Chicligasta:** Pueblo Viejo, 4000 m, 24 Mar 1924, *S. Venturi 6559* (US); Estancia Santa Rosa, 4200 m, 18 Feb 1925, *S. Venturi 9693* (US); Estancia Las Pavas, 3200 m, 19 Jan 1926, *S. Venturi s.n.* (US). **Tafi:** Sierra del Cajon, 4000 m, 17 Feb 1924, *S. Venturi 10096* (US); Sierra del Cajon, 3000 m, 03 Feb 1926, *S. Venturi 6569* (US); Cerro el Negrito, 4000 m, 09 Nov 1958, *H.A. Fabris 1436* (F); s.l. 2800 m, 29 Jan 1950, *H. Sleumer 205* (B). **BOLIVIA.** LA PAZ: **Pacajes:** Andino Region, Charaña, 4000 m, 01 Mar 1921, *E. Asplund 12390* (US); Caquiaviri, 3900 m, 09 Feb1927, *Frole 3070* (B). POTOSI: Miraflor, *A.D. d'Orbigny 1342* (P barcodes P00648955, P00648956, P00648957). TARIJA: **Cercado:** Sama, 21°29'S, 64°53'W, 3500 m, 23 Mar 1986, *R. Ehrich 202* (B**). CHILE.** Sine loc., Dec. 1873, *E.C. Reed s.n.* (K); Trans-Andean crossing, Cachenta, 15 Nov 1927, *E. Clarence 275* (E). ANTOFAGASTA: **Antofagasta:** Parque Nacional Llullaillaco, Quebrada Zorritas, 3983 m, 02 Dec 2008, *R. Baines et al. 262* (E). **El Loa:** San Pedro de Atacama, 4200 m, 03 Feb 2001, *Ackermann 2* (B); San Pedro de Atacama, 3012 m, 30 Nov 2008, *R. Baines et al. 238* (E); **Taltal:** Cord. Volcán Lullaillaco, 4000 m, Feb 1926, *E. Werdermann 1006* (B, E, SI). ARICA Y PARINACOTA: **Parinacota:** 17.99333°S, 69.64°W, 3650 m, 18 Mar 2015, A. *Moreira & F. Luebert AM2410* (BONN); 18.25083°S, 69.17694°W, 4500 m. 20 Mar 2015, *A. Moreira & F. Luebert AM2440* (BONN). **Putre:** road Ch11 at kilometre 102, 3500 m, 15 Dec 1999, *Gardner & Knees 6263* (E). ATACAMA: **Vallenar:** Vicinity of Laguna Grande, 3150 m, 05 Jan 1926, *I.M. Johnston 5905* (US). COQUIMBO: **Elqui:** Baños del Toro, 3500 m, Dec 1923, *E.*

Werdermann 196 (E, F, U, SI); 3300 m, 5 Feb 1939, *T.H. Goodspeed 17264* (F); 3300 m, 05 Feb 1939, *J.L. Morrison 17264* (MO, SI). TARAPACÁ: **Iquique:** Huara, 3406 m, 18 Feb 2003, *Gardner & Knees 6517* (E); Huara, 3847 m, 11 Dec 2008, *R. Baines et al. 362* (E). **Tamarugal:** Sine loc., 1888, *Philippi s.n.* (K, US). **PERU.** "In rupibus Cordiller. pr. San Antonio", Jan 1854, *W. Lechler 1801* (P barcode P00648954, F barcode F0041234F, K barcode K000612633); Sine loc., 27 Nov 1929, *F.L. Herrera 2206* (B). AREQUIPA: **Arequipa:** 3000 m, 2002, *Enero 2870* (B); Mirador del Colca Chivay, 3950 m, 18 Mar 2006, *Caceres & al. 3350* (B); Caylloma, Chivay, 3700–3800 m, 30 Apr 2000, *Quipuscoa &. al. 1873* (HUSA); Tocra, 17 Mar 2006, *Caceres & al. 3350* (HUSA); Chivay 15°26'8''S, 71°36'17''W, 3600–3800m, 21 Apr 2006, *Van Der Werff & al. 20824* (HUSA). AYACUCHO: **Lucanas:** Entre Puqio y Nasca, arriba de Puqio Pedregoso. 5500 m, 20 Mar 1949, *R. Ferreyra 5500* (USM, BONN). MOQUEGUA: **Mariscal Nieto**: Cocutea, 3420 m, Sep 1997, *J. Alban & N. Malca 10146* (BONN); 17°01'34,7''S, 070°51'04,1''W, 2616 m, 21 Apr 2006, *Weigend, Caceres & Schwarzer 8413* (B); 16°47'26,0''S, 70°54'23,2''W, 1600 m, 21 Apr 2006, *Weigend, Caceres & Schwarzer 8381* (B); Apr 2009, *Weigend 9192* (B). PUNO: Titicaca See, 3822 m, 22 Jan 1975, *Schwabe s.n.* (B). **Azángaro:** Tequena, 3900 m, 19 Feb 1948, R. Ferreyra 405 (USM, BONN). **Roman:** 15°41,336'S, 070°48,688'W, 4196 m, 11 Apr 2004, *Weigend & Schwarzer 7828* (B). TACNA: **Tarata:** Tarata, 2780 m, 10 Nov 2006, *Caceres 5838* (HUSA).

Key to South-American species of *Phacelia* sect. *Glandulosae*

1. Leaves long petiolate bipinnatisect with 3–5 pairs of obovate pinnules, 2–4 lobed and a larger, blunt terminal one. Occasionally intercalated one-lobed pinnae between the pinnules. Corolla tubular (rarely campanulate), white (rarely lilac), 2–5 mm long, with scales next to the insertion point of the stamens. Stamens and style equaling the corolla. Style glabrous. Capsules spherical. Seeds reticulate, dorsally slightly echinate, ventrally rugose……...***P. artemisioides***

1*. Leaves (oblong) lanceolate, pinnatifid to pinnatisect (rarely bipinnatisect, or lobed with lobes either sinuate or obtuse, or blunt at apex and base), (sessile to) petiolate to long petiolate, with (3)4–9(12) pairs of pinnae (toothed sometimes lobed or trilobate) of variable sizes and intercalated segments of variable lengths. Corolla campanulate, lilac to purple (bluish), 4–8 mm long, with semi-ovate to longitudinal scales next to the insertion point of the stamens, these sometimes minute or absent. Stamens and style included or exserted. Style pilose–glandular from base to midpoint. Capsule widely ovoidal. Testa reticulate on both sides, margins more or less rugose……***P. pinnatifida***

3.7 ACKNOWLEDGMENTS

We are grateful to the Herbaria of B, BONN, HUSA, LP, SDSU, SI for allowing access to their collections, as well as to all the collectors. We thank Laura Iharlegui, Melisa Mulero, Michael Simpson, Federico Luebert, Gisela Sancho, for providing access to crucial plant material for this study or information and photos relevant to the collections, as well as the Instituto de Botánica Darwinion for providing most of the photographs of Figure 3.2 and the respective photographers. We also thank Thomas Joßberger, Nicole Schmandt, Claudia Schütte and Yaron Malkowsky for their technical assistance during this study, Karuna Verma for her contributions to the data analysis and Julius Jeiter for his support and valuable comments throughout this work.

3.8 LITERATURE CITED

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., & Hohenlohe, P.A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet* 17: 81–92.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. 2012. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/accessed.

Antonelli, A. 2015. Multiple origins of mountain life. *Nature* 524: 300–301.

Antonelli, A., Zizka, A., Carvalho, F.A., Scharn, R., Bacon, C.D., Silvestro, D., & Condamine, F.L. 2018. Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences* 115: 6034–6039.

Atwood, N.D. 1972. New species in the *Phacelia Crenulata* group (Hydrophyllaceae). *Rhodora* 74: 451–468.

Atwood, N.D. 1975. A revision of the *Phacelia Crenulatae* group (Hydrophyllaceae) for North America. *The Great Basin Naturalist* 35: 127–190.

Baldini, R.M., Cota-Sánchez, J.H., & Aedo, C. 2021. Is the demise of plant taxonomy in sight? Maybe yes, maybe no…. *Webbia* 76: 3–10.

Beier, B.-A., Chase, M.W., & Thulin, M. 2003. Phylogenetic relationships and taxonomy of subfamily Zygophylloideae (Zygophyllaceae) based on molecular and morphological data. *Plant Syst. Evol.* 240: 11–39.

Böhnert, T., Luebert, F., Merklinger, F.F., Harpke, D., Stoll, A., Schneider, J.V., Blattner, F.R., Quandt, D., & Weigend, M. 2022. Plant migration under long-lasting hyperaridity – phylogenomics unravels recent biogeographic history in one of the oldest deserts on Earth. *New Phytologist* 234: 1863–1875.

Böhnert, T., Neumann, M., Quandt, D., & Weigend, M. 2023. Phylogeny based generic reclassification of *Muscari* sensu lato (Asparagaceae) using plastid and genomic DNA. *TAXON* n/a.

Böhnert, T., Weigend, M., Merklinger, F., Quandt, D., & Luebert, F. 2020. Historical assembly of Zygophyllaceae in the Atacama Desert. *Frontiers of Biogeography* 12 (3): e45197

Brand, A. 1913. Hydrophyllaceae. In: Engler, A. (ed.), *Das Pflanzenreich*, IV.251 (Heft 59). Leipzig: Engelmann.

Cai, L., Kreft, H., Taylor, A., Denelle, P., Schrader, J., Essl, F., van Kleunen, M., Pergl, J., Pyšek, P., Stein, A., Winter, M., Barcelona, J.F., Fuentes, N., Inderjit, Karger, D.N., Kartesz, J., Kuprijanov, A., Nishino, M., Nickrent, D., Nowak, A., Patzelt, A., Pelser, P.B., Singh, P., Wieringa, J.J., & Weigelt, P. (2022). Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytologist* 237: 1432–1445.

Chifman, J., & Kubatko, L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.

Constance, L. 1949. A Revision of *Phacelia* subgenus *Cosmanthus* (Hydrophyllaceae). *Contributions from the Gray Herbarium of Harvard University* 1–48.

Constance, L. 1950. Some interspecific relationships in *Phacelia* subgenus *Cosmanthus*. *Proceedings of the American Academy of Arts and Sciences* 78: 135–147.

Constance, L. 1952. *Howellanthus*, a new subgenus of *Phacelia*. *Madroño* 11: 198–203.

Constance, L. 1963. Chromosome number and classification in Hydrophyllaceae. *Brittonia* 15: 273–285.

Crisci, J.V., Katinas, L., Apodaca, M.J., & Hoch, P.C. 2020. The end of botany. *Trends in Plant Science* 25: 1173–1176.

Damasco, G., Daly, D.C., Vicentini, A., & Fine, P.V.A. 2019. Reestablishment of *Protium cordatum* (Burseraceae) based on integrative taxonomy. *TAXON* 68: 34–46.

Deginani, N.B. 1982. Revisión de las especies argentinas del género *Phacelia* (Hydrophyllaceae). *Darwiniana* 24: 405–435.

Doyle, J.J., & Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem.bull, 19: 11–15.

Eaton, D.A.R. 2020. Toytree: A minimalist tree visualization and manipulation library for Python. *Methods in Ecology and Evolution* 11: 187–191.

Eaton, D.A.R., & Overcast, I. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36: 2592–2594.

Eaton, D.A.R., Spriggs, E.L., Park, B., & Donoghue, M.J. 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology* 66: 399–412.

Ebach, M.C., Valdecasas, A.G., & Wheeler, Q.D. 2011. Impediments to taxonomy and users of taxonomy: accessibility and impact evaluation. *Cladistics* 27: 550–557.

Evanno, G., Regnaut, S., & Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620.

Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: Finding maximum likelihood estimates. *Evolution* 35: 1229–1242.

Ferguson, D.M. 1998. Phylogenetic analysis and relationships in Hydrophyllaceae based on *ndhF* sequence data. *Systematic Botany* 23: 253–268.

Fjeldså, J., Bowie, R.C.K., & Rahbek, C. 2012. The role of mountain ranges in the diversification of birds. *Annual Review of Ecology, Evolution, and Systematics* 43: 249–265.

GBIF. 2021. Global Biodiversity Information Facility. https://www.gbif.org/ (accessed Sep 2022)

Gilbert, C., Dempcy, J., Ganong, C., Patterson, R., & Spicer, G.S. 2005. Phylogenetic relationships within *Phacelia* subgenus *Phacelia* (Hydrophyllaceae) inferred from nuclear rDNA ITS Sequence Data. *Systematic Botany* 30: 627–634.

Gillett, G.W. 1960. A systematic treatment of the *Phacelia franklinii* group. *Rhodora* 62: 205–222.

Glass, P.M., & Levy, F. 2011. Correspondence of morphology, phylogeny and reproductive barriers in *Phacelia* subgenus *cosmanthus* (Hydrophyllaceae)1. *Tbot* 138: 341–352.

Grenier, S., Barre, P., & Litrico, I. 2016. Phenotypic plasticity and selection: Nonexclusive mechanisms of adaptation. *Scientifica* 2016: e7021701.

Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N., & Hickey, D.A. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* 23: 167–172.

Hansen, D.R., Spicer, G.S., & Patterson, R. 2009. Phylogenetic relationships between and within *Phacelia* Sections *Whitlavia* and *Gymnobythus* (Boraginaceae). *Systematic Botany* 34: 737–746.

Heckard, L.R. 1960. Taxonomic studies in the *Phacelia magellanica* polyploid complex: with special reference to the California members. *Univ. Calif. Publ. Bot.* 32: 1–126.

Heckard, L.R. 1963. The Hydrophyllaceae. *The Quarterly Review of Biology* 38.

Helmstetter, A.J., Oztolan-Erol, N., Lucas, S.J., & Buggs, R.J.A. 2020. Genetic diversity and domestication of hazelnut (*Corylus avellana* L.) in Turkey. *PLANTS, PEOPLE, PLANET* 2: 326–339.

Hofmann, M. 1999. Flower and fruit development in the genus *Phacelia* (Phacelieae, Hydrophyllaceae): Characters of systematic value. *Systematics and Geography of Plants* 68: 203–212.

Hofmann, M., Walden, G.K., Hilger, H.H., & Weigend, M. 2016. Hydrophyllaceae. Pp. 221–238 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*. Springer, Heidelberg, Germany.

Hoorn, C., Mosbrugger, V., Mulch, A., & Antonelli, A. 2013. Biodiversity from mountain building. *Nature Geosci* 6: 154–154.

Howell, J.T. 1946. A revision of *Phacelia* sect. *Euglypta*. *The American Midland Naturalist* 36: 381–411.

Huang, H., & Knowles, L.L. 2016. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. *Systematic Biology* 65: 357–365.

Hughes, C., & Eastwood, R. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences* 103: 10334–10339.

Jackson, N.D., Carstens, B.C., Morales, A.E., & O'Meara, B.C. 2017. Species delimitation with gene flow. *Systematic Biology* 66: 799–812.

Jordon-Thaden, I.E., Beck, J.B., Rushworth, C.A., Windham, M.D., Diaz, N., Cantley, J.T., Martine, C.T., & Rothfels, C.J. 2020. A basic ddRADseq two-enzyme protocol performs well with herbarium and silica-dried tissues across four genera. *Applications in Plant Sciences* 8: e11344.

JSTOR. 2021 (continuously updated). JSTOR Global Plants. https:// plants.jstor.org/ (accessed Aug 2022)

Kim, B.Y., Wei, X., Fitz-Gibbon, S., Lohmueller, K.E., Ortego, J., Gugger, P.F., & Sork, V.L. 2018. RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae). *Molecular Ecology* 27: 4556–4571.

Lagomarsino, L.P., Condamine, F.L., Antonelli, A., Mulch, A., & Davis, C.C. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* 210: 1430–1442.

Lang, P.L.M., Weiß, C.L., Kersten, S., Latorre, S.M., Nagel, S., Nickel, B., Meyer, M., & Burbano, H.A. 2020. Hybridization ddRAD-sequencing for population genomics of nonmodel plants using highly degraded historical specimen DNA. *Molecular Ecology Resources* 20: 1228–1247.

Lemmon, E.M., & Lemmon, A.R. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44: 99–121.

Levy, F. 2021. Evaluation of the taxonomy of *Phacelia purshii* (Hydrophyllaceae). *The Journal of the Torrey Botanical Society* 148(1): 1-16.

Levy, F., Hou, Z.H., & Taylor-Bennetts, S. 2021. A new variety of *Phacelia dubia* (Hydrophyllaceae) and a summary of variation and biogeography of the *P. dubia* complex. *The Journal of the Torrey Botanical Society* 148(4): 266-284.

Linder, H.P. 2008. Plant species radiations: where, when, why? *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 3097–3105.

Lipson, M. 2020. Applying f4-statistics and admixture graphs: Theory and examples. *Molecular Ecology Resources* 20: 1658–1667.

Lipson, M., Loh, P.-R., Levin, A., Reich, D., Patterson, N., & Berger, B. 2013. Efficient moment-based inference of admixture parameters and sources of gene flow. *Molecular Biology and Evolution* 30: 1788–1802.

Liu, J.-X., Zhou, M.-Y., Yang, G.-Q., Zhang, Y.-X., Ma, P.-F., Guo, C., Vorontsova, M.S., & Li, D.-Z. 2020. ddRAD analyses reveal a credible phylogenetic relationship of the four main genera of *Bambusa-Dendrocalamus-Gigantochloa* complex (Poaceae: Bambusoideae). *Molecular Phylogenetics and Evolution* 146: 106758.

Lörch, M., Mutke, J., Weigend, M., & Luebert, F. 2021. Historical biogeography and climatic differentiation of the *Fulcaldea-Archidasyphyllum-Arnaldoa* clade of Barnadesioideae (Asteraceae) suggest a Miocene, aridity-mediated Andean disjunction associated with climatic niche shifts. *Global and Planetary Change* 201: 103495

Luebert, F., Cecchi, L., Frohlich, M.W., Gottschling, M., Guilliams, C.M., Hasenstab-Lehman, K.E., Hilger, H.H., Miller, J.S., Mittelbach, M., Nazaire, M., Nepi, M., Nocentini, D., Ober, D., Olmstead, R.G., Selvi, F., Simpson, M.G., Sutorý, K., Valdés, B., Walden, G.K., & Weigend, M. 2016. Familial classification of the Boraginales. *TAXON* 65: 502–522.

Madriñán, S., Cortés, A., & Richardson, J. 2013. Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Frontiers in Genetics* 4.

Martynov, A.V. 2012. Ontogenetic systematics: The synthesis of taxonomy, phylogenetics, and evolutionary developmental biology. *Paleontol. J.* 46: 833–864.

Mirra, F.B., O'Leary, N., & Moroni, P. 2021. Typification of names linked to *Phacelia* (Hydrophyllaceae) for Flora Argentina. *Darwiniana, Nueva Serie* 9: 209–216.

Mizutani, M., & Kanaoka, M.M. 2018. Environmental sensing and morphological plasticity in plants. *Seminars in Cell & Developmental Biology* 83: 69–77.

Murillo-A., J., Valencia-D., J., Orozco, C.I., Parra-O., C., & Neubig, K.M. (2022). Incomplete lineage sorting and reticulate evolution mask species relationships in Brunelliaceae, an Andean family with rapid, recent diversification. *American Journal of Botany* 109: 1139–1156

Mutke, J., & Barthlott, W. 2005. Patterns of vascular plant diversity at continental to global scale. *Biologiske Skrifter* 55: 521–537.

Nazareno, A.G., Dick, C.W., & Lohmann, L.G. 2017. Wide but not impermeable: Testing the riverine barrier hypothesis for an Amazonian plant species. *Molecular Ecology* 26: 3636–3648.

Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F., & van Kleunen, M. 2010. Plant phenotypic plasticity in a changing climate. *Trends in Plant Science* 15: 684–692.

O'Leary N., & al. 2022. Hydrophyllaceae. Pp 194-204 in: F. O. Zuloaga & C. A. Zanotti (eds.), *Flora Argentina* 19(3). Instituto Multidisciplinario de Biología Vegetal (CONICET-UNC)

Otero, A., Vargas, P., Fernández-Mazuecos, M., Jiménez-Mejías, P., Valcárcel, V., Villa-Machío, I., & Hipp, A.L. 2022. A snapshot of progenitor–derivative speciation in *Iberodes* (Boraginaceae). *Molecular Ecology* 31: 3192–3209.

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.C., & Samadi, S. 2015. Use of RAD sequencing for delimiting species. *Heredity* 114: 450–459.

Pearman, P.B., Alioto, T.S., Trotta, J.-R.P., & Columbus, J.T. 2021. Genotyping-by-sequencing resolves relationships in Polygonaceae tribe Eriogoneae. *TAXON* 70: 826–841.

Pennington, R.T. 1996. Molecular and morphological data provide phylogenetic resolution at different hierarchical levels in *Andira*. *Systematic Biology* 45: 496–515.

Pennington, R.T., Lavin, M., Särkinen, T., Lewis, G.P., Klitgaard, B.B., & Hughes, C.E. 2010. Contrasting plant diversification histories within the Andean biodiversity hotspot. *Proceedings of the National Academy of Sciences* 107: 13783–13787.

Pérez-Escobar, O.A., Zizka, A., Bermúdez, M.A., Meseguer, A.S., Condamine, F.L., Hoorn, C., Hooghiemstra, H., Pu, Y., Bogarín, D., Boschman, L.M., Pennington, R.T., Antonelli, A., & Chomicki, G. 2022. The Andes through time: Evolution and distribution of Andean floras. *Trends in Plant Science* 27: 364–378.

Peter, B.M. 2016. Admixture, population structure, and F-statistics. *Genetics* 202: 1485–1501.

Pickrell, J.K., & Pritchard, J.K. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics* 8: e1002967.

Pritchard, J.K., Stephens, P., Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155, 945 – 959.

Pritchard, J.K., Wen, X. & Falush, D. 2010. Documentation for structure software: Version 2.3.

Rouhan, G., & Gaudeul, M. 2021. Plant Taxonomy: A historical perspective, current challenges, and perspectives. Pp. 1–38 in: P. Besse (ed.), *Molecular plant taxonomy: Methods and protocols*. Methods in Molecular Biology. New York, NY: Springer US.

Särkinen, T., Pennington, R.T., Lavin, M., Simon, M.F., & Hughes, C.E. 2012. Evolutionary islands in the Andes: persistence and isolation explain high endemism in Andean dry tropical forests. *Journal of Biogeography* 39: 884–900.

Sassone, A.B., Arroyo, M.T.K., Arroyo-Leuenberger, S.C., García, N., & Román, M.J. 2021. One species with a disjunct distribution or two with convergent evolution? Taxonomy of two South American garlics. *TAXON* 70: 842–853.

Scherson, R.A., Vidal, R., & Sanderson, M.J. 2008. Phylogeny, biogeography, and rates of diversification of New World *Astragalus* (Leguminosae) with an emphasis on South American radiations. *American Journal of Botany* 95: 1030–1039.

Schilling, M.P., Gompert, Z., Li, F.-W., Windham, M.D., & Wolf, P.G. 2018. Admixture, evolution, and variation in reproductive isolation in the *Boechera puberula* clade. *BMC Evolutionary Biology* 18: 61.

Scotland, R.W., Olmstead, R.G., & Bennett, J.R. 2003. Phylogeny reconstruction: The role of morphology. *Systematic Biology* 52: 539–548.

Simpson, M.G., Johnson, L.A., Villaverde, T., & Guilliams, C.M. 2017. American amphitropical disjuncts: Perspectives from vascular plant analyses and prospects for future research. *American Journal of Botany* 104: 1600–1650.

Smith, L.M., & Carstens, C.B. 2022. Species delimitation using molecular data. P. 16 in: *Species Problems and Beyond*. CRC Press.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Sultan, S.E. 2000. Phenotypic plasticity for plant development, function and life history. *Trends in Plant Science* 5: 537–542.

Tripp, E.A., Tsai, Y.-H.E., Zhuang, Y., & Dexter, K.G. 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution* 7: 7920–7936.

Vargas, O.M., Ortiz, E.M., & Simpson, B.B. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist* 214: 1736–1750.

Vasile, M.-A., Jeiter, J., Weigend, M., & Luebert, F. 2020. Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia*. *Systematics and Biodiversity* 18: 757–770.

Vasile, M.-A., Luebert, F., Jeiter, J., & Weigend, M. 2021. Fruit evolution in Hydrophyllaceae. *American Journal of Botany* 108: 925–945.

Vaux, F., Dutoit, L., Fraser, C.I., & Waters, J.M. 2023. Genotyping-by-sequencing for biogeography. *Journal of Biogeography* 50: 262–281.

Voss, J.W. 1935. A revisional study of the *Phacelia hispida* group. *Bulletin of the Southern California Academy of Sciences* 33: 169–178.

Voss, J.W. 1937. A Revision of the *Phacelia crenulata* group for North America. *Bulletin of the Torrey Botanical Club* 64: 81–96.

Walden, G.K., Garrison, L.M., & Patterson, R. 2016. Nomenclatural adjustments in *Phacelia* sect. *Glandulosae* (Hydrophyllaceae, Boraginales). *Wnan* 76: 287–290.

Walden, G.K., Garrison, L.M., Spicer, G.S., Cipriano, F.W., & Patterson, R. 2014. Phylogenies and chromosome evolution of *Phacelia* (Boraginaceae: Hydrophylloideae) inferred from nuclear ribosomal and chloroplast sequence data. *Madr* 61: 16–47.

Walden, G.K., & Patterson, R. 2011. Nomenclatural kankedorts in *Phacelia* (Boraginaceae: Hydrophylloideae). *Madr* 58: 267–272.

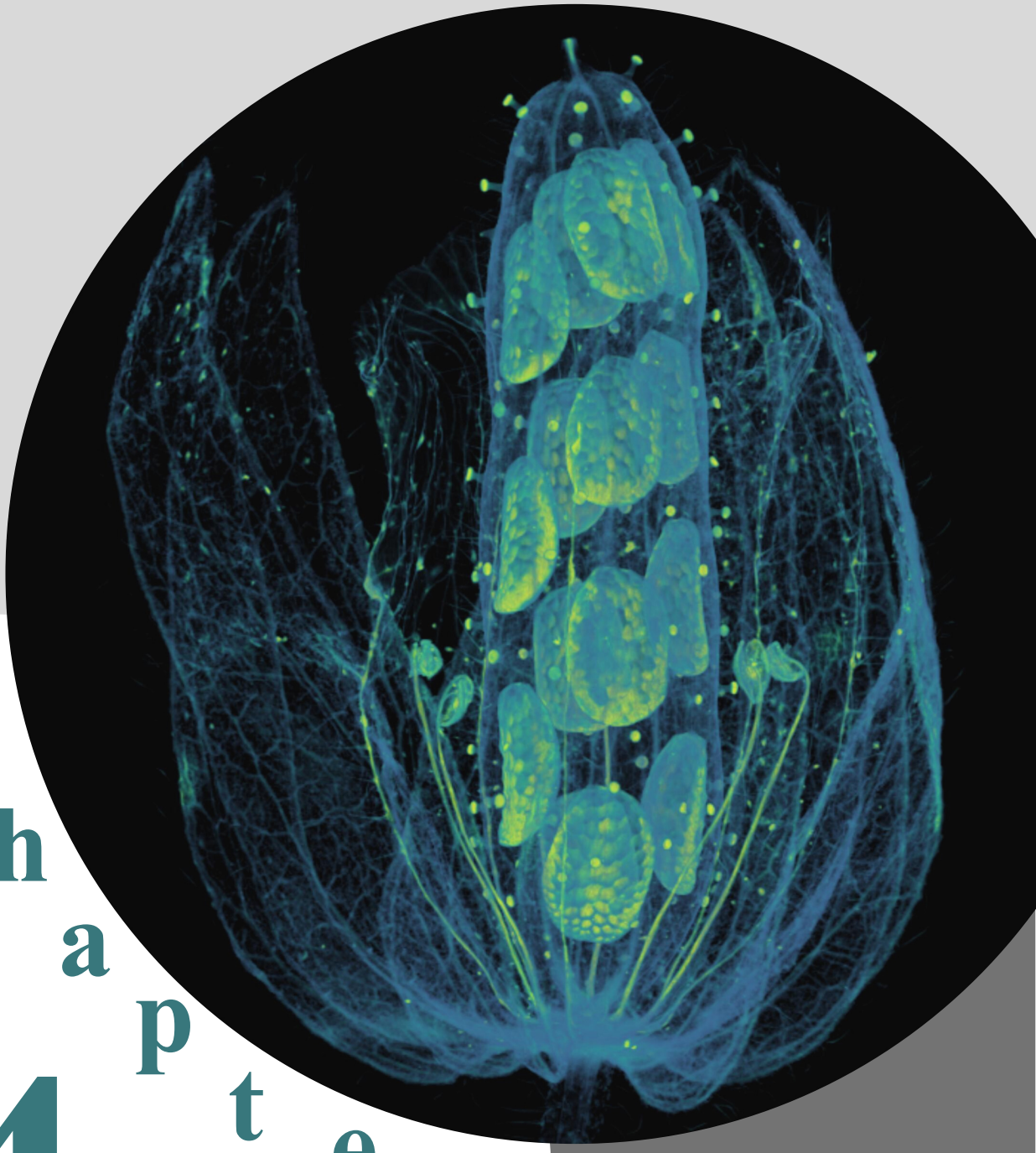
Walden, G.K., & Patterson, R. 2012. Nomenclature of subdivisions within *Phacelia* (Boraginaceae: Hydrophylloideae). *Madr* 59: 211–222.

Weddell, H.A. 1857. Chloris Andina: essai d'une flore de la région alpine des Cordillères de l'Amérique du Sud. P. 86 in: vol. 2. Paris: P. Bertrand.

Wilkins, J.S., Zachos, F.E., & Pavlinov, I.Y. (eds.). 2022. *Species Problems and Beyond*, 1st ed. Boca Raton: CRC Press.

Zachos, F.E., Christidis, L., & Garnett, S.T. 2020. Mammalian species and the twofold nature of taxonomy: a comment on Taylor et al. 2019. *Mammalia* 84: 1–5.

Flower and fruit development of three species from Hydrophylleae sheds new light on flower evolution in Hydrophyllaceae



Chapter 4

Micro–computed tomography–based volume rendering showing seed arrangement of a mature capsule of *Emmenanthe penduliflora* Benth. (Hydrophyllaceae) collected in Bonn Botanical Gardens.

Vasile MA, Weigend M, Ziegler A., Jeiter J. 2022. Flower and fruit development of three species of Hydrophylleae sheds new light on flower evolution in Hydrophyllaceae. *International Journal of Plant Sciences* 183 (5): 331–347. https://doi.org/10.1086/719907

FLOWER AND FRUIT DEVELOPMENT OF THREE SPECIES OF HYDROPHYLLEAE SHEDS NEW LIGHT ON FLOWER EVOLUTION IN HYDROPHYLLACEAE

Maria-Anna Vasile,[1,*] Maximilian Weigend,* Alexander Ziegler,† and Julius Jeiter*

*Nees-Institut für Biodiversität der Pflanzen, Rheinische Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany; and †Institut für Evolutionsbiologie und Ökologie, Rheinische Friedrich-Wilhelms-Universität Bonn, An der Immenburg 1, 53121 Bonn, Germany

Editor: Maria von Balthazar

Premise of research. Hydrophyllaceae are characterized by simple floral organization contrasting with a peculiar floral architecture resulting from complex compartments formed by stamen–corolla tube modifications. Additionally, the internal ovary architecture of Hydrophyllaceae shows significant variation, while the developmental trajectory of the gynoecium is relatively conserved. Despite insights from recent studies of the flower and fruit evolution of the family, there are only a few ontogenetic studies of Hydrophyllaceae, and a complete understanding of the underlying processes has not yet been achieved.

Methodology. Here, we use scanning electron microscopy and micro–computed tomography to investigate the flower and fruit ontogeny of two genera of Hydrophylleae, *Emmenanthe* and *Pholistoma*, with a particular focus on the gynoecium and modifications of the stamen–corolla tube.

Pivotal results. Our results complement two previously published data sets, broadening our understanding of Hydrophylleae evolution. Hydrophylleae comprise only a few monotypic or small genera, but their floral evolution appears to be remarkably complex, in terms of both gynoecial structure and perianth modifications. We find 10 stamen–corolla tube modifications, although these may be rudimentary in species previously considered as lacking them altogether (e.g., *Emmenanthe*). The relative conservation of perianth architecture contrasts with the highly variable internal ovary architecture of Hydrophylleae. There is considerable divergence in ovule/seed number, as well as in the details of (parietal) placentation and septation, and we propose a hypothetical evolutionary series for the internal ovary architectural diversity of Hydrophyllaceae.

Conclusions. We propose that—starting from a fairly conserved floral organization—minor heterochronous shifts in both perianth and ovary development can explain most of the morphological diversity found in the flowers and fruits of Hydrophylleae.

Keywords: Boraginales, gynoecium, heterochrony, internal ovary architecture, placentation, stamen–corolla tube modifications.

Introduction

Boraginales are divided into two major clades: Boraginales I and II (Weigend et al. 2014). Boraginales I comprise Codonaceae, Wellstediaceae, and Boraginaceae, the most species-rich family of the entire order. Boraginales II comprise eight families: Hydrophyllaceae, Namaceae, Ehretiaceae, Lennoaceae, Heliotropiaceae, Cordiaceae, Coldeniaceae, and Hoplestigmataceae (Luebert et al. 2016). Fruit morphology has traditionally been considered an important character complex for their familial classification. In both major clades, the early-diverging lineages (Codonaceae/Wellstediaceae and Hydrophyllaceae/Namaceae, respectively) bear capsular fruits, while the remaining lineages, with the sole exception of Lennoaceae (Luebert et al. 2016), have either mericarpids (i.e., dispersal units of schizocarpic or eremocarpic fruits [nutlets] formed by part of a carpel; Hilger 2014; Jeiter et al. 2018) or drupes. Details of fruit morphology, development, and evolution have been the focus of several recent studies of Boraginales (Gottschling 2004; Gottschling et al. 2014; Luebert et al. 2016; Jeiter et al. 2018; Heigl et al. 2020; Vasile et al. 2021). Indehiscent fruits have been the primary focus of these studies, while capsular fruits have largely been neglected until recently (Jeiter et al. 2016; Vasile et al. 2021). The phylogenetic results showing the independent evolutionary trends from

[1] Author for correspondence; email: m.vasile@uni-bonn.de.

Manuscript received January 2022; revised manuscript received March 2022; electronically published April 25, 2022.

International Journal of Plant Sciences, volume 183, number 5, June 2022. © 2022 The University of Chicago. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits non-commercial reuse of the work with attribution. For commercial use, contact journalpermissions@press.uchicago.edu. Published by The University of Chicago Press. https://doi.org/10.1086/719907

capsules to nutlets in both major clades (Weigend et al. 2014) have raised questions about the mechanisms for the reduction and subsequent stabilization of low seed numbers in capsular fruits. Hydrophyllaceae, as the largest capsule-bearing family within Boraginales, was identified as an ideal object for studying such evolutionary trends.

Hydrophyllaceae comprise some 240–260 species in 12 genera (Luebert et al. 2016) distributed in North and South America (Vasile et al. 2020). The family is divided into three clades. The largest and morphologically most diverse clade is the Romanzoffieae, with the genera *Phacelia* Juss. and *Romanzoffia* Cham. The second-largest clade is the Hydrophylleae, comprising the genera *Hydrophyllum* L., *Nemophila* Nutt. ex W.P.C.Barton, *Pholistoma* Lilja, *Eucrypta* Nutt, *Emmenanthe* Benth., and *Ellisia* L. A third, unnamed clade comprises the genera *Draperia* Torr., *Howellanthus* (Constance) Walden & R.Patt., *Hesperochiron* S. Watson, and *Tricardia* Torr. ex S.Watson (Luebert et al. 2016; Vasile et al. 2020).

In a recent study by Vasile et al. (2021), the internal ovary architecture of Hydrophyllaceae and its evolution were studied, but the study included only three representatives of Hydrophylleae. The focus was on the reduction versus the increase of ovule numbers across the family, and it revealed a surprising diversity of placentation types (i.e., placenta form and size and the way ovules are attached to the placentae). Placentation was found to be particularly divergent in the three species of Hydrophylleae studied. Placentation appeared to correlate with the high variability of ovule and seed number in the clade. In the present study, we expand the sampling to explore this aspect in greater detail.

Another aspect of Hydrophyllaceae that is still incompletely studied is modifications of the stamen–corolla tube that are present in the form of often scale-shaped outgrowths, protuberances, or invaginations. They are variously termed corolla scales, faucal scales, and basal scales. Such modifications are found in several families of Boraginales I but are absent in the majority of the families of Boraginales II (Bittrich 2016; Gottschling et al. 2016). Major exceptions in Boraginales II are Hydrophyllaceae and Namaceae (Hofmann et al. 2016). Stamen–corolla tube modifications were recently studied in Hydrophyllaceae, but only three genera of Hydrophylleae were included in that study (Jeiter and Weigend 2018). These authors demonstrated an evolutionary trend toward increasing size, complexity, and degree of functional integration for the flowers in the clade.

Modifications of the corolla (fused or free) or the stamen–corolla tube are a widespread phenomenon among angiosperms. These modifications evolved independently in various plant groups other than Boraginales—such as the coronas or paracorollas of many Sileneae, Caryophyllaceae (Arber 1939; Thomson 1942), the often brightly colored scales that enclose the nectar in *Nasa* Weigend, Loasaceae (Weigend and Gottschling 2006), and the corolla appendages in Schwenckieae, Solanaceae (Paucar et al. 2020). These modifications have significant repercussions for floral architecture (sensu Endress 1996). In Hydrophyllaceae, stamen–corolla tube modifications have been shown to be integrated with the gynoecial nectary disk located at the base of the ovary. In several Hydrophyllaceae, especially some Hydrophylleae, the modifications compartmentalize the flower, partitioning the nectar and leading to revolver architecture (Jeiter and Weigend 2018).

The present study investigates the flower and fruit development of three additional species of Hydrophylleae (Hydrophyllaceae), the monotypic genus *Emmenanthe* and two of the three species of the genus *Pholistoma*, which have not previously been studied in detail. We address the following questions: (i) How do the flower and fruit of these three species develop? (ii) How do the stamen–corolla tube modifications and the nectaries integrate into the general floral architecture? (iii) How does placentation in this clade evolve, and can something be learned from this aspect with regard to the fruit evolution of Boraginales?

Material and Methods

Plant Material

Floral buds of various developmental stages, anthetic and postanthetic flowers, and fruits of three Hydrophyllaceae species of Hydrophylleae, namely, *Emmenanthe penduliflora* Benth. (US-0-BONN-38156), *Pholistoma auritum* (Lindl.) Lilja (US-0-BONN-38190), and *Pholistoma membranaceum* (Benth.) Constance (US-0-BONN-36930), were collected from cultivated plants at Bonn University Botanic Gardens (fig. 1). The seeds were obtained from the California Botanic Garden (formerly Rancho Santa Ana Botanic Garden). The samples were fixed in formaldehyde–acetic acid–ethanol (FAA; formaldehyde, ~2%; acetic acid, 4%; ethanol, 70%) for at least 1 wk, and the samples were subsequently prepared for scanning electron microscopy (SEM) and micro–computed tomography (μCT). Vouchers are stored at the herbarium of the Nees-Institut für Biodiversität der Pflanzen (BONN).

Scanning Electron Microscopy

Fixed samples were rinsed with ethanol (70%) three times, dissected under a stereomicroscope, and stored in ethanol (70%). For chemical dehydration, subsets of the samples were transferred into ethanol (70%) and acetic acid (4%) for at least 1 h or overnight. The samples were sequentially transferred into formaldehyde dimethyl acetal (FDA) for 2 h, FDA and acetone (1:1) for 1 h, and finally acetone for 1 h. The last step was repeated twice, and the samples were finally stored in acetone. Flowers representing different developmental stages were critical-point dried with CO_2 in a critical-point dryer (CPD 020, Balzers Union, Balzers, Liechtenstein) following standard protocols. Dried specimens were mounted on glass slide–covered aluminum stubs using conductive carbon cement (Leit-C, PLANO, Wetzlar, Germany). Dried corollas were directly mounted onto aluminum stubs using conductive carbon tape (25-mm Leit Tabs, PLANO) and conductive carbon cement. The mounted specimens were sputter-coated with gold (SCD 040, Balzers Union) for 1.5–3 min at 30 mA, depending on their topological complexity. The SEM images were taken at 15 kV using a Stereoscan 200 electron microscope (CAMBRIDGE, Cambridge, UK). The contrast and brightness of the images were linearly adjusted when needed using Affinity Photo 1.8 (Serif, Nottingham, UK).

Micro–computed Tomography

Samples for micro–computed tomography (μCT) were prepared following Staedler et al. (2013). Four different FAA-fixed

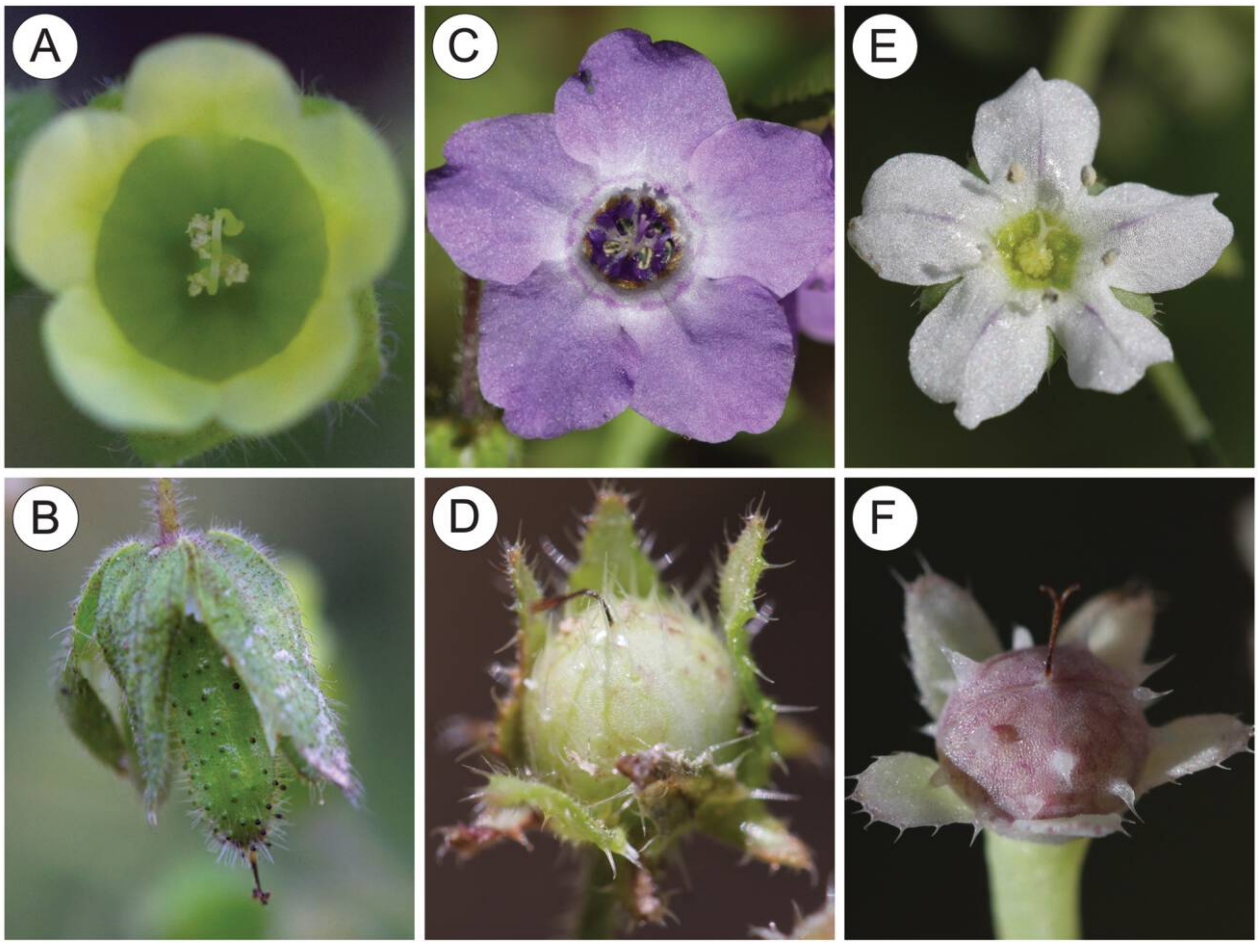


Fig. 1 Flowers and fruits of the studied species. *A*, *B*, *Emmenanthe penduliflora*. *A*, Anthetic flower. *B*, Capsule; persistent corolla and androecium removed. *C*, *D*, *Pholistoma auritum*. *C*, Anthetic flower. *D*, Capsule. *E*, *F*, *Pholistoma membranaceum*. *E*, Anthetic flower. *F*, Capsule; the fruits turn purple as they mature.

developmental stages for each species (table 1) were transferred for infiltration into a solution consisting of ethanol (70%) and phosphotungstic acid (PTA; 1%) for at least 2 wk. The infiltration medium was exchanged every second day. Subsequently, the samples were dehydrated in an ascending ethanol series (80%, 90%, and 99%), each step containing PTA (1%), and finally were transferred into a solution of acetone and PTA (1%) that was exchanged three times. The samples were then critical-point dried from acetone in CO_2 (CPD 020, Balzers Union) and individually mounted on thin aluminum rods (diameter: 3 mm) using two-component epoxy glue (UHU PLUS Sofortfest, UHU, Bühl, Germany).

Analyses were carried out using a SKYSCAN 1272 μCT scanner (Bruker, Kontich, Belgium) equipped with an L11871-20 microfocus X-ray source (Hamamatsu Photonics, Hamamatsu City, Japan) and a Ximea xiRAY16 camera (Ximea, Münster, Germany). Scan settings are given in table 1. Reconstructions were performed using the InstaRecon engine version 2 (InstaRecon, Champaign, IL). The resulting image stacks were visualized using Dragonfly 2020.1 (Object Research Systems 2020). The histogram for each scanned object was manually adjusted between 5000 and 50,000. For visualization purposes, in the volume view of the reconstructed data, functions such as "shapes" and "clipping box" were used to hide distracting features, such as parts of the calyx, corolla, or anthers. "Shapes" or the "clip" function was also used to separate the individual flowers of an inflorescence tip. Screenshots of virtual longitudinal and cross sections were taken after manually adjusting the yaw, pitch, and role of the image plane. To visualize the placentae, ovules, and seeds, the "segmentation" module of the software Dragonfly was used. Regions of interest (ROIs) were segmented using a combination of histogram-based and manual segmentation. Segmented volumes were refined using the function "process islands," with the voxel count set to 50. To further remove artifacts, the following morphological operations were applied where necessary: 2 × "dilate," 1 × "smooth," and 1 × "erode." Surfaces were rendered with the "generate contour mesh" tool, and the threshold values for the ROIs were adjusted to between 5 and 50.

Results

General Development and Morphology

Flowers of *Emmenanthe* and *Pholistoma* are tetracyclic and pentamerous with a bicarpellate, syncarpous gynoecium. Flowers are arranged in terminal scorpioid cymes, although complex

Table 1

Scanned Samples and Scan Settings

Taxon, developmental stage	Source voltage (kV)	Source current (μA)	Exposure time (ms)	Frame averaging	Pictures per sample	Camera binning	Isotropic voxel size (μm)	Figure(s)
Emmenanthe penduliflora:								
Inflorescence tip	45	165	444	6	1472	2 × 2	4.5	5*A*–5*D*
Preanthetic flower	45	165	1166	6	1438	2 × 2	1.5	6*A*, 6*B*
Anthetic flower	45	165	444	6	1544	2 × 2	6.5	6*C*, 6*D*, 9*A*–9*C*
Capsule	45	165	444	6	1537	2 × 2	7	7*B*, 7*C*
Pholistoma auritum:								
Inflorescence tip	45	165	755	6	1401	2 × 2	4	5*E*–5*H*
Preanthetic flower	45	165	1166	6	1532	2 × 2	2.7	6*E*, 6*F*
Anthetic flower	45	165	444	6	909	2 × 2	7.5	6*G*, 6*H*, 9*D*–9*F*
Capsule	45	165	444	6	1557	2 × 2	6	7*E*, 7*F*
Pholistoma membranaceum:								
Inflorescence tip	45	165	1166	6	1485	2 × 2	3	5*I*–5*L*
Preanthetic flower	45	165	1166	6	1478	2 × 2	2.5	6*I*, 6*J*
Anthetic flower	45	165	1166	6	1240	2 × 2	3	6*K*, 6*L*, 9*G*–9*I*
Capsule	45	165	444	6	1472	2 × 2	4.5	7*H*, 7*I*

synflorescence types are usually formed. Flowers are pedicellate and are subtended by a bract and a pair of prophylls. The inflorescence tip is slightly rolled, and the flowers develop on the outer side of the spiral. Flower development is associated with an elongation and unrolling of the inflorescence axis. The flower buds develop acropetally in a zigzag line.

At the onset of floral ontogeny, soon after the hemispherical floral apex arises from the inflorescence axis, it develops into a pentagonal receptacular plate. Subsequently, the calyx, corolla, and androecium develop in centripetal order (figs. 2, 3). The sepals develop in a 2/5 spiral order (not shown) surrounding the pentagonal plate. The calyx reaches its final size before anthesis. Sepal estivation is always apert. The corolla and androecium are initiated sequentially and independently of each other (fig. 2). The petal primordia are initiated subsimultaneously as rounded edges of the pentagonal receptacular plate alternating with the sepals. They grow to almost the same size in the early stages of development (fig. 2*B*, 2*F*, 2*J*). The staminal primordia appear subsimultaneously after the petal primordia in antesepalous positions, and they either grow to the same size (fig. 2*C*) or have profound differences in size (fig. 2*G*, 2*K*). The petals and stamens develop at the same rate until they become basally fused and grow on a common basis (fig. 2*D*, 2*H*, 2*L*). Following basal fusion, the petal and stamen primordia flatten, and a first distinction between the anthers and filaments becomes apparent (fig. 3). The elongation of the corolla lobes occurs late during ontogeny, after the stamen–corolla tube is completely formed.

The sepals of *Emmenanthe penduliflora* elongate faster than the inner floral organs (fig. 2*A*–2*D*) and protect them up to anthesis. The calyx lobes are lanceolate to ovate, pubescent, and glandular. The corolla lobes show an imbricate-quincuncial estivation (fig. 3*C*). At anthesis, the anthers of *E. penduliflora* are included in the corolla tube and are shorter than the style branches. Flowers are nodding.

In *Pholistoma auritum*, the sepals do not enclose the remaining floral organs (fig. 2*E*–2*H*). The androecium and gynoecium are instead protected by dense trichomes until the corolla closes (fig. 3*D*–3*F*). The calyx lobes of *P. auritum* are lanceolate, with distinct sinus appendages and large trichomes. The corolla lobes show contorted estivation (fig. 3*F*). The stamens are exserted and are the same height as the style. Flowers are erect.

Pholistoma membranaceum (figs. 2*I*–2*L*, 3*G*–3*I*) shows only minor differences compared with *P. auritum*. These are the only differences: at anthesis, *P. membranaceum* has small oblong calyx lobes with large trichomes and without sinus appendages. The stamens are exserted and surpass the style.

Gynoecium and Nectary

The gynoecium is initiated when the petals and stamens begin to bend over the floral apex. At this point, a depression is formed at the center of the floral apex (fig. 2*D*, 2*H*, 2*L*). The walls of the two young horseshoe-shaped carpels form a ring at the beginning of gynoecium development (fig. 3*A*, 3*B*, 3*D*, 3*G*). The carpel primordia are fused from the onset of their development and grow apically to form a gynoecial cone (fig. 4*A*, 4*E*, 4*I*). The growth of the tips of the carpels slightly exceeds the growth of the fused walls so that the resulting gynoecial cone is bilobed at the tip.

After formation of the gynoecial cone, the two carpel lobes elongate to form a bifid style with two separate stigmata (fig. 4*B*, 4*F*, 4*J*). Simultaneously, the ovary is differentiating, and its basal part starts to flatten. The flattening of the ovary base demarcates the onset of nectary development. Specifically, at the very bottom part of the ovary, five small protrusions appear, usually with one nectarostoma in an antepetalous position. These five protrusions gradually develop into nectary lobes, and more nectarostomata appear on their surface (fig. 4*B*, 4*F*, 4*J*).

In *E. penduliflora*, the gynoecial cone is oblong and exceeds the developing anthers in height (figs. 4*A*, 5*B*). During the preanthetic developmental stages, the ovary is ovoid and elongates to gradually become oblong at anthesis (fig. 4*C*, 4*D*). Large glandular and filiform trichomes appear on the ovary surface during development (fig. 4*C*). The nectary disk enlarges and becomes a distinct hypogynous disk at anthesis (fig. 4*C*, 4*D*). It slightly exceeds the circumference of the ovary at its widest point (figs. 4*D*, 6*D*). The style splits into two short branches in the

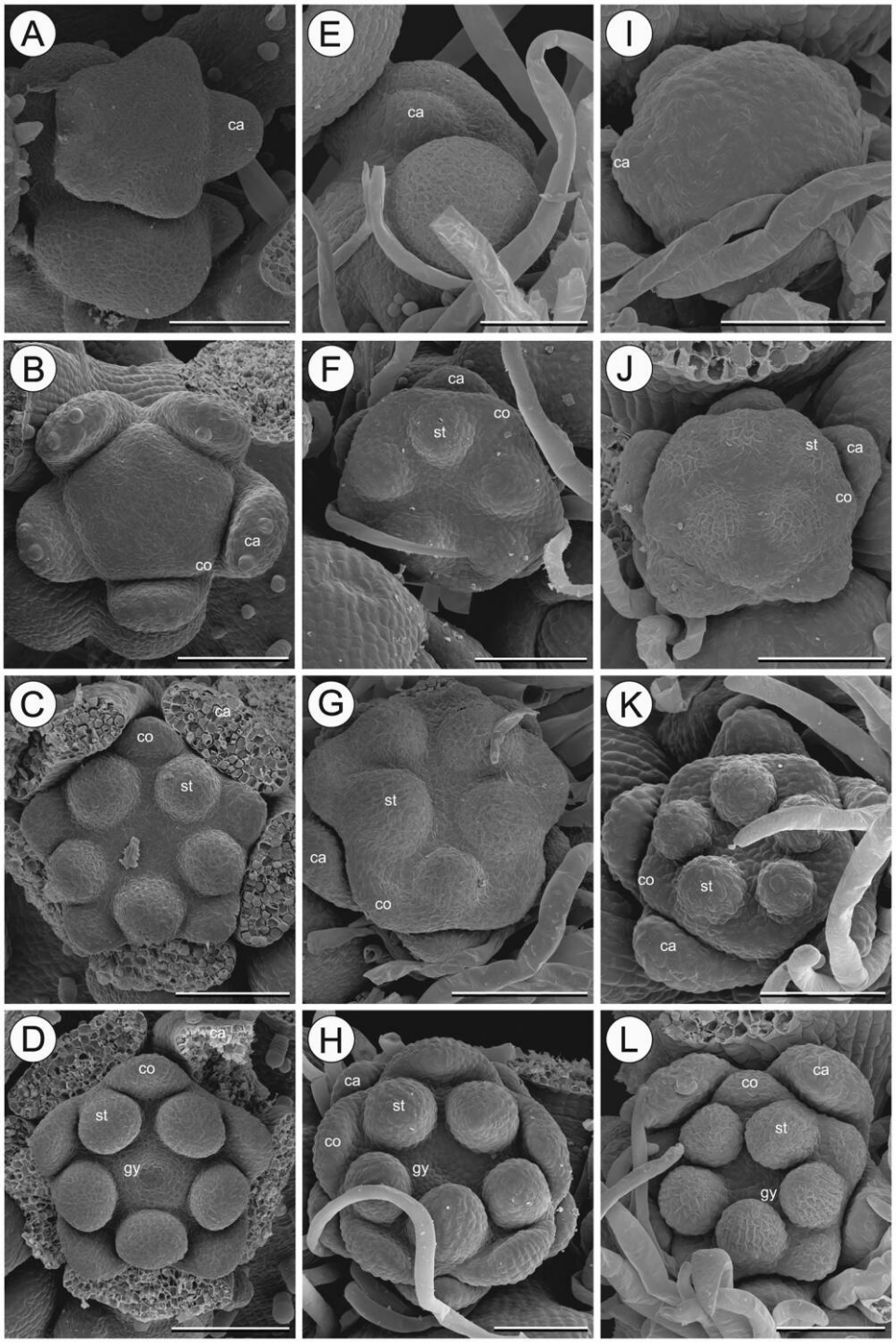


Fig. 2 Early flower development shown in successive developmental stages of the studied species, analyzed using scanning electron microscopy. *A–D*, *Emmenanthe penduliflora*. *A*, Calyx initiation. *B*, Corolla initiation. *C*, Androecium initiation. *D*, Gynoecium initiation. *E–H*, *Pholistoma auritum*. *E*, Calyx initiation. *F*, Sequential corolla and androecium initiation. *G*, Petal and staminal primordia develop subsimultaneously. *H*, Gynoecium initiation. *I–L*, *Pholistoma membranaceum*. *I*, Calyx initiation. *J*, Sequential corolla and androecium initiation. *K*, Petal and staminal primordia develop subsimultaneously. *L*, Gynoecium initiation. ca = calyx; co = corolla; gy = gynoecium; st = stamen. Scale bars = 100 μm (*A–H*), 80 μm (*I–L*).

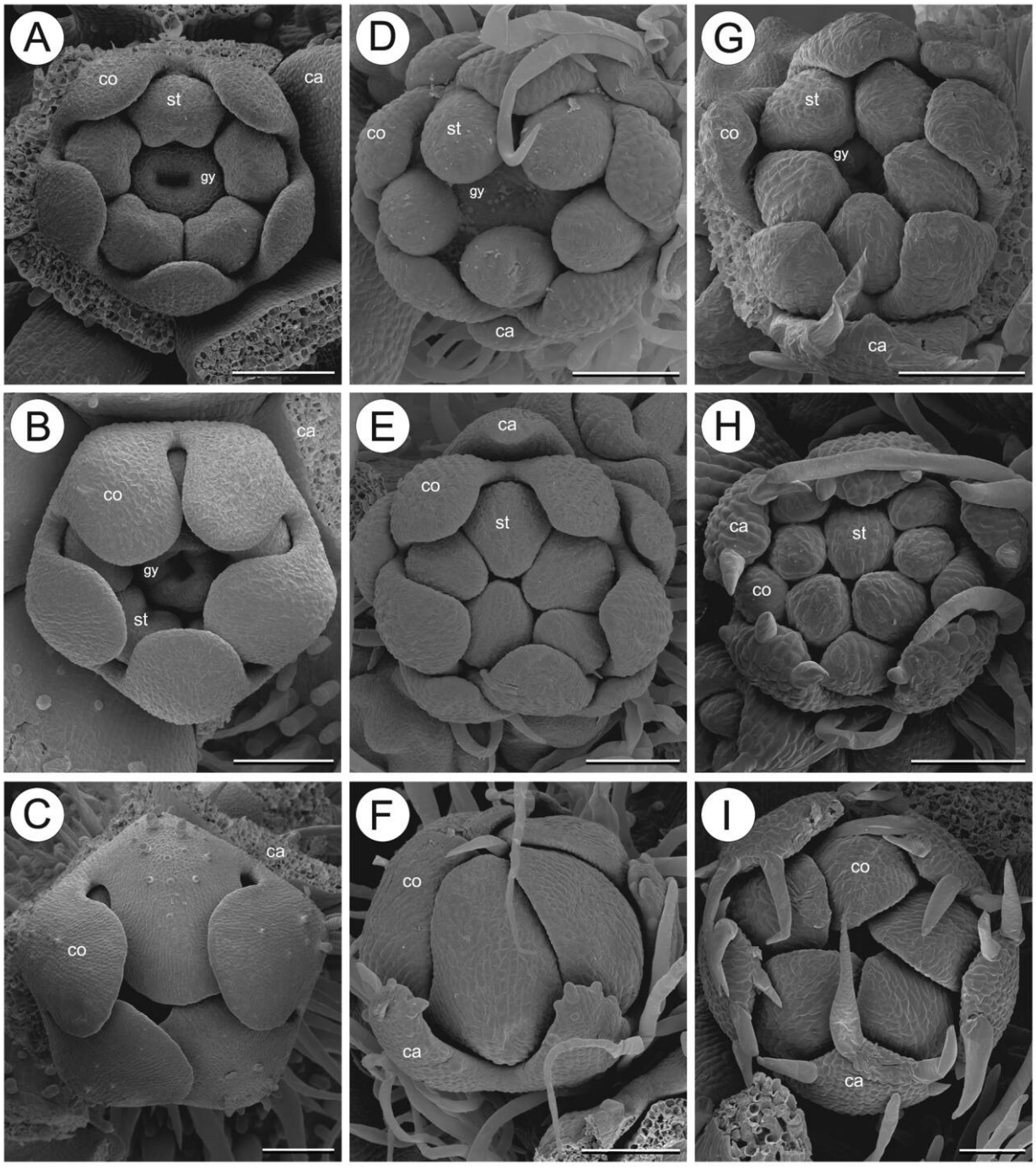


Fig. 3 Flower development shown in successive developmental stages of the studied species, analyzed using scanning electron microscopy. *A–C*, *Emmenanthe penduliflora*. *A*, Anthers and gynoecial cone initiation. *B*, Corolla elongation. *C*, Corolla closes, and corolla lobe estivation is visible. *D–F*, *Pholistoma auritum*. *D*, Gynoecial cone initiation. *E*, Gynoecium is covered by the stamens. *F*, Corolla closes, and corolla lobe estivation is visible. *G–I*, *Pholistoma membranaceum*. *G*, Gynoecial cone initiation. *H*, Gynoecium is covered by the stamens. *I*, Corolla closes, and corolla lobe estivation is visible. ca = calyx; co = corolla; gy = gynoecium; st = stamen. Scale bars = 100 μm.

early developmental stages (fig. 4*B*). As they elongate, they bend away from each other (fig. 4*C*). The distinct subglobose stigmata with a papillate surface develop relatively late (fig. 4*C*, 4*D*). At anthesis, the style branches are one-third of the style length, and scattered filiform trichomes cover the style (fig. 4*D*).

Pholistoma auritum has a gynoecial cone half the size of the developing anthers (figs. 4*E*, 5*F*). The conical ovary increases in diameter and retains its shape throughout development. Large filiform trichomes first develop at the upper part of the ovary (fig. 4*F*, 4*G*), and later, shorter trichomes densely cover the

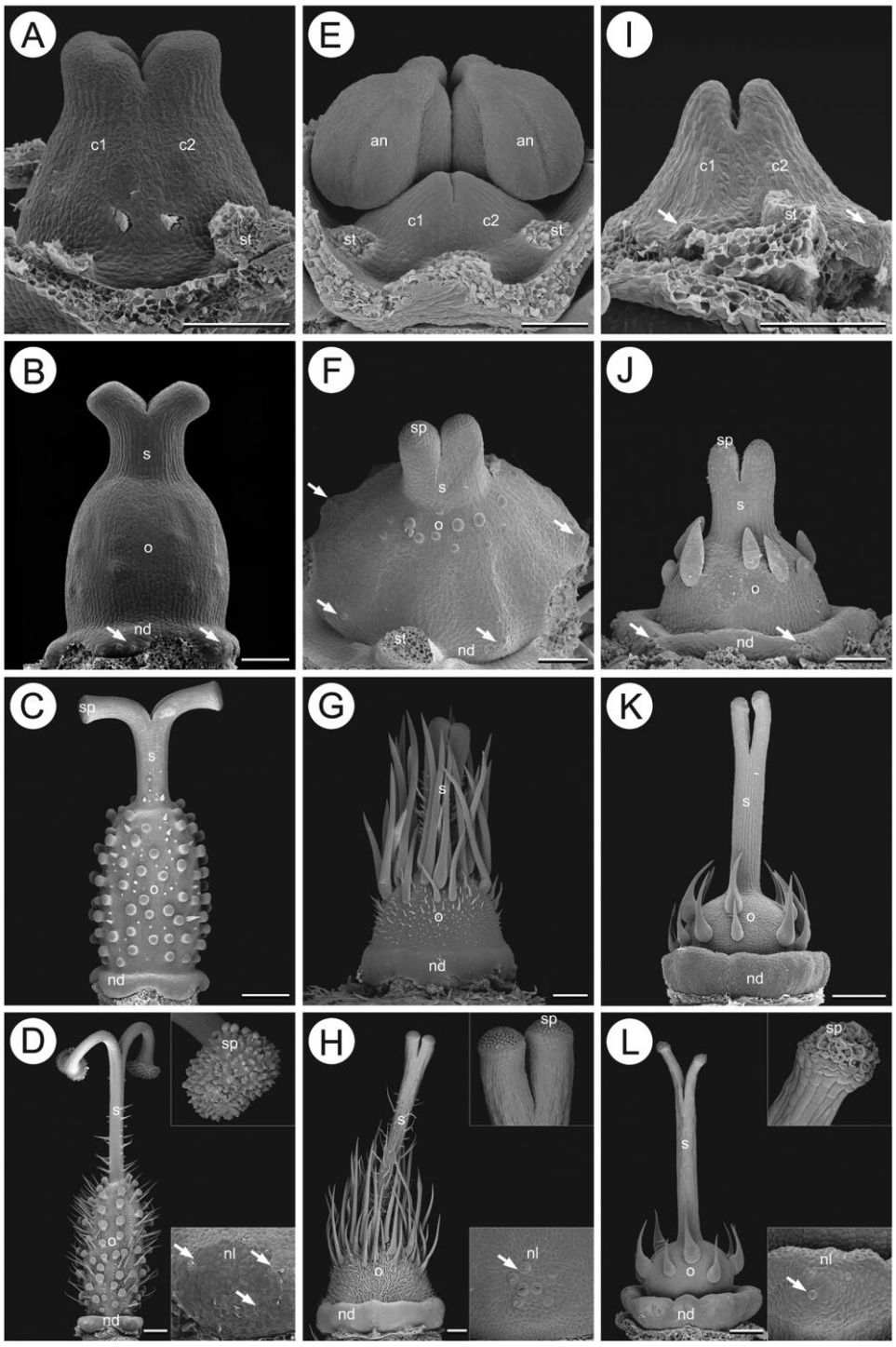


Fig. 4 Gynoecium development shown in successive developmental stages of the studied species analyzed using scanning electron microscopy. *A–D*, *Emmenanthe penduliflora*. *A*, Gynoecial cone. *B*, Ovary formation and nectary disk initiation. *C*, Ovary enlargement, style elongation, stigmatic papillae, and nectary disk development. *D*, Anthetic gynoecium; close-ups of the stigmatic head (*top right*) and of a nectary lobe (*bottom right*). *E–H*, *Pholistoma auritum*. *E*, Gynoecial cone. *F*, Ovary formation and nectary lobe initiation. *G*, Ovary enlargement, style elongation, stigmatic papillae, and nectary ring development. *H*, Anthetic gynoecium; close-ups of the stigmatic head (*top right*) and of a nectary lobe (*bottom right*). *I–L*, *Pholistoma membranaceum*. *I*, Gynoecial cone. *J*, Ovary formation and nectary disk initiation. *K*, Ovary enlargement, style elongation, stigmatic papillae, and nectary disk development. *L*, Anthetic gynoecium; close-ups of the stigmatic head (*top right*) and of a nectary lobe (*bottom right*). The arrows point to nectarostomata. an = anther; c1 = carpel 1; c2 = carpel 2; nd = nectary disk; nl = nectary lobe; o = ovary; s = style; sp = stigmatic papillae; st = stamen. Scale bars = 100 μm (*A*, *B*, *E*, *F*, *I*, *J*), 200 μm (*C*, *D*, *G*, *H*, *K*, *L*).

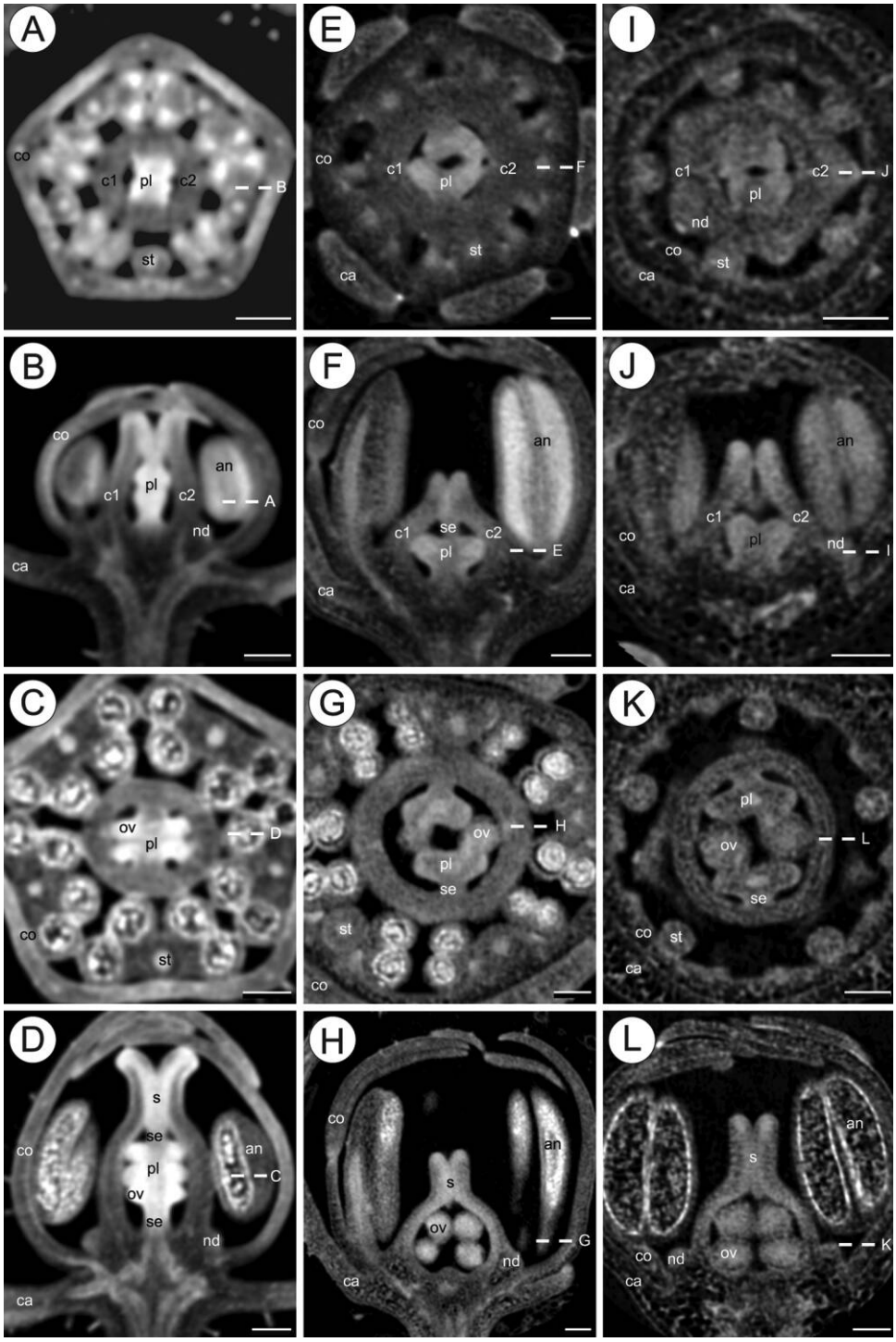


Fig. 5 Inner gynoecium development of the studied species, analyzed using micro–computed tomography, very early stages. *A–D*, *Emmenanthe penduliflora*. *A*, *B*, Gynoecial cone formation; onset of placenta and ovule development. *A*, Virtual cross section. *B*, Virtual longitudinal section. *C*, *D*, Ovule development. *C*, Virtual cross section. *D*, Virtual longitudinal section. *E–H*, *Pholistoma auritum*. *E*, *F*, Gynoecial cone formation; onset of placenta and ovule development. *E*, Virtual cross section. *F*, Virtual longitudinal section. *G*, *H*, Ovule development. *G*, Virtual cross section. *H*, Virtual longitudinal section. *I–L*, *Pholistoma membranaceum*. *I*, *J*, Gynoecial cone formation; onset of placenta and ovule development. *I*, Virtual cross section. *J*, Virtual longitudinal section. *K*, *L*, Ovule development. *K*, Virtual cross section. *L*, Virtual longitudinal section. Dashed lines with letters indicate the corresponding virtual sections in other panels of this figure. an = anther; c1 = carpel 1; c2 = carpel 2; ca = calyx; co = corolla; nd = nectary disk; ov = ovule; pl = placenta; s = style; se = septum; st = stamen. Scale bars = 100 μm.

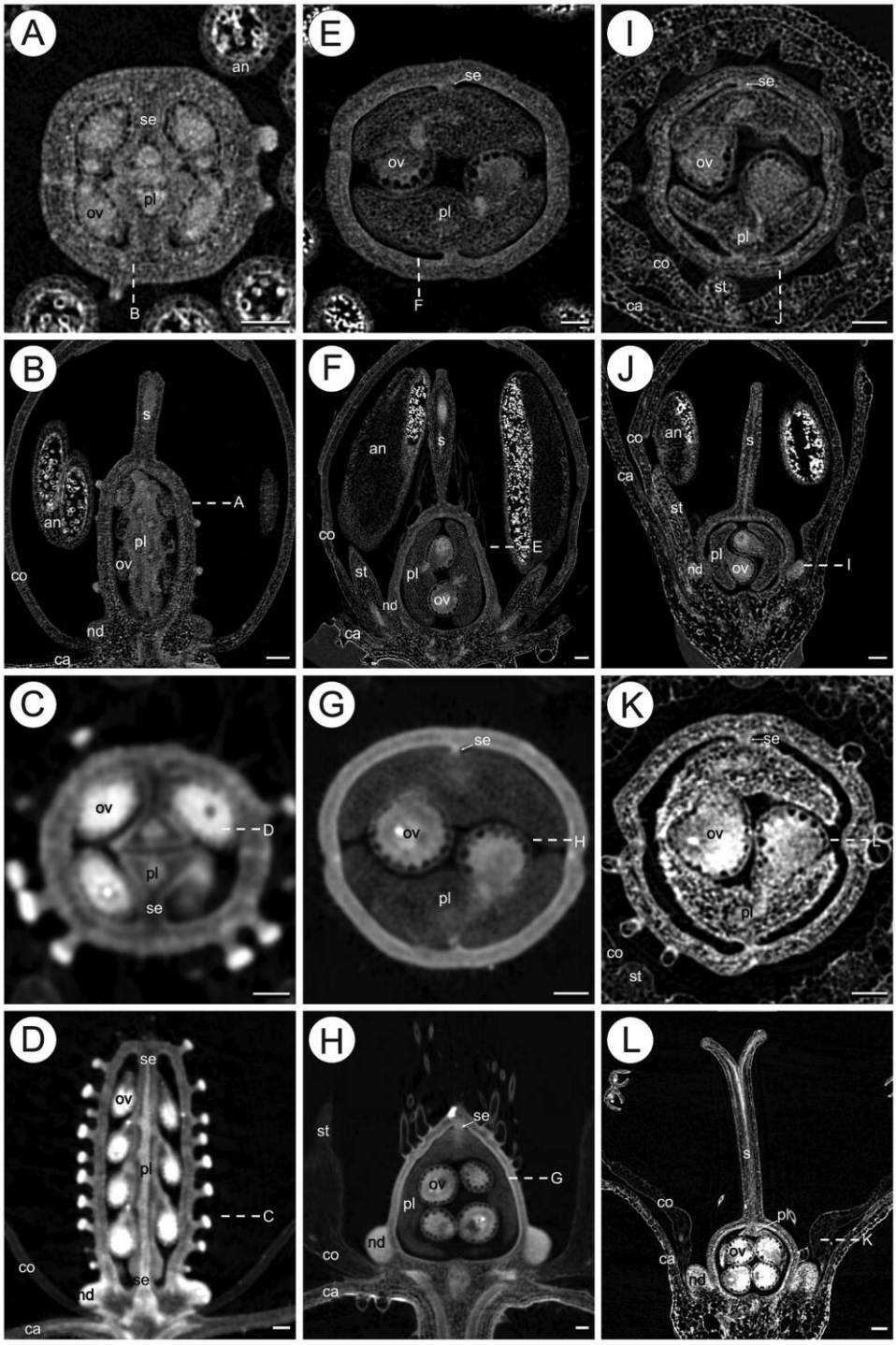


Fig. 6 Inner gynoecium development of the studied species, analyzed using micro–computed tomography, preanthetic and anthetic developmental stages. *A–D*, *Emmenanthe penduliflora*. *A*, *B*, Preanthetic gynoecium. *A*, Virtual cross section. *B*, Median virtual longitudinal section through the carpels. *C*, *D*, Anthetic gynoecium. *C*, Virtual cross section. *D*, Virtual longitudinal section. *E–H*, *Pholistoma auritum*. *E*, *F*, Preanthetic gynoecium. *E*, Virtual cross section. *F*, Median virtual longitudinal section through the carpels. *G*, *H*, Anthetic gynoecium. *G*, Virtual cross section. *H*, Virtual longitudinal section. *I–L*, *Pholistoma membranaceum*. *I*, *J*, Preanthetic gynoecium. *I*, Virtual cross section. *J*, Median virtual longitudinal section through the carpels. *K*, *L*, Anthetic gynoecium. *K*, Virtual cross section. *L*, Virtual longitudinal section through the carpels. Dashed lines with letters indicate the corresponding virtual sections in other panels of this figure. an = anther; ca = calyx; co = corolla; nd = nectary disk; ov = ovule; pl = placenta; s = style; se = septum; st = stamen. Scale bars = 100 μm.

remaining surface and the base of the style. As the ovary enlarges, a ring-shaped nectary disk is formed and becomes distinct at anthesis (figs. 4*G*, 4*H*, 6*H*). The five lobes slightly protrude at anthesis and can be distinguished only by the presence of the nectarostomata, which are limited to the tip of each lobe (fig. 4*H*). The subglobose papillate receptive surface of the stigmata appears early during development (fig. 4*F*). The style of *P. auritum* significantly elongates shortly before anthesis. At this stage, it is covered with scattered trichomes. The style cleft is in the upper third-quarter of the style, and the two stigmata slightly separate from each other at anthesis (fig. 4*H*).

Pholistoma membranaceum has a gynoecial cone almost the same size as the developing anthers (figs. 4*I*, 5*J*). In some specimens of *P. membranaceum*, one carpel is slightly larger than the other one, indicating potential sequential insertion, but with a very short plastochron. This difference in size remains visible until anthesis, as very often one style branch is shorter than the other one (fig. 4*K*, 4*L*). The initially conical ovary (fig. 4*I*) increases its diameter while it continues to grow to reach its final subglobose shape (fig. 4*J*–4*L*). At anthesis, the ovary wall significantly bulges at the line of carpel unification. Scattered, large, scabrid trichomes with an extensive base gradually develop on the ovary surface (fig. 4*J*–4*L*). The nectary disk is initially flattened and has a notably wider diameter than the rest of the young ovary (fig. 4*J*). It grows until it forms a cup-shaped disk with an undulate pentagonal circumference (fig. 4*K*, 4*L*). The stigma formation on the carpel apexes starts early with the development of papillate tissue (fig. 4*J*) and later forms a subglobose stigmatic head. The bifid style mostly elongates at later developmental stages, and the style cleft is formed in the upper third of the style's length (fig. 4*F*, 4*K*, 4*L*).

Internal Ovary Architecture

At the onset of gynoecial cone development, two placentae develop through contraction of the septa to the periphery (fig. 5*A*, 5*B*, 5*E*, 5*F*, 5*I*, 5*J*). Placentation is parietal. Subsequently, internal ovary architecture development is largely affected by ovule number and placentation. A compitum is always present in the style.

In *E. penduliflora* (figs. 5*A*–5*D*, 6*A*–6*D*), after the gynoecial cone closes, the placentae and a basal as well as an apical septum are differentiated, and the first ovule primordia become apparent (fig. 5*A*, 5*B*). This species is pluriovulate (i.e., it has more than four ovules), and the intrusive parietal placentae with true septa form two compartments in the ovary as they develop. At anthesis, there are usually three and sometimes four layers of ovules, all of which develop almost synchronously (figs. 5*B*, 5*D*, 6*B*, 6*D*). Each placenta ends up having two rows of ovules and in total six to eight ovules arranged in three or four layers, respectively (fig. 6*C*, 6*D*). The ovules of one layer are not at the exact same level, resulting in a zigzag arrangement of the two rows of ovules of each placenta (fig. 6*D*). The mature placentae have a triangular shape in cross section (fig. 6*D*). Ovule orientation is epitropous-ventral.

In *P. auritum* (figs. 5*E*–5*H*, 6*E*–6*H*), the placentae as well as short basal and apical septa are already differentiated once the gynoecial cone closes (fig. 5*E*, 5*F*). The lateral parts of the young placentae flatten right after initiation and, together with the septa, form an initially T-shaped structure in cross section. As the ovary enlarges, four round ovules, two per placenta, start to develop (fig. 5*E*–5*H*). The ovules of each placenta alternate as a result of mutual displacement and a lack of space in the central compartment; however, all four ovules belong to a single layer. This ovule alternation results in a different ovule orientation (fig. 6*F*). The lowermost ovule of each placenta becomes epitropous-ventral, whereas the uppermost becomes hypotropous-ventral. The placentae gradually enlarge, create one central compartment, start to line the ovary wall, expand around and between the ovules, and almost completely fill the ovary at anthesis (fig. 6*E*–6*H*).

Pholistoma membranaceum (figs. 5*I*–5*L*, 6*I*–6*L*) has a developmental trajectory identical to that of *P. auritum*. The main differences are that the placentae are smaller, creating, as a result, a slightly larger central compartment, and that the ovules are not perfectly rounded.

Fruit

After anthesis, the ovary, placentae, and fertilized ovules gradually enlarge until fruit maturity. The fruit is a loculicidal capsule with two locules. The style persists on one valve of the opened capsule.

Emmenanthe penduliflora (figs. 1*B*, 7*A*–7*C*) retains all of its floral organs, including the corolla and stamens, until fruit maturity. The pedicel doubles in length, and the withering, papery, marcescent corolla encloses the capsule. The nectary disk retains its shape intact at fruit maturity (fig. 7*A*). The two placentae enlarge and occupy most of the internal volume of the oblong capsule. The seeds are flat and slightly concave. They are spirally arranged around the placentae (fig. 7*B*, 7*C*). Fruits are nodding at maturity (fig. 1*B*).

In *P. auritum* (figs. 1*D*, 7*D*, 7*E*), the entire stamen–corolla tube wilts and falls off after anthesis. The nectary disk becomes part of the pericarp (fig. 7*D*). The placentae completely fill the fruit (fig. 7*E*). Fruits are globose to ovoid and are enclosed by the persistent accrescent calyx (figs. 1*D*, 7*D*, 7*E*). The capsules have a light green color with red spots (fig. 1*D*) until they are completely dry and open. Seeds are brown and round in shape, and the seed coat surface is moderately reticulate with polygonal cavities. There are usually four or fewer mature seeds.

The fruit development and morphology of *P. membranaceum* (figs. 1*F*, 7*F*–7*H*) are similar to that of *P. auritum*. The large nectary disk creates a platform at the fruit base. The accrescent calyx persists but does not enclose the mature fruit (fig. 1*F*). Fruits are initially light green in color but turn purple as they mature (fig. 1*F*). The dry opened capsules retain a deep purple color. Seeds are ovate in shape, and the seed coat surface is moderately reticulate with polygonal cavities. There are four or, usually, fewer mature seeds.

Stamen–Corolla Tube Modifications and Floral Synorganization

Ten modifications develop on the adaxial side of the stamen–corolla tube. Their development starts during late flower ontogeny, after the differentiation of the stamens into well-defined anthers and filaments and of the gynoecium into young stigmata, style, and ovary with a nectary disk. The modifications, the corolla tube itself, the filaments, and the style subsequently

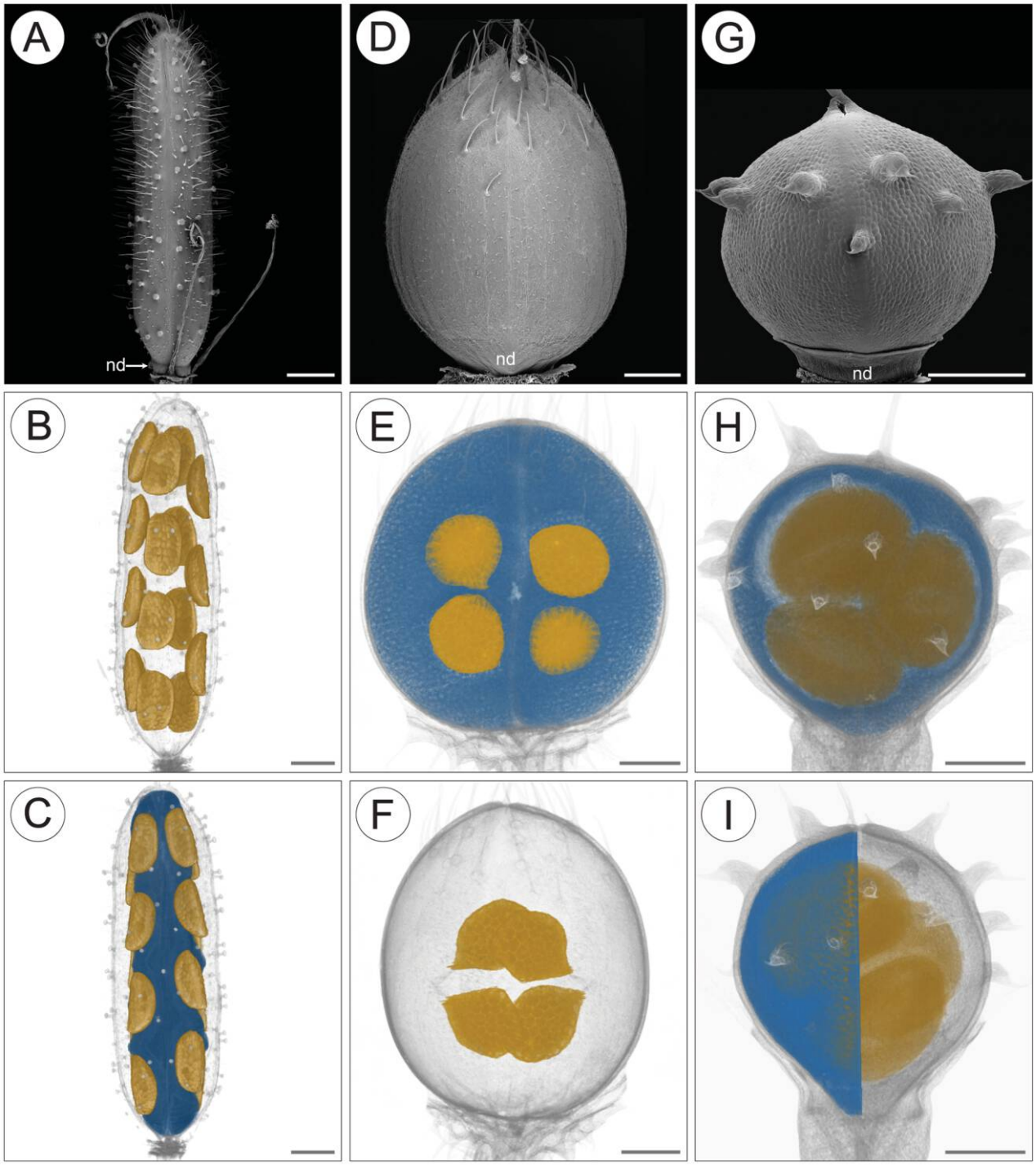


Fig. 7 Mature capsules of the studied species. *A–C*, *Emmenanthe penduliflora*. *A*, Scanning electron microscopy (SEM) image; outer view of a closed capsule, dehiscence line visible. *B*, Micro–computed tomography (μCT)–based volume rendering showing seed arrangement; lateral view (looking at the line of carpel fusion). *C*, μCT-based volume rendering showing seeds and placentae; view of one carpel (looking at the dehiscence line). *D–F*, *Pholistoma auritum*. *D*, SEM image; outer view of a closed capsule, dehiscence line visible. *E*, μCT-based volume rendering showing seeds and placentae; only half of the placenta is displayed because it completely lines the capsule and the ovules would otherwise be obscured; lateral view. *F*, μCT-based volume rendering showing seed arrangement; vascular bundles connecting seeds to placentae visible; view of one carpel. *G–I*, *Pholistoma membranaceum*. *G*, SEM image; outer view of a closed capsule; dehiscence line visible. *H*, μCT-based volume rendering showing seeds and placentae; only half of the placenta is displayed because it completely lines the capsule and the ovules would otherwise be obscured; lateral view. *I*, μCT-based volume rendering showing seeds and placentae; only half of the placenta is displayed because it completely lines the capsule and the ovules would otherwise be obscured; view of one carpel. In *B*, *C*, *E*, *F*, *H*, and *I*, the pericarp is reconstructed and displayed with low opacity and solidity. Seeds are displayed in yellow and placentae in blue. nd = nectary disk. Scale bars = 1 mm.

elongate and are elaborated in parallel. This pattern is found across all three studied species.

In *E. penduliflora* (figs. 1*A*, 8*A*–8*C*, 9*A*–9*C*), the modifications are initiated as elongated protrusions neighboring the base of the filaments (fig. 8*A*). They retain their overall shape and only slightly increase in size during development and the elongation of the basal part of the corolla tube (fig. 8*B*). At the stage when the filaments have reached their final size, the modifications

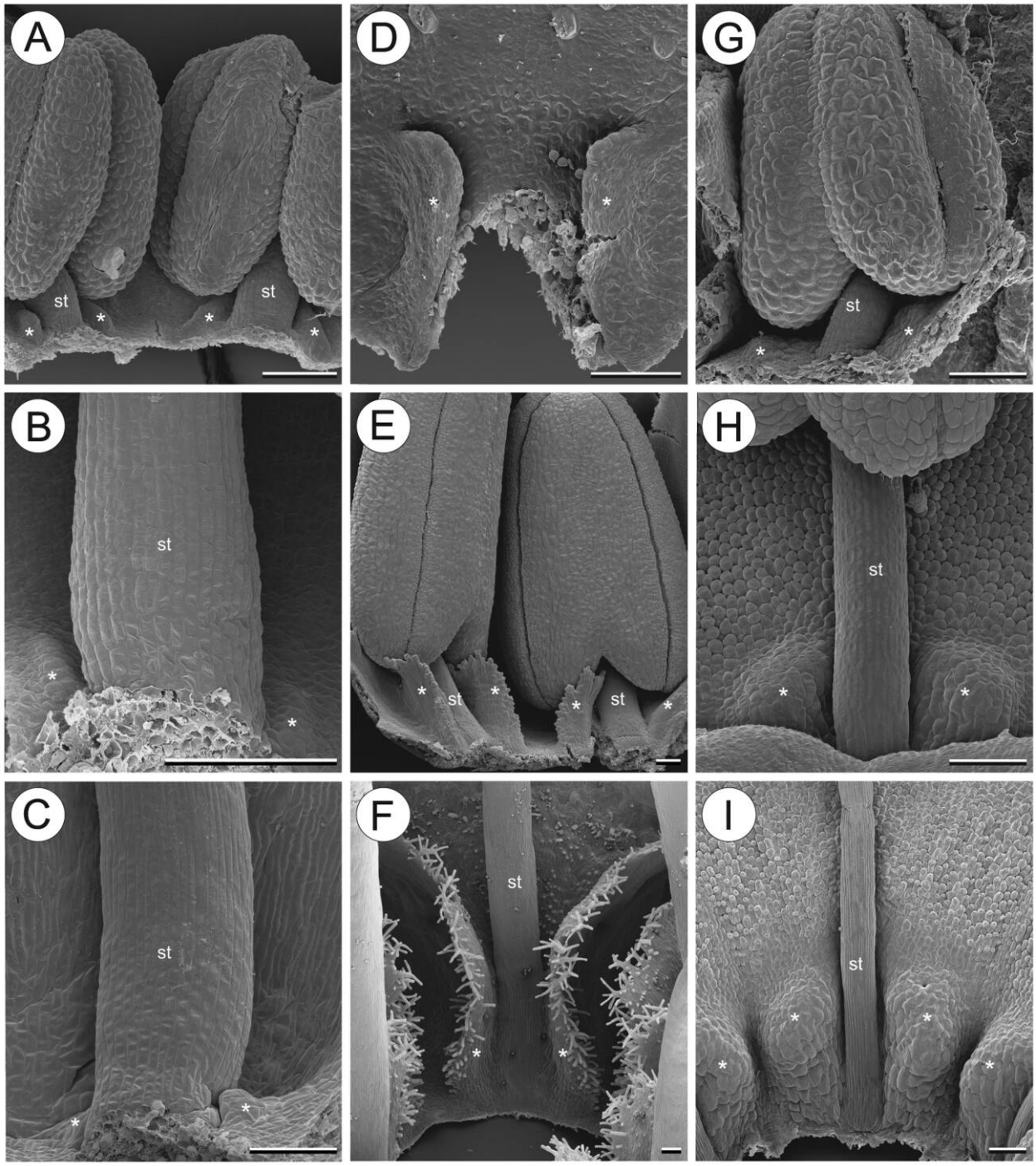


Fig. 8 Development of stamen–corolla tube modifications of the studied species, analyzed using scanning electron microscopy. *A–C*, *Emmenanthe penduliflora*. *A*, Very young flower. *B*, Preanthetic flower. *C*, Anthetic flower. *D–F*, *Pholistoma auritum*. *D*, Very young flower; stamen removed. *E*, Preanthetic flower. *F*, Anthetic flower. *G–I*, *Pholistoma membranaceum*. *G*, Very young flower. *H*, Preanthetic flower. *I*, Anthetic flower. Asterisks indicate the modifications. st = stamen. Scale bars = 100 μm.

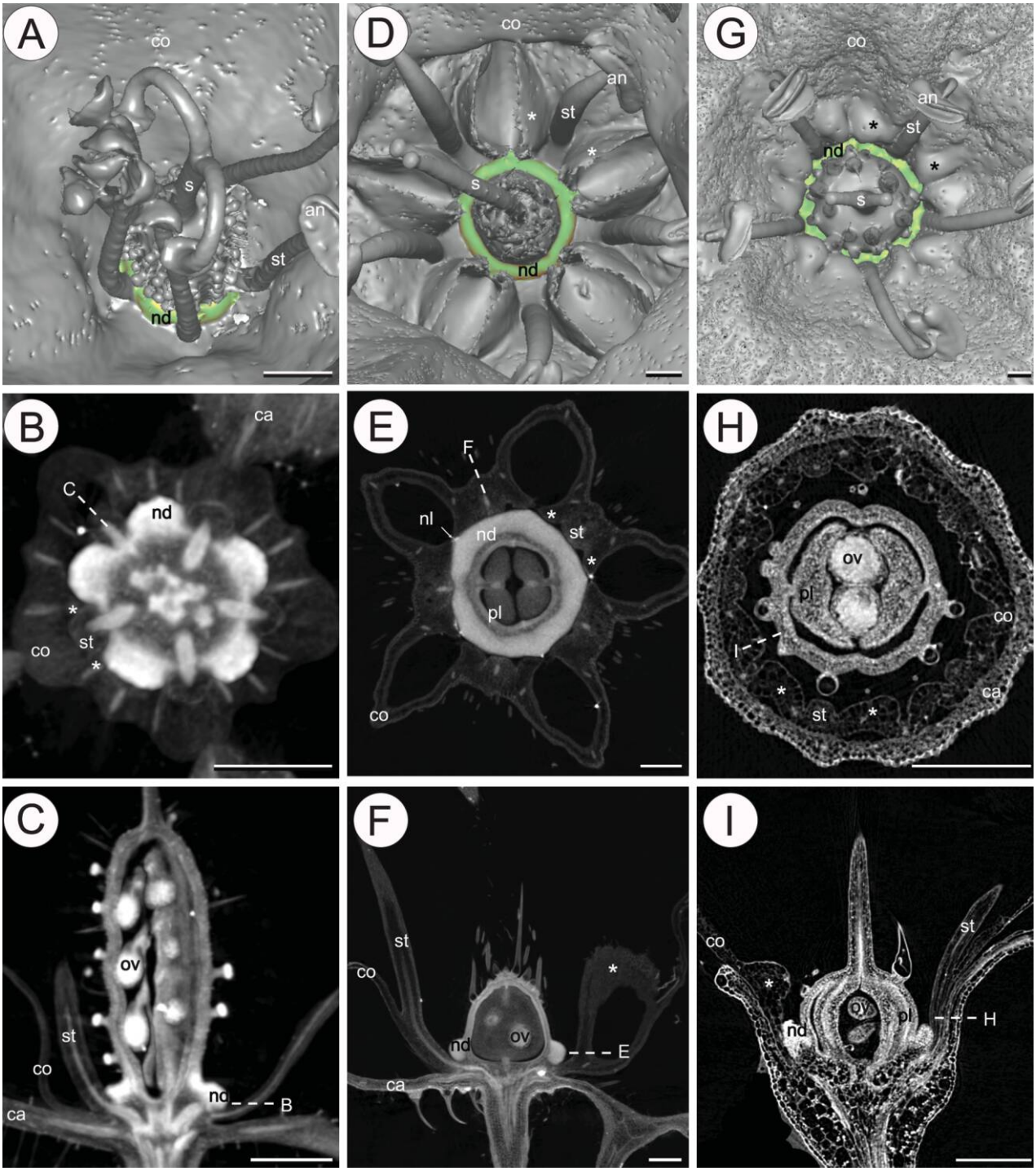


Fig. 9 Stamen–corolla tube modifications of the studied species, micro–computed tomography–based surface renderings and virtual sections. *A–C*, *Emmenanthe penduliflora*. *A*, Surface rendering of the flower viewed from the top and slightly tilted, nectary disk in green. *B*, Virtual cross section at the level of the nectary disk. *C*, Virtual longitudinal section. *D–F*, *Pholistoma auritum*. *D*, Surface rendering of the flower viewed from the top, nectary disk in green. *E*, Virtual cross section at the level of the nectary disk. *F*, Virtual longitudinal section. *G–I*, *Pholistoma membranaceum*. *G*, Surface rendering of the flower viewed from the top, nectary disk in green. *H*, Virtual cross section. *I*, Virtual longitudinal section. Dashed lines with letters indicate the corresponding reconstructed sections in other panels of this figure. Asterisks indicate the modifications. an = anther; ca = calyx; co = corolla; nd = nectary disk; nl = nectary lobe; ov = ovule; pl = placenta; s = style; st = stamen. Scale bars = 500 μm.

protrude toward and touch the base of the filaments (fig. 8*C*). At anthesis, the stamen–corolla tube modifications are small and rudimentary and do not form any compartments (fig. 9*A*–9*C*). They are found at the level where the nectary disk initiates (fig. 9*B*). The nectary lobes are found in an antepetalous position and protrude between the rudimentary modifications. The filaments are fused to the base of the corolla tube and between the lobes of the nectaries (fig. 9*C*). The anthetic corolla is white to yellow and tubular to campanulate (fig. 1*A*). The corolla tube is three times the length of the free corolla lobes. Trichomes are found only across the midvein on the abaxial side of the petals.

Development of the modifications of *P. auritum* starts as elongated ridges. In the beginning, they are associated with the base of the neighboring filament (fig. 9*D*). As the modifications elaborate, simple trichomes emerge along their margins (fig. 9*E*) and later also on the corolla tube surface between the modifications, behind the filaments (fig. 7*F*). The apical part of the modifications, where the filaments diverge from the corolla tube, gradually widens more than the basal part. At anthesis, each pair of modifications is present in the form of wing-shaped protuberances (figs. 8*F*, 9*D*, 9*E*). This shape results in the enclosure of a broad and deep groove between neighboring modifications. The narrowest part of the groove is where it touches the nectary disk (figs. 8*F*, 9*D*–9*F*). This groove is in an antepetalous position, the same position as the nectary lobes. The anthetic corolla is purple, and a distinct dark purple ring is formed at the pale entrance to the corolla tube (fig. 1*C*). The stamen–corolla tube modifications are deep purple with a yellowish line on top. The white trichomes of the neighboring modifications are intertwined and, as a consequence, completely cover the grooves between the modifications. The corolla is cyathiform, and the lobes are twice as long as the corolla tube.

In *P. membranaceum*, the modifications are initiated as two shallow lobes above the insertion point of the filaments (fig. 8*G*). As they develop, they elongate and swell at the same stage that the filaments become free (figs. 8*H*, 8*I*, 9*I*). During development, the lower part of the corolla tube invaginates in the position of the modifications. Additional protrusions on the invaginations are absent (fig. 9*H*). At anthesis, neighboring modifications touch or almost touch each other (fig. 8*G*, 8*H*), thus forming a narrow canal above the nectary lobes. The irregular shape of the large nectary disk is due to the modifications touching the disc. The shape of the anthetic corolla is similar to that of *P. auritum* but smaller overall (fig. 1*E*). The white corolla has purple midveins on its lobes that lead toward the narrow canals created by the modifications (fig. 1*E*).

Discussion

In the present study, we describe the general flower and fruit development of *Emmenanthe* and *Pholistoma*, two genera of Hydrophylleae within Hydrophyllaceae. This is the first ontogenetic survey of overall flower development in this family. There have been few floral ontogenetic studies of Hydrophyllaceae, and they have largely concentrated on gynoecial development (Payer 1857; Hofmann 1999; Vasile et al. 2021) or on stamen–corolla tube modifications (Hofmann 1999; Jeiter and Weigend 2018). Other studies have focused on the floral morphology and anatomy of various species of Hydrophyllaceae (e.g., Berg 1985, 2009; Chuang and Constance 1992; Di Fulvio et al. 1999). The floral ontogeny of *Emmenanthe* and *Pholistoma* seems to be similar to that of other species in the family (Hofmann 1999; Vasile et al. 2021). There are minor differences among the species studied here, in particular with regard to sepal and gynoecium development and morphology, and there are also major differences in the gynoecial nectary disk and stamen–corolla tube modifications.

Gynoecial nectary disks of Hydrophylleae often form separate glands (e.g., *Hydrophyllum*, *Nemophila*; Jeiter and Weigend 2018; Vasile et al. 2021). The species studied here have fully developed nectary disks distinct from the base of the ovary. However, nectarostomata are restricted to lobes of the disk, as is typical of Hydrophyllaceae (Jeiter and Weigend 2018). Stamen–corolla tube modifications in Hydrophyllaceae share a highly conserved developmental trajectory (Jeiter and Weigend 2018), and the species studied here are no exception to this. Modifications develop late during floral ontogeny, and they originate from both the corolla tube and the filament bases; their number always appears to be 10. In the present study, we were able to document the presence of modifications in *Emmenanthe penduliflora*, which had previously been described as lacking modifications (Hofmann et al. 2016). The pattern of initiation and the subsequent truncation of their development have been observed in other species supposedly lacking modifications (e.g., *Draperia systyla*; Jeiter and Weigend 2018). The presence of 10 modifications of the stamen–corolla tube appears to be the plesiomorphic condition for the family, and crucial aspects of the divergent floral morphologies of the family appear to go back at least partly to the truncation of scale development—that is, heterochronous shifts.

The predominant type of stamen–corolla tube modifications of Hydrophyllaceae are protuberances (Hofmann 1999; Jeiter and Weigend 2018). However, surprisingly, in *Pholistoma membranaceum* invaginations are found where modifications are initiated, as in all other species, but further development is dominated by the invagination of the basal part of the corolla tube.

The flower architecture of Hydrophyllaceae can thus be shown to follow broadly uniform patterns, with minor developmental shifts explaining different mature morphologies. Conversely, our study demonstrates surprisingly divergent internal ovary architectures. As shown previously, seed numbers are highly variable in Hydrophyllaceae and largely depend on changes in internal gynoecium development (Vasile et al. 2021). Here, we focus on another highly variable aspect: placentation. Placentation in Hydrophyllaceae is originally parietal or intrusively parietal. However, various changes in developmental trajectories result in profoundly different internal ovary architectures as well as changes in placentation. These changes also co-occur with changes in ovule and seed number, shape, size, and orientation. Figure 10 summarizes our current understanding of the internal ovary architecture of Hydrophyllaceae (Di Fulvio et al. 1999; Hofmann 1999; Vasile et al. 2021). On the basis of the ancestral character state reconstruction of Vasile et al. (2021), we propose hypothetical character states for each node of the phylogeny (fig. 10). An ovary with four rows of ovules (two per placenta) in several tiers can be found in all three clades of the family (fig. 10; type I or II). We propose that this represents the ancestral character state of the family. We also suggest that the multiplication of ovule rows per placenta has independently taken place at least

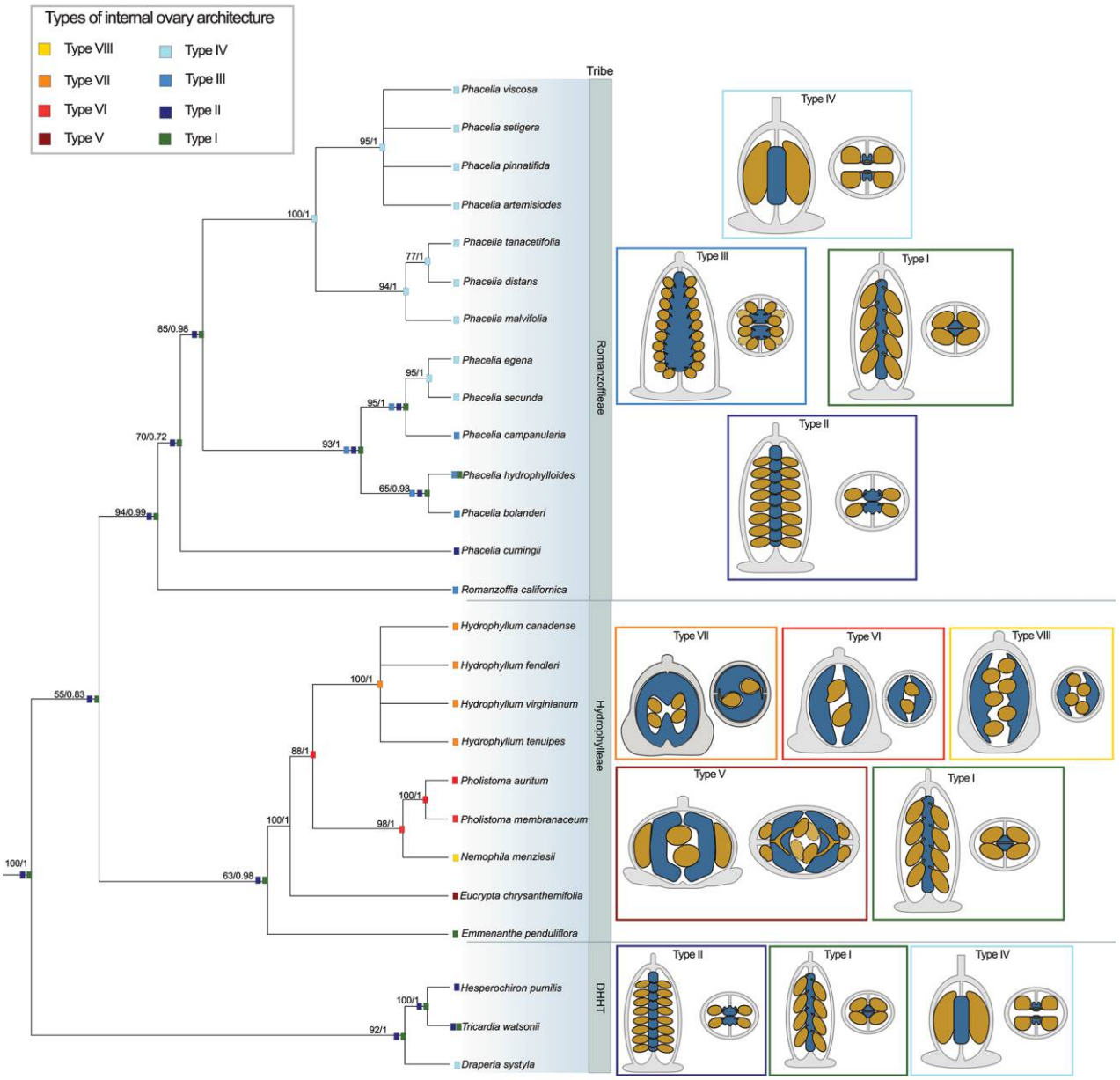


Fig. 10 Cladogram of Hydrophyllaceae with character mapping and hypothetical evolutionary series of different internal ovary architectures using a recently published phylogeny (Vasile et al. 2020) and ancestral character state reconstruction (Vasile et al. 2021). Corresponding architectural types are illustrated next to each clade and have frames of different colors that correspond to the colors of the squares. Nodes with uncertain character states have more than one potential type, or potential types are absent. DHHT = *Draperia* + *Howellanthus* + *Hesperochiron* + *Tricardia* clade.

twice within Romanzoffieae (fig. 10; type III). Accordingly, the simple truncation of ovary development following the initiation of the first layer of ovules (Vasile et al. 2021) must have resulted in descendants with only the four ovules of the first layer (paedomorphosis; fig. 10; type IV), probably explaining the multiple transitions in Hydrophylleae toward a reduction in ovule number to four (Vasile et al. 2021). In Hydrophylleae, this is also associated with the development of unilocular ovaries and the inclusion of the ovules between the enlarged placentae. Specifically, in *Hydrophyllum*, the placentae completely fill the ovary (fig. 10; type VII), leaving no free space. Conversely, the placentae of *Pholistoma* and *Nemophila* leave more free space in the central locule (fig. 10; types VI, VIII). This probably permits higher plasticity and opens the pathway to secondary multiplications of the ovule layers during gynoecium development. In *Pholistoma*, as well as *Nemophila*, more than four ovules are common (Berg

1985; Chuang and Constance 1992; Vasile et al. 2021; fig. 10; type VIII). The key factor for the evolution of unilocular ovaries seems to be their heterochronous development: lateral parts of the placentae develop at very early stages, whereas ovule initiation is slightly delayed, as our data on *Pholistoma* demonstrate (fig. 5*F*, 5*G*, 5*J*, 5*K*). As a result, ovule positions shift from the position observed in *Emmenanthe* to the center of the ovary. The suppression of the septa and the significant enlargement of the placentae lining the ovary walls further facilitate the creation of a single central compartment (fig. 10; types VI, VII, VIII). The formation of flattened placentae on which ovules develop on both surfaces in the closely related *Eucrypta chrysanthemifolia* (Benth.) Greene (Hofmann et al. 2016; Vasile et al. 2021) thus must have evolved only once within the family (fig. 10; type V), and whether this species is a descendant of a four-ovulate or pluriovulate ancestor remains enigmatic.

In addition to the reduction of ovule numbers and a fixed number of four ovules per ovary, placentation is another crucial factor in the evolution of mericarpic and endomericarpic fruits in Boraginales. All capsular fruited families across Boraginales have parietal or intrusive parietal placentation (Weigend and Hilger 2010; Hilger and Weigend 2016; Hofmann et al. 2016; Jeiter et al. 2016), whereas the remaining noncapsular fruited families show basal to axile placentation. Here, the ontogenetic data on the internal ovary architecture show that changes in developmental rate or timing (heterochrony) can cause profound changes in otherwise very conserved morphologies. Heterochrony has also been suggested as the main mechanism that led to the reduction to four ovules (Vasile et al. 2021). The data presented here underscore a surprisingly high degree of conservation in the flower architecture of Hydrophyllaceae—despite the obvious differences in flower morphology across the family. In this taxon, functional floral diversity is realized by simple shifts in the relative development of different organs. At a first glance, this is in stark contrast with the considerable morphological diversity of the nectaries, internal and external ovary architecture, and ovule numbers and arrangements observed in the Hydrophyllaceae. These characters are largely conserved across most families in Boraginales and indeed are sufficiently conserved across many groups of angiosperms to have been classically used in higher-order classification. Flower and fruit ontogeny and evolution in Boraginales are thus surprisingly complex phenomena, and recent progress with contrast-enhanced μCT, 3D visualization, and an expanded sampling provides novel insights into flower evolution across this order. Our data, however, indicate that minor heterochronous shifts in individual aspects of the developmental trajectories—placentation, ovule initiation, ovary expansion—are able to explain the diversification of the internal and external ovary architectures of Hydrophyllaceae. An expansion of these studies to the few morphologically aberrant lineages that so far have not been studied—for example, Hoplestigmataceae and Lennoaceae—promises a deeper understanding of flower and fruit evolution across the entire order's lineage.

Acknowledgments

We thank A. Menidzabal, Y. Malkowsky, and H.-J. Ensikat for technical support during SEM and μCT preparations. We are grateful to C. Löhne, T. Joßberger, and the staff of the Bonn University Botanic Gardens for maintaining the collections. Funding for the μCT scanner was provided by the Deutsche Forschungsgemeinschaft (INST 217/849-1 FUGG). We want to express our sincere gratitude to California Botanic Garden (formerly Rancho Santa Ana Botanic Garden) for providing the seeds for all three species investigated in this study.

Literature Cited

Arber A 1939 Studies in flower structure. V. On the interpretation of the petal and "corona" in *Lychnis*. Ann Bot 3:337–346.

Berg RY 1985 Gynoecium and development of embryo sac, endosperm, and seed in *Pholistoma* (Hydrophyllaceae) relative to taxonomy. Am J Bot 72:1775–1787.

——— 2009 Embryo sac, endosperm, and seed of *Nemophila* (Boraginaceae) relative to taxonomy, with a remark on embryogeny in *Pholistoma*. Am J Bot 96:565–579.

Bittrich V 2016 Lennoaceae. Pages 257–261 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Chuang TI, L Constance 1992 Seeds and systematics in Hydrophyllaceae: tribe Hydrophylleae. Am J Bot 79:257–264.

Di Fulvio TE, MT Cosa, N Dottori 1999 Morfología y vascularización floral de *Draperia*, *Emmenanthe*, *Hesperochiron*, *Romanzofia* y *Tricardia* (Phacelieae, Hydrophyllaceae). Kurtziana 27: 187–209.

Endress PK 1996 Homoplasy in angiosperm flowers. Pages 303–325 *in* MJ Sanderson, L Hufford, eds. Homoplasy: the recurrence of similarity in evolution. Academic Press, San Diego, CA.

Gottschling M 2004 Floral ontogeny in *Bourreria* (Ehretiaceae, Boraginales). Flora 199:409–423.

Gottschling M, S Nagelmüller, HH Hilger 2014 Generative ontogeny in *Tiquilia* (Ehretiaceae: Boraginales) and phylogenetic implications. Biol J Linn Soc 112:520–534.

Gottschling M, M Weigend, HH Hilger 2016 Ehretiaceae. Pages 165–178 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Heigl HML, J Kretschmann, HH Hilger, M Gottschling 2020 Flower and fruit anatomy of *Cordia nodosa* Lam. and *Varronia bonplandii* Desv. (Cordiaceae, Boraginales) with phylogenetic implications. Org Divers Evol 20:9–24.

Hilger HH 2014 Ontogeny, morphology, and systematic significance of glochidiate and winged fruits of Cynoglosseae and Eritrichieae (Boraginaceae). Plant Divers Evol 131:167–214.

Hilger HH, M Weigend 2016 Wellstediaceae. Pages 403–406 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Hofmann M 1999 Flower and fruit development in the genus *Phacelia* (Phacelieae, Hydrophyllaceae): characters of systematic value. Syst Geogr Plants 68:203–212.

Hofmann M, GK Walden, HH Hilger, M Weigend 2016 Hydrophyllaceae. Pages 221–238 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Jeiter J, F Danisch, HH Hilger 2016 Polymery and nectary chambers in *Codon* (Codonaceae): flower and fruit development in a small, capsule-bearing family of Boraginales. Flora 220:94–102.

Jeiter J, YM Staedler, J Schönenberger, M Weigend, F Luebert 2018 Gynoecium and fruit development in *Heliotropium* sect. *Heliothamnus* (Heliotropiaceae). Int J Plant Sci 179:275–286.

Jeiter J, M Weigend 2018 Simple scales make complex compartments: ontogeny and morphology of stamen–corolla tube modifications in Hydrophyllaceae (Boraginales). Biol J Linn Soc 125:802–820.

Luebert F, L Cecchi, MW Frohlich, M Gottschling, CM Guilliams, KE Hasenstab-Lehman, HH Hilger, et al 2016 Familial classification of the Boraginales. Taxon 65:502–522.

Object Research Systems 2020 Dragonfly. Object Research Systems, Montreal. http://www.theobjects.com/dragonfly.

Paucar JOA, RMS Isaias, JR Stehmann 2020 Unravelling the structure and function of the petal appendages in the tribe Schwenckieae (Solanaceae). Plant Biol 22:146–156.

Payer J-B 1857 Traité d'organogénie comparée de la fleur. V. Masson, Paris.

Staedler YM, D Masson, J Schönenberger 2013 Plant tissues in 3D via x-ray tomography: simple contrasting methods allow high resolution imaging. PLoS ONE 8:e75295.

Thomson BF 1942 The floral morphology of the Caryophyllaceae. Am J Bot 29:333–349.

Vasile M-A, J Jeiter, M Weigend, F Luebert 2020 Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia*. Syst Biodivers 18:757–770.

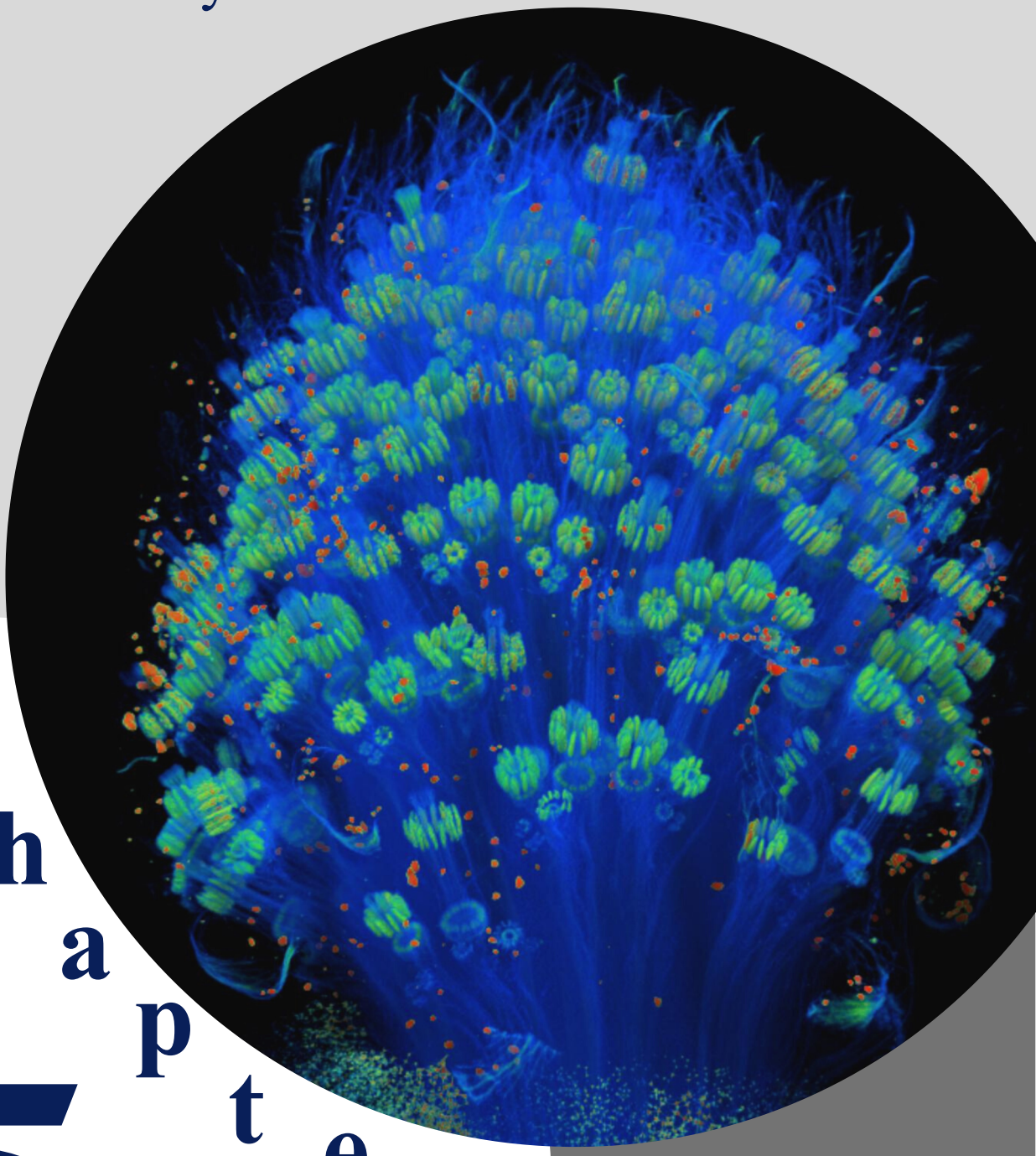
Vasile M-A, F Luebert, J Jeiter, M Weigend 2021 Fruit evolution in Hydrophyllaceae. Am J Bot 108:925–945.

Weigend M, M Gottschling 2006 Evolution of funnel-revolver flowers and ornithophily in *Nasa* (Loasaceae). Plant Biol 8:120–142.

Weigend M, HH Hilger 2010 Codonaceae: a newly required family name in Boraginales. Phytotaxa 10:26–30.

Weigend M, F Luebert, M Gottschling, TLP Couvreur, HH Hilger, JS Miller 2014 From capsules to nutlets—phylogenetic relationships in the Boraginales. Cladistics 30:508–518.

The odd one out: a comparison of flower and fruit development in holoparasitic *Pholisma arenarium* (Lennoaceae, Boraginales) to that in closely related Ehretiaceae.



Chapter 5

Micro–computed tomography–based volume rendering showing an inflorescence of *Pholisma arenarium* Nutt. (Lennoaceae) collected by B. Lee & D. Grossenbacher in California, USA. Image by J. Jeiter.

Jeiter J, Vasile MA, Lewin EM, Weigend M. 2023. The odd one out: a comparison of flower and fruit development in holoparasitic *Pholisma arenarium* (Lennoaceae, Boraginales) to that in closely related Ehretiaceae. *International Journal of Plant Sciences* 184 (1):1–18. https://doi.org/10.1086/722593

THE ODD ONE OUT: A COMPARISON OF FLOWER AND FRUIT DEVELOPMENT IN HOLOPARASITIC *PHOLISMA ARENARIUM* (LENNOACEAE, BORAGINALES) TO THAT IN CLOSELY RELATED EHRETIACEAE

Julius Jeiter,[1,2,*] Maria-Anna Vasile,[2,*] Emma M. Lewin,* and Maximilian Weigend*

*Nees-Institute for Biodiversity of Plants, University of Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany

Editor: Maria von Balthazar

Premise of research. Lennoaceae is a small root-parasitic plant family comprising four species in two genera. They are included in Boraginales, but their position within the order has long been questionable. Only recent phylogenomic studies have placed Lennoaceae within Ehretiaceae with highest support. Because of their holoparasitic lifestyle, Lennoaceae are morphologically aberrant, and their morphological evolution is not fully understood. The position of Lennoaceae within Ehretiaceae invites a study of their development and morphology in comparison to closely related species of that family.

Methodology. We investigate the flower and fruit development of *Pholisma arenarium* from the Lennoaceae and two species from Ehretiaceae (*Bourreria succulenta* and *Ehretia dicksonii*) using scanning electron microscopy and micro–computed tomography.

Pivotal results. The three species share a number of common morphological and architectural features. Most noteworthy are the absence of a stylar compitum and the formation of false septa, which are involved in pyrene formation. *Pholisma arenarium* and *B. succulenta* share more similarities with each other than with *E. dicksonii*.

Conclusions. Our findings are in line with recent phylogenomic analysis and underscore that Lennoaceae should not be treated as a separate family. Aberrant fruit morphology of *P. arenarium* appears to be driven by its holoparasitic lifestyle, but fruit anatomy clearly reflects a position in Ehretiaceae. Increased seed production is ensured by carpel multiplication as an obvious escape route to the evolutionary dead end of seed and pericarp integration. Changes in merosity of the outer floral organs are a result of an increased floral apex size.

Keywords: Boraginales, root parasite, parasitism, ontogeny, micro–computed tomography, evolution.

Online enhancement: table.

Introduction

Parasitic plants independently evolved several times within angiosperms (Nickrent 2020). In several cases small clades of parasitic plants evolved within clades of otherwise autotrophic clades. These parasites are often morphologically highly divergent, both vegetatively and generatively. Examples include the Apodanthaceae (Curcubitales; Bellot and Renner 2013), Cynomoriaceae (Saxifragales; Nickrent et al. 2005), and Lennoaceae (Boraginales; Luebert et al. 2016). It seems obvious that a parasitic lifestyle has massive repercussions on reproductive biology and hence flower and fruit morphology, particularly in holoparasitism as the most specialized type. Holoparasitic plants are entirely dependent on their host plant, since they are unable to photosynthesize and acquire water and all nutrients from their host (Teixeira-Costa and Davis 2021). The aberrant morphology of parasites has traditionally complicated their systematic placement. Even with molecular phylogenetic techniques, their placement often remained uncertain because of the reliance of most early phylogenetic studies on plastid markers. These markers are considered unreliable, since the plastid genome is typically degenerated in holoparasites (Wicke and Naumann 2018). Progress in molecular phylogenetics, especially the use of nuclear markers and whole-genome sequencing, increasingly permits the resolution of these phylogenetic issues (e.g., Nickrent et al. 2005; Zhang et al. 2020). Questions of morphological evolution can now be addressed according to well-resolved relationships.

[1] Author for correspondence; email: jjeiter@uni-bonn.de.
[2] These authors contributed equally to this work.

Manuscript received August 2022; revised manuscript received September 2022; electronically published November 11, 2022.

International Journal of Plant Sciences, volume 184, number 1, January 2023. © 2022 The University of Chicago. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits non-commercial reuse of the work with attribution. For commercial use, contact journalpermissions@press.uchicago.edu. Published by The University of Chicago Press. https://doi.org/10.1086/722593

We investigated a representative of the holoparasitic family Lennoaceae in comparison to two closely related taxa from the Ehretiaceae. Lennoaceae comprise four species in two genera: *Pholisma* Nutt. ex Hook., comprising three species, and the monotypic *Lennoa* Lex. (Yatskievych and Mason 1986; Bittrich 2016). All Lennoaceae are perennial or annual root parasites, depending on the life history of their host plants (Bittrich 2016). Ehretiaceae and Lennoaceae belong to the Boraginales and are nested in the Boraginales II clade (Luebert et al. 2016). Molecular studies retrieved a close relationship between these two families (Gottschling et al. 2014; Weigend et al. 2014; Stull et al. 2015; Schneider et al. 2018), but the exact phylogenetic placement of Lennoaceae was resolved only recently (Zhang et al. 2020). The study by Zhang et al. (2020) indicates that Lennoaceae are nested within Ehretiaceae and more closely related to *Bourreria* P.Browne than *Ehretia* P.Browne. Even though the two families are represented by few taxa in the phylogeny of Zhang et al. (2020), the position of Lennoaceae within Ehretiaceae is highly supported (100% on all six trees).

Independent of their placement within Boraginales, Lennoaceae are distinguished from the rest of the order by their unusual morphology. As root holoparasites, Lennoaceae are reduced to a callus connected to their respective hosts via haustoria in their vegetative phase and appear aboveground only when they produce their massive inflorescences (fig. 1*A*, 1*B*, 1*E*, 1*F*; Suessenguth 1927; Copeland 1935; Yatskievych and Mason 1986; Bittrich 2016; Carlquist and Guilliams 2017). Their flowers are similar to those of the other Boraginales: they are tetracyclic, actinomorphic, and hermaphroditic (Bittrich 2016). However, they differ strikingly from other Boraginales in their polymerous (vs. pentamerous) flowers. The most striking difference is the increase in carpel number, which is otherwise fixed to two in Boraginales (Luebert et al. 2016), even in genera with otherwise polymerous flowers, such as *Codon* L. (Weigend and Hilger 2010; Jeiter et al. 2016; Weigend and Hilger 2016) or some *Cordia* L. (Gottschling et al. 2016). The mechanisms and evolutionary drivers underlying the increase in organ number in Lennoaceae so far have not been studied. To date, there are only two studies on flower and fruit development in Lennoaceae (Suessenguth 1927; Copeland 1935). Both were conducted in the context of a nowadays outdated phylogenetic framework and before modern imaging techniques were available.

We used scanning electron microscopy (SEM) in combination with micro–computed tomography (μCT) to study the development of flowers and fruits in *Pholisma arenarium* Nutt. in comparison with two species of Ehretiaceae, *Bourreria succulenta*

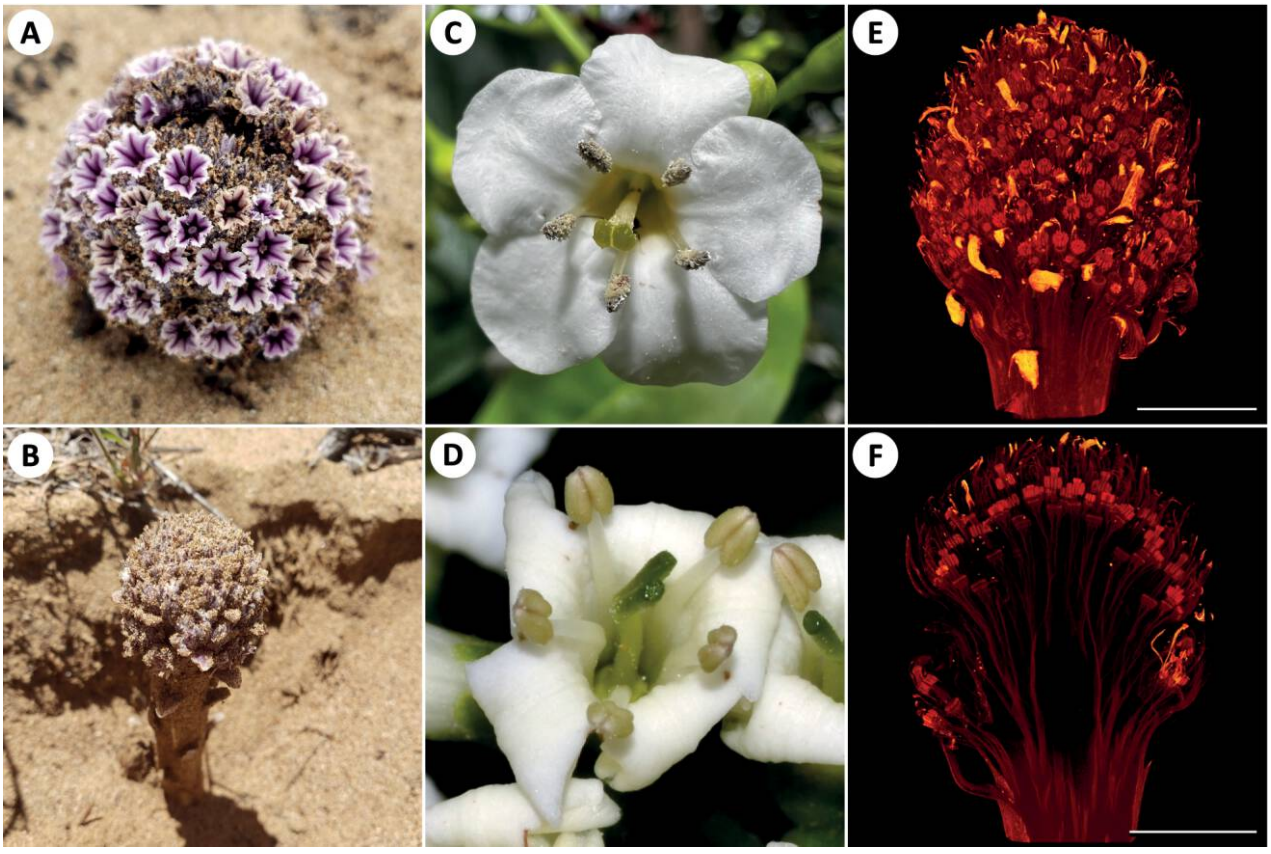


Fig. 1 Overview of flowers and inflorescences of *Bourreria succulenta*, *Ehretia dicksonii*, and *Pholisma arenarium*. *A*, *B*, *E*, *F*, *Pholisma arenarium*. *C*, *Bourreria succulenta*. *D*, *Ehretia dicksonii*. *A*, Inflorescence in full bloom. *B*, Young inflorescence before flowering. *E*, Volume rendering of a micro–computed tomography scan of an inflorescence in full bloom. *F*, Virtual section through the volume rendering of *E*. Photographs in *A* and *B* by Bridget Lee. Scale bars = 1 cm.

Table 1

Plant Material Used in This Study and Voucher Placement

Species	Collector name and no.	IPEN no.	Material	Herbarium
Pholisma arenarium Nutt.	Bridget Lee and D. Grossenbacher 1	...	Flowers, inflorescences, mature fruits, infructescences	BONN/OBI
P. arenarium Nutt.	Hartmut H. Hilger 62/1992	...	Young fruits	BSB/B
Bourreria succulenta Jacq.	...	xx-0-BONN-35343	Flowers, fruits	BONN
Ehretia dicksonii Hance	...	xx-0-BONN-12602	Flowers, fruits	BONN

Note. IPEN = International Plant Exchange Network.

Jacq. and *Ehretia dicksonii* Hance, all included in the recent phylogeny from Zhang et al. (2020). We address the following questions: (i) how do the flower and fruit of *P. arenarium* develop compared with closely related members of Ehretiaceae and (ii) does the flower and fruit development of *P. arenarium* provide additional insights into the exact relationships of Lennoaceae to the remaining Boraginales?

Material and Methods

Plant material. Entire inflorescences of *Pholisma arenarium* were collected in San Luis Obispo County, California (by Bridget Lee and D. Grossenbacher 1) and fixed in EtOH (70%; fig. 1*A*, 1*B*). Additionally, older inflorescences and infructescences were collected in California (by Hartmut H. Hilger 62/1992) and fixed in formaldehyde–acetic acid–ethanol. For *Bourreria succulenta* (fig. 1*C*) and *Ehretia dicksonii* (fig. 1*D*), flowers and fruits of different developmental stages were collected from the living collection of Bonn University Botanic Gardens. For details on material and voucher placement, see table 1.

Scanning electron microscopy. The fixed material was kept in the respective fixative for at least 1 wk and rinsed three times with EtOH (70%) before storage in EtOH (70%). The material was dissected submerged in EtOH (70%) under a stereomicroscope. Bracts, stamen–corolla tubes, and parts of the calyx were removed. The dissected samples were transferred into EtOH (70%) plus acetic acid (4%) for approximately 2 h. The samples were dehydrated using formaldehyde dimethyl acetal (100%) and gradually transferred into acetone. The samples were finally stored in acetone. Samples were critical-point dried from acetone in an automatic critical-point dryer (Leica CPD 300, Leica Microsystems, Wetzlar, Germany) using 18–20 exchange cycles and gas outlet set to 85% slow. Dried samples were mounted with conductive carbon cement (Plano, Wetzlar, Germany) on aluminium stubs covered with glass slides. Mounted samples were further dissected, and structures such as sepals and larger trichomes were removed. Subsequently, mounted samples were sputter-coated with gold or palladium (Balzers SCD 400, Balzers Union, Liechtenstein) for 1.5–3 min at 30 mA depending on sample complexity. Samples were observed in a Stereoscan 200 electron microscope (Cambridge, UK). After image acquisition, contrast and brightness were improved using standard image-editing software.

Micro–computed tomography. The fixed samples were transferred into a staining solution of ethanol (70%) and phosphotungstic acid (PTA; 1%) following Staedler et al. (2013). Samples were kept in the staining solution for at least 40 d following the recommendations by Duncan et al. (2022). The staining solution was exchanged approximately every second day. Stained samples were dehydrated using increasing concentrations of ethanol and acetone. All solutions contained PTA (1%). The samples were critical-point dried (Leica CPD 300, Leica Microsystems) and mounted on different stubs, such as aluminium rods or toothpicks, depending on their size. Samples were mounted with either hot glue or two-component epoxy resin (UHU Plus Sofortfest, Rotterdam, Netherlands).

Samples were scanned in a Skyscan 1272 (Bruker, Billerics, MA) equipped with a Hamamatsu L11871-20 source (Hamamatsu, Japan) and a Ximea xiRAY16 camera (Ximea, Münster, Germany). The entire inflorescences were scanned in a Skyscan 1173 equipped with a Hamamatsu 130/300 X-ray source and a flat panel sensor as a camera. Scanner settings can be found in table A1. Image stacks were reconstructed using the InstaRecon Engine version 2 (NRecon, Champaign, IL). Image stacks were visualized in Dragonfly versions 2021.1 and 2021.3 (Object Research Systems, Montreal).

Results

The focus of this article is the flower and fruit development of *Pholisma arenarium* (fig. 1*A*, 1*B*). For this reason, we present the development of this species separately and in greater detail. *Bourreria succulenta* (fig. 1*C*) and *Ehretia dicksonii* (fig. 1*D*) are presented in a separate section.

Pholisma arenarium

Early flower development. The flowers develop in irregular double cymes in the axil of a scalelike bract. Each flower develops in the axil of a scalelike bract. All flowers are fused to the main axis of the capitulum-like inflorescence (fig. 1*E*, 1*F*). The branches of the cymose partial inflorescences develop from top to bottom, with the oldest flower towering over the younger flowers. Flowers sit on short pedicels and are almost sessile throughout their entire development.

The floral apex is dome-shaped and slightly compressed in its median axis between the bract and the peduncle (fig. 2*A*). The first sepal is initiated in abaxial position (fig. 2*B*), but its exact position can vary. The remaining sepals develop in a 2/5 spiral pattern (fig. 2*C*, 2*D*). A sixth sepal primordium is usually inserted between the second and fourth sepal primordia (fig. 2*D*). The number of sepals varies between five and six, with six being the most common. The floral apex develops into a pentagonal

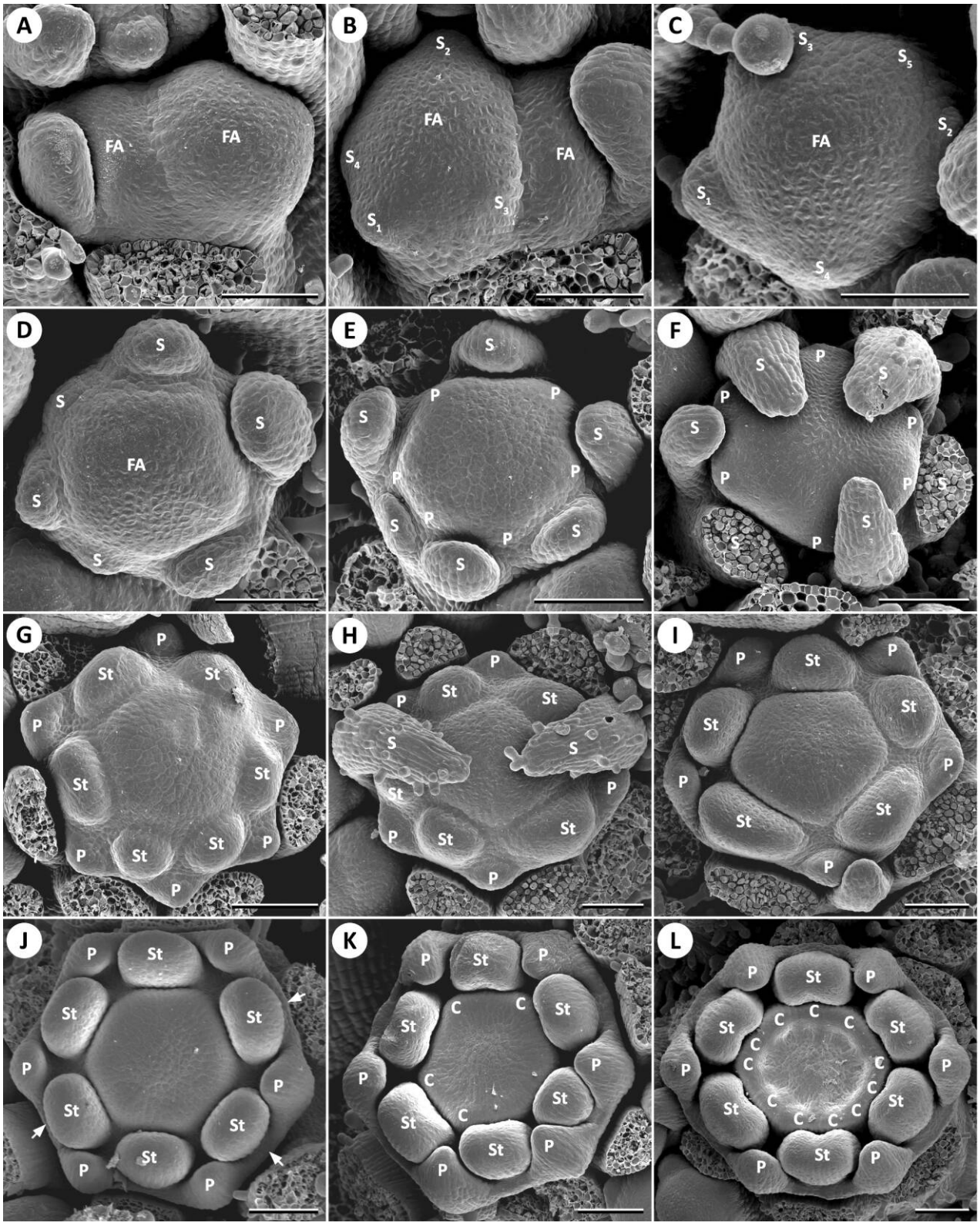


Fig. 2 *Pholisma arenarium* early flower development. *A*, Flower initiation. *B*, Onset of calyx development. *C*, Calyx development. *D*, Calyx development before petal primordia initiation. *E*, Petal primordia initiation. *F*, Petal development. *G*, Stamen primordia initiation. *H*, Stamen development. *I*, Stamen and petal development. *J*, Flattening of the floral apex before carpel primordia initiation and onset of corolla and stamen–corolla tube development (arrows). *K*, Initiation of carpel primordia. *L*, Differentiation of stamen primordia into anthers and filaments and fusion of the carpel primordia. C = carpel; FA = floral apex; P = petal; S = sepal; St = stamen. Scale bars = 100 μm.

or hexagonal shape, while the sepal primordia flatten (fig. 2*D*, 2*E*). The number of sepals and the number of following floral organs are only loosely related. The petal primordia demarcate from the corners of the pentagonal or hexagonal floral apex. They initiate subsequently, without any noticeable plastochrone in between (fig. 2*E*, 2*F*).

The floral apex begins to flatten after petal primordia initiation, and a constriction is formed below the points where the petals are inserted. At this point the young sepals start to enclose the floral apex. They remain slender throughout their entire development and are apert with only their tips overlapping, showing more or less imbricate aestivation. They also start to develop glandular trichomes on multicellular stalks covering the entire outer surface of the sepals (fig. 2*F*, 2*H*). Staminal primordia arise in alternipetalous positions. They are initiated subsimultaneously and correlate in their number with the petal primordia. Staminal primordia are variable in size. Their sizes correlate with the sizes of the gaps between the petal primordia in antesepalous position (fig. 2*G*–2*I*). The size difference between the staminal primordia is lost during their development (fig. 2*J*, 2*K*). The anthers start to differentiate when the carpel primordia are initiated (fig. 2*K*). Shortly after initiation of the staminal primordia, the petal primordia are shifted upward from their marginal position on the pentagonal or hexagonal floral apex. From this point onward, petal and staminal primordia are associated during their development (fig. 2*H*, 2*I*). Petal primordia start to emerge on a common ring that then fuses with the staminal primordia (fig. 2*J*). The stamen–corolla tube is initiated. At anther initiation, the tips of the now clearly differentiated corolla lobes are bent inward and alternate with the anthers. The sinuses in between the corolla lobes remain broad, and the anthers are only basally and marginally covered. After staminal primordia are differentiated from the floral apex, the floral apex starts to take a hexagonal or pentagonal shape depending on the merosity of the flower (fig. 2*G*). The margins of the floral apex start to rise, forming a shallow depression. In the center of this depression a shallow dome-like structure remains (fig. 2*I*–2*L*). On the raised margins of this polygonal structure, the carpel primordia start to emerge in an irregular pattern (fig. 2*K*).

External gynoecium development. Gynoecium development is initiated before the anthers start to demarcate from the staminal primordia (fig. 3*A*). Carpels are initiated in an irregular pattern. Carpel primordia number and position do not appear to be correlated to the merosity of the outer floral organs (fig. 3*B*). The most common number of carpel primordia we observed was 11. From carpel initiation onward, the carpel primordia form a common ring with a wide, shallowly dome-shaped space in the center of the gynoecium (fig. 3*B*). The individual carpel primordia are recognizable as shallow bulges on the common ring. Septa start to develop at this early ring stage (fig. 3*B*). The carpels continue to grow apically, and the base of the young gynoecium widens (fig. 3*C*–3*E*). The tips of the carpels are clearly recognizable at this stage, and sinuses form in the position of the septa (fig. 3*D*). The overall outer shape of the gynoecium remains pentagonal or hexagonal, and the corners slightly bulge in between the stamens (fig. 3*E*). The entire young gynoecium and the individual carpels are still open at this stage (fig. 3*D*, 3*E*). With the formation of an almost closed ovary and the formation of the open gynoecial cone, the development of the nectary disk is initiated and externally visible as first nectarostomata appear on its surface (fig. 3*F*). The carpel tips grow and fill the opening of the stylar channel, and a constriction forms that demarcates the transition from the ovary into the style (fig. 3*F*, 3*G*). In the final phase of external gynoecium development, the ovary widens and increases in overall size, transitioning abruptly into the style (fig. 3*H*–3*J*). The style continues to elongate, and the stigma starts to form (fig. 3*H*). In the anthetic flower the entire style has approximately the same circumference, and ovary and stigma are clearly demarcated from it (fig. 3*J*). The stigma is open in the center and slightly lobed at the very tip, with each lobe corresponding to one carpel. The stigmatic surfaces touch each other and appear to form a common receptive surface (fig. 3*J*). The gynoecial nectary disk is not distinct from the ovary and recognizable only by a small number of nectarostomata at the very base of the ovary. Ovary and gynoecial nectary disk are shallowly lobed (fig. 3*J*).

Internal gynoecium development. After carpel initiation and the formation of the gynoecial cone, the distal parts of the septa start to develop into the ovules (fig. 4*A*–4*D*). Two ovules develop per carpel. In cross section, the septa develop in an arrowhead shape, with the ovules inserted at the base of the arrowheads (fig. 4*C*, 4*F*, 4*I*, 4*K*). With the onset of ovule development, false septa are initiated at the distal walls from the base of the ovary and the distal walls of the carpels at the position of the dorsal carpellary bundles (fig. 4*C*, 4*F*). False septa initiation coincides with the differentiation of the gynoecial cone into ovary and style (fig. 4*F*, 4*G*). The false septa continue to grow synchronously with the ovules (fig. 4*F*, 4*I*). The two ovules of each carpel are completely separated and enclosed in separate compartments in the gynoecium of the anthetic flower (fig. 4*L*). The arrowhead-shaped septa in between the carpels enlarge and fill most of the ovarian cavity (fig. 4*F*, 4*I*, 4*L*). The ovarian cavity at the base of the stylar channel has an irregular star-shape with narrow channels branching off in each direction (fig. 4*L*). The number of channels corresponds to the number of carpels. Each channel is subdivided in the end by the false septa (fig. 4*L*).

The system of stylar channel, ovarian cavity, and connected channels is largely separate from the pollen tube transmitting tissue (fig. 4*M*). Each carpel has its own duct of pollen tube transmitting tissue running from the stigmatic tissue to the ovary (fig. 4*E*, 4*G*, 4*H*, 4*J*, 4*K*, 4*M*). The strands of pollen tube transmitting tissue are completely separate (fig. 4*E*, 4*G*, 4*H*, 4*J*, 4*K*, 4*M*). They are terminated at the channels branching off from the ovarian cavity (fig. 4*G*, 4*J*, 4*M*). There is no clear compitum formed. The only connection between the pollen tube transmitting tissues is at the level of the ovarian cavity through the network of channels (fig. 4*L*). However, a PTA-accumulating glandular or mucous tissue is not identifiable inside the ovarian cavity (fig. 4*L*, 4*M*). The stigmatic tissue of the stigmatic head is connected and covered in a mucus layer indicating its function as an external compitum (figs. 3*J*, 4*M*). The gynoecium remains open throughout gynoecium development (fig. 4*D*, 4*G*, 4*J*). Only at late preanthetic stages, the stylar channel is loosely closed through irregular growth of the inner epidermis (fig. 4*J*, 4*M*).

Fruit development. After anthesis the ovary and the base of the style start to widen, resulting in a conical appearance of the postanthetic ovary (fig. 5*A*). Stigma and style wilt continuously from the top toward the base throughout flower development but continue to slightly elongate (fig. 5*A*–5*D*). As the ovary continues to increase its diameter, the base of the style

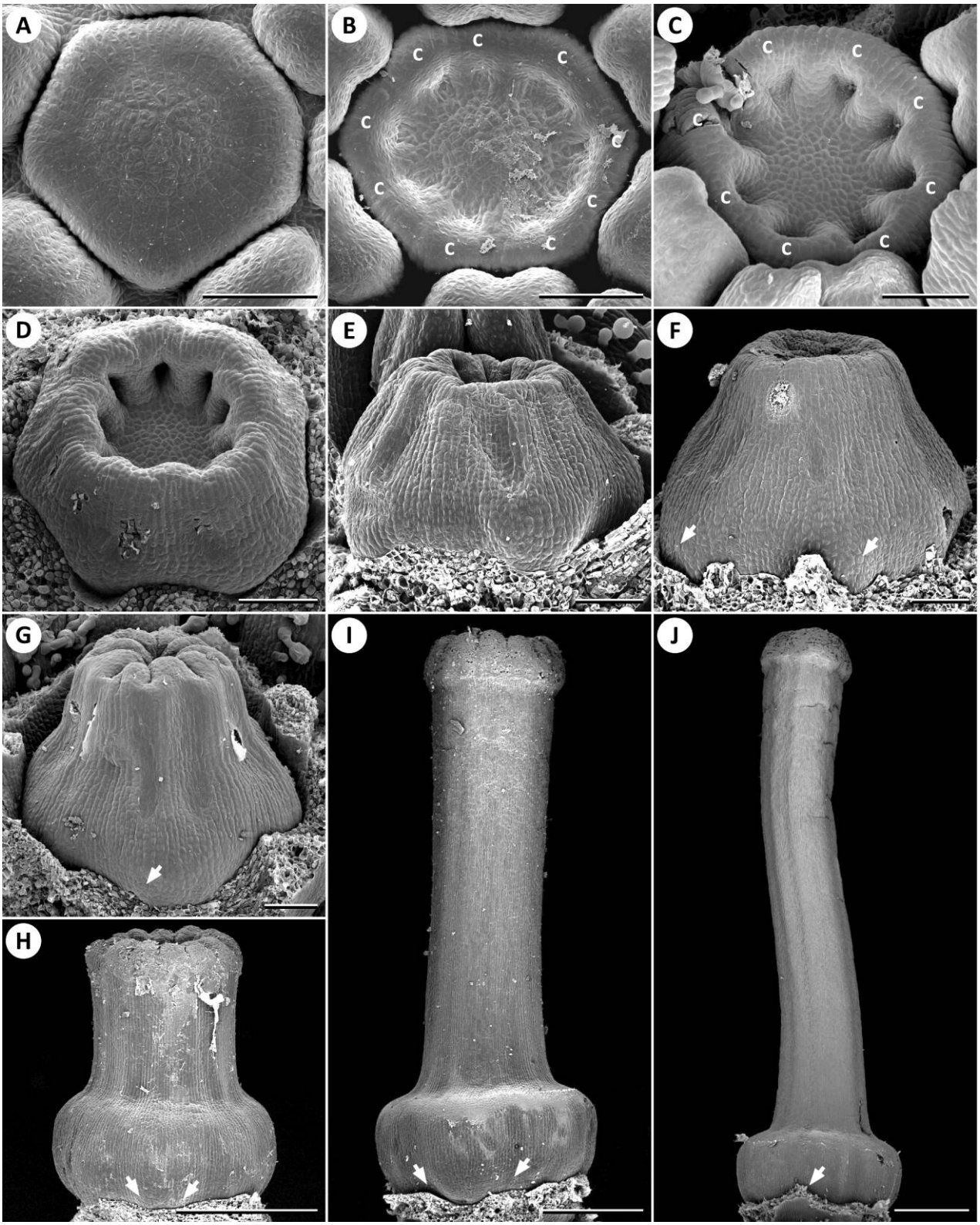


Fig. 3 *Pholisma arenarium* external gynoecium development. *A*, Floral apex before carpel initiation. *B*, Carpel initiation. Note the onset of septa development. *C*, Carpels develop as a common structure with distinct septa between the carpels. *D*, Carpels begin to close. Note the open spacious ovarian cavity in the center. *E*, Further closure of the gynoecial cone. *F*, Constriction of the gynoecial cone, indicating the onset of differentiation into style and ovary. Stomata at the base of the ovary (arrows). *G*, Onset of differentiation of the stigma. Stomata (arrow). *H*, Young gynoecium differentiated into ovary with stomata at the base, style, and stigma. *I*, Gynoecium of a preanthetic flower and differentiation of the capitate stigma. Stomata (arrows). *J*, Gynoecium of an anthetic flower. Stomata (arrow). C = carpel. Scale bars = 100 μm (*A–G*), 500 μm (*H–J*).

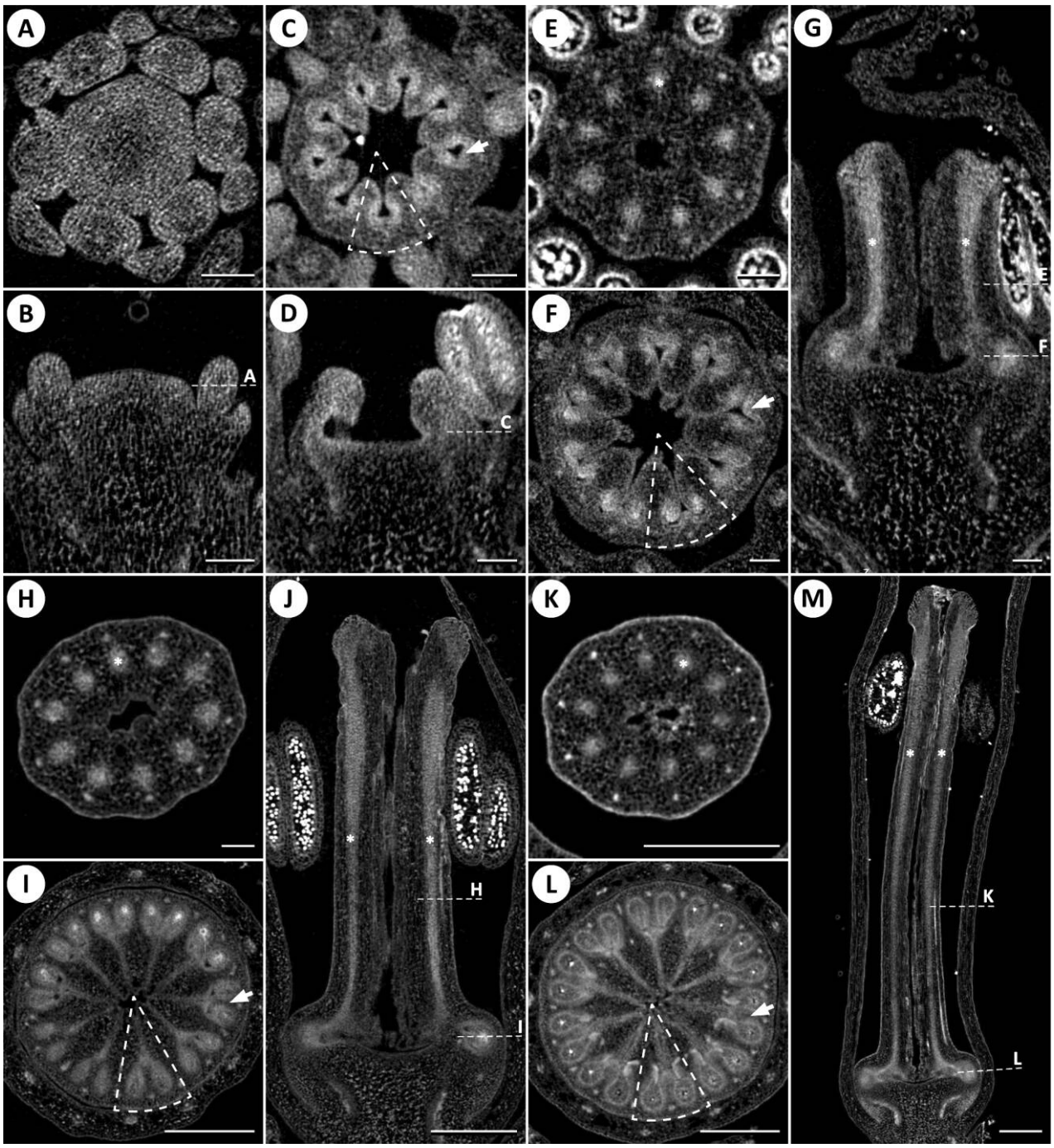


Fig. 4 *Pholisma arenarium* internal gynoecium development. Virtual longitudinal and cross sections based on micro–computed tomography scans. *A*, *B*, Gynoecium before carpel initiation. *A*, Cross section. *B*, Longitudinal section. *C*, *D*, Gynoecial cone. *C*, Cross section. Note the tissue with stronger contrast inside each carpel, which indicates the onset of ovule development. Beginning of false septa development (arrow). *D*, Longitudinal section. *E–G*, Young, differentiated gynoecium. *E*, Cross section through the style. *F*, Cross section through the ovary. Developing false septum (arrow). *G*, Longitudinal section. *H–J*, Gynoecium of a preanthetic flower. *H*, Cross section through the style. *I*, Cross section through the ovary. False septum (arrow). *J*, Longitudinal section. *K–M*, Gynoecium of an anthetic flower. *K*, Cross section through the style. *L*, Cross section through the ovary. False septum (arrow). *M*, Longitudinal section. Asterisks indicate the pollen tube transmitting tissue. The section of a circle with a dashed outline highlights a carpel. Dashed lines labeled with an uppercase letter refer to the corresponding image of this figure. Scale bars = 100 μm (*A–H*), 500 μm (*I–M*).

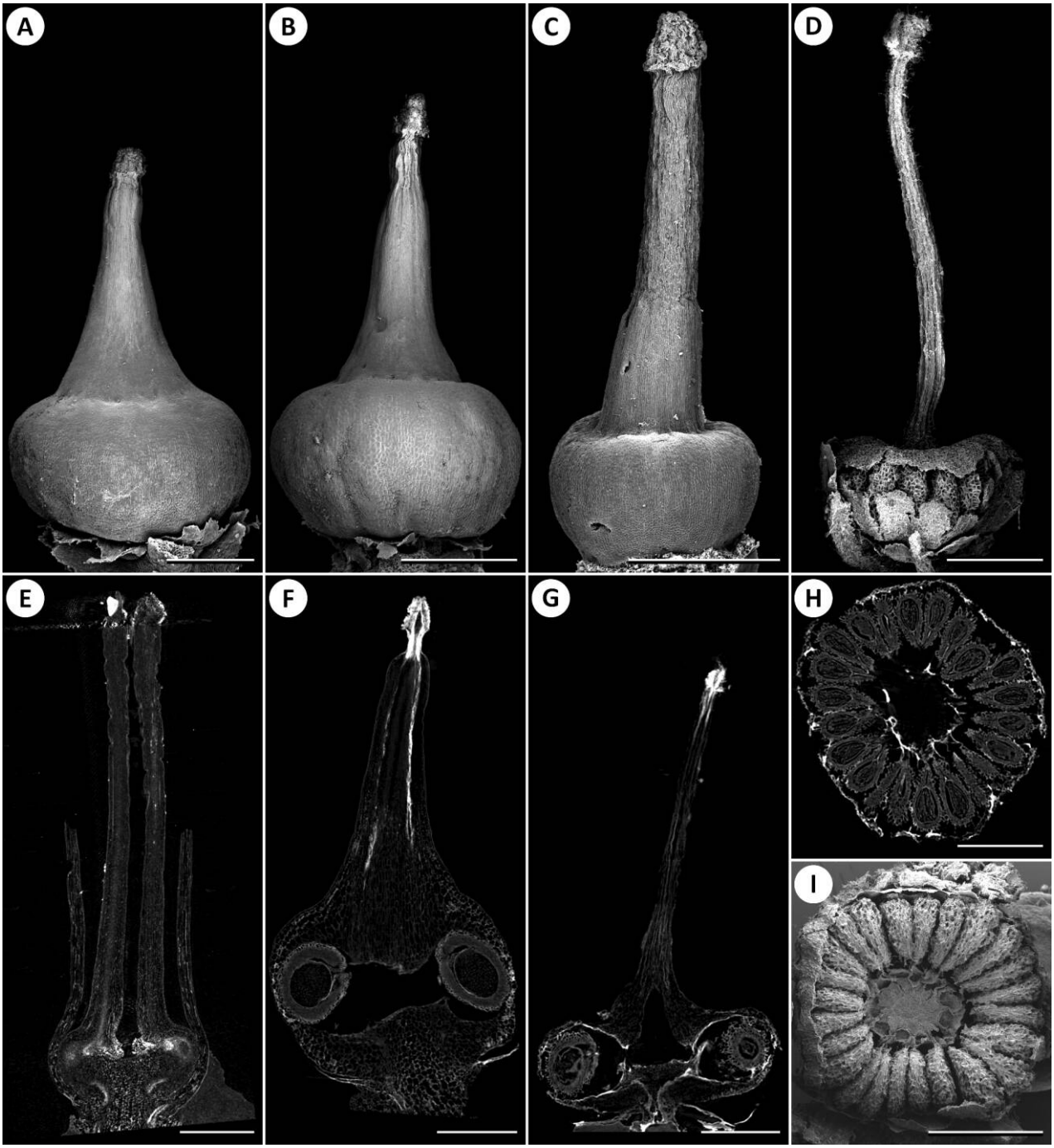


Fig. 5 *Pholisma arenarium* fruit development. *A–D*, External development of the gynoecium after anthesis to mature fruit. *A*, After anthesis. *B*, Older postanthetic stage. *C*, Fruit close to maturity. *D*, Mature fruit. Note the irregular opening of the ovary. *E–G*, Virtual longitudinal sections based on micro–computed tomography scans. *E*, Postanthetic gynoecium. *F*, Older postanthetic gynoecium. *G*, Mature fruit. *H*, Virtual cross section through the ovary of a mature fruit. *I*, Mature fruit. Lid of the capsule removed. Scale bars = 1 mm.

stops its development (fig. 5*A–5D*). The ovary becomes distinguishably flat with a noticeable depression in the center where the style is attached (fig. 5*C*, 5*D*). At fruit maturity the style and the outer layers of the pericarp are completely dry. The pericarp breaks irregularly at the widest point of the ovary (fig. 5*D*). A clear dehiscence line or a band of specialized dehiscence tissue is not visible on the surface of the ovary or in virtual sections according to μCT scans (fig. 5*D*, 5*F*, 5*G*).

The ovary encloses numerous single-seeded endomericarpids (fig. 5*D*, 5*F–5I*). The number of endomericarpids is variable and related to the number of carpels, with each carpel developing into two endomericarpids. The walls of the endomericarpids are

formed by the adjacent walls of the false and true septa separating the individual ovules in the gynoecium of the anthetic flower (fig. 5*E*–5*H*). The seeds are enclosed by a dense tissue (fig. 5*H*). The outer cells of the endomericarpids are alveolate and are formed through breakdown of the mesocarp tissue (fig. 5*I*).

Stamen–corolla tube development. Stamens and petals start to grow on a common ring at the developmental stage where anthers and filaments demarcate from each other and the septa in between the carpels start to develop (fig. 2*K*, 2*L*). At this state the sinuses between the flattened and slightly hood-shaped corolla lobes are relatively broad, and often the anthers are not covered by the lobes (fig. 2*K*, 2*L*). In the following developmental stages, the upper part of the corolla tube and the corolla lobes start to overgrow the developing anthers (fig. 6). The margins become irregular in shape. When the gynoecial cone is fully formed, the corolla lobes become conduplicate, and their margins overlap in an irregularly imbricate pattern (fig. 6*A*, 6*D*). In some of the sinuses, irregular appendages are formed. On the outside around the midvein of the petals, glandular trichomes develop (fig. 6*D*). The corolla lobes now fully cover the anthers, but the center of the flower is still open. In later development and with the onset of style development, the anthers are nested between the slightly enlarged, capitate stigma and the compressed ovary (fig. 6*D*). The anthers are externally fully developed and inserted on short filaments (fig. 6*A*, 6*D*). There is a considerable gap between the style and the corolla lobes. Up until anthesis,

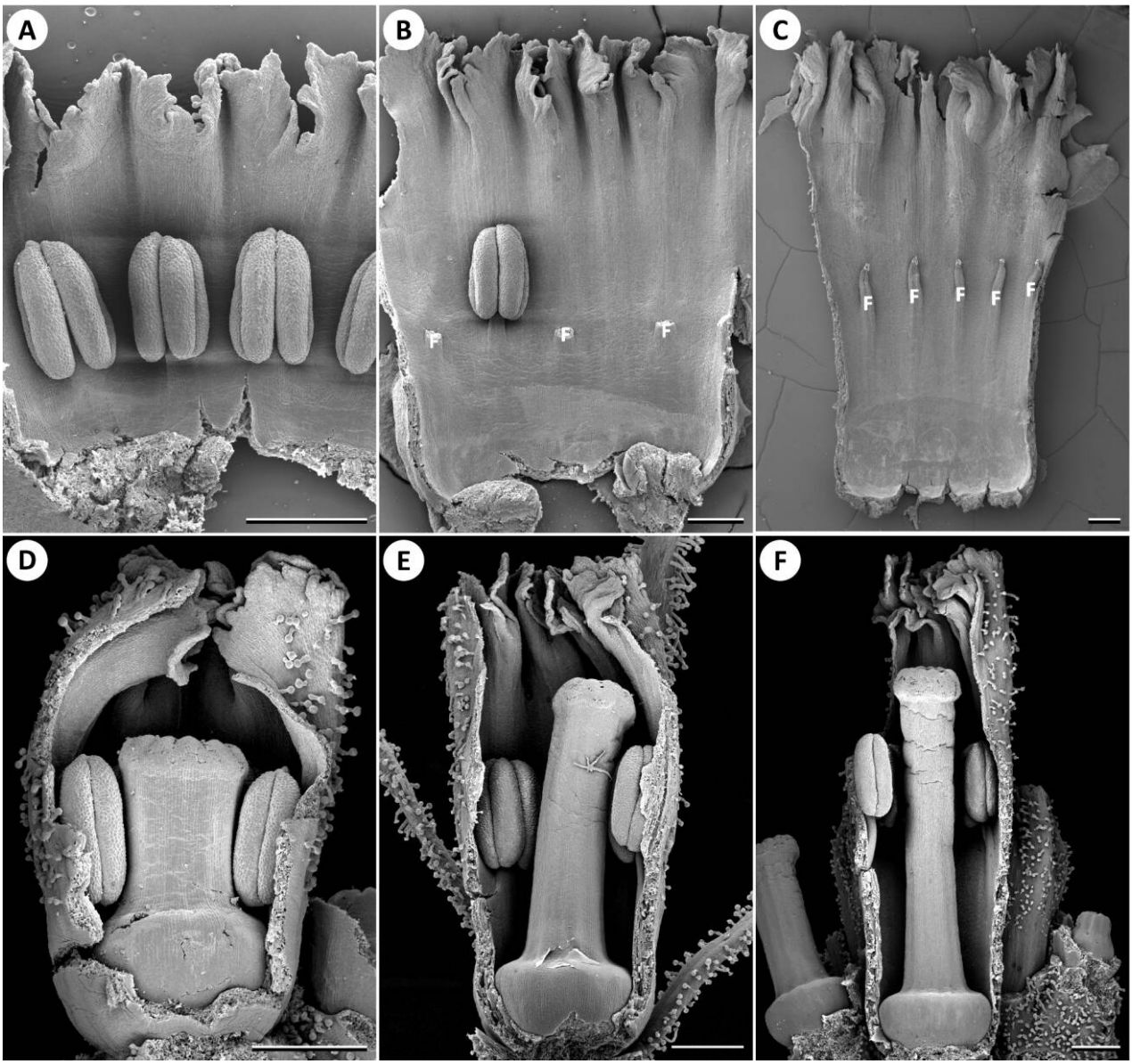


Fig. 6 *Pholisma arenarium* stamen–corolla tube development. *A*–*C*, Stamen–corolla tube development. *A*, Intermediate developmental state. *B*, Stamen–corolla tube before anthesis. *C*, Stamen–corolla tube of an anthetic flower. Anthers removed. *D*–*F*, Dissected flowers illustrating the relation between anthers and stigma. *D*, Young flower. *E*, Preanthetic flower. *F*, Anthetic flower. F = filament. Scale bars = 500 μm.

the elongation of the style exceeds the elongation of the corolla tube below the insertion points of the anthers (fig. 6*B*, 6*E*). At anthesis, the anthers sit on short filaments and are positioned roughly at the upper third of the style (fig. 6*C*, 6*F*). The corolla lobes are erect to spreading (figs. 1*C*, 6*F*).

Bourreria succulenta *and* Ehretia dicksonii

Early flower development. Both species show similar developmental trajectories (fig. 7). Sepals are initiated in a 2/5 spiral pattern (fig. 7*A*, 7*G*). Sepal initiation depends on flower position on the inflorescence. In *B. succulenta* the floral apex flattens, and the sepal primordia are initiated on a shallow rim surrounding the floral apex (fig. 7*A*). Petals are initiated subsimultaneously, alternating with the sepals (fig. 7*B*, 7*H*, 7*I*). In *B. succulenta* the developing sepals cover the floral apex at petal primordia initiation (fig. 7*B*), while petals in *E. dicksonii* are initiated without cover from the sepals (fig. 7*H*, 7*I*). Both species have five stamens. Stamen primordia are initiated shortly after the petals (fig. 7*C*, 7*I*, 7*J*). All five stamens are initiated subsimultaneously (fig. 7*C*, 7*I*, 7*J*). In both species slight variation in the size of the stamen primordia can be observed (fig. 7*J*). A fixed size relationship between stamen primordia position and flower position on the inflorescence or in relation to other floral organs was not observed. Before carpel primordia initiation, stamen primordia and petals cover the floral apex (fig. 7*D*, 7*E*, 7*J*, 7*K*). At this stage petal primordia fuse, and the stamen–corolla tube is formed (fig. 7*E*, 7*K*). The two carpel primordia are initiated as a ring-shaped structure (fig. 7*F*, 7*L*). In some flowers of both species, slight size differences between the carpel primordia were observed (fig. 7*F*). Minor variations in merism occur in both species but are more frequent in *E. dicksonii*. Most commonly, the number of perianth elements is increased, and rarely, the androecium or gynoecium are affected. Usually, organ number increases to six in the affected whorls and very rarely a third carpel is inserted. In *E. dicksonii*, occasionally tetramerous flowers with a regular, bicarpellate gynoecium occur.

External and internal gynoecium development. Early gynoecium development is similar in both species (figs. 8, 9). The two carpels grow as one structure after initiation of the ring-shaped primordium (fig. 8*A*, 8*G*). In both species one carpel is slightly larger than the other, indicating a sequential development but with a short plastochrone (fig. 8*A*, 8*G*). In some flowers the size differences are carried throughout the entire flower development. Shortly after carpel initiation, the gynoecial cone forms (fig. 8*B*, 8*H*). In *B. succulenta*, the gynoecial cone is slightly sunken into the receptacle (fig. 8*B*). The gynoecial cone is completely closed in both species. The tips of both carpels grow, and a cleft forms between them (fig. 8*C*, 8*I*). The tips form the stigmatic plates in *B. succulenta* and the short stylodia in *E. dicksonii* (fig. 8*C*, 8*D*, 8*I*, 8*J*). At the stage where the cleft is formed, the gynoecium of *B. succulenta* gradually tapers from the base to the tips (fig. 8*C*). The gynoecium of *E. dicksonii* already shows a differentiation into ovary and stigma with a constriction in the position where the style is formed (fig. 8*I*). The first nectarostomata appear at the base of the ovary after differentiation into ovary, style, and stigma (fig. 8*D*, 8*J*). The gynoecium of *B. succulenta* develops two connected, disk-shaped stigmata on very short stylodia (sometimes absent; fig. 8*E*, 8*F*). The style elongates relatively quickly in comparison to the rest of the flower, leading its folding (fig. 8*E*). At anthesis it unfolds, exserting the stigma from the flower (figs. 1*A*, 8*F*). In *E. dicksonii* the style elongates in parallel with the stamen–corolla tube; as a result, it is straight throughout development (fig. 8*J*–8*L*). The two stylodia are terminated by truncate stigmata and are approximately one-fifth the length of the style (fig. 8*K*, 8*L*). *Bourreria succulenta* has a prominent nectary disk at the base of the ovary (fig. 8*F*). A distinct nectary disk is absent in *E. dicksonii*, and stomata cover the entire ovary (fig. 8*L*). There is a higher density of stomata at the base of the ovary.

Internally, the two septa between the carpels that separate two locules are formed in the gynoecial cone state (fig. 9*A*, 9*B*, 9*H*, 9*I*). From each septum two ovules are formed. In *B. succulenta* false septa are initiated in the gynoecial cone (fig. 9*A*). The false septa continue to grow throughout gynoecium development and separate the ovules of each carpel, resulting in an ovary with four locules (fig. 9*A*, 9*C*, 9*F*). The false septa in *E. dicksonii* develop later and are less pronounced (fig. 9*J*, 9*M*). The ovary at anthesis has four locules, but separation of the locules is mainly realized through growth of the true septa (fig. 9*M*). Placentation is parietal. In both species the ovarian cavity and the stylar channel are reduced to narrow channels (fig. 9*F*, 9*G*, 9*M*, 9*N*). Each carpel has a separate pollen tube transmitting tissues in the style (fig. 9*D*, 9*E*, 9*G*, 9*K*, 9*L*, 9*N*). In *B. succulenta* an extragynoecial compitum is realized through fusion of the discoid stigmata, and a true compitum is present in the closed ovarian cavity (fig. 9*F*, 9*G*). Narrow channels from the compitum in the ovarian cavity to the four separate locules serve as pollen tube transmitting ducts (fig. 9*F*). Compitum and ovules are on the same level in the upper third of the ovary (fig. 9*F*, 9*G*). In *E. dicksonii* an extragynoecial compitum is absent. A compitum is formed in the closed ovarian cavity (fig. 9*M*, 9*N*). From the compitum, four pollen tube conducting ducts run toward the ovules (fig. 9*M*). Ovules are positioned below the compitum and in the center of the ovary (fig. 9*N*).

Fruit development. The fruits of both species are drupes with four endomericarpids (fig. 10). Each seed is enclosed by part of the hardened endocarp. In *B. succulenta* the endocarp of the false septa also hardens (fig. 10*A*–10*D*). Between the endocarp of the false septa and the endocarp enclosing the embryo, an additional chamber is formed (fig. 10*A*, 10*C*). Ridges of hardened tissue run from each minor vascular bundle in the mesocarp to the hardened tissue of the endocarp (fig. 10*A*, 10*C*). The drupe abruptly transitions into the wilting style (fig. 10*E*). In *E. dicksonii* the endomericarpids are without any ridges (fig. 10*F*–10*I*). The false septa are not involved in endomericarpid formation. The drupe is ovoid and gradually tapers into the wilting style (fig. 10*J*).

Stamen–corolla tube development. Stamen–corolla tube development in both species follows similar developmental trajectories. After fusion between stamens and corolla and subsequent differentiation of the stamen primordia into anthers and filaments, the corolla lobes enlarge and enclose the inner floral organs. Aestivation is imbricate-quincuncial in both species. The stamen–corolla tube elongates before anthesis (fig. 11*A*, 11*D*). At this phase it starts to differentiate into a tube below the point of filament attachment and a tube between filament attachment points and corolla lobes (fig. 11*B*, 11*E*). Elongation of the tubular parts and filaments are correlated. In the anthetic flowers the filaments are only slightly longer than the upper part of the corolla tube (fig. 11*C*, 11*F*). There are no modifications of the

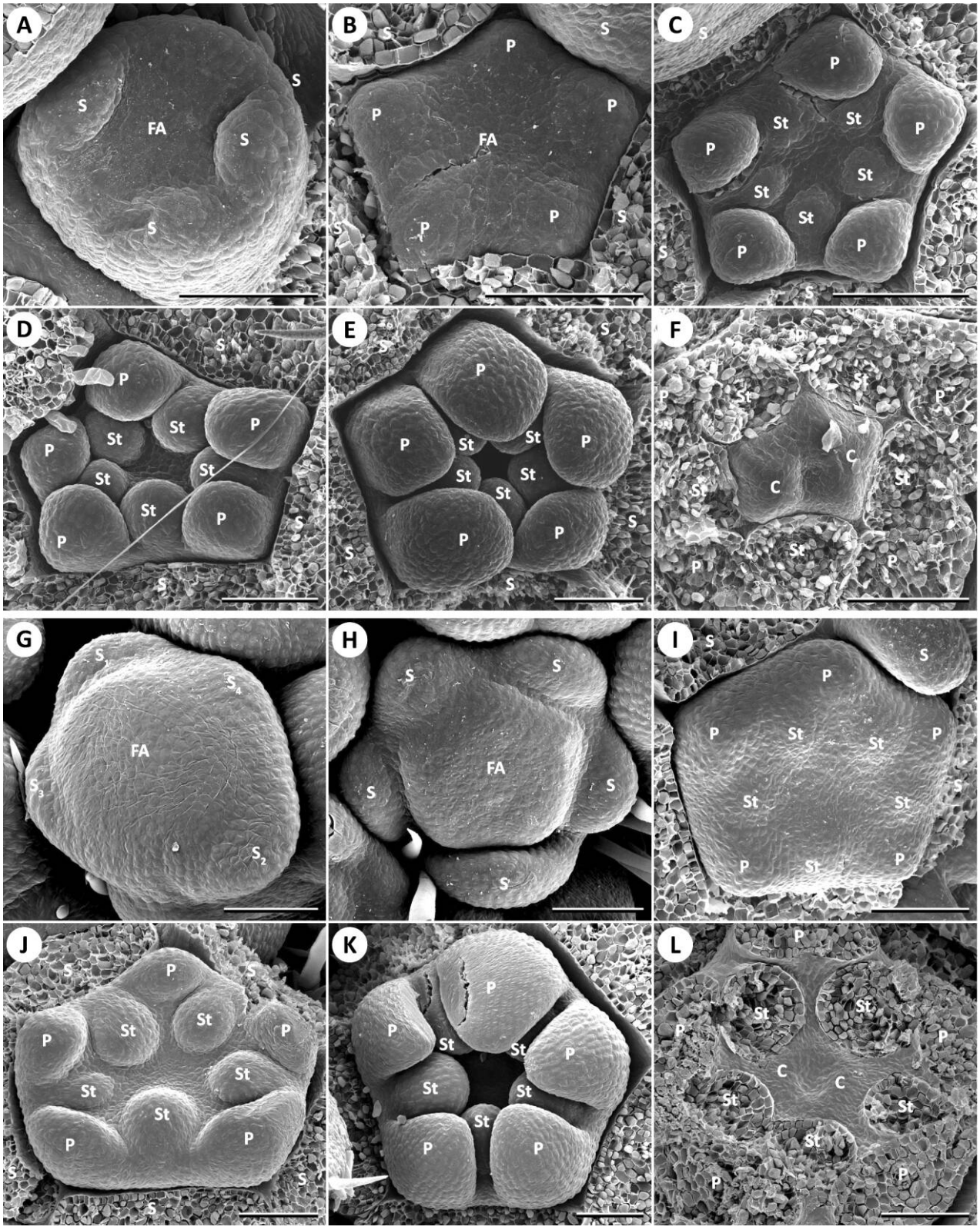


Fig. 7 *Bourreria succulenta* and *Ehretia dicksonii* early flower development. *A–F*, *Bourreria succulenta*. *A*, Sepal primordia initiation. *B*, Petal primordia initiation. *C*, Stamen primordia initiation. *D*, Early corolla and androecium development. *E*, Corolla and androecium shortly before formation of the stamen–corolla tube. *F*, Carpel primordia initiation. *G–L*, *Ehretia dicksonii*. *G*, Sepal primordia initiation. *H*, Calyx development before petal primordia initiation. *I*, Petal primordia and stamen primordia initiation. *J*, Early corolla and androecium development. *K*, Formation of the stamen–corolla tube. *L*, Carpel primordia initiation. C = carpel; FA = floral apex; P = petal; S = sepal; St = stamen. Scale bars = 100 μm.

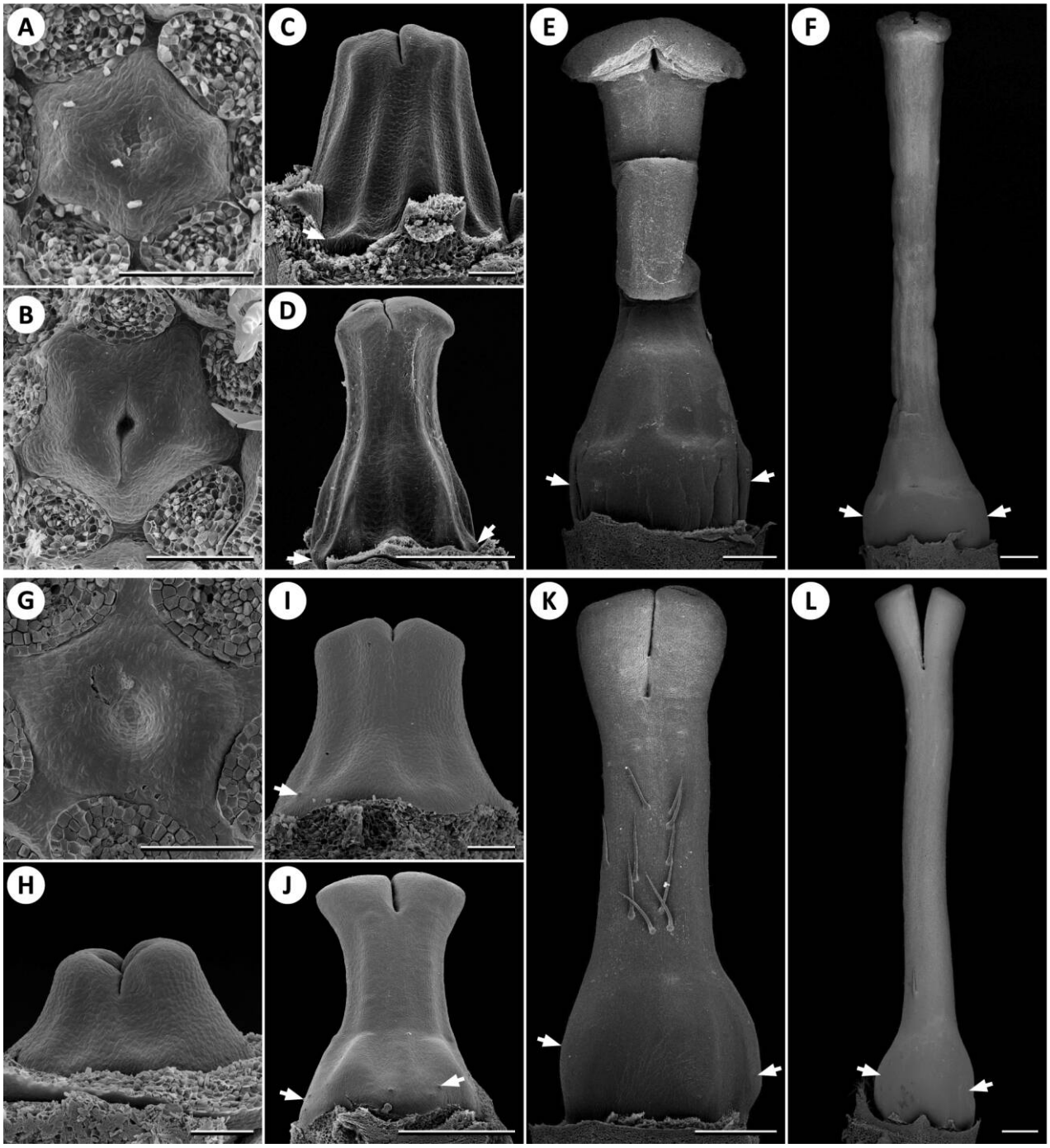


Fig. 8 *Bourreria succulenta* and *Ehretia dicksonii* external gynoecium development. *A–F*, *Bourreria succulenta*. *A*, Gynoecium after carpel initiation. *B*, Gynoecial cone. *C*, Gynoecial cone before differentiation into style and ovary. *D*, Young gynoecium differentiated into stigma, style, and ovary. *E*, Gynoecium of a preanthetic flower. Note the fused stigmatic surfaces and the buckled style. *F*, Gynoecium of an anthetic flower. *G–L*, *Ehretia dicksonii*. *G*, Gynoecium after carpel initiation. *H*, Gynoecial cone. *I*, Gynoecial cone before differentiation into style and ovary. *J*, Young gynoecium differentiated into stigma, style, and ovary. Note the initiation of stylodia development. *K*, Gynoecium of a preanthetic flower. *L*, Gynoecium of an anthetic flower. Arrows point at nectarostomata. Scale bars = 100 μm (*A–C*, *G–I*), 500 μm (*D–F*, *J–L*).

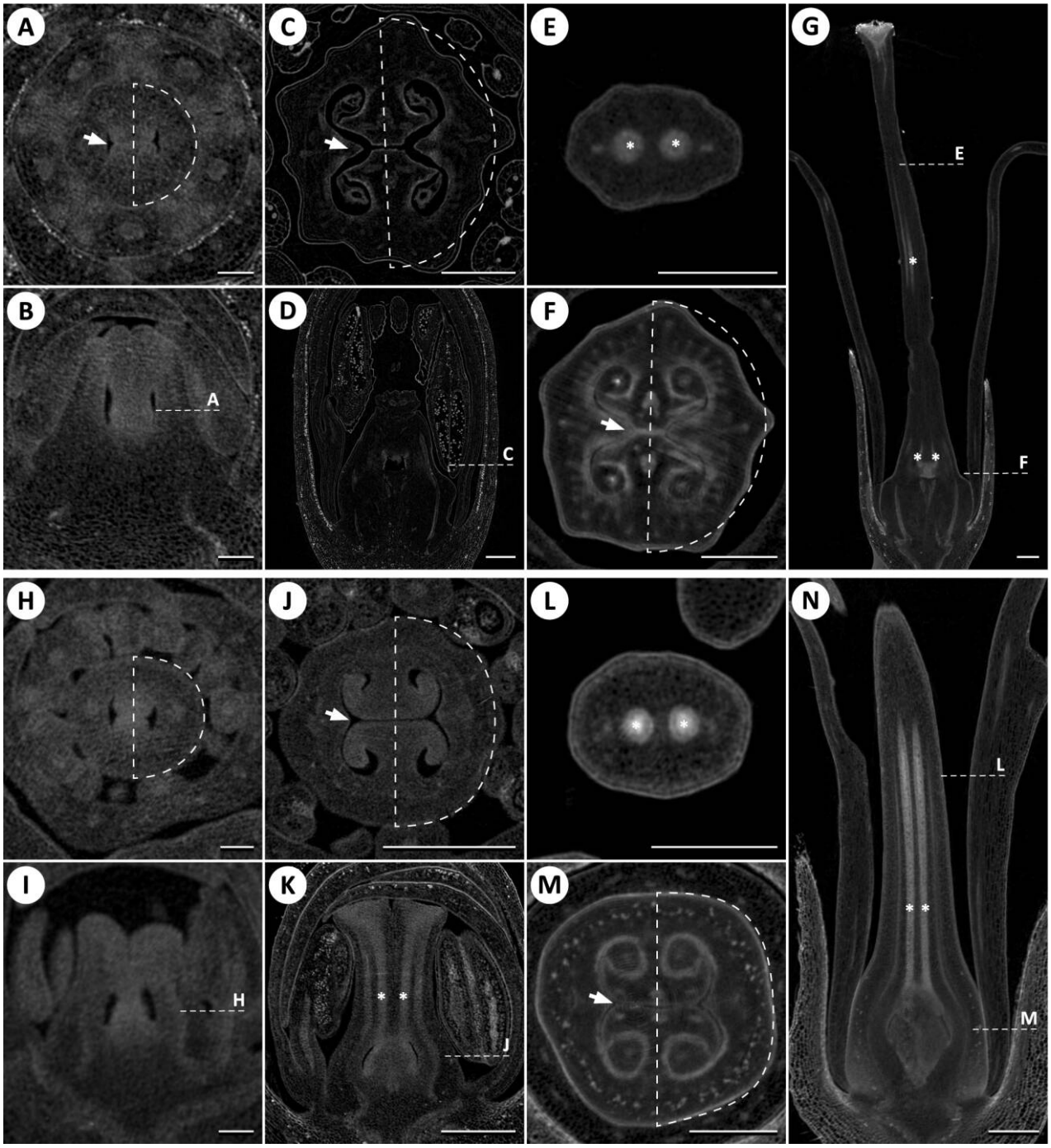


Fig. 9 *Bourreria succulenta* and *Ehretia dicksonii* internal gynoecium development. Virtual longitudinal and cross sections based on micro–computed tomography scans. *A–G*, *Bourreria succulenta*. *A*, *B*, Gynoecial cone. *A*, Cross section at the level of ovule initiation. *B*, Longitudinal section. *C*, *D*, Gynoecium of a preanthetic flower. *C*, Cross section. *D*, Longitudinal section. *E–G*, Gynoecium of an anthetic flower. *E*, Cross section through the style. *F*, Cross section through the gynoecium at the level of the ovules. *G*, Longitudinal section. *H–N*, *Ehretia dicksonii*. *H*, *I*, Gynoecial cone. *H*, Cross section at the level of ovule initiation. *I*, Longitudinal section. *J*, *K*, Gynoecium of a preanthetic flower after differentiation into stigma, style, and ovary. *J*, Cross section. *K*, Longitudinal section. *L–N*, Gynoecium of an anthetic flower. *L*, Cross section through the style. *M*, Cross section through the gynoecium at the level of the ovules. *N*, Longitudinal section. Asterisks indicate the pollen tube transmitting tissue. Arrows point at false septa. The section of a circle with a dashed outline highlights a carpel. Dashed lines labeled with an uppercase letter refer to the corresponding image of this figure. Scale bars = 100 μm (*A*, *B*, *H*, *I*), 500 (*C–G*, *J–N*).

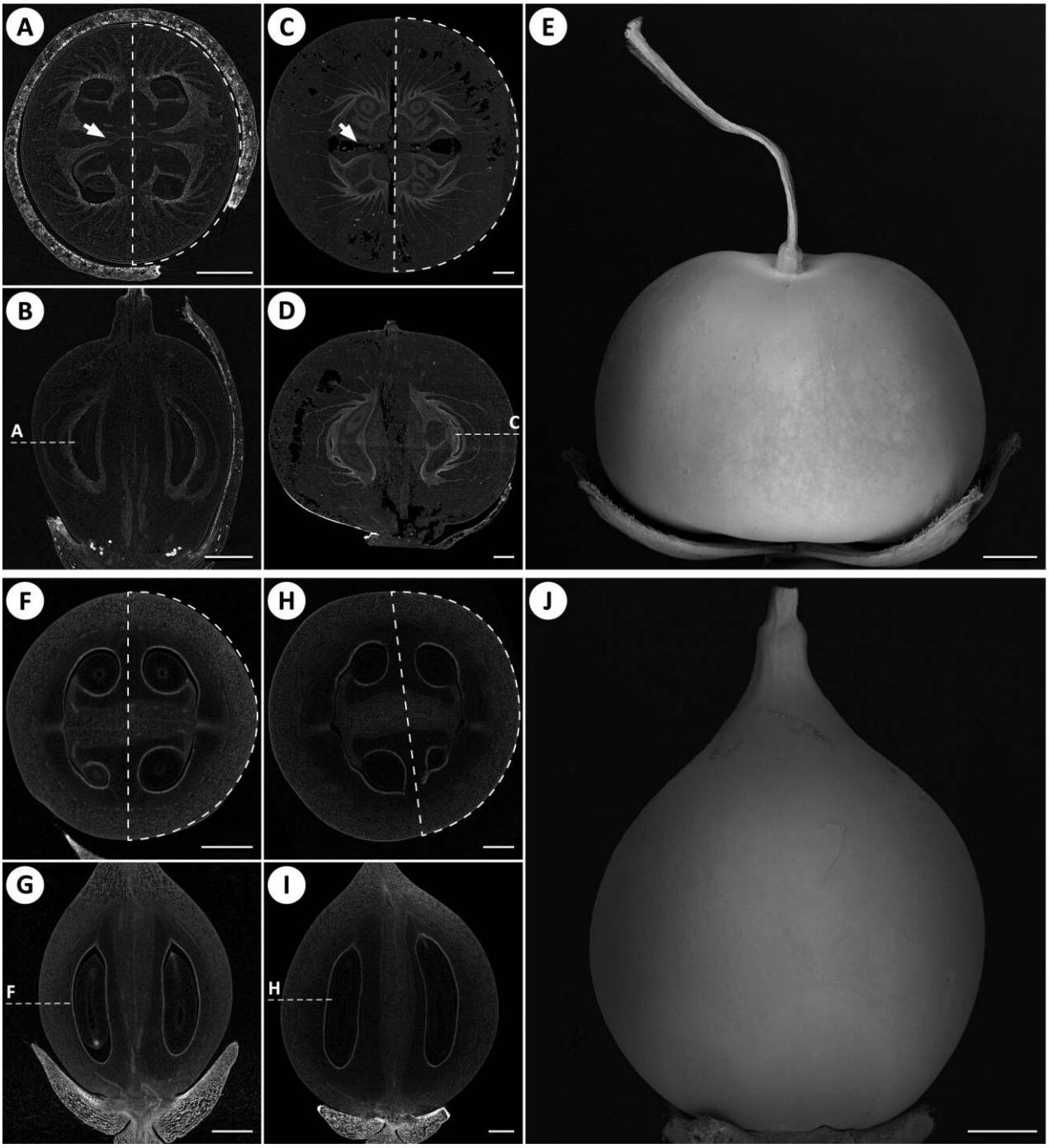


Fig. 10 *Bourreria succulenta* and *Ehretia dicksonii* fruit development. Virtual longitudinal and cross sections based on micro–computed tomography scans. *A–E*, *Bourreria succulenta*. *A*, *B*, Young fruit. *A*, Cross section. *B*, Longitudinal section. *C*, *D*, Mature fruit. *C*, Cross section. *D*, Longitudinal section. *E*, SEM micrograph of a mature fruit. *F–J*, *Ehretia dicksonii*. *F*, *G*, Young fruit. *F*, Cross section. *G*, Longitudinal section. *H*, *I*, Mature fruit. *H*, Cross section. *I*, Longitudinal section. *J*, SEM micrograph of a mature fruit. Arrows point at false septa. The section of a circle with a dashed outline highlights a carpel. Dashed lines labeled with an uppercase letter refer to the corresponding image of this figure. Scale bars = 1 mm.

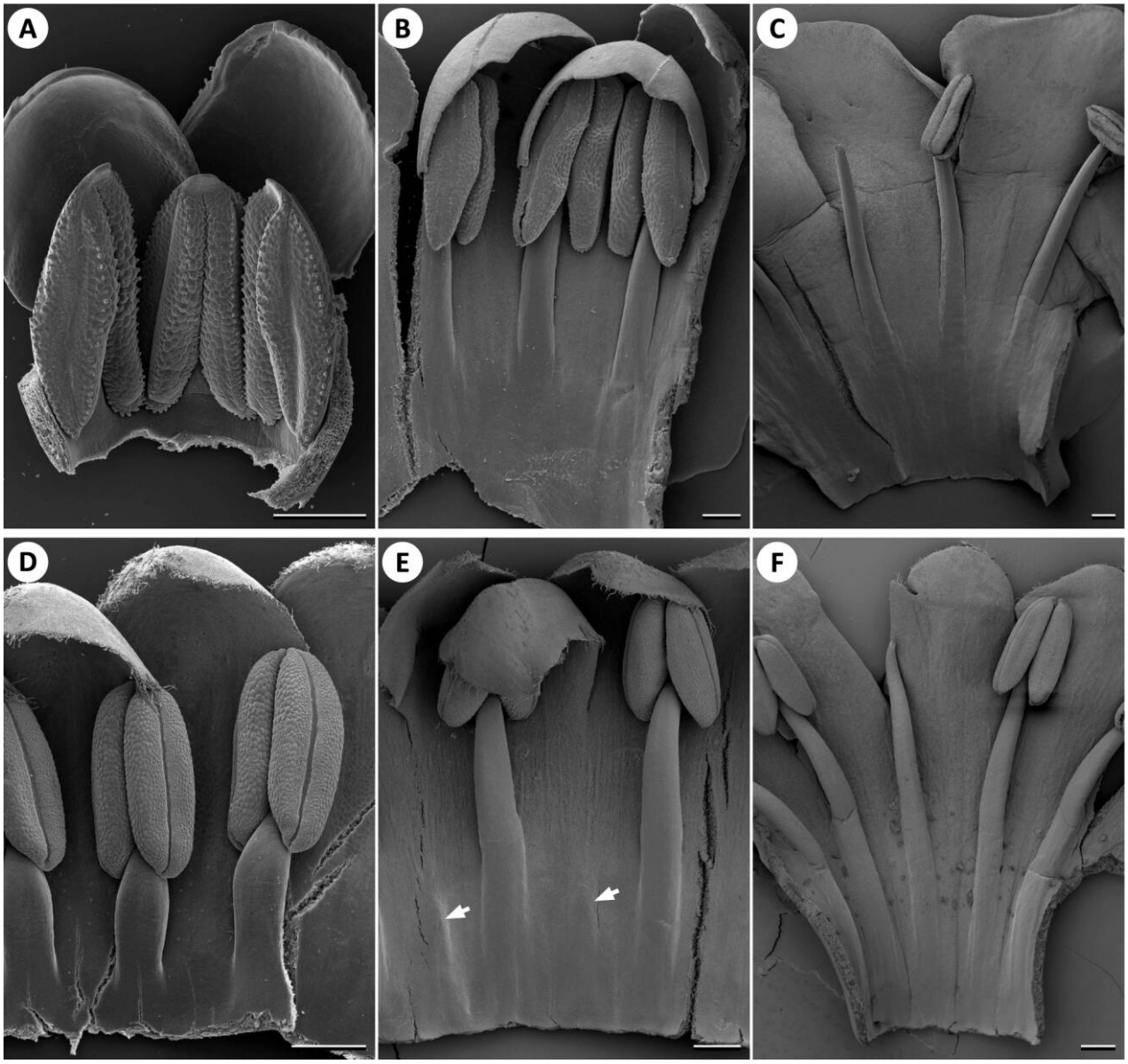


Fig. 11 *Bourreria succulenta* and *Ehretia dicksonii* stamen–corolla tube development. *A*–*C*, *Bourreria succulenta*. *A*, Young stamen–corolla tube. *B*, Stamen–corolla tube of a preanthetic flower. *C*, Stamen–corolla tube at anthesis. *D*–*F*, *Ehretia dicksonii*. *D*, Young stamen–corolla tube. *E*, Stamen–corolla tube of a preanthetic flower. *F*, Stamen–corolla tube at anthesis. Arrows point at protrusions. Scale bars = 500 μm.

stamen–corolla tube. Only *E. dicksonii* shows shallow ridges alternating with the stamens in intermediate stages of flower development (fig. 11*E*). Anthers are introrse to latrorse in *B. succulent* and latrorse in *E. dicksonii*. They are exserted at anthesis.

Discussion

General flower and fruit development. The three species studied here follow similar developmental trajectories in line with observations of other representatives of Ehretiaceae and Lennoaceae (Suessenguth 1927; Copeland 1935; Gottschling 2004; Gottschling et al. 2014). The flowers are relatively simple, and no specialized structures or floral architectures have been observed. The ovaries of the three species share the raised position of the ovules, which is most prominent in *Pholisma arenarium* and *Bourreria succulenta*, where the ovules can be found in the upper third of the ovary, while the nectary is located in the lower half. This phenomenon reaches an extreme expression in the genus *Tiquilia* where the ovule-carrying part of the ovary is lifted by a stalk (gynophore) with a gynoecial nectary disk at its base (Gottschling et al. 2014). Modifications of the stamen–corolla tube similar to those of other families of the Boraginales (e.g., Boraginaceae, Codonaceae, Hydrophyllaceae, Namaceae; Hofmann 1999; Hofmann et al. 2016; Jeiter et al. 2016, 2020; Jeiter and Weigend 2018; Vasile et al. 2022) have

not been observed in the three taxa studied here. Only shallow protrusions alternating with the stamens are observed during flower development in *Ehretia dicksonii*, which are again lost during further development. In some species of *Bourreria* different types of simple stamen–corolla tube modifications have been reported (Gottschling and Miller 2018), but these appear not to be structurally equivalent to modifications in other Boraginales. The overall floral architecture (sensu Endress 1996; definition modified by Jeiter et al. 2017; Jeiter and Weigend 2018) of the species studied here can be considered simple and nonspecialized. Flowers of the investigated species can be regarded as simple with infundibuliform to campanulate corollas, in terms of floral modifications and architecture (sensu Endress 1996; Jeiter et al. 2017; Jeiter and Weigend 2018).

Most noteworthy are the variations in merism, which are frequently observed in flowers of all three species. Variation within one plant was observed in all three species and between different plants in *P. arenarium*. This might indicate variability resulting from an increased floral apex size (Ronse De Craene 2016). The variation in *P. arenarium* is likely related to an increase in carpel number. Since the order of organ initiation is centripetal, we believe that there is a strong selection for an increase in carpel number and consequently ovule number, while the increase of the outer floral organs is not strongly selected for or against and thus is more variable. The gynoecium consists of two carpels in most Boraginales, even if the outer whorls show lower (*Wellstedia*; Hilger and Weigend 2016) or higher (some *Cordia* species [Gottschling et al. 2016]; *Codon* [Jeiter et al. 2016]) organ numbers. There are only two taxa in Boraginales known to routinely have more than two carpels: Lennoaceae (Suessenguth 1927; Copeland 1935; Yatskievych and Mason 1986; Bittrich 2016) and some species of the enigmatic genus *Trigonotis*, deeply nested in Boraginaceae (Weigend et al. 2016). We believe that the increase of seeds per flower in Lennoaceae is related to the parasitic lifestyle in the family.

Fruit evolution in Lennoaceae. Fruits of *Ehretia*, *Bourreria*, and *Pholisma* are not that different. The main differences are the number of carpels, the structure of the pyrene wall, and the actual dispersal unit. The probably endochorous fruits of *E. dicksonii* and *B. succulenta* are very similar and differ only in the structure of the pyrene wall, which is smooth in *E. dicksonii* and finned in *B. succulenta*. The fruit of *P. arenarium*, on the other hand, consists of numerous carpels and twice as many pyrenes as carpels. In contrast to *E. dicksonii* and *B. succulenta*, the fruit dries and splits open irregularly around its equator (fig. 5*D*), where the pyrenes are located. The lid of this fruit detaches only if some kind of force is involved (e.g., wind or ants). While the lid detaches, some pyrenes may be dispersed with the lid or fall out, but usually, most or all of the pyrenes remain fixed to the basal part of the fruit (fig. 5*I*), which is lignified and remains firmly attached on the infructescence. This fruit type is unique within Ehretiaceae and, to our knowledge to date, also within Boraginales. All other Ehretiaceae have drupes with one to four pyrenes, or schizocarps (Gottschling et al. 2016). Schizocarp development appears to be similar to drupe development until anthesis and in early postanthetic developmental stages. The development of the schizocarp seems then to be associated with a drying of the exo- and mesocarp enclosing the endomericarpids (Gottschling et al. 2014). In the greenhouse we occasionally observed fruits of *B. succulenta* where the pulp around the pyrenes had been consumed (probably by blackbirds or robins) and only the finned pyrenes remained on the infructescence. This indicates that the pyrenes, despite their enlarged, finned surface, are not firmly connected to the fleshy mesocarp.

Holoparasitic plants entirely depend on a host for their development. Strategies to increase the chances of host infection are to take a generalist approach in host choice, to develop highly specialized dispersal mechanisms (e.g., Viscaceae; Watson and Rawsthorne 2013), or to drastically increase diaspore number (Mescher et al. 2009). Of course, combinations of these strategies exist. This is clearly the case in *P. arenarium*, which has a relatively broad host range (four host species mentioned by Yatskievych [1985] and at least eight host species documented by Yatskievych and Mason [1986]) and many very small diaspores. Baskin and Baskin (2022) classify the diaspores of *P. arenarium* as "dust seeds" with an under- or undifferentiated embryo and a seed size of less than 2 mm but usually smaller than 1 mm. The massive increase of flower number in the inflorescences of *P. arenarium* and the increase in seed number per flower compared with the other Ehretiaceae studied can be considered strategies to greatly increase diaspore number as an adaptation to holoparasitism. Dispersal by ants has been suggested (Yatskievych 1985), and a separation of the complete infructescence and a tumbleweed mechanism have been reported for the related *Pholisma sonorae* (Bittrich 2016), but the relative incidence of these dispersal modes and their absolute importance remain unknown. The infructescences of *P. arenarium* are often found more or less intact even when the next vegetation period has started and may represent an aboveground seed bank, with dispersal possibly delayed until the onset of winter rains.

The integration of pericarp tissue and seed coat, as found in most Boraginales (e.g., Diane et al. 2016; Weigend et al. 2016; Jeiter et al. 2018), likely represents a (highly successful) evolutionary dead end: an increase of ovule or seed number per carpel is not known in any of the nutlet or drupe-producing taxa of Boraginales. It is observed only in those early-branching families of the order with capsular fruits (Vasile et al. 2021). Apparently, given a sufficiently strong selective force such as holoparasitism, one possible departure from this dead end is an increase in carpel number but at the cost of an increased floral apex size and concomitant variability of the floral organ number. In capsule-fruited Boraginales, such as Hydrophyllaceae, evolutionary changes of ovule number per carpel are related to prolonged or truncated ovary development and the insertion (or not) of additional tiers of seeds in the ovary (Vasile et al. 2021). This path to diaspore multiplication was not available to Ehretiaceae because of the integration of seeds and pericarp in the biovulate carpels.

Systematic implications of flower and fruit development. The phylogenetic placement of Lennoaceae has been controversially discussed, and the well-supported position of Lennoaceae within Boraginales has been elusive (Gottschling et al. 2014; Weigend et al. 2014; Stull et al. 2015). A recent study using transcriptome data finally placed Lennoaceae (represented by *P. arenarium* and *Lennoa madreporoides*) in Ehretiaceae (represented by *E. dicksonii*, *Ehretia microphylla* Lam., *Ehretia tsangii* I.M.Johnst., and *Bourreria ovata* [syn. *B. succulenta*]) with maximal statistical support (Zhang et al. 2020).

The presence of separate strands of transmitting tissue in the style and the apparent absence of a stylar compitum are characters shared by all species studied here and—to our knowledge—present in all Ehretiaceae that have previously been investigated (Gottschling 2004; Gottschling et al. 2014). The separate strands of pollen tube transmitting tissue were described from Lennoaceae but not recognized as such (Suessenguth 1927; Copeland 1935). A compitum appears to be present in the ovarian cavity, which is reduced to a narrow slit in *E. dicksonii* and *B. succulenta*. It remains unclear whether a functional compitum is present in *P. arenarium*. However, our data find a striking similarity in floral morphology between *B. succulenta* and *P. arenarium*, while *E. dicksonii* seems more divergent. The size and shape of the false septa and the location of the ovules in the upper third of the ovary, as well as the fused stigmatic surfaces and thus the formation of an extragynoecial compitum, underscore a particularly close relationship between *Pholisma* and *Bourreria*. Clearly, a wider sampling in Ehretiaceae and Lennoaceae would be highly desirable to support this finding, but surprisingly, the data presented here provide morphological, anatomical, and developmental support for the idea that Lennoaceae is indeed be nested within Ehretiaceae.

Conclusions

We propose that the main selective pressure in fruit evolution in Lennoaceae is their holoparasitic lifestyle, favoring an increase of diaspore number and adjustment of fruit type. We propose that the increased carpel number has a profound effect on flower morphology, leading to an increased and variable number of floral organs, changes that may be additionally modified because of the condensation and resulting mechanical forces on development of the many-flowered inflorescences (Bull-Hereñu et al. 2022).

Furthermore, our findings regarding flower and fruit development and morphology are in line with the molecular phylogenetic placement of Lennoaceae within Ehretiaceae. Especially the absence of a stylar compitum and the formation of endomericarpids tie both taxa together. We here demonstrate the importance of investigating the morphology, anatomy, and ontogeny of characters in the context of a molecular phylogenetic framework. Investigations like this might not only provide additional support for a surprising phylogenetic placement of a given taxon but also help to understand the evolution of aberrant morphologies (e.g., in relation to parasitism within otherwise morphologically conservative taxa).

Acknowledgments

We thank B. Lee and D. Grossenbacher for contributing the samples of *Pholisma arenarium* and E. LoPresti for establishing the contact. B. Lee also took the images of *P. arenarium* displayed in figure 1. We thank E. Smets for valuable input on the manuscript. We are grateful to T. Joßberger, Y. Malkowsky, J. Vehof, and D. Wenzel for technical support. Funding for the Skyscan 1272 micro–computed tomography was provided by the Deutsche Forschungsgemeinschaft (INST 217/849-1 FUGG). The entire plant of *P. arenarium* was scanned at the Zoological Research Museum Alexander König, Bonn, with permission of B. Wipfler. We thank the staff of Bonn University Botanic Gardens for cultivating *Bourreria succulenta* and *Ehretia dicksonii*.

Literature Cited

Baskin JM, CC Baskin 2022 Germination and seed/embryo size in holoparasitic flowering plants with "dust seeds" and an undifferentiated embryo. Bot Rev 88:1–49.

Bellot S, SS Renner 2013 Pollination and mating systems of Apodanthaceae and the distribution of reproductive traits in parasitic angiosperms. Am J Bot 100:1083–1094.

Bittrich V 2016 Lennoaceae. Pages 257–261 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Bull-Hereñu K, P dos Santos, JFG Toni, JHL El Ottra, P Thaowetsuwan, J Jeiter, LP Ronse De Craene, A Iwamoto 2022 Mechanical forces in floral development. Plants 11:661.

Carlquist S, CM Guilliams 2017 Distinctive wood anatomy of the root-parasitic family Lennoaceae (Boraginales). IAWA J 38:3–12.

Copeland HF 1935 The structure of the flower of *Pholisma arenarium*. Am J Bot 22:366–383.

Diane N, HH Hilger, H Förther, M Weigend, F Luebert 2016 Heliotropiaceae. Pages 203–211 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Duncan KE, KJ Czymmek, N Jiang, AC Thies, CN Topp 2022 X-ray microscopy enables multiscale high-resolution 3D imaging of plant cells, tissues, and organs. Plant Physiol 188:831–845.

Endress PK 1996 Homoplasy in angiosperm flowers. Pages 303–325 *in* MJ Sanderson, L Hufford, eds. Homoplasy: the recurrence of similarity in evolution. Academic Press, San Diego.

Gottschling M 2004 Floral ontogeny in *Bourreria* (Ehretiaceae, Boraginales). Flora 199:409–423.

Gottschling M, F Luebert, HH Hilger, JS Miller 2014 Molecular delimitations in the Ehretiaceae (Boraginales). Mol Phylogenet Evol 72:1–6.

Gottschling M, JS Miller 2018 A taxonomic revision of the New World species of *Bourreria* (Ehretiaceae, Boraginales). Syst Bot Monogr 103:1–100.

Gottschling M, S Nagelmüller, HH Hilger 2014 Generative ontogeny in *Tiquilia* (Ehretiaceae: Boraginales) and phylogenetic implications. Biol J Linn Soc 112:520–534.

Gottschling M, M Weigend, HH Hilger 2016 Ehretiaceae. Pages 165–178 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Hilger HH, M Weigend 2016 Wellstediaceae. Pages 403–406 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Hofmann M 1999 Flower and fruit development in the genus *Phacelia* (Phacelieae, Hydrophyllaceae): characters of systematic value. Syst Geogr Plants 68:203–212.

Hofmann M, GK Walden, HH Hilger, M Weigend 2016 Hydrophyllaceae. Pages 221–238 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Jeiter J, F Danisch, HH Hilger 2016 Polymery and nectary chambers in *Codon* (Codonaceae)—flower and fruit development in a small, capsule-bearing family of Boraginales. Flora 220:94–102.

Jeiter J, HH Hilger, EF Smets, M Weigend 2017 The relationship between nectaries and floral architecture: a case study in Geraniaceae and Hypseocharitaceae. Ann Bot 120:791–803.

Jeiter J, S Langecker, M Weigend 2020 Towards an integrative understanding of stamen-corolla tube modifications and floral architecture in Boraginaceae s.s. (Boraginales). Bot J Linn Soc 193:100–124.

Jeiter J, YM Staedler, J Schönenberger, M Weigend, F Luebert 2018 Gynoecium and fruit development in *Heliotropium* sect. *Heliothamnus* (Heliotropiaceae). Int J Plant Sci 179:275–286.

Jeiter J, M Weigend 2018 Simple scales make complex compartments—ontogeny and morphology of stamen-corolla tube modifications in Hydrophyllaceae (Boraginales). Biol J Linn Soc 125:802–820.

Luebert F, L Cecchi, MW Frohlich, M Gottschling, CM Guilliams, KE Hasenstab-Lehman, HH Hilger, et al. 2016 Familial classification of the Boraginales. Taxon 65:502–522.

Mescher MC, J Smith, CM De Moraes 2009 Host location and selection by holoparasitic plants. Pages 101–118 *in* F Baluška, ed. Plant-environment interactions: signaling and communication in plants. Springer, Berlin.

Nickrent DL 2020 Parasitic angiosperms: how often and how many? Taxon 69:5–27.

Nickrent DL, JP Der, FE Anderson 2005 Discovery of the photosynthetic relatives of the "Maltese mushroom" *Cynomorium*. BMC Evol Biol 5:38.

Ronse De Craene L 2016 Meristic changes in flowering plants: how flowers play with numbers. Flora 221:22–37.

Schneider AC, T Braukmann, A Banerjee, S Stefanović 2018 Convergent plastome evolution and gene loss in holoparasitic Lennoaceae. Genome Biol Evol 10:2663–2670.

Staedler YM, D Masson, J Schönenberger 2013 Plant tissues in 3D via X-ray tomography: simple contrasting methods allow high resolution imaging. PLoS ONE 8:e75295.

Stull GW, RD de Stefano, DE Soltis, PS Soltis 2015 Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. Am J Bot 102:1794–1813.

Suessenguth K 1927 Über die Gattung *Lennoa*: Ein Beitrag zur Kenntnis exotischer Parasiten. Flora Allg Bot Ztg 122:264–305.

Teixeira-Costa L, CC Davis 2021 Life history, diversity, and distribution in parasitic flowering plants. Plant Physiol 187:32–51.

Vasile M-A, F Luebert, J Jeiter, M Weigend 2021 Fruit evolution in Hydrophyllaceae. Am J Bot 108:925–945.

Vasile M-A, M Weigend, A Ziegler, J Jeiter 2022 Flower and fruit development of three species of Hydrophylleae sheds new light on flower evolution in Hydrophyllaceae. Int J Plant Sci 183:331–347.

Watson DM, J Rawsthorne 2013 Mistletoe specialist frugivores: latter-day "Johnny Appleseeds" or self-serving market gardeners? Oecologia 172:925–932.

Weigend M, HH Hilger 2010 Codonaceae—a newly required name in Boraginales. Phytotaxa 10:26–30.

——— 2016 Codonaceae. Pages 137–141 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Weigend M, F Luebert, M Gottschling, TLP Couvreur, HH Hilger, JS Miller 2014 From capsules to nutlets—phylogenetic relationships in the Boraginales. Cladistics 30:508–518.

Weigend M, F Selvi, DC Thomas, HH Hilger 2016 Boraginaceae. Pages 41–102 *in* JW Kadereit, V Bittrich, eds. Families and genera of vascular plants. Vol 14. Springer, Heidelberg.

Wicke S, J Naumann 2018 Molecular evolution of plastid genomes in parasitic flowering plants. Pages 315–347 *in* S-M Chaw, RK Jansen, eds. Advances in botanical research. Vol 85. Elsevier, Amsterdam.

Yatskievych G 1985 Notes on the biology of the Lennoaceae. Cactus Succul J 57:73–79.

Yatskievych G, CT Mason 1986 A revision of the Lennoaceae. Syst Bot 11:531–548.

Zhang C, T Zhang, F Luebert, Y Xiang, C-H Huang, Y Hu, M Rees, et al. 2020 Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. Mol Biol Evol 37:3188–3210.

Unravelling Boraginales: an updated phylogeny based on the Angiosperms353 probe set



Chapter 6

Trichodesma boissieri Post (Boraginaceae) in cultivation in Bonn Botanical Gardens.

Vasile MA, Böhnert T., Jeiter J., Weigend M. Unravelling Boraginales: an updated phylogeny based on the Angiosperms353probe set. To be submitted.

Unravelling Boraginales: an updated phylogeny based on the Angiosperms353 probe set

Maria-Anna Vasile[1], Tim Böhnert[1], Julius Jeiter[1], Maximilian Weigend[1]

[1] *Nees Institute for Biodiversity of Plants, University of Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany*

Address for correspondence: *Maria-Anna Vasile, m.vasile@uni-bonn.de*

Abstract

Boraginales, a cosmopolitan and highly diverse order comprising 11 families and about 2700 species, was studied utilizing the Angiosperms353 probe set to reconstruct the phylogenetic relationships. We obtained phylogenomic data from 208 samples representing 10 Boraginales families and 79 % of ca. 133 accepted genera. The study confirms and expands upon previous findings, revealing well-resolved species trees with consistently high support. The topology of previous studies is largely recovered, but there are several novelties: Ehretiaceae are identified as paraphyletic, and we therefore propose the inclusion of Lennoaceae; Codonaceae are shown for the first time to be sister to the remaining families of Boraginales II when trees are reconstructed using the exons and the introns, while the exonic dataset supports their sister relationship to Wellstediaceae and Boraginaceae (Boraginales I). Some uncertainties remain regarding the sister relationship of Namaceae and Hydrophyllaceae and within Boraginaceae regarding the intergeneric relationships in Cynoglosseae. These are likely to be resolved by strategically expanding sampling. Overall, the results highlight the effectiveness of the target enrichment probe set for understanding the evolutionary history of Boraginales and pave the way for detailed studies on the evolution of group-specific morphological traits.

Keywords: Asterids; Hyb- Seq; Lamiids; molecular phylogenetics; systematics; target sequence capture; tree of life.

6.1 INTRODUCTION

Boraginales are a medium size order of angiosperms within Lamiids (Zhang & al., 2020) that is fascinating for its diversity. It comprises 11 families with ca. 125 genera and 2700 species (Luebert & al., 2016). Predominantly found in seasonally arid habitats across temperate and tropical regions, this order has an otherwise worldwide distribution and shows remarkable morphological diversity (Fig. 6.1; Weigend & al., 2014; Jeiter, 2020). Similar to other representatives of asterids, Boraginales have generally conserved floral organization (i.e. merosity of the different organ whorls and their arrangement; Jeiter, 2020) with mostly tetracyclic, sympetalous, and pentamerous flowers, mainly actinomorphic and often arranged in scorpioid inflorescences (Buys & Hilger, 2003). However, slight modifications of floral organs lead to variable and frequently complex floral architecture, suggesting an array of pollination modes. For instance, buzz pollination is evident in some species of Boraginaceae (Teppner, 2011) and Ehretiaceae (Holstein & Gottschling, 2018), whereas others exhibit revolver architecture (Jeiter & al., 2016; Jeiter & Weigend, 2018; Vasile & al., 2022). Their fruit morphology, varying from simple seeds to intricate structures often exhibiting different degree of seed and pericarp integration

(mericarpids, endomericarpids, synaptospermy), has also attracted significant attention from researchers both in phylogenetics and systematics (Diane & al., 2002; Gottschling & Hilger, 2004; Selvi & al., 2006; Weigend & Hilger, 2010; Hasenstab-Lehman & Simpson, 2012; Weigend & al., 2013; Gottschling & al., 2014a; Weigend & al., 2014; Diane & al., 2016; Luebert & al., 2016; Otero & al., 2019) and in morphology (Hilger, 1987, 1992; Di Fulvio & al., 1997, 1999; Gottschling & al., 2014b; Hilger, 2014, 2014; Jeiter & al., 2018; Heigl & al., 2020; Vasile & al., 2021; Jeiter & al., 2023).

Historically, Boraginales classification has been in flux, largely due to molecular data's influence on our understanding of morphological characters, especially regarding the multiple fruit type transitions within both major clades of the order (Boraginales I and II; Fig. 6.2). The taxonomic voyage of Boraginales has seen it oscillate from being classified as a family (Boraginaceae) to its subsequent elevation to an order. Earlier classifications often shifted groups between family and subfamily levels (De Candolle, 1845, 1846; Bentham & Hooker, 1876; Gürke, 1893; Takhtajan, 1987, 1997), while taxonomists consistently working with Boraginales have been treating the subdivisions of the order as families (e.g. Brand, 1913; Atwood, 1975; Hilger & Diane, 2003; Gottschling & al., 2004; Selvi & al., 2006; Luebert & Wen, 2008; Chacón & al., 2016). Recently APG IV (2016) formally recognized Boraginales as an order, but still treats them as one family (Boraginaceae s.l) with several subfamilies. Yet, disagreements persist, with some taxonomists proposing a finer subdivision of the order into distinct families, which can accommodate most of the fruit diversity (Luebert & al., 2016). Current divisions identify two main clades based on chloroplast data (Weigend & al., 2014), Boraginales I and II (Fig. 6.2). Each clade comprises a basal grade of two small to medium-sized families having capsules (Boraginales I: Codonaceae and Wellstediaceae; Boraginales II: Hydrophyllaceae and Namaceae). The majority of the remaining families of each clade (Boraginales I: Boraginaceae; Boraginales II: Ehretiaceae, Lennoaceae, Heliotropiaceae, Cordiaceae, Coldeniaceae, Hoplestigmataceae) appear to have a reduced ovule number to four and drupes or nutlets of either eremocarpic or schizocarpic origin. Boraginales, with an age estimate for the stem node of ca. 109 Ma, likely originated in West Gondwana during the Early Cretaceous, and its two major clades diverged during the Late Cretaceous (Luebert & al., 2017). Africa and South America are the most likely ancestral areas for Boraginales and an initial West Gondwana vicariance event as an explanation for the origin of the two major clades has been inferred (Luebert & al., 2017).

The past century, Boraginales systematics and consequently their classification was heavily reliant on morphological traits, most notably gynoecium and fruit morphology. Such classifications, while invaluable, have often presented challenges due to the dynamic nature of the group, the overlapping of some defining traits or the convergent evolution of similar traits in different groups. The interfamilial dynamics within Boraginales, though not as convoluted, present their own set of challenges. The demarcation of boundaries between families, their evolutionary trajectories, and their placement within the order are areas still rife with questions. Presently, while there is a general consensus about Boraginales standing as an order, the finer details about its systematics, especially at the family and subfamily levels, remain a subject of debate and investigation.

Systematic relationships within Boraginales I

The interfamilial relationships of Boraginales I are resolved based on chloroplast data, with Codonaceae being successive sister to Wellstediaceae and Boraginaceae (Fig. 6.2; Weigend & al., 2014), even though Codonaceae used be considered more closely related to Hydrophyllacae s. l. based on morphology (Brand, 1913; Weigend & Hilger, 2010). An updated infrafamilial classification for Boraginaceae as the largest family with about 1600-1700 species (Luebert & al., 2016) was recently provided by



Figure 6.1 Morphological diversity in Boraginales. A, Codonaceae (*Codon royenii* L., photo: J. Jeiter); B, Wellstediaceae (*Wellstedia cf. dinteri* Pilg, photo and collection: Q. Dietmar & J. Bechteler, SFB963 Namibia); C-N, Boraginaceae, C, Echiochiloideae (*Echiochilon fruticosum* Desf., xx-0-BONN-15043); Borageineae (*Anchusa azurea* Mill., photo: M. Vasile, Greece); E, Lithospermeae (*Echium arenarium* Guss., photo: M. Vasile, Greece); F,

Lithospermeae (*Cerinthe major* L., photo: M. Vasile, Greece); G, Trichodesmeae (*Trichodesma scottii* Balf.f., YE-0-BONN-14450); H, Omphalodeae (*Omphalodes nitida* (Willd.) Hoffmanns. & Link, xx-0-BONN-14659); I, Rochelieae, Heterocaryinae (*Heterocaryum szovitsianum* (Fisch. & C.A.Mey.) A.DC., xx-0-BONN-15128); J, Rochelieae, Eritrichiinae (*Lappula consanguinea* (Fisch. & C.A.Mey.) Gürke., xx-0-BONN-39580); K, Myosotideae (*Trigonotis formosana* Hayata, xx-0-BONN- 29917); L, Cynoglossinae (*Cynoglossum creticum* Mill., photo: M. Vasile, Greece); M, Cynoglossinae (*Lindelofia longiflora* (DC.) Baill. xx-0-BONN-27824); N, Amsinckiinae (*Plagiobothrys chorisianus var. hickmanii* (Greene) I.M.Johnst., xx-0-BONN-27824); O, Namaceae (*Wigandia urens* (Ruiz & Pav.) Kunth , photo and collection: M. Weigend COL-06, Kolumbien); P, Hydrophyllaceae (*Phacelia bolanderi* A.Grey, xx-0-BONN-40976); Q, Ehretiaceae (*Ehretia dicksonii* Hance, xx-0-BONN-12602); R, Lennoaceae (*Pholisma arenarium* Nutt., photo and collection: B. Lee & D. Grossenbacher 1, California USA); S, Heliotropiaceae (*Heliotropium europaeum* L., DE-0-BONN-26153); T, Cordiaceae (*Cordia sebestena* L., photo: M. Ackermann, Cuba).

Chacón et al. (2016), however, some relationships remain unresolved. Nuclear alignments (ITS1 and ITS2 regions) resulted in a poorly resolved backbone, with Core-Cynoglosseae retrieved as an unresolved polytomy (Chacón & al., 2016). On the contrary, chloroplast data yielded an overall well-resolved backbone, but not always with maximum support, either at subfamily level (i.e. the sister relationship of Cynoglossoideae to Boraginoideae) or at tribal and subtribal level (Chacón & al., 2016). Within the expansive realm of Boraginales, the family Boraginaceae stands out, both in terms of its size and complexity. It comprises the lion's share of genera and species in the order. Systematic relationships within the family are characterized by a tapestry of clades, each with its own unique evolutionary tale. Yet, relationships between these clades and their placements within Boraginaceae remain incompletely resolved (see previous studies; Hilger & al., 2004; Selvi & al., 2006; Thomas & al., 2008; Weigend & al., 2009; Hasenstab-Lehman & Simpson, 2012; Nazaire & Hufford, 2012; Weigend & al., 2014; Chacón & al., 2016, 2019), with some clades consistently problematic in phylogenetic analyses (e.g. tribal and subtribal classification of Cynoglossoideae) and some genus limits completely uncertain (e.g. *Cynoglossum* L.).

Systematic relationships within Boraginales II

At a higher level, the bulk of the open questions with regards to classification are found in Boraginales II, with three main phylogenetic problems. Starting from the basal grade of Boraginales II, Namaceae and Hydrophyllaceae were only recently resolved as sister groups with maximum support based on whole transcriptome sequencing data, but with rather very limited sampling and without including *Nama* L. (Zhang & al., 2020). Traditionally, Namaceae was treated as clade II of Hydrophyllaceae s.l. and based on chloroplast data, they were resolved as sister or successive sister groups but with low or moderate support (Ferguson, 1998; Nazaire & Hufford, 2012; Refulio-Rodriguez & Olmstead, 2014; Weigend & al., 2014). Despite the fact that neither the molecular nor the morphological data provide strong support for their separation, they were recognized as two seperate families (Luebert & al., 2016). The second uncertainty concerns the position of Heliotropiaceae. Only recently, they were retrieved as sister to the Cordiaceae + Coldeniaceae clade, while Ehretiaceae were retrieved as sister to Heliotropiaceae + Cordiaceae + Coldeniaceae, albeit with very limited sampling (Zhang & al., 2020). Previously, based on either nuclear or chloroplast markers, Heliotropiaceae were retrieved as sister to Lennoaceae + Ehretiaceae + Coldeniaceae + Hoplestigmataceae + Cordiaceae with weak support (Gottschling & al., 2001; Refulio-Rodriguez & Olmstead, 2014; Weigend & al., 2014). The third doubtful phylogenetic relationship is that of Lennoaceae and Ehretiaceae. Opinions on the relationships of the Lennoaceae on the angiosperms tree since their first description by Solms-Laubach (1870), have been largely variable (see Yatskievych & Mason, 1986), until molecular data revealed that they are closely related to Boraginales II (Smith & al., 2000; Olmstead & Ferguson, 2001). Later studies retrieved

Lennoaceae and Ehretiaceae as sister groups based on nuclear markers (Gottschling & al., 2001), but otherwise Lennoaceae has been found nested in Ehretiaceae (Nazaire & Hufford, 2012; Weigend & al., 2014; Zhang & al., 2020). Lennoaceae have been historically treated as separate family, due to them being parasites with highly derived morphology (Bittrich, 2016; Luebert & al., 2016). Infrafamilial classification of all Boraginales II families has been addressed in several studies (Hydrophyllaceae and Namaceae; Ferguson, 1998; Walden & al., 2014; Vasile & al., 2020; Heliotropiaceae; Hilger & Diane, 2003; Luebert & al., 2011a; Luebert, 2013; Frohlich & al., 2022; Cordiaceae and Ehretiaceae; Gottschling & al., 2005, 2014a), however, there are still a number of unanswered questions.

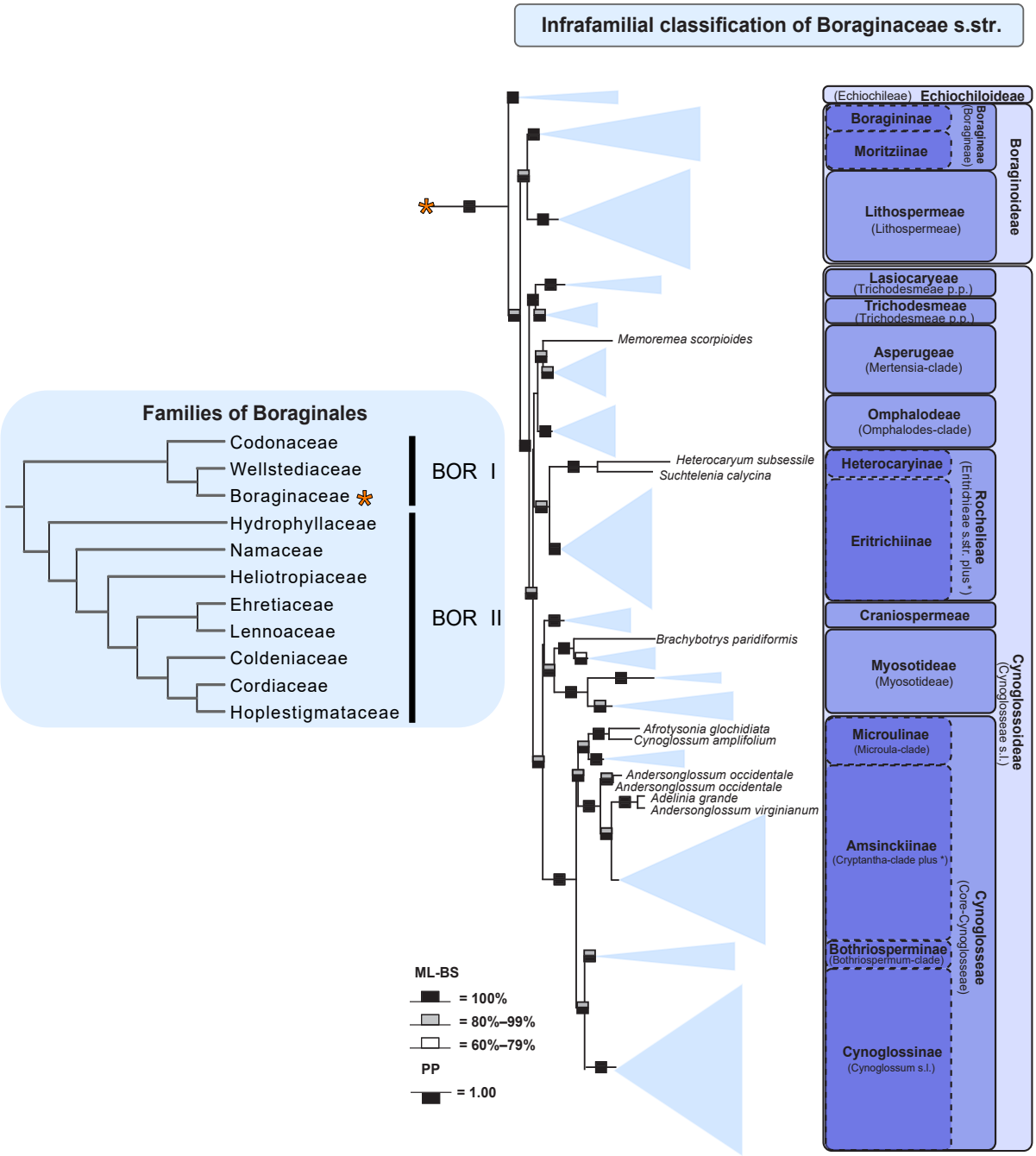


Figure 6.2 Familial classification of Boraginales modified after Luebert & al. 2016 (left) and infrafamilial classification of Boraginaceae modified after Chacón & al. 2016 (right).

A target enrichment method for gathering phylogenetic information

Establishing a systematic framework that is robust, comprehensive, and able to withstand the test of new data is a high priority. Much progress has been made in understanding the phylogeny of Boraginales, but many questions remain, especially with conflicting results from different molecular data sets. Recent progress in molecular techniques has dramatically improved our understanding of the evolutionary tree of life. The recent development of the Angiosperms353 probe set (Johnson & al., 2019) provides a promising tool for systematists working with angiosperms (Baker & al., 2021) as it enables the generation and exploitation of hundreds of nuclear loci for any angiosperm lineage. However, target enrichment remains almost completely unexplored in Boraginales and it has been only used at genus level so far (Cohen, 2022). This powerful tool in the phylogenetic toolbox allows for selective sequencing of genomic regions of interest (Pezzini & al., 2023). In the context of Boraginales, this method promises better resolution and higher support for problematic clades, enabling researchers to tackle the unresolved and controversial relationships head-on. By focusing on specific regions, target enrichment offers a potential pathway to cut through the phylogenetic noise and yield more consistent and robust insights into the systematic relationships of the order.

Aims of the study

In an era where a phylogenetic framework for navigating and interpreting the species, trait and functional diversity of angiosperms has never been more urgent (Baker & al., 2022), we aim to provide new insights into the evolution of Boraginales. To achieve this, we test the effectiveness of the Angiosperm353 target enrichment probe-set at different taxonomic levels in order to evaluate previously published hypotheses for relationships of Boraginales. Our primary goals are: (1) address and resolve the open issues that have long hampered the order's classification; (2) offer an updated classification at the family level that not only reflects current understanding, but also provides a solid foundation for subsequent studies; (3) present a well-resolved phylogeny, which will not just elucidate the evolutionary relationships within Boraginales but also serve as a benchmark for future explorations in the domain of plant evolutionary biology.

6.2 MATERIALS AND METHODS

Taxon sampling

Our taxon sampling comprises 208 accessions of Boraginales, representing ten families (i.e. all currently accepted families according to Luebert & al., 2016, except for Hoplestigmataceae). Boraginaceae is the largest family in the order, encompassing approx. 1700 currently accepted species, and is therefore sampled more intensively, including 134 species (138 accessions), representing 76 % of its genus diversity (74 out of approx. 97 genera). Those 74 genera represent all three subfamilies (i.e. Echiochiloideae, Boraginoideae, Cynoglossoideae), all 10 tribes and all 8 subtribes as described in Chacón & al. (2016). Detailed taxon coverage at tribal and subtribal level is given in table 6.1. Boraginales I clade further includes the monogeneric Codonaceae (1 genus, 2 species: both species here included) and Wellstediaceae (1 genus, 6 species: represented by two species). The remaining boraginalean families, forming the Boraginales II clade, are also well represented with 30 of the accepted 35 genera. This includes Hydrophyllaceae (12 genera, ~ 260 species: here 14 accessions representing 9 genera and 12 species), Namaceae (4 genera, ~ 75 species: here 7 accessions representing 4 genera and 7 species), Helitropiaceae (4 genera, ~ 450 species: here 22 accessions representing 4 genera and 21 species and all four major clades within the genus Heliotropium; Luebert & al., 2011b), Ehretiaceae (7 genera, ~

150 species: here 11 accessions representing 5 genera and 10 species), Lennoaceae (2 genera, ~ 4 species: here 3 accessions representing 2 genera and 2 species), Cordiaceae (2 genera, ~ 400 species: here 5 accessions representing 2 genera and 5 species), and finally the monotypic Coldeniaceae (1 species, *Coldenia procumbens* L., here 3 accessions). Ninety-eight samples were prepared for sequencing at Nees-Institute for Biodiversity of plants. Leaf tissue preserved in silica gel, mainly collected from the field or botanical gardens, or tissue samples from herbarium specimens, were provided by the leaf tissue collection of Herbarium BONN. Sequences of the remaining 110 Boraginales samples and of 4 additional samples from the closely related orders Solanales and Gentianales used as outgroup (1 Solanaceae, 1 Convolvulaceae, 1 Gentianaceae, 1 Rubiaceae) were downloaded via Tree of Life Explorer (Baker & al., 2022). The outgroup selection was made according to Baker & al. (2022) who retrieved Gentianales sister to Boraginales and Solanales sister to both. All voucher information can be found in Appendix D (Table S1).

Table 6.1 Taxon coverage at genus level and number of accessions per taxonomic group. Number of accepted genera is given according to Luebert & al. 2016 and Chacón & al. 2016. The single asterisk indicates that only two genera are currently accepted *Microparacaryum* (Popov ex Riedl) Hilger & Podlech and *Cynoglossum* L. s.l. The double asterisk indicates that *Coldenia* L. is monotypic**.**

Family	Subfamily	Tribe	Subtribe	Nr. of accepted genera	Nr. of genera in this study	Nr. of accessions
Boraginaceae	Echiochiloideae	Echiochileae		3	2	4
Boraginaceae	Boraginoideae	Borageineae	Boragininae	15	12	16
Boraginaceae	Boraginoideae	Borageineae	Mortziinae	2	2	4
Boraginaceae	Boraginoideae	Lithospermeae		25	18	25
Boraginaceae	Cynoglossoideae	Trichodesmeae		2	2	4
Boraginaceae	Cynoglossoideae	Lasiocaryeae		3	3	5
Boraginaceae	Cynoglossoideae	Asperugeae		4	4	6
Boraginaceae	Cynoglossoideae	Omphalodeae		6	2	2
Boraginaceae	Cynoglossoideae	Rochelieae	Eritrichiinae	7	6	11
Boraginaceae	Cynoglossoideae	Rochelieae	Heterocaryinae	2	1	1
Boraginaceae	Cynoglossoideae	Craniospermeae		1	1	3
Boraginaceae	Cynoglossoideae	Myosotideae		4 (5)	4	9
Boraginaceae	Cynoglossoideae	Cynoglosseae	Cynoglossinae	2*	2	17
Boraginaceae	Cynoglossoideae	Cynoglosseae	Bothriosperminae	5	2	3
Boraginaceae	Cynoglossoideae	Cynoglosseae	Microulinae	3	2	2
Boraginaceae	Cynoglossoideae	Cynoglosseae	Amsinckiinae	ca. 14	12	26
Codonaceae				1	1	3
Wellstediaceae				1	1	2
Hydrophyllaceae				12	9	14
Namaceae				4	4	7
Heliotropiaceae				4	4	22
Ehretiaceae				7	5	11
Lennoaceae				2	2	3
Cordiaceae				2	2	5
Coldeniaceae				1**	1	3
Hoplestigmataceae				1	0	0

DNA extraction, library preparation and sequencing

Extraction of genomic DNA was done according to a slightly modified CTAB protocol (Doyle & Doyle, 1987), by increasing the incubation times. Initially, a quality control was carried out using electrophoresis on 1 % agarose gels using Lonza GelStar Nucleic Acid Gel Stain (100x; Lonza Bioscience, Basel,

Switzerland) and by including lambda DNA (250µg, Promega, Madison, Wisconsin, U.S.A.) the DNA concentration was approximately calculated. Subsequently, samples were quantified using Quantus Fluorometer (Promega, Madison, Wisconsin, U.S.A.). Ninety-eight samples (ca. 1 µg genomic DNA per sample) were sent for library preparation and sequencing to a sequencing service (LGC Genomics GmbH, Biosearch Technologies, Berlin, Germany). An Illumina NextSeq 500/550 v2 (Illumina, San Diego, California, U.S.A.) was used for sequencing, generating short paired-end reads of 150 bp and a total of 100 million reads. Raw sequencing data will be deposited in the European Nucleotide Archive (ENA).

Data processing and assembly

Demultiplexing of the sequencing libraries was done with the Illumina bcl2fastq v2.20 conversion software (Illumina, San Diego, California, U.S.A.). One or two mismatches or Ns were allowed in the barcode read when the barcode distances between all libraries on the lane allowed for it. FastQC v0.11.9 (Andrews & al., 2012) was used for evaluating raw reads quality, checking adapters remnants and producing FastQC reports for all FASTQ files. Sequencing reads were further processed using Trimmomatic v0.39 (Bolger & al., 2014) to remove leading and trailing low-quality bases with a Phred-scaled quality score below 20. With this setting, bases were cut when the average quality within a 4-bp sliding window droped below 20, and reads shorter than 30bp were excluded (LEADING: 20, TRAILING: 20, SLIDING WINDOW: 4:20, MINLEN: 30). Only cleaned reads in which both read pairs passing this quality control were kept for subsequently assembly and analyses. For sequence assembly and extraction, the HybPiper v2.1.1 (Johnson & al., 2016) pipeline was used. For the HybPiper assembly, the mega353 target file (McLay & al., 2021) was used instead of the default Angio353 target file (Johnson & al., 2019). The mega353 fasta file was filtered for Boraginales (15 targets) but also for sister orders of Boraginales: Gentianales (17 targets) and Solanales (26 targets). The filtered mega353 target file included at the end sequences from 58 taxa in addition to the standard 353 target sequences. It was checked for sequences with low-complexity regions prior to running 'hybpiper assemble' command, by applying the command 'hybpiper check_targetfile' and then fixed with the *fix_targetfile.py* script available with HybPiper. The trimmed paired reads were mapped to the filtered mega353 targets using the Burrows–Wheeler alignment (BWA) option (Li & Durbin, 2009) and then assembled de novo with SPAdes (Bankevich & al., 2012). The exonic sequences as well as the supercontigs (exon + flanking intron regions) for each locus were recovered by providing the --run_intronerate flag while running the exonerate_hits.py script available with HybPiper. Both exon and supercontigs were retained for downstream analyses. To summarize recovery efficiency, the *hybpiper_stats.py* script available with HybPiper was used and visualization of the exon coverage was done by implementing the *gene_recovery_heatmap.py* script (output files are available in Appendix D, supplementary material S1).

Evaluation of paralogy

During the Hybpiper assembly, if multiple contigs with a > 10× coverage depth in a sample mapped to the same target gene with >75% sequence identity, this target gene is marked for presence of paralogs in this sample. Hybpiper selects the sequence with the greatest percent identity to the reference if all competing long contigs have similar depth. The *paralog_retriever.py* script available with HybPiper was used to further investigate the distribution of paralogs within our samples. This script produces, among other files, a report table with the number of sequences for each gene and sample and a heatmap visualisation of this table (see Appendix D, supplementary material S2). If many samples have more than one copy for many genes, it may indicate an ancient gene duplication. If one sample tends to have many copies, it may indicate polyploidy. For evaluating the effect of the paralogous genes on the

phylogenetic reconstruction, genes which had paralog warnings for more than 10% of the samples were eliminated, creating a filtered dataset in order to address issues caused in the downstream analyses by potential comparison of non-orthologous sequences.

Sequence Alignment

The two datasets of exons and supercontigs without genes excluded and the filtered exon-dataset with potentially paralogous genes excluded were first aligned using MAFFT v7.515 (Katoh & Standley, 2013). The multiple sequence alignments (MSAs) produced with MAFFT were further processed with CIAlign v1.1.0 (Tumescheit & al., 2022), a command line tool to clean, interpret and visualize MSAs. By using the parameter option all, all available functions were run with the default parameters. The main functionalities of CIAlign include removing regions of low coverage due to insertions, removing gaps, cropping poorly aligned sequence ends and removing sequences that are too divergent or too short (Tumescheit & al., 2022). The fine-tuned MSAs were concatenated using the R package TOAST v0.0.1.5000 (Transcriptome Ortholog Alignment Sequence Tools; available in github.com/ziyili20/TOAST) in R version 4.3.1 (R Core Team, 2020). TOAST utilizes the ortholog searches based on the software benchmarking Universal Single-Copy Orthologs (BUSCO) to assemble multiple sequence alignments of orthologous loci. In addition to the construction of concatenated alignments, TOAST generated a partition block and assessed overall missing data patterns across the alignments (see Appendix D, supplementary material S3).

Phylogeny reconstruction

Concatenation methods can be inconsistent under ILS and can estimate incorrect species trees as the number of genes increases (Xiong & al., 2022), even with high support under some conditions (Warnow, 2015). Most importantly, Roch and Steel (2015) proved that concatenation using unpartitioned maximum likelihood analysis can be statistically inconsistent and even positively misleading in the presence of ILS. In the present study, both a multi-locus species tree estimation approach and locus concatenation approach were implemented. However, for the aforementioned reason, more attention will be given to the multi locus species tree. The cleaned by CIAlign MSAs of exons and supercontigs were used as input in IQ-TREE 2.0.3 (Nguyen & al., 2015) which provides the option to compute individual locus trees given a directory and simultaneously run ModelFinder (Kalyaanamoorthy & al., 2017) to find the best fitting model for each locus. The ultrafast bootstrap option (Hoang & al., 2018) was used to generate 1000 bootstrap replicates for each gene tree. To account for ILS, an unrooted species tree was estimated with ASTRAL-III v.5.7.8 (Zhang et al., 2018) and the branches on the species tree were scored using local posterior probabilities (-t 3 option in ASTRAL). For the concatenated alignments assembled by TOAST for both exons and supercontigs, a partitioned analysis was conducted in IQ-TREE 2 in order to reduce statistical inconsistency. The analysis was performed with model finder identifying the best model + partitioning strategy (MFP+MERGE) and with the ultrafast bootstrap feature to generate 1000 bootstrap replicates. Trees were visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The downstream analyses are summarized in Figure 6.3.

6.3 RESULTS

Sequence recovery

All of the Angersperms353 genes were included in our analyses, however, among the 211 samples used to reconstruct the Boraginales phylogeny, target gene recovery ranged from 39 in *Johnstonella*

angustifolia (Torr.) Hasenstab & M.G.Simpson to 350 genes (See Appendix D, supplementary material S1, S3). The average gene recovery per sample was 307 genes. The evaluation of paralogy yielded that out of the 353 Angiosperm genes, 95 had paralog warnings for at least one out of the 211 samples, but only 26 genes had paralog warnings for at least 10% of the samples. Therefore, only these 26 were filtered out from the exonic dataset in order to test the effect of paralogy on the phylogenetic reconstruction. The distribution of paralogs in our data ranged from 0 to 77 paralogs per sample with most of samples having low paralog representation (average 11 paralog warnings). There were no distinct patterns regarding the distribution of paralogs across clades. Overall missing data assessment with TOAST after further cleaning the multiple sequence alignments with CIAlign, showed that only 14 samples had more than 50% missing data.

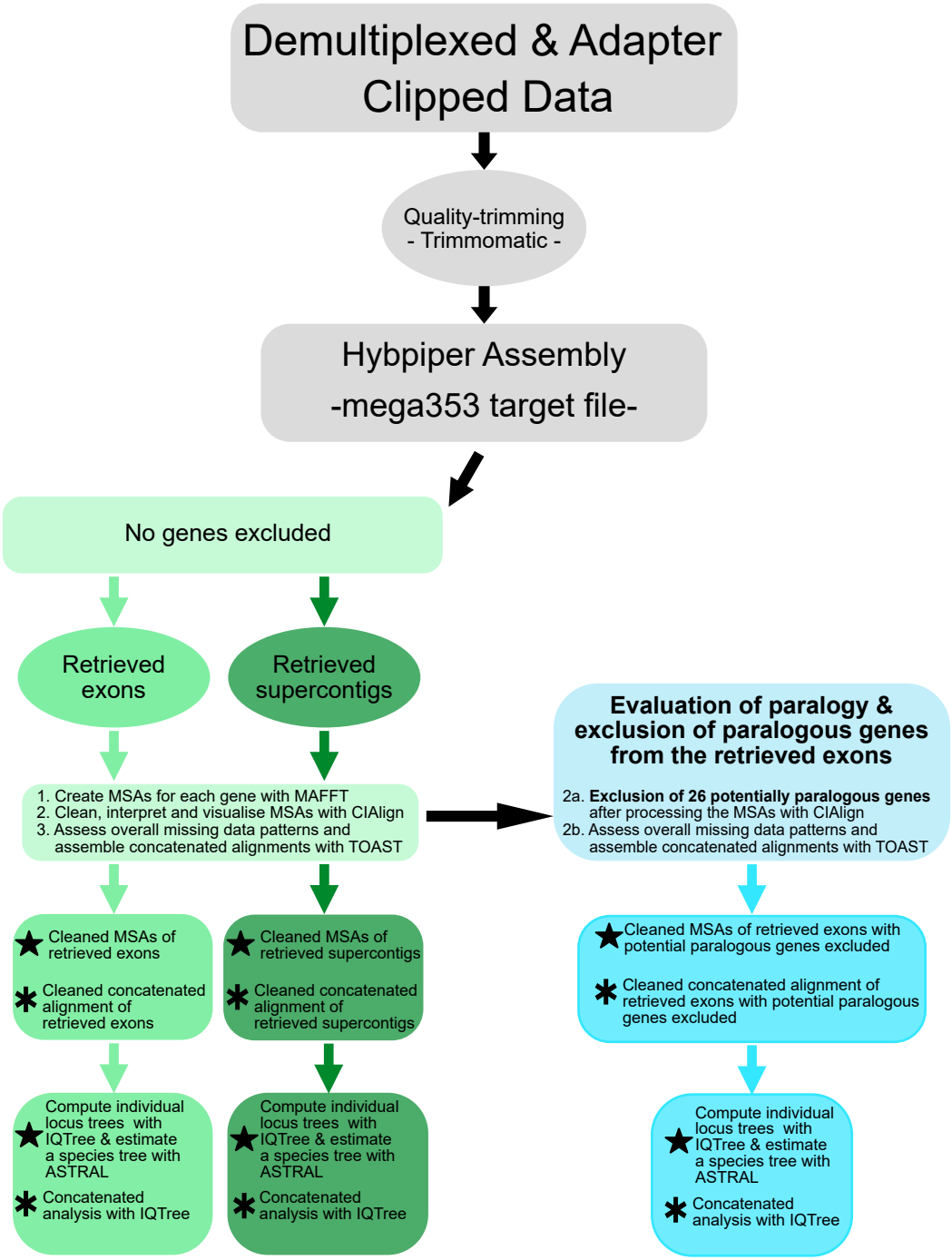


Figure 6.3 Summary of methodology and downstream analyses.

Phylogenetic analyses

Both the concatenation analysis and multi-locus species tree estimation approach yielded well resolved trees with overall congruent topologies. However, as mentioned before, the concatenation methods can be inconsistent and therefore we emphasize here the results of the multi-locus species tree (ASTRAL-III species tree) of the exonic dataset and compare them with those of the multi-locus species tree of the supercontig dataset. Nevertheless, significant incongruences between the concatenation analyses and multi-locus species trees will be also mentioned. Support values of the ASTRAL-III species tree are given as local posterior probabilities (LPP), while those of the concatenation analyses as ultrafast bootstrap approximation (UFBoot). All the trees that are not presented in the results can be found as supplementary figures (S1-S4) in Appendix D.

According to the ASTRAL-III species tree of the exonic dataset, Gentianales are retrieved as sister to Boraginales, and Solanales sister to these two with maximum support. The tree shows overall well-resolved phylogenetic relationships among the families of both Boraginales I and II clades (Figs. 6.4 & 6.5). Within Boraginales I (Fig. 6.4), Codonaceae is sister to Wellstediaceae and Boraginaceae with maximum support. The infrafamilial relationships of Boraginaceae are also well resolved. Echiochiloideae is sister to Boraginoideae and Cynoglossoideae with maximum support. At a tribal level within Boraginoideae, the two tribes Boragineae and Lithospermeae are sister to each other with maximum support. The backbone of Cynoglossoideae is well-resolved and only in few cases the tribal relationships are not resolved with maximum support. Trichodesmeae is successive sister group to *Lasiocaryeae* (LPP = 1, UFBoot = 100), which is then successive sister with maximum support to a clade comprising Omphalodeae and Asperugeae. These two tribes are sister to each other with moderate support in the ASTRAL-III species tree, but with high support in the species trees of the concatenation analysis (LPP = 0.82, UFBoot = 92) and they form a clade successive sister to the remaining Cynoglossoideae with maximum support. The remaining Cynoglossoideae consists of four tribes: Rochelieae is successive sister to Craniospermeae (LPP = 1, UFBoot = 100), Craniospermeae is successive sister to a clade comprising Myosotideae and Cynoglosseae with maximum support. Myosotideae and Cynoglosseae are sister to each other with high support (LPP = 0.95, UFBoot = 100). At subtribal level, Mortziinae and Boragininae of the tribe Boragineae are sister groups with maximum support. The same goes for the Heterocaryinae and Eritrichiinae of the tribe Rochelieae. Within the large tribe Cynoglosseae there are some ambiguities. Bothriosperminae is sister to Cynoglossinae with maximum support and together form a clade sister to Microulinae and Amsinckiinae with maximum support. Microulinae is sister to Amsinckiinae with very low support for the ASTRAL-III species tree (LPP = 0.33, UFBoot = 85). Low resolution for the ASTRAL-III species tree is recovered regarding the intergeneric relationships within the tribe Lithospermeae and within the subtribes Boragininae, Cynoglossinae and Amsinckiinae. Nothworthy is that within Cynoglossinae, *Microparacaryum* is sister to the remaining species (*Cynoglossum s.l.*) with moderate support for the ASTRAL-III species tree (LPP = 0.63, UFBoot = 97). Moderate or low support values within these clades in Figure 6.4, often coincide with incongruences between the ASTRAL-III species tree and the concatenation analysis species tree. Nevertheless, overall the intergeneric relationships of the concatenation analysis species tree are well resolved, with only exception the Amsinckiinae.

The ASTRAL-III species tree of the exonic dataset shows well resolved phylogenetic relationships among the families of Boraginales II clade as well (Fig. 6.5). The basal grade of this clade consists of Namaceae and Hydrophyllaceae. Although Hydrophyllaceae are monophyletic with maximum support, the family is not recovered as direct sister group of the Namaceae, since *Nama dichotoma* (Ruiz & Pav.)

Choisy is successive sister with maximum support to a clade including the remaining Namaceae as sister to the Hydrophyllaceae (LPP = 0.99, UFBoot = 100). Thus, *Nama* L., even though it is represented by only two accessions, is retrieved as polyphyletic. *Wigandia* Kunth is sister with maximum support to a clade comprising *Nama rothrockii* A.Gray as is sister to *Eriodictyon* Benth. and *Turricula parryi* (A.Gray) J.F.Macbr (LPP = 1, UFBoot = 100). In Hydrophyllaceae, all relationships are fully resolved

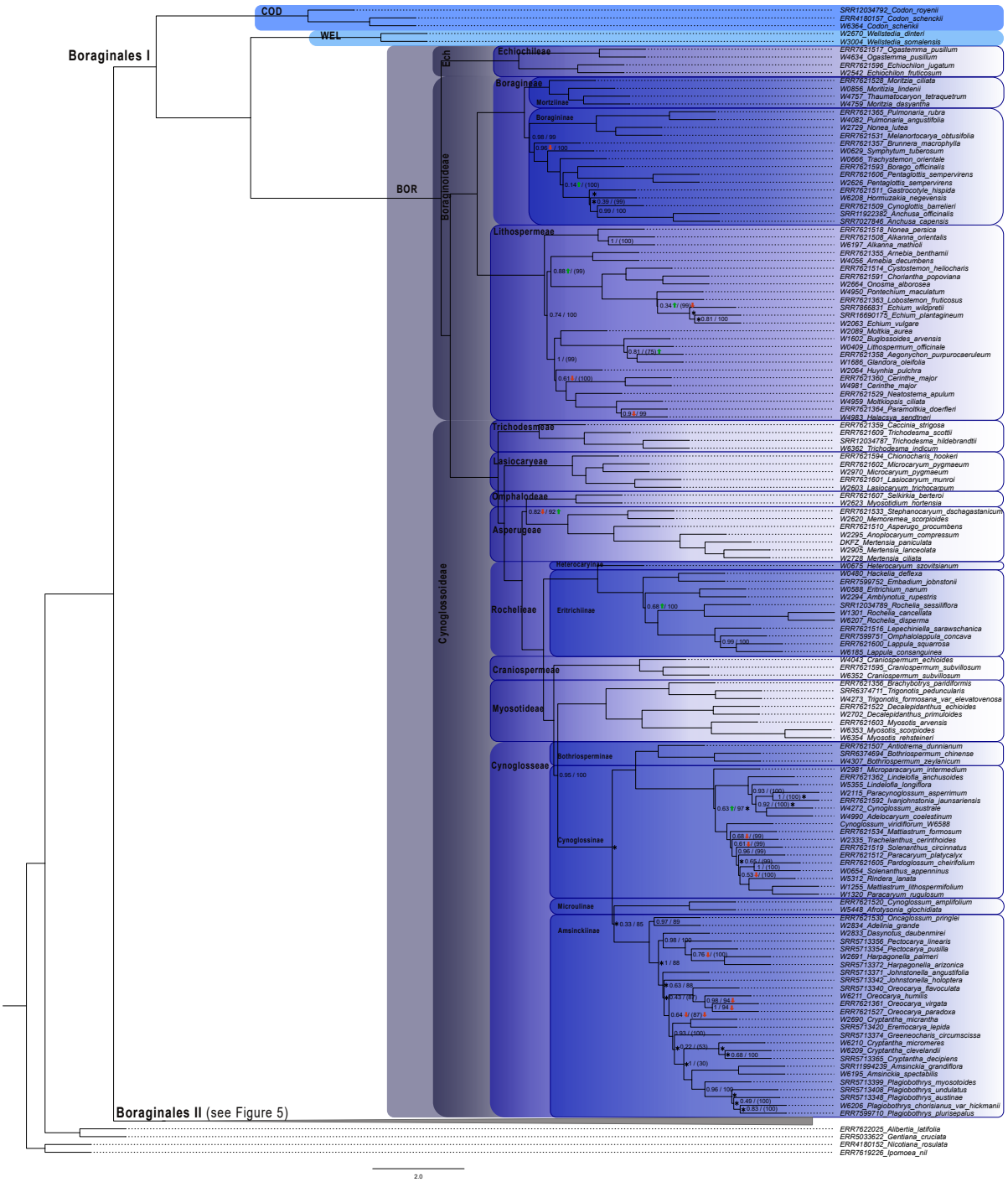


Figure 6.4 Phylogenetic relationships within Boraginales I. ASTRAL-III species tree generated using the retrieved exonic regions of 353 nuclear loci from the Angiosperms353 probe set. Each of the three families — Codonaceae (COD), Wellstediacae (WEL), Boraginaceae (BOR)— is shown in different colour. Within Boraginaceae the subfamilies (Echiochiloideae (Ech), Boraginoideae, Cynoglossoideae), tribes (Echiochileae,

Boragineae, Lithospermeae, Trichodesmeae, Lasiocaryeae, Omphalodeae, Asperugeae, Rochelieae, Craniospermeae, Myosotideae, Cynoglosseae) and subtribes (Mortziinae, Boragininae, Heterocaryinae, Eritrichiinae, Bothriosperminae, Cynoglossinae, Microulinae, Amsinckiinae) are also annotated with frames of different colour depending on the taxonomic rank. Numerical values denote, the local posterior probability (LPP) for the ASTRAL-III species tree (left), and the ultrafast bootstrap approximation (UFBoot) from the concatenation analysis (right). UFBoot values in parentheses indicated incongruences in the topology between the coalescence and the concatenation analyses. Branches without support values indicate full support and no topological incongruences for both analyses (LPP/ UFBoot = 1/100). Green arrows (⬆) indicate increased support values after paralogs exclusion by at least 0.05 for LPP and/or 5 by UFBoot, while red arrows (⬇) show decreased support values. The arrows are placed next to the support values that were affected. Asterisks placed at the right side of a node denote topological changes for both trees after paralogs exclusion. If there is topological change in only one of the trees the asterisk is placed next to the respective support value.

and supported. *Draperia* Torr. is sister to Romanzoffieae and Hydrophylleae clade. Romanzoffieae comprises *Romanzoffia* Cham. as sister to *Phacelia* Juss. Hydrophylleae comprises the genera *Emmenanthe* Benth., *Eucrypta* Nutt., *Ellisia* L., *Hydrophyllum* L., *Nemophila* Nutt. ex W.P.C.Barton and *Pholistoma* Lilja as successive sister groups. The basal grade of Boraginales II (Namaceae and Hydrophyllaceae) is sister to the remaining families of the clade with maximum support. Ehretiaceae comprises two major clades sister to each other (LPP = 1, UFBoot = 100). The first clade includes *Halgania* Gaudich. as sister to *Ehretia* P.Browne clade with maximum support. The *Ehretia* clade also includes *Rotula aquatica* Lour. The phylogenetic relatioships of the second clade are not that well resolved. *Tiquilia* Pers. is successive sister to *Lepidocordia* Ducke with moderate support (LPP = 0.87, UFBoot = 81), which is then successive sister to *Bourreria* P.Browne (LPP = 0.74). *Bourreria* is sister to Lennoaceae with low support (LPP=0.3). There are topological differences in the second Ehretiaceae clade between the ASTRAL-III species tree and the concatenation analysis species tree. In the concatenation analysis, Lennoaceae are retrieved as sister to *Tiquilia* (UFBoot = 98), *Lepidocordia* as sister to *Bourreria* (UFBoot = 69) and these two clades as sister to each other (UFBoot=81). However, in both cases Lennoaceae are clearly nested in Ehretiaceae. Ehretiaceae are sister with maximum support to a clade including Coldeniaceae, Cordiaceae and Heliotropiaceae. The monotypic Coldeniaceae is sister to Cordiaceae with maximum support. Within Cordiaceae, *Cordia* L. is not monophyletic, since *Cordia sonorae* Rose is retrieved as sister to a clade including *Varronia* P.Browne sister to two other *Cordia* species. The clade comprising Coldeniaceae and Cordiaceae is retrieved as sister to Heliotropiaceae (LPP = 1, UFBoot = 96). The crown node of Heliotropiaceae is fully supported in both trees, however there are topological incongruences due to the different placement of *Ixorhea* Fenzl between the ASTRAL-III and the concatenation analysis species tree. In the ASTRAL-III species tree *Ixorhea* is retrieved as sister to the *Heliotropium* clade (LPP = 0.72), while in the concatenation analysis species tree is retrieved as successive sister to the *Euploca* Nutt. and *Myriopus* Small clade (UFBoot=95). Phylogenetic relationship within the *Heliotropium* Tourn. ex. L. clade are fully resolved and supported. Specifically, *H. arborescencs* (sect. *Heliothamnus*) is successive sister to *Heliotropium* II clade, which is sister to sect. *Cochranea* + sect. *Tournefortia*. *Euploca* and *Myriopus* are sister clades with maximum support. Within *Euploca* clade, even though the phylogenetic relationships in both trees are well resolved and supported, there are some topological incongruences.

The phylogenetic trees of the exonic and the supercontig dataset are largely congruent, but with some major differences (refer to fig. 6.6 for direct comparison). Codonaceae are retrieved with maximum support sister to the remaining families of Boraginales II instead of Boraginales I in both ASTRAL-III and the concatenation analysis species trees of the supercontig dataset. In Boraginales I, there are some noteworthy differences within Boraginaceae. The subfamily Echiochiloideae is direct sister group of

Boraginoideae and together are sister to Cynoglossoideae in both trees of the supercontig dataset. Within Cynoglossoideae, the sister relationship of Omphalodeae and Asperugeae is fully supported in both trees of the supercontig dataset. Myosotideae is successive sister to Craniospermeae and not the other way around, but only in ASTRAL-III species tree of the supercontigs and not in that of the concatenation analysis. Furthermore, *Microulinae* are better supported as sister to Amsinckiinae (LPP = 0.87, UFBoot = 100). In Boraginales II clade there some topological incongruences regarding Ehretiaceae and Heliotropiaceae. Within Ehretiaceae, *Tiquilia* is successive sister to *Lepidocordia* which is successive sister to the *Ehretia* + *Halgania* clade and the *Bourreria* + Lennoaceae clade. These two clades are sister to each other with low support in the ASTRAL-III species tree (LPP = 0.57), and *Bourreria* is sister to Lennoaceae also with low support (LPP = 0.63). However, all these relationships are fully supported in the concatenation analysis tree. Within Heliotropiaceae, *Ixorhea* is fully supported as sister to *Heliotropium* clade in both ASTRAL-III and the concatenation analysis species trees of the supercontig dataset.



Figure 6.5 Phylogenetic relationships within Boraginales II. ASTRAL-III species tree generated using the retrieved exonic regions of 353 nuclear loci from the Angiosperms353 probe set. Each of the six families — Namaceae (NAM), Hydrophyllaceae (HYD), Ehretiaceae (EHR), Coldeniaceae (COL), Cordiaceae (COR), Heliotropiaceae (HEL)— is shown in different colour, while Lennoaceae (LEN) is shown with a dashed line box. Numerical values denote the local posterior probability (LPP) for the ASTRAL-III species tree (left), and the ultrafast bootstrap approximation (UFBoot) from the concatenation analysis (right). UFBoot values in parentheses indicated incongruences in the topology between the coalescence and the concatenation analyses. Branches without support values indicate full support and no topological incongruences for both analyses (LPP/ UFBoot = 1/100). Green arrows (⬆) indicate increased support values after paralogs exclusion by at least 0.05 for LPP and/or 5 by UFBoot, while red arrows (⬇) show decreased support values. The arrows are placed next to the support values that were affected.

Effect of paralogy on phylogenetic reconstruction

In order to test the effect of paralogy on the phylogenetic reconstruction, only the filtered dataset of exons was used. The 26 potential paralogous genes which were excluded seem to have no major effect on the topology except for few cases (see Figs. 6.4 & 6.5). Overall the backbone of Boraginales I clade remained unaffected. The support values of the internal nodes of the individual tribes and subtribes were slightly affected, mostly negatively, when removing potential paralogs. However, those nodes that were fully supported, remained thus. Some cases are noteworthy. The position of Microulinae is better resolved in the ASTRAL-III species tree removing some paralogous genes and it is retrieved as successive sister group to Bothriospermae + Cynoglossae (LPP increased to 0.86 from 0.33). These three successive sister group are retrieved with maximum support as sister to Amsinckiinae. However, the concatenation analysis species tree has higher support for a sister relationship of Microulinae to Amsinckiinae (UFBoot decreased to 60 from 85 when excluding paralogs). Furthermore, *Microparacaryum* is not supported as sister to *Cynoglossum s.l.* in the concatenation analysis species tree, but it is retrieved as sister to *Lindelofia* Lehm. with low support (UFBoot = 47). Within the Amsinckiinae there are a lot of topological differences. The effect of potential paralog exclusion on Boraginales II clades is even less compared to Boraginales I. In the ASTRAL-III species tree *Ixorhea* is better supported as sister to *Heliotropium* clade (LPP increased to 0.9 from 0.72). Within Ehretiaceae, the sister relationship of *Bourreria* and Lennoaceae is also better supported in the ASTRAL-III species tree (LPP increased to 0.82 from 0.3), while the position of *Lepidocordia* as their sister is less supported (LPP decreased to 0.22 from 0.74). The concatenation analysis species tree had only minor changes in support values. Both trees of Boraginales II had no topological incongruences after paralogs exclusion.

6.4 DISCUSSION

This is the first study using the Angiosperms353 probe set in Boraginales, and in general the first study to use an NGS approach to explore phylogenetic relationships across the order. Our results show the effectiveness of Angiosperms353 probe set for deep divergences, but also at different taxonomic ranks, which is in line with several recent studies across asterids, utilizing the Angio353 in particular, or other target enrichment approaches in general (Bagley & al., 2020; Antonelli & al., 2021; Ottenlips & al., 2021; Thomas & al., 2021b). We were able to retrieve the first multilocus (avg. 307 loci recovered per sample) phylogeny of Boraginales, with considerable taxonomic coverage at family and genus level (10 out of 11 families, ~79% of genus diversity). The phylogenetic relationships here recovered, are largely congruent with those of previous analyses (Ferguson, 1998; Gottschling & al., 2005; Moore & Jansen, 2006; Luebert & al., 2011a; Hasenstab-Lehman & Simpson, 2012; Weigend & al., 2013, 2014; Chacón & al., 2016; Hasenstab-Lehman, 2017; Vasile & al., 2020; Zhang & al., 2020; Cohen, 2022). Single-

copy nuclear genes are, however, very clearly shown to be more useful for resolving deep relationships compared to chloroplast or nuclear (ITS) markers. Only two previous studies adopted a phylogenomic approach, but the first (Zhang & al., 2020) dealt with asterids, a very large group encompassing about 17 orders and ca. 110 families, and therefore had limited sampling within Boraginales, while the second was focused on the genus *Lithospermum* L. (Cohen, 2022) and only had restricted sampling across remaining Boraginaceae, used as outgroup. The very extensive sampling in this study combined with a phylogenomic approach has consequently led to some topological incongruences of particular interest, compared to previous studies. These are mentioned in detail in the sections about systematics of Boraginales I and II below.

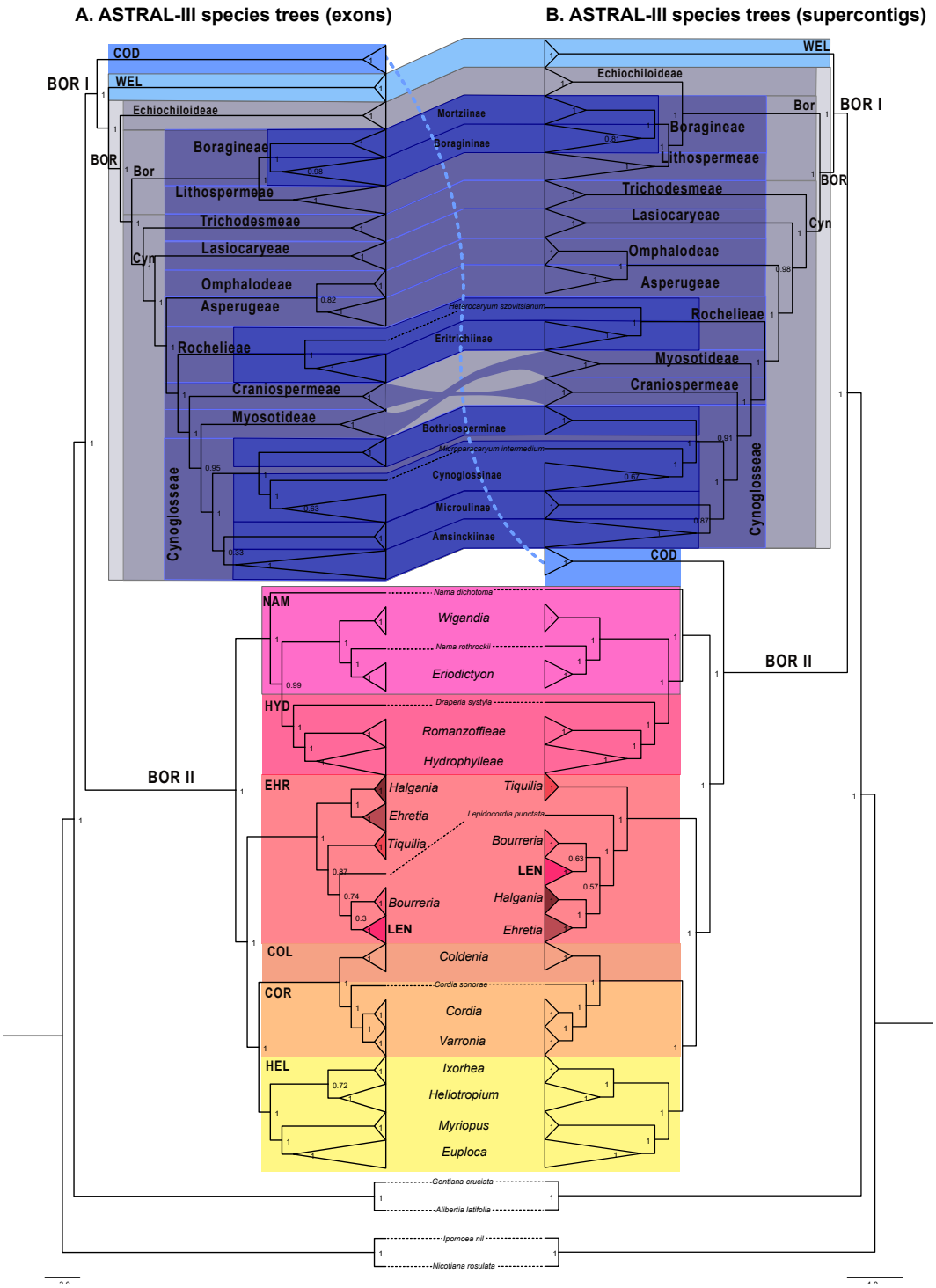


Figure 6.6 Cophylo plot showing a comparison between ASTRAL-III species tree of the exons (A) and supercontigs (B) inferred using IQ-TREE. Families —Codonaceae (COD), Wellstediaceae (WEL), Boraginaceae (BOR), Namaceae (NAM), Hydrophyllaceae (HYD), Ehretiaceae (EHR), Coldeniaceae (COL), Cordiaceae (COR), Heliotropiaceae (HEL)— is shown in different colour. Lennoaceae (LEN) is nested within Ehretiaceae. Within Boraginaceae the subfamilies, tribes and subtribes are also annotated with frames of different colour depending on the taxonomic rank. Numerical values denote the local posterior probability (LPP).

Both our exonic dataset and the longer, more variable supercontig dataset, performed equally well. We did not observe dramatic differences between the reconstructed topologies, although sometimes supercontigs have been shown to generate more robust phylogenies with higher support values along internal nodes (Bagley & al., 2020; Murphy & al., 2020; Thomas & al., 2021a). Concatenation and coalescent approaches (including summary gene tree methods like ASTRAL; Zhang & al., 2018) are radically different, but they often generate similar topologies for data sets with overall strong phylogenetic signal (as for example in Apiaceae, Clarkson & al., 2021; Gentianales, Antonelli & al., 2021; Myrtales, Maurin & al., 2021; Ochnaceae, Shah & al., 2021; and Sapindaceae, Buerki & al., 2021). Smith et al. (2020) proposed that a "data-centric middle-way" between concatenation and coalescent approaches might be the way to go in order to identify relationships that require more attention and to better understand the evolutionary processes underlying the tree of life. Since there is no fixed recipe for how to deal with biological heterogeneity, we decided to emphasize the results of the ASTRAL species trees because of concatenation methods being statistically inconsistent and prone to ILS (Roch & Steel, 2015; Mendes & Hahn, 2018; Bryant & Hahn, 2020), but also compare them to the concatenated ML species trees. In our case, the coalescent-based species tree method and the concatenation approach performed equally well, yet with a few topological differences and often higher support values in concatenated ML trees. Incongruences and low support values localized in a few clades, especially in the Amsinckiinae and Cynoglossinae, can be partially attributed to missing data (see recovery heatmap, supplementary material S1 in Appendix D), but might also be the outcome of rapid radiations or undetected paralogy in general. Exclusion of potential paralogs detected by Hybpiper, did, however, not have a major effect on topology. Nevertheless, Boraginales have at least three whole genome duplications (WGDs) detected, one suprafamilial at the crown node of Boraginales II, and two familial, one in Boraginaceae and one in Codonaceae (Zhang & al., 2020). Consequently, more sophisticated approaches for paralogy characterization such as HybPhaser v1 (Nauheimer & al., 2021) or PPD (Zhou & al., 2022) should be considered in future studies.

Systematic relationships within Boraginales I

Boraginales I clade has been shown so far to include three families: Codonaceae sister to Wellstediaceae + Boraginaceae (Ferguson, 1998; Nazaire & Hufford, 2012; Weigend & al., 2013, 2014; Refulio-Rodriguez & Olmstead, 2014). Our exonic dataset confirms this relationship, however the longer and more variable dataset of supercontigs retrieved Codonaceae as sister to the remaining Boraginales II (Fig. 6.6). Codonaceae are unique within Boraginales I. Their polymerous flowers, pluri-seeded capsular fruits, endospermous seeds and unique spines are characters that clearly set them apart from Wellstediaceae (1(2)-seeded capsular fruits, tetramerous flowers; Hilger & Weigend, 2016) and Boraginaceae (4 nutlets, pentamerous flowers and exendospermous seeds; Weigend & al., 2016). Prior to the era of molecular phylogenetics, *Codon* was considered to be closely related to Hydrophyllaceae in the traditional sense (Brand, 1913), due to their similar pollen, gynoecium, fruit and seed morphology. After molecular data indicated that *Codon* is closely allied to Boraginales I, these shared character states with Hydrophyllaceae were considered as plausibly plesiomorphic (Weigend & Hilger, 2010). Even though, a sister relationship of Codonaceae to Hydrophyllaceae and Namaceae is supported by a set of (floral)

morphological characters (especially gynoecium, fruit and seed morphology and development), the topological contrast of different datasets in our study requires more attention. Long and variable sequences are more difficult to align compared to the short and more conserved exons. We observed that CIAlign, significantly shortened the supercontig alignments by reducing those regions containing many and large gaps, and this might lead to alignment ambiguity. Before making any conclusion, more MSA tools should be tested, as it has been previously shown that their performance can vary significantly, impacting the ancestral sequence reconstruction (Vialle & al., 2018).

Our phylogeny provides the best resolution and highest support so far for the infrafamilial relationships of Boraginaceae, which are identified as largely congruent with previous molecular studies (Weigend & al., 2013; Chacón & al., 2016). Previously unresolved or ambiguous relationships in the backbone are here addressed. Noteworthy differences compared to the extensive study by Chacón & al. (2016) are the following: we retrieved Trichodesmeae as sister to the remaining Cynoglossoideae, while previously they were sister to Lasiocaryeae and together sister to the rest; we were able to provide a higher support for a sister relationship of Asperugeae + Omphalodeae, however, expanded sampling is required; Asperugeae + Omphalodeae are found as successive sister to Rochelieae, while before a sister relationship of these two clades was unresolved; Craniospermeae are successive sister to Myosotideae and then to Cynoglossae with maximum support, while previously they were retrieved as sister to each other with low support; finally, Microulinae is here not strongly supported as sister to Amsinckiinae, but this is probably due to restricted taxon sampling, which should be expanded in the future. Interestingly, the two main topological differences between the exon and supercontig datasets are not that aberrant. Previous nuclear datasets (Nazaire & Hufford, 2012) retrieved Echiochiloideae as sister to Boraginoideae and not to Boraginoideae + Cynoglossoideae, a relationship that we also retrieved when we used the supercontig dataset. The successive sister relationship of Craniospermeae to Myosotideae in the exonic dataset is reversed in the supercontig dataset (Myosotideae as successive sister to Craniospermeae), revealing a complex evolutionary history. This could partially justify why the sister relationship of these two tribes was previously completely unsupported when based on only a few loci (Chacón & al., 2016).

Systematic relationships within Boraginales II

The Boraginales II clade is a natural group including Namaceae, Hydrophyllaceae, Ehretiaceae, Lennoaceae, Coldeniaceae, Hoplestigmataceae (not included in our sampling), Cordiaceae, and Heliotropiaceae (Luebert & al., 2016), which is congruent with our results and previous studies (Weigend & al., 2014; Zhang & al., 2020). Nevertheless, there are some differences in this study. The exact relationships of the basal grade of Boraginales II (Hydrophyllaceae and Namaceae) were ambiguous (Ferguson, 1998; Nazaire & Hufford, 2012; Refulio-Rodriguez & Olmstead, 2014; Weigend & al., 2014), until recently, when Zhang & al. (2020) retrieved them as sister to each other. This result is based, however, on four Hydrophyllaceae and two Namaceae species, but *Nama* itself, the largest genus in the family (with 53 of the approximately 75 species in the family), is missing. This sister relationship is disturbed in the present study by the inclusion of *Nama dichotoma*, a species that has not been included in previous studies addressing the relationships of Boraginales, and which is recovered as successive sister to Namaceae + Hydrophyllaceae with high or even full support. *Nama dichotoma* is very widespread, amphitropically disjunct in the Americas (Simpson & al., 2017). It is an ephemeral plant with small flowers and few distinguishing morphological characters that would permit a study of its relationships to other species of the genus. The single extensive molecular phylogenetic work available on Namaceae, shows *N. dichotoma* nested in the *Nama* clade, but with conflict between the chloroplast

and the ITS data (Taylor, 2012). *Nama* is obviously underrepresented in our sampling (two species included). Before providing any familial or infrafamilial classification changes, inclusion of more species of *Nama* is needed and could likely shed more light on the relationships of Namaceae and Hydrophyllaceae. Previous studies based on chloroplast data have already shown the polyphyly of *Nama* as well as the need to finally re-include *Turricula parryi* to *Eriodictyon* (Vasile & al., 2020). These findings are further confirmed here.

Similarly, in Ehretiaceae we cannot state the exact intergeneric relationships (Figs. 6.5 & 6.6), but we confirm the inclusion of Lennoaceae in Ehretiaceae as suggested by the phylogenomic study of Zhang & al. (2020) and by recent and detailed morphological data (Jeiter & al., 2023). Lennoaceae have been traditionally treated as separate family due to their parasitic lifestyle (Bittrich, 2016) and their unique morphological traits (Yatskievych & Mason, 1986; Jeiter & al., 2023). One extensive phylogenetic work on Ehretiaceae unfortunately did not include Lennoaceae since they lack the *ndhF* region (Moore & Jansen, 2006). Most phylogenetic studies in Boraginales addressing the position of Lennoaceae suggest long branch attraction for this lineage (Nazaire & Hufford, 2012; Weigend & al., 2014; Hasenstab-Lehman, 2017), while our data do not. Furthermore, all previous phylogenetic works are based almost exclusively on plastome data, which is fairly problematic, due to the fact that Lennoaceae exhibit sequence length reduction and loss in gene content in their plastomes compared to their autotrophous relatives (Schneider & al., 2018). Long branch attraction on plastid phylogenies is rather common for parasitic (Su & al., 2015) or mycoheterotrophic lineages (Feng & al., 2016). Before re-evaluating the intergeneric relationships of Ehretiaceae, sampling expansion in order to better cover species diversity and inclusion of small but morphological distinct genera are crucial steps, especially *Rochefortia* Sw. and *Keraunea* Cheek & Sim.-Bianch., which was only recently recognized as member of Boraginales (Almeida & al., 2023; Moonlight & Cardoso, 2023).

Historically, Cordiaceae included only one broadly defined genus, *Cordia*, with 5 to 8 sections or subgenera (Johnston, 1930). Borhidi & al. (1988) proposed a re-evaluation of morphological data and the separation of three genera: *Cordia*, *Varronia* and *Gerascanthus* P. Browne. The actual resurrection of the genus *Varronia* and the necessary nomenclatural changes have happened only later, based on molecular evidence (Gottschling & al., 2005; Miller & Gottschling, 2007; Weeks & al., 2010). Our taxon sampling might not represent all different and morphologically distinct clades of *Cordia* and *Varronia* formerly described, but indicate the paraphyly of *Cordia*. Inclusion of representatives from all the different subdivisions of both genera would likely help to resolve genus limits.

Heliotropiaceae taxonomy has been in flux for decades (Frohlich & al., 2022). Currently, four genera are recognized ((*Heliotropium* (*Ixorhea* (*Euploca*, *Myriopus*))); Luebert & al., 2016), all included in this study and all supported as monophyletic. A novel sister relationship of *Ixorhea* to *Heliotropium* is supported by the majority of our trees (exons and supercontigs, see fig. 6.6). Only the concatenated ML tree of exons supports *Ixorhea* as sister to *Euploca* and *Myriopus*, as found by previous studies based on chloroplast data (Luebert & al., 2011b; Weigend & al., 2014). The *Heliotropium* clade comprises four major clades: *Heliotropium* sect. *Heliothamnus* I.M.Johnst., Old World *Heliotropium*, *Heliotropium* sect. *Cochranea* (Miers) Post & Kuntze, and the *Tournefortia*-clade, the latter comprising *Tournefortia* sect. *Tournefortia* and all remaining New World species of *Heliotropium* (Hilger & Diane, 2003; Luebert & Wen, 2008; Luebert & al., 2011b). These four clades are here retrieved as monophyletic and their relationships in the same order as in previous studies.

6.5 CONCLUSIONS

The Angiosperms353 probe set is a promising tool which can be successfully implemented for phylogenetic reconstruction at many different taxonomic levels in Boraginales. Crucial next steps are to strategically fill gaps in the sampling (e.g., *Hoplestigma* Pierre from Hoplestigmataceae, *Rochefortia* and *Keraunea* form Ehretiaceae, individual taxa in Omphalodeae, Microuleae etc.) in order to balance genus representation and species diversity across the order. A custom probe set for the order should also be considered for future studies addressing clade-specific questions. Overall, this approach revealed that conflicting evolutionary histories of different genes can complicate reconstructing the species tree. Topological incongruences between different datasets (exons and supercontigs) and between different methods (coalescent-based species tree and concatenation) should definitely get more attention in the future in order to understand what causes conflicting topologies between the different analyses (e.g., Ehretiaceae). Unquestionably, the present study is able to resolve a number of long-standing questions in the phylogeny of Boraginales. This, and future expanded datasets, will open up possibilities to critically study the morphological evolution and patterns of diversification and historical biogeography of Boraginales.

6.6 ACKNOWLEDGMENTS

We thank the Bonn University Botanic Gardens and the scientific curator Cornelia Löhne for maintaining the research collection and providing plant material for this study. We gratefully acknowledge all the plant collectors and herbaria who have contributed to the Boraginales sampling of the Nees Institute over the years. We also thank T. Joßberger, N. Schmandt and Claudia Schütte for their technical assistance during this study. The author Maria-Anna Vasile wishes to further thank all the instructors and participants of the EMBO practical course in computational molecular evolution (2023) for all the valuable discussions on efficiently carrying out molecular evolutionary analyses on sequence data.

6.7 LITERATURE CITED

Almeida, R.F. de, Pellegrini, M.O.O., Morais, I.L. de, Simão-Bianchini, R., Rattanakrajang, P., Cheek, M., & Simões, A.R.G. 2023. Barking up the wrong tree: the dangers of taxonomic misidentification in molecular phylogenetic studies. *Plant Ecology and Evolution* 156: 146–159.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. 2012. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/accessed.

Antonelli, A., Clarkson, J.J., Kainulainen, K., Maurin, O., Brewer, G.E., Davis, A.P., Epitawalage, N., Goyder, D.J., Livshultz, T., Persson, C., Pokorny, L., Straub, S.C.K., Struwe, L., Zuntini, A.R., Forest, F., & Baker, W.J. 2021. Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. *American Journal of Botany* 108: 1143–1165.

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.

Atwood, N.D. 1975. A revision of the *Phacelia crenulata* group (Hydrophyllaceae) for North America. *The Great Basin Naturalist* 35: 127–190.

Bagley, J.C., Uribe-Convers, S., Carlsen, M.M., & Muchhala, N. 2020. Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical *Burmeistera* bellflowers as a case study. *Molecular Phylogenetics and Evolution* 152: 106769.

Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers, T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B., Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., Mcginnie, C., Pokorny, L., Roy, S., Stone, M.,

Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J., & Forest, F. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* 71: 301–319.

Baker, W.J., Dodsworth, S., Forest, F., Graham, S.W., Johnson, M.G., McDonnell, A., Pokorny, L., Tate, J.A., Wicke, S., & Wickett, N.J. 2021. Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants. *American Journal of Botany* 108: 1059–1065.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., & Pevzner, P.A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19: 455–477.

Bentham, G., & Hooker, J.D. 1876. Borragineae. Pp. 832–865 in: *Genera Plantarum (Ad Exemplaria Imprimis in Herbariis Kewensibus Servata Definita)*, vol. 2. London: Lovell Reeve, William & Norgate. Reprint 1965, Weinheim: Cramer.

Bittrich, V. 2016. Lennoaceae. Pp. 257–261 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*, vol. XIV. Heidelberg: Springer

Bolger, A.M., Lohse, M., & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Borhidi, A., Gondár, E., & Orosz-Kovács, Z.S. 1988. The re-consideration of the genus *Cordia* L. 34: 375–423.

Brand, A. 1913. *Hydrophyllaceae*, Das Pflanzenreich, (Heft 59) IV. 251. Leipzig: Wilhelm Engelmann: Engler.

Bryant, D., & Hahn, M.W. 2020. The concatenation question. P. 3.4:1-3.4:23 in: C. Scornavacca, F. Delsuc, & N. Galtier (eds.), *Phylogenetics in the Genomic Era*. No commercial publisher | Authors open access book. https://hal.science/hal-02535651

Buerki, S., Callmander, M.W., Acevedo-Rodriguez, P., Lowry, P.P., Munzinger, J., Bailey, P., Maurin, O., Brewer, G.E., Epitawalage, N., Baker, W.J., & Forest, F. 2021. An updated infra-familial classification of Sapindaceae based on targeted enrichment data. *Am J Bot* 108: 1234–1251.

Buys, M.H., & Hilger, H.H. 2003. Boraginaceae cymes are exclusively scorpioid and not helicoid. *TAXON* 52: 719–724.

Chacón, J., Luebert, F., Hilger, H.H., Ovchinnikova, S., Selvi, F., Cecchi, L., Guilliams, C.M., Hasenstab-Lehman, K., Sutorý, K., Simpson, M.G., & Weigend, M. 2016. The borage family (Boraginaceae s.str.): A revised infrafamilial classification based on new phylogenetic evidence, with emphasis on the placement of some enigmatic genera. *TAXON* 65: 523–546.

Chacón, J., Luebert, F., Selvi, F., Cecchi, L., & Weigend, M. 2019. Phylogeny and historical biogeography of Lithospermeae (Boraginaceae): Disentangling the possible causes of Miocene diversifications. *Molecular Phylogenetics and Evolution* 141: 106626.

Clarkson, J.J., Zuntini, A.R., Maurin, O., Downie, S.R., Plunkett, G.M., Nicolas, A.N., Smith, J.F., Feist, M.A.E., Gutierrez, K., Malakasi, P., Bailey, P., Brewer, G.E., Epitawalage, N., Zmarzty, S., Forest, F., & Baker, W.J. 2021. A higher-level nuclear phylogenomic study of the carrot family (Apiaceae). *Am J Bot* 108: 1252–1269.

Cohen, J. 2022. Phylogenomics, floral evolution, and biogeography of Lithospermum L. (Boraginaceae). *Molecular Phylogenetics and Evolution* 166: 107317.

De Candolle, A.P. 1845. Borragineae. Pp. 466–566 in: *De Candolle, A.L.P.P., (ed.), Prodromus Systematis Naturalis Regni Vegetabilis*, vol. 9. Paris: Treuttel & Würtz.

De Candolle, A.P. 1846. Borragineae. Pp. 1–178 in: *De Candolle, A.L.P.P., (ed.), Prodromus Systematis Naturalis Regni Vegetabilis*, vol. 10. Paris: Treuttel & Würtz.

Di Fulvio, T.E., Cosa, M.T., & Dottori, N. 1997. Morfología y vascularización floral en *Turricula*, *Eriodictyon* y *Wigandia* (Phacelieae; Hydrophyllaceae) con relación a la taxonomía. *Kurtziana* 25: 47–66.

Di Fulvio, T.E., Cosa, M.T., & Dottori, N. 1999. Morfología y vascularización floral de *Draperia*, *Emmenanthe*, *Hesperochiron*, *Romanzoffia* y *Tricardia* (Phacelieae, Hydrophyllaceae). *Kurtziana* 27: 187–209.

Diane, N., Förther, H., & Hilger, H.H. 2002. A systematic analysis of *Heliotropium*, *Tournefortia*, and allied taxa of the Heliotropiaceae (Boraginales) based on ITS1 sequences and morphological data. *American Journal of Botany* 89: 287–295.

Diane, N., Hilger, H.H., Förther, M., Weigend, M., & Luebert, F. 2016. Heliotropiaceae. Pp. 203–211 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*, vol. XIV. Heidelberg: Springer

Doyle, J.J., & Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem.bull*, 19: 11–15.

Feng, Y.-L., Wicke, S., Li, J.-W., Han, Y., Lin, C.-S., Li, D.-Z., Zhou, T.-T., Huang, W.-C., Huang, L.-Q., & Jin, X.-H. 2016. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biology and Evolution* 8: 2164–2175.

Ferguson, D.M. 1998. Phylogenetic analysis and relationships in Hydrophyllaceae based on *ndhF* sequence data. *Systematic Botany* 23: 253–268.

Frohlich, M.W., Sage, R.F., Craven, L.A., Schuster, S., Gigot, G., Hilger, H.H., Akhani, H., Mahdavi, P., Luebert, F., Weigend, M., Thulin, M., Doyle, J.J., Doyle, J.L., Vogan, P., Forrest, A., Fulcher, T.K., Devey, D.S., & Chase, M.W. 2022. Molecular phylogenetics of *Euploca* (Boraginaceae): homoplasy in many characters, including the C4 photosynthetic pathway. *Botanical Journal of the Linnean Society* 199: 497–537. https://doi.org/10.1093/botlinnean/boab082

Gottschling, M., Diane, N., Hilger, H.H., & Weigend, M. 2004. Testing hypotheses on disjunctions present in the primarily woody Boraginales: Ehretiaceae, Cordiaceae, and Heliotropiaceae, inferred from ITS1 sequence data. *International Journal of Plant Sciences* 165: S123–S135.

Gottschling, M., & Hilger, H.H. 2004. Characterisation of a novel fruit type found in *Ehretia* (Ehretiaceae, Boraginales). *Blumea - Biodiversity, Evolution and Biogeography of Plants* 49: 145–153.

Gottschling, M., Hilger, H.H., Wolf, M., & Diane, N. 2001. Secondary structure of the ITS1 transcript and its application in a reconstruction of the phylogeny of Boraginales. *Plant Biol (Stuttg)* 3: 629–636.

Gottschling, M., Luebert, F., Hilger, H.H., & Miller, J.S. 2014a. Molecular delimitations in the Ehretiaceae (Boraginales). *Molecular Phylogenetics and Evolution* 72: 1–6.

Gottschling, M., Miller, J.S., Weigend, M., & Hilger, H.H. 2005. Congruence of a phylogeny of Cordiaceae (Boraginales) inferred from ITS1 sequence data with morphology, ecology, and biogeography. *Annals of the Missouri Botanical Garden* 92: 425–437.

Gottschling, M., Nagelmüller, S., & Hilger, H.H. 2014b. Generative ontogeny in *Tiquilia* (Ehretiaceae: Boraginales) and phylogenetic implications. *Biol J Linn Soc* 112: 520–534. https://doi.org/10.1111/bij.12266

Gürke, M. 1893. Borraginaceae (Asperifoliaceae). Pp. 71–131 in: Engler, A., Prantl, K. (eds.), *Die Natürlichen Pflanzenfamilien*, vol. IV, 3a. Leipzig: Engelmann.

Hasenstab-Lehman, K. 2017. Phylogenetics of the Borage family: delimiting Boraginales and assessing closest relatives. *Aliso* 35: 41–49.

Hasenstab-Lehman, K.E., & Simpson, M.G. 2012. Cat's Eyes and Popcorn Flowers: Phylogenetic Systematics of the Genus Cryptantha s. l. (Boraginaceae). *Systematic Botany* 37: 738–757.

Heigl, H.M.L., Kretschmann, J., Hilger, H.H., & Gottschling, M. 2020. Flower and fruit anatomy of *Cordia nodosa* Lam. and *Varronia bonplandii* Desv. (Cordiaceae, Boraginales) with phylogenetic implications. *Org Divers Evol* 20: 9–24.

Hilger, H., & Diane, N. 2003. A systematic analysis of Heliotropiaceae (Boraginales) based on *trnL* and ITS1 sequence data. *Botanische Jahrbücher* 125: 19–51.

Hilger, H.H. 1987. Flower and fruit development in *Wigandia caracasana* (Hydrophyllaceae). *American Journal of Botany* 74: 250–259.

Hilger, H.H. 1992. Morphology of *Heliotropium* (Boraginaceae) dispersal units. *Botanica Acta* 105: 387–393.

Hilger, H.H. 2014. Ontogeny, morphology, and systematic significance of glochidiate and winged fruits of Cynoglosseae and Eritrichieae (Boraginaceae). *Plant Diversity and Evolution* 131: 167–214.

Hilger, H.H., Selvi, F., Papini, A., & Bigazzi, M. 2004. Molecular systematics of Boraginaceae tribe Boragineae based on ITS1 and *trnL* sequences, with special reference to *Anchusa* s.l. *Annals of Botany* 94: 201–212.

Hilger, H.H., & Weigend, M. 2016. Wellstediaceae. In: Kadereit JW, Bittrich V, eds. Pp. 403–406 in: *Families and genera of vascular plants.* Heidelberg: Springer.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., & Vinh, L.S. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35: 518–522.

Holstein, N., & Gottschling, M. 2018. Flowers of *Halgania* (Ehretiaceae, Boraginales) are set up for being buzzed and the role of intertwining anther trichomes. *Flora* 240: 7–15.

Jeiter, J. 2020. *Evolution of floral organisation and architecture in Boraginales and Geraniales*. Dissertation, University of Bonn, Bonn, Germany.

Jeiter, J., Danisch, F., & Hilger, H.H. 2016. Polymery and nectary chambers in *Codon* (Codonaceae): Flower and fruit development in a small, capsule-bearing family of Boraginales. *Flora - Morphology, Distribution, Functional Ecology of Plants* 220: 94–102.

Jeiter, J., Staedler, Y.M., Schönenberger, J., Weigend, M., & Luebert, F. 2018. Gynoecium and fruit development in *Heliotropium* sect. *Heliothamnus* (Heliotropiaceae). *International Journal of Plant Sciences* 179: 275–286.

Jeiter, J., Vasile, M.-A., Lewin, E.M., & Weigend, M. 2023. The odd one out: a comparison of flower and fruit development in holoparasitic *Pholisma arenarium* (Lennoaceae, Boraginales) to that in closely related Ehretiaceae. *International Journal of Plant Sciences* 184: 1–18.

Jeiter, J., & Weigend, M. 2018. Simple scales make complex compartments: ontogeny and morphology of stamen–corolla tube modifications in Hydrophyllaceae (Boraginales). *Biol J Linn Soc* 125: 802–820.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C., & Wickett, N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment1. *Appl Plant Sci* 4: apps.1600016.

Johnson, M.G., Pokorny, L., Dodsworth, S., Botigué, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., Forest, F., Kim, J.T., Leebens-Mack, J.H., Leitch, I.J., Maurin, O., Soltis, D.E., Soltis, P.S., Wong, G.K., Baker, W.J., & Wickett, N.J. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-nedoids clustering. *Systematic Biology* 68: 594–606.

Johnston, I.M. 1930. Studies in the Boraginaceae VIII. *Contr. Gray Herb.* 1–95.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., & Jermiin, L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14: 587–589.

Katoh, K., & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30: 772–780.

Li, H., & Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.

Luebert, F. 2013. A revision of *Heliotropium* sect. *Cochranea* (Heliotropiaceae). *Kew Bull* 68: 1–54.

Luebert, F., Brokamp, G., Wen, J., Weigend, M., & Hilger, H.H. 2011a. Phylogenetic relationships and morphological diversity in Neotropical *Heliotropium* (Heliotropiaceae). *TAXON* 60: 663–680.

Luebert, F., Cecchi, L., Frohlich, M.W., Gottschling, M., Guilliams, C.M., Hasenstab-Lehman, K.E., Hilger, H.H., Miller, J.S., Mittelbach, M., Nazaire, M., Nepi, M., Nocentini, D., Ober, D., Olmstead, R.G., Selvi, F., Simpson, M.G., Sutorý, K., Valdés, B., Walden, G.K., & Weigend, M. 2016. Familial classification of the Boraginales. *TAXON* 65: 502–522.

Luebert, F., Couvreur, T.L.P., Gottschling, M., Hilger, H.H., Miller, J.S., & Weigend, M. 2017. Historical biogeography of Boraginales: West Gondwanan vicariance followed by long-distance dispersal? *Journal of Biogeography* 44: 158–169.

Luebert, F., Hilger, H.H., & Weigend, M. 2011b. Diversification in the Andes: age and origins of South American *Heliotropium* lineages (Heliotropiaceae, Boraginales). *Molecular Phylogenetics and Evolution* 61: 90–102.

Luebert, F., & Wen, J. 2008. Phylogenetic analysis and evolutionary diversification of *Heliotropium* sect. *Cochranea* (Heliotropiaceae) in the Atacama Desert. *Systematic Botany*. 32: 390-–402

Maurin, O., Anest, A., Bellot, S., Biffin, E., Brewer, G., Charles-Dominique, T., Cowan, R.S., Dodsworth, S., Epitawalage, N., Gallego, B., Giaretta, A., Goldenberg, R., Gonçalves, D.J.P., Graham, S., Hoch, P., Mazine, F., Low, Y.W., McGinnie, C., Michelangeli, F.A., Morris, S., Penneys, D.S., Pérez Escobar, O.A., Pillon, Y., Pokorny, L., Shimizu, G., Staggemeier, V.G., Thornhill, A.H., Tomlinson, K.W., Turner, I.M., Vasconcelos, T., Wilson, P.G., Zuntini, A.R., Baker, W.J., Forest, F., & Lucas, E. 2021.

A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set. *American Journal of Botany* 108: 1087–1111.

McLay, T.G.B., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, E.M., Simpson, L., Schmidt-Lebuhn, A.N., Baker, W.J., Forest, F., & Jackson, C.J. 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* 9.

Mendes, F.K., & Hahn, M.W. 2018. Why concatenation fails near the anomaly zone. *Systematic Biology* 67: 158–169.

Miller, J.S., & Gottschling, M. 2007. Generic classification in the Cordiaceae (Boraginales): resurrection of the genus *Varronia* P. Br. *TAXON* 56: 163–169.

Moonlight, P.W., & Cardoso, D.B.O.S. 2023. A taxonomic revision of *Keraunea*, including three new species and its phylogenetic realignment with Ehretiaceae (Boraginales). *PhytoKeys* 219: 145–170.

Moore, M.J., & Jansen, R.K. 2006. Molecular evidence for the age, origin, and evolutionary history of the American desert plant genus *Tiquilia* (Boraginaceae). *Molecular Phylogenetics and Evolution* 39: 668–687.

Murphy, B., Forest, F., Barraclough, T., Rosindell, J., Bellot, S., Cowan, R., Golos, M., Jebb, M., & Cheek, M. 2020. A phylogenomic analysis of Nepenthes (Nepenthaceae). *Mol Phylogenet Evol* 144: 106668.

Nauheimer, L., Weigner, N., Joyce, E., Crayn, D., Clarke, C., & Nargar, K. 2021. HybPhaser: A workflow for the detection and phasing of hybrids in target capture data sets. *Applications in Plant Sciences* 9.

Nazaire, M., & Hufford, L. 2012. A broad phylogenetic analysis of Boraginaceae: implications for the relationships of *Mertensia*. *Systematic Botany* 37: 758–783.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., & Minh, B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32: 268–274.

Olmstead, R., & Ferguson, D. 2001. A molecular phylogeny of the Boraginaceae/Hydrophyllaceae., vol. systematic section, abstract 521. Presented at the BSA Congress Botany, Albuquerque (New Mexico).

Otero, A., Jiménez-Mejías, P., Valcárcel, V., & Vargas, P. 2019. Being in the right place at the right time? Parallel diversification bursts favored by the persistence of ancient epizoochorous traits and hidden factors in Cynoglossoideae. *American Journal of Botany* 106: 438–452.

Ottenlips, M.V., Mansfield, D.H., Buerki, S., Feist, M.A.E., Downie, S.R., Dodsworth, S., Forest, F., Plunkett, G.M., & Smith, J.F. 2021. Resolving species boundaries in a recent radiation with the Angiosperms353 probe set: the *Lomatium packardiae*/*L. anomalum* clade of the *L. triternatum* (Apiaceae) complex. *American Journal of Botany* 108: 1217–1233.

Pezzini, F.F., Ferrari, G., Forrest, L.L., Hart, M.L., Nishii, K., & Kidner, C.A. 2023. Target capture and genome skimming for plant diversity studies. *Applications in Plant Sciences* 11: e11537.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Refulio-Rodriguez, N.F., & Olmstead, R.G. 2014. Phylogeny of Lamiidae. *American Journal of Botany* 101: 287–299.

Roch, S., & Steel, M. 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theor Popul Biol* 100C: 56–62.

Schneider, A.C., Braukmann, T., Banerjee, A., & Stefanović, S. 2018. Convergent plastome evolution and gene loss in holoparasitic Lennoaceae. *Genome Biol Evol* 10: 2663–2670.

Selvi, F., Bigazzi, M., Hilger, H.H., & Papini, A. 2006. Molecular phylogeny, morphology and taxonomic recircumscription of the generic complex *Nonea*/*Elizaldia*/*Pulmonaria*/*Paraskevia* (Boraginaceae-Borragineae). *TAXON* 55: 907–918.

Shah, T., Schneider, J.V., Zizka, G., Maurin, O., Baker, W., Forest, F., Brewer, G.E., Savolainen, V., Darbyshire, I., & Larridon, I. 2021. Joining forces in Ochnaceae phylogenomics: a tale of two targeted sequencing probe kits. *Am J Bot* 108: 1201–1216.

Simpson, M.G., Johnson, L.A., Villaverde, T., & Guilliams, C.M. 2017. American amphitropical disjuncts: Perspectives from vascular plant analyses and prospects for future research. *American Journal of Botany* 104: 1600–1650.

Smith, R.A., Ferguson, D.M., Barkman, T.J., & Depamphilis, C.W. 2000. Molecular phylogenetic evidence for the origin of Lennoaceae: a case of adelphoparasitism in the angiosperms. *Am. J. Bot. S* 87: 158.

Smith, S.A., Walker-Hale, N., Walker, J.F., & Brown, J.W. 2020. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Systematic Biology* 69: 579–592.

Solms-Laubach, H.G. 1870. *Die Familie der Lennoaceen*. Halle: Schmidt.

Su, H.-J., Hu, J.-M., Anderson, F.E., Der, J.P., & Nickrent, D.L. 2015. Phylogenetic relationships of Santalales with insights into the origins of holoparasitic Balanophoraceae. *TAXON* 64: 491–506.

Takhtajan, A. 1987. *Systema Magnoliophytorum.* Leningrad: Soviet Sciences Press.

Takhtajan, A. 1997. *Diversity and Classification of Flowering Plants.* New York: Columbia University Press.

Taylor, S.E. 2012. *Molecular systematics and the origins of gypsophily in* Nama *L. (Boraginaceae)*. Dissertation, University of Texas, Austin, Texas, U.S.A.

Teppner, H. 2011. Flowers of Boraginaceae (*Symphytum*, *Onosma*, *Cerinthe*) and *Andrena symphyti* (Hymenoptera-Andrenidae) – Morphology, pollen portioning, vibratory pollen collection, nectar robbing. *Phyton (Horn, Austria)* 50: 145–328.

Thomas, A.E., Igea, J., Meudt, H.M., Albach, D.C., Lee, W.G., & Tanentzap, A.J. 2021a. Using target sequence capture to improve the phylogenetic resolution of a rapid radiation in New Zealand *Veronica*. *American Journal of Botany* 108: 1289–1306.

Thomas, D.C., Weigend, M., & Hilger, H.H. 2008. Phylogeny and systematics of *Lithodora* (Boraginaceae—Lithospermeae) and its affinities to the monotypic genera *Mairetis*, *Halacsya* and *Paramoltkia* based on ITS1 and trnLUAA-sequence data and morphology. *TAXON* 57: 79–97.

Thomas, S.K., Liu, X., Du, Z.-Y., Dong, Y., Cummings, A., Pokorny, L., Xiang, Q.-Y. (Jenny), & Leebens-Mack, J.H. 2021b. Comprehending Cornales: phylogenetic reconstruction of the order using the Angiosperms353 probe set. *American Journal of Botany* 108: 1112–1121.

Tumescheit, C., Firth, A.E., & Brown, K. 2022. CIAlign: A highly customisable command line tool to clean, interpret and visualise multiple sequence alignments. *PeerJ* 10: e12983.

Vasile, M.-A., Jeiter, J., Weigend, M., & Luebert, F. 2020. Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia*. *Systematics and Biodiversity* 18: 757–770.

Vasile, M.-A., Luebert, F., Jeiter, J., & Weigend, M. 2021. Fruit evolution in Hydrophyllaceae. *American Journal of Botany* 108: 925–945. https://doi.org/10.1002/ajb2.1691

Vasile, M.-A., Weigend, M., Ziegler, A., & Jeiter, J. 2022. Flower and fruit development of three species of hydrophylleae sheds new light on flower evolution in Hydrophyllaceae. *International Journal of Plant Sciences* 183: 331–347.

Vialle, R.A., Tamuri, A.U., & Goldman, N. 2018. Alignment modulates ancestral sequence reconstruction accuracy. *Molecular Biology and Evolution* 35: 1783–1797.

Walden, G.K., Garrison, L.M., Spicer, G.S., Cipriano, F.W., & Patterson, R. 2014. Phylogenies and chromosome evolution of *Phacelia* (Boraginaceae: Hydrophylloideae) inferred from nuclear ribosomal and chloroplast sequence data. *Madr* 61: 16–47.

Warnow, T. 2015. Concatenation analyses in the presence of incomplete lineage sorting. *PLoS Curr* 7: ecurrents.currents.tol.8d41ac0f13d1abedf4c4a59f5d17b1f7.

Weeks, A., Baird, K.E., & McMullen, C.K. 2010. Origin and evolution of endemic Galápagos *Varronia* species (Cordiaceae). *Molecular Phylogenetics and Evolution* 57: 948–954.

Weigend, M., Gottschling, M., Selvi, F., & Hilger, H.H. 2009. Marbleseeds are gromwells – Systematics and evolution of *Lithospermum* and allies (Boraginaceae tribe Lithospermeae) based on molecular and morphological data. *Molecular Phylogenetics and Evolution* 52: 755–768.

Weigend, M., & Hilger, H.H. 2010. Codonaceae-a newly required family name in Boraginales. *Phytotaxa* 10: 26–30.

Weigend, M., Luebert, F., Gottschling, M., Couvreur, T.L.P., Hilger, H.H., & Miller, J.S. 2014. From capsules to nutlets—phylogenetic relationships in the Boraginales. *Cladistics* 30: 508–518.

Weigend, M., Luebert, F., Selvi, F., Brokamp, G., & Hilger, H.H. 2013. Multiple origins for Hound's tongues (*Cynoglossum* L.) and Navel seeds (*Omphalodes* Mill.) – The phylogeny of the borage family (Boraginaceae s.str.). *Molecular Phylogenetics and Evolution* 68: 604–618.

Weigend, M., Selvi, F., Thomas, D.C., & Hilger, H.H. 2016. Boraginaceae. Pp. 41–102 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*. Heidelberg: Springer.

Xiong, H., Wang, D., Shao, C., Yang, X., Yang, J., Ma, T., Davis, C.C., Liu, L., & Xi, Z. 2022. Species tree estimation and the impact of gene loss following whole-genome duplication. *Systematic Biology* 71: 1348–1361.

Yatskievych, G., & Mason, C.T. 1986. A revision of the Lennoaceae. *Systematic Botany* 11: 531–548.

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.

Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., Rees, M., Frohlich, M.W., Qi, J., Weigend, M., & Ma, H. 2020. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Molecular Biology and Evolution* 37: 3188–3210.

Zhou, W., Soghigian, J., & Xiang, Q.-Y. (J.). 2022. A new pipeline for removing paralogs in target enrichment data. *Systematic Biology* 71: 410–425.

General conclusions



Chapter 7

Omphalodes nitida (Willd.) Hoffmanns. & Link (Boraginaceae) in cultivation in Bonn Botanical Gardens.

"We are always slow in admitting any great change of which we do not see the intermediate steps"

Charles Darwin, On Natural Selection

7. General Conclusions

The following discussion attempts to address the three hypotheses formulated in chapter 1. Section 1 addresses the first hypothesis (H1) and the related questions (Q1-Q4) by discussing the findings of chapters 2 and 3, but also 6 in order to give a comprehensive overview of the evolution of the basal grade of Boraginales II. Section 2 deals with the second hypothesis (H2) and the related questions (Q5-Q7) which were addressed in chapters 4 and 5, aiming to provide concrete answers about significant processes and drivers of fruit evolution in Boraginales II. Section 3 refers to third hypothesis (H3) and the related questions (Q8, Q9) and not only discusses the results of chapter 6, but also combines all the findings of this thesis to propose a comprehensive evolutionary series of boraginalean fruits. Finally, section 4 provides an overview of future research opportunities following this work.

7.1. Insights into the evolutionary history of the basal grade of Boraginales II

The basal grade of Boraginales II consist of two primarily herbaceous families (Hydrophyllaceae and Namaceae) which differ from the remaining lineages of the clade mainly in having capsular fruits with four to many seeds rather than four-seeded schizocarps or drupes. However, delimitation of Hydrophyllaceae from Namaceae only based on morphology, has been difficult. There is no strong morphological evidence neither for the separation of the two families nor for their treatment as one family (Luebert & al., 2016). Both families share very similar flower and fruit traits, with only distinct differences being that Namaceae lack almost always corolla scales (Brand, 1913) and have exclusively pluri-ovulate ovaries, while Hydrophyllaceae often possess corolla scales and can have either pluri- or four-ovulate ovaries. Four-ovulate ovaries arose several times independently and also transitioned back to pluri-ovulate ovaries (Vasile & al., 2021). Nevertheless, Vasile & al. (2021) overlooked that anatomical studies of Di Fulvio (1991) showed that *Nama californica* (A.Gray) J.D.Bacon has a unilocular ovary with four ovules, indicating that at least one transition to four ovules occurred also in Namaceae. What is clearly variable and often characteristic for individual groups of the two families, is the habit and the leaf morphology (Brand, 1913). Hydrophyllaceae consist of annual or perennial herbs of various sizes, with very variable leaf morphology (lamina can be widely ovate, narrowly or triangular-ovate, obovate; leaf margins can be entire to pinnatisect or even palmatifid; Brand, 1913; Hofmann & al., 2016). Namaceae, on the other hand, present less variable leaf morphology (usually entire leaf margins or crenate), but more distinct habit for individual groups: *Nama* L. is herbaceous (or subshrub); *Eriodictyon* Benth. (incl. *Turricula* J.F.Macbr.) includes perennial herbs, subshrubs and shrubs; *Wigandia* Kunth. can be a tall shrub or a small tree (Hitchcock, 1933a, 1933b; Hofmann & al., 2016). By and large, the vast majority of morphological characters of Hydrophyllaceae and Namaceae are pleisiomorphic (Ferguson, 1998b) and often highly plastic, and this is reflected in taxonomy, especially with the frequent occurrence of species complexes (e.g., Heckard, 1960; Levy & al., 2021). It is therefore reasonable that Hydrophyllaceae and Namaceae were traditionally treated as one family (Brand, 1913; Heckard, 1963; Bacon, 1987; Di Fulvio & al., 1997; Hofmann & al., 2016), until the advent of molecular data. Ferguson (1998b) was the first to address the relationships of Hydrophyllaceae and Namaceae using the chloroplast gene *ndhF* and the nuclear ITS region, and she did not find sufficient evidence to support their separation. However, she was able to identify that characters useful for delimitation are highly homoplasious and only very few, such as placentation, are less so. Subsequent phylogenetic studies were more or less congruent to these results (Nazaire & Hufford, 2012; Weigend & al., 2014). Although this thesis does not explicitly support the treatment of Hydrophyllaceae and Namaceae as separate families, it confirms most of the aforementioned findings and provides clear and new insights into their evolutionary history, which appears to be more complex than expected.

7.1.1. Phylogenetic relationships and historical biogeography

Delimitation of different taxonomic groups (families, tribes, genera, species) within the basal grade of Boraginales II based on morphology alone, is clearly challenging and has long been an impediment to a modern taxonomic revision. The problem also persists due to the scarcity of phylogenetic data, most of which, suffer from either limited taxon sampling, or low resolution, or both (Ferguson, 1998a, 1998b; Hansen & al., 2009; Nazaire & Hufford, 2012; Walden & al., 2014; Weigend & al., 2014). By increasing taxonomic and geographic representation of sampling across Hydrophyllaceae and Namaceae when using standard plastid (*ndhF*) or nuclear (ITS) markers, increased phylogenetic resolution of interfamilial, infrafamilial and intergeneric relationships was expected (see H1 in chapter 1). Indeed, the extended sampling of the two families, analyzed together with an extended sampling of the entire order (see the complete *ndhF* tree; fig. S1 in Appendix A) significantly improved the topology recovered by Ferguson (1998a). Ferguson retrieved Hydrophyllaceae and Namaceae as sister to each other with very low support, while here, the two clades were retrieved as successive sister, exactly as in the combined plastid dataset (*ndhF*, *rbcL*, *rps16*, and *trnL–trnF*) of Weigend & al. (2014), where Namaceae is successive sister to the remaining lineages of Boraginales II. The infrafamilial phylogenetic relationships based on the *ndhF* gene, are highly congruent with previous studies (Ferguson, 1998a; Walden & al., 2014), but with increased resolution and better support values in most cases. The three main clades recovered within Hydrophyllaceae and the two within Namaceae (see the detailed discussion of chapter 2) are congruent with the recent classification provided by Luebert & al. (2016) and reflect the heterogeneity in habit and leaf morphology, but also in placentation (see section 7.3.2). However, interspecific relationships, especially in recently diversified clades, remained largely unresolved. Nuclear and plastid markers are significantly heterogeneous in Boraginales (Ferguson, 1998b; Moore & Jansen, 2006; Nazaire & Hufford, 2012; Taylor, 2012; Walden & al., 2014; Chacón & al., 2016), and therefore were not combined as done elsewhere. The nuclear marker (ITS) resulted in similar topologies with the plastid marker (*ndhF*), but with weaker support for the deep nodes and often better resolution for the shallower internal nodes which is in agreement with previous studies (Hansen & al., 2009; Walden & al., 2014). Namaceae were retrieved as sister to Hydrophyllaceae with high support (see figs. S2, S3 in Appendix A), but this result should be treated with caution due to the use of one marker and a very limited outgroup (two samples).

The biogeographic analysis was based on the *ndhF* dataset because, as its completeness allowed the estimation of divergence times in Boraginales using multiple fossil calibrations (see fig. S4 in Appendix A), but also because the ITS alignment across Boraginales can be unreliable due to extensive sequence variation. The analysis indicates that both Hydrophyllaceae and Namaceae originated in North America. This finding is congruent with the previous hypothesis of Luebert & al. (2017) suggesting that a colonization of North America happened early in the history of both families, even before the divergence of these lineages from the other clades of Boraginales II. A similar hypothesis was formulated by Ferguson (1998b), who suggested that the origin of the basal grade of Boraginales II coincided with a migration event from the tropics to the temperate zone, primarily into the California Floristic Province, where about half of the species are native. She further underlined that this fits a trend in angiosperm evolution called the tropical/temperate family-pairs hypothesis. Interestingly, in this study (chapter 2) age estimates for the crown nodes of Hydrophyllaceae (66.1 (95% HPD 51.7–83.1 Ma)) and Namaceae (66.8 (95% HPD 46–88 Ma)) were pushed back compared to Luebert & al. (2017), which can be justified by the increased ingroup sampling (Soares & Schrago, 2015). Although all these findings shed new light on the evolutionary history of the basal grade of Boraginales II, the timing of the colonization of North America, the parallel diversification, the independent dispersal events to South America, as well as the

ancestral area reconstruction do not provide clear support for treating the basal grade as two separate families.

The phylogenomic dataset of Zhang & al. (2020) succeeding this study (chapter 2), presented a sister relationship of the Namaceae and Hydrophyllaceae with maximum support. Therefore, the sister relationship of the two clades and their phylogenetic position within the order was tested in the present thesis with a target capture sequencing approach using 353 loci (chapter 6). The taxon sampling across the order was extensive, while within the basal grade of Boraginales II, 13 out of 16 genera were represented, at least by one accession, in contrast to the more limited taxon sampling of Zhang & al. (2020). The inclusion of *Nama dichotoma* (Ruiz & Pav.) Choisy as a representative of the *Nama* clade and its sister relationship to the Hydrophyllaceae + Namaceae clade, confirmed once again that we cannot draw solid conclusions about the relationship of the two groups. Formerly, each group separately was always strongly supported as monophyletic (Ferguson, 1998a; Gottschling & al., 2001; Moore & Jansen, 2006; Luebert & Wen, 2008; Nazaire & Hufford, 2012; Weigend & al., 2014). Now, chapter 6 also questions their monophyly. An urgent next step would be to significantly expand the sampling, especially within the diverse and understudied genus *Nama*, while implementing a phylogenomic approach.

7.1.2. Species delimitation in absence of distinct morphological characters.

Interspecific relationships, particularly those of recently diversified clades, often remain unresolved on phylogenetic trees reconstructed based on a few plastid or nuclear markers (Spangler & al., 1999; Givnish & al., 2007; Mitsui & al., 2008; Wen & al., 2008). As expected, although ITS performed better at resolving shallower relationships compare to *ndhF*, both markers are not adequate for species delimitation (see H1 in chapter 1). This issue is more severe in younger clades. Some problematic groups are the sub-clades of *Phacelia* sect. *Phacelia* and *P.* sect. *Glandulosae* which include amphitropically disjunct species, but also the *Eriodictyon*-clade. Subtropical *Wigandia* might have started diversifying already by the end of the Paleogene, but apparently has a very complex biogeographical history (see AAR in chapter 2) and likely high genetic admixture, which also obstructs species delimitation based on few loci. All these groups have nonetheless one thing in common: they often lack distinct morphological characters at species level. As a result, the establishment of a stable classification system has been problematic.

In chapter 3, a ddRADseq approach for species delimitation was tested for one of the aforementioned clades, the South American *Phacelia* sect. *Glandulosae*. The phylogenetic results were combined with a methodology adopted in any classical taxonomic revision, since taxonomy still is a phenotypic science. Defining distinct morphotypes was extremely difficult due to phenotypic overlap in regards to all the potentially diagnostic characters (corolla scales, gynoecium and seed morphology), but also due to the existence of specimens with intermediate phenotypes between the previously defined morphospeices. The underlying hypothesis of chapter 3 was that these intermediate forms represent new species and that different morphotypes within *P.* sect. *Glandulosae* will fall into monophyla on the ddRAD phylogeny. This hypothesis is clearly rejected. There is clear ecogeographical differentiation, but the morphospecies do not represent defined monophyla and we therefore suggested the recognition of only two species in South American sect. *Glandulosae*: the polymorphic species *P. pinnatifida* Griseb. ex Wedd. and its sister species *P. artemisioides* Griseb.

Our approach in chapter 3 highlights prominent issues of modern taxonomy, which are often a reason for conflict among taxonomists. One issue is species inflation, which has huge implications for understanding, for example, the phylogenetic relationships and biogeographic patterns or for estimating the

diversification rates and the ecological niches, and of course for conservation assessments (Bonneuil, 2002; Isaac & al., 2004; Knapp & al., 2005; Karremans & al., 2020). Rapid radiations and incomplete lineage shortage, phenotypic plasticity due to ecogeographical factors, polyploidization, hybridization and genetic drift can indeed complicate species delimitation (Naciri & Linder, 2015; Boluda & al., 2021; Karbstein & al., 2022; Oberprieler & al., 2023) and all these phenomena are more than common across Boraginales (Britton, 1951; Walden & al., 2014; Kook & al., 2015; Weigend & al., 2016; Zhang & al., 2020), but also across angiosperms (Ren & al., 2018; Hörandl, 2022). With the advent of novel molecular phylogenetic techniques all these evolutionary events are identifiable and should be considered for providing new and useful classifications. Another issue is the dangers/impact of taxonomic misidentifications due to fine, non-diagnostic morphological differences (e.g. the inclusion of *P. nana* Wedd. in *P.* sect. *Glandulosae*, previously assigned to *P.* sect. *Euglypta*; chapter 3), but also due to misleading homoplasious characters. For example, *Codon* L. (Codonaceae, Boraginales) and *Hydrolea* L.(Hydroleaceae, Solanales) were for over 100 years erroneously placed in Hydrophyllaceae (Boraginales) due to their gynoecial morphology and capsular fruits (Brand, 1913; Weigend & Hilger, 2010; Refulio-Rodriguez & Olmstead, 2014). Similarly, *Keraunea* Cheek & Simão-Bianchini was uncomfortably placed in Convolvulaceae until recently due to an enlarged and adnate fruit bract, while it has now been shown to belong to Ehretiaceae (Almeida & al., 2023; Moonlight & Cardoso, 2023). These misidentifications are sometimes of minor importance, like in the case of *P. nana*, but can often be a major obstacle to biological research and conservation efforts as for example in the case of the neotropical *Vanilla* Plumier ex Miller (Karremans & al., 2020). Identifying species-level biological diversity is the basis of any kind of biodiversity research, therefore, integrative taxonomic efforts as the one presented in chapter 3 are highly needed in future research in plant systematics.

7.2. Underlying processes of fruit evolution in Boraginales II

Among the two clades of Boraginales, most of the fruit diversity is concentrated in Boraginales II (Luebert & al., 2016; Jeiter & al., 2018). Extensive data have been published in the past decades on fruit morphology in order to elucidate trends in fruit evolution of Boraginales II (Berg, 1985; Bacon, 1987; Hilger, 1987; Chuang & Constance, 1992; Hilger, 1992; Di Fulvio & al., 1997; Di Fulvio, 1998; Di Fulvio & al., 1999; Hofmann, 1999; Gottschling, 2004; Gottschling & Hilger, 2004; Berg, 2009; Gottschling & al., 2014b; Holstein & Gottschling, 2018; Jeiter & al., 2018; Heigl & al., 2020; Vasile & al., 2021). A couple of these authors have identified that the development of various secondary septa subdividing the two carpels of the gynoecium is the main parameter leading to a complex fruit architecture. However, the underlining processes involved in fruit type transitions have been overlooked, and were only recently addressed in the capsular fruits of Hydrophyllaceae (Vasile & al., 2021). Vasile et al. (2021), specifically propose that reductions and multiplications of seed number are readily reversible as long as fruits remain dehiscent, and that paedomorphosis is likely a crucial process leading to a reduction of the ovule number to four. This thesis places a focus on studying the ontogeny of clade Hydrophylleae (chapter 4), one of the most diverse clades of Hydrophyllaceae and apparently of the whole order, in terms of placentation and internal ovary architecture. One of the very few lineages presenting carpel multiplication (Lennoaceae; chapter 5) is also studied. Combining the results of these two studies, reveals underlying processes involved in fruit evolution of Boraginales II and confirms the second hypothesis (H2) of this thesis.

At least two processes, heterochrony and heterotopy, have resulted in various modifications of individual structures (ovules, septa, placentation, ovary walls, ovarian cavities, compitum and transmitting tissues), and via different processes combinations and synorganization of these structures lead to

complex internal ovary architectures. Chapter 4, clearly shows how ―from the same starting point― minor heterochronous shifts in individual aspects of the developmental trajectories ―placentation, ovule initiation, ovary expansion― can lead to very different mature forms. The results of chapter 4 and 5, combined with those of Vasile et al. (2021), reveal that heterotopy is also highly common. For example, the relocation of placentae to an apical or subapical position through elongation of the basal septa (e.g. in *Tiquilia* Pers.; Gottschling & al., 2014; and in *Draperia systyla* (A.Gray) Torr.; Vasile & al., 2021), also changes ovules position, orientation and shape, as all the minor or major changes in placentation do (see chapter 4). The bifacial placentae of *Eucrypta chrysanthemifolia* (Benth.) Greene appear to be an extreme case of heterotopy combined with heterochrony. Heterotopy was identified also in Ehretiaceae and Lennoaceae. The relocation of ovules into the upper third of the ovary as well as the diminishing of the stylar cleft in such a way that the stigmata fuse and create an extragynoecial compitum in both *Pholisma arenarium* Nutt. and *Bourreria succulenta* Jacq., not only indicate heterotopy, but also a particularly close relationship between these two genera, while *Ehretia dicksonii* Hance lacks these traits and seems more divergent.

Throughout all the elements involved in the formation of complex internal ovary architecture in Boraginales II, placentation and its plasticity require more attention. Shivaprakash and Bawa (2022) suggest derived evolution of parietal placentation from axile placentation by retraction of the septa, as reported by others (Lindsey, 1940; Wilkinson, 1944). However, they do not reject the opposite scenario, such as intruding parietal placentae turning into axile placentation. This phenomenon is evident in Boraginales II. Intrusive parietal placentae, found across the basal grade, are likely the ancestral state of the crown node of Boraginales II. Within the basal grade, clades with parietal placentation emerged (Hydrophylleae; chapter 4). A transition to axile placentation most probably occurred at the crown node of Heliotropiaceae + Lennoacea + Ehretiaceae + Coldeniaceae + Hoplestigmataceae + Cordiaceae and constitutes a key character for the diversification and the evolution of various fruit types of all these lineages. Axile placentation, which is dominant in angiosperms, isolates the ovules in well-separated locules, thereby reducing the harmful effects of competition (Shivaprakash & Bawa, 2022). The isolating of one ovule per locule appears to open the way for various levels of seed and pericarp integration and the formation of a range of different dispersal units (endomericarpid, endomericarp, synendomericarpid, mericarpid, mericarp, synmericarpid; Box 1 in chapter 1). Furthermore, Shivaprakash and Bawa (2022) demonstrate that the evolution of placentation is associated with an increase or decrease of ovule number; in many families with fixed carpel numbers (e.g. Scrophulariaceae to Gesneriaceae; Ericaceae to Pyrolaceae/Monotropaceae) there is a transition from axile/apical to parietal placentation with increasing ovule number per locule. This is in line with all previous hypotheses for the evolution of seed number in Boraginales II (Weigend & al., 2014; Luebert & al., 2016; Vasile & al., 2021)

Most likely, important developmental processes that have led to novel fruit types (see Box 1 in chapter 1), such as carpel bugling (as described for Boraginales by Endress, 2011) during ovary or fruit development in some Heliotropiaceae (Hilger, 1992; Jeiter & al., 2018), in *Tiquilia* (Ehretiaceae; Gottschling & al., 2014b) and *Coldenia* L. (Coldeniaceae), and loss of dehiscence in all lineages apart from the basal grade, have taken place subsequent to changes regarding placentation and internal ovary architecture in general. The reversal to dehiscent fruits and an increase in seed number to more than four through carpel multiplication in the fleshy capsules of the parasitic Lennoaceae (chapter 5), is the ultimate proof that we should be very cautious when talking about evolutionary dead ends. Strong evolutionary drivers, such as the loss of photosynthesis and the transition to parasitism, are able to create a way out of otherwise very conserved morphologies. The assumption that parasitism itself is an evolutionary dead-end

is common, but has not been conclusively demonstrated. On the contrary, expansion of parasites' host range in order to prevent an evolutionary dead-end when a host species becomes locally extinct is possible (Nelson, 2021).

7.3. Evolution of Boraginales based on an updated phylogeny.

7.3.1. Phylogenetic relationships and systematics

Chapter 6 confirms the utility of the Angiosperms353 bait kit for obtaining a reliable Boraginales phylogeny and therefore the third hypothesis (H3) of this thesis. To a great extent, previous phylogenetic hypotheses (Ferguson, 1998a; Gottschling & al., 2005; Moore & Jansen, 2006; Luebert & al., 2011; Hasenstab-Lehman & Simpson, 2012; Nazaire & Hufford, 2012; Weigend & al., 2013, 2014; Chacón & al., 2016; Hasenstab-Lehman, 2017; Vasile & al., 2020; Zhang & al., 2020; Cohen, 2022) across different taxonomic levels find clear support, but several novelties are detected. In Boraginales I, the backbone of the most species-rich Boraginaceae, is now better resolved and supported compare to the most recent study by Chacón & al. (2016). Most of the persisting ambiguous relationships at tribal (e.g., Asperugeae + Omphalodeae) or subtribal level (mainly in Cynoglosseae) could be elucidated by strategically expanding the sampling and/or by stricter data filtering during the downstream analyses. Indication of a sister relationship of Codonaceae to Hydrophyllaceae + Namaceae by the supercontigs dataset raises new questions and this placement requires more attention in future studies. Possibly, a detailed morphological study of Wellstediaceae, still very poorly understood, could provide additional morphological evidence for a closer relationship to Codonaceae. In Boraginales II, as discussed in detail in section 7.1, the exact relationships within the basal grade is still a fuzzy point, which may be elucidated with an expanded sampling of *Nama*. Chapter 6 was altogether able to provide a very robust backbone for Boraginales II and finally, strong evidence that Lennoaceae are deeply nested within Ehretiaceae and should no longer be recognized as a separate family. Remaining questions regarding the phylogeny of Boraginales II can likely be resolved by expanding sampling across Cordiaceae, Heliotropiaceae and especially Ehretiaceae where the most incongruences between different datasets were identified. Inclusion of all the absent taxa in this study (e.g., *Hoplestigma* Pierre from Hoplestigmataceae, *Rochefortia* Sw. and *Keraunea* form Ehretiaceae, individual taxa in Omphalodeae, Microuleae etc.) will likely resolve most of the remaining issues. Undisputedly, the novel relationships here retrieved are highly supported and offer the opportunity to revisit topics such as historical biogeography and divergence times estimates of Boraginales, and test the accuracy of previously published hypotheses. Spatiotemporal studies in other plant groups based on target enrichment sequencing data have often demonstrated considerable differences to previous divergence time estimates and biogeographical scenarios (Villaverde & al., 2018; Joyce & al., 2023).

The point of reconstructing a highly supported phylogeny for Boraginales, as done in chapter 6, was not only to improve statistical support values and resolution. As mentioned in Chapter 1, in evolutionary biology, interest comes from the phenotype and function of taxa combined with knowledge of their relationships. Chapter 5 and 6 prove how powerful this combination is: the ontogenetic studies of *Pholisma arenarium* (Lennoaceae)*, Bourreria succulenta* and *Ehretia dicksonii* (Ehretiaceae) provided strong support of a closer relationship of *Pholisma* Nutt. Ex Hook. to *Bourreria* P.Browne than to *Ehretia* P.Browne, which was further supported by the phylogenomic analysis in chapter 6 and that of Zhang & al. (2020). This approach should be adopted in future studies of Boraginales and angiosperms in general. The phylogeny presented in chapter 6 paves the way for detailed studies on the evolution of group-specific morphological traits.

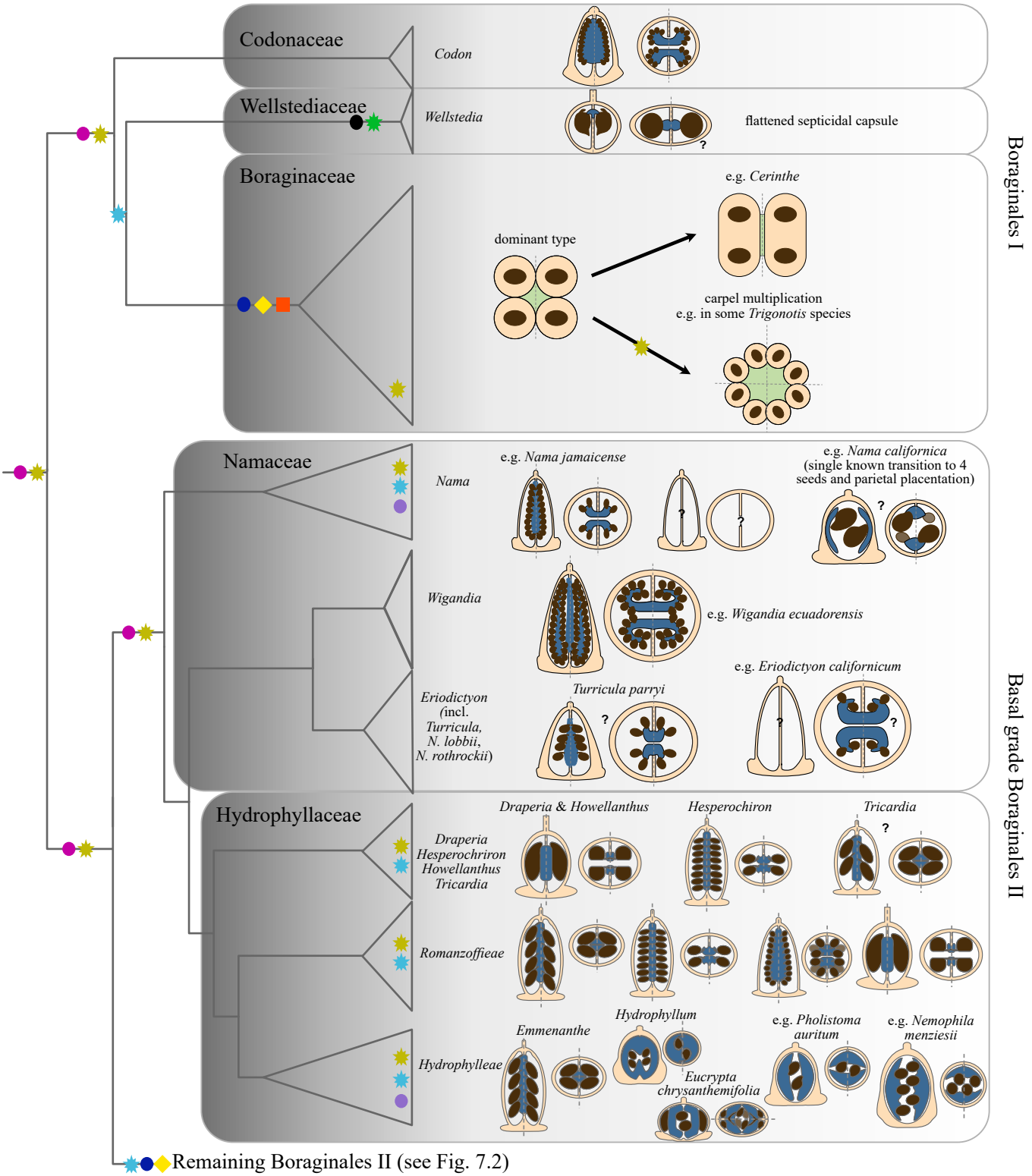
7.3.2. Fruit evolution

Most lamiids present similar trends in fruit evolution. Even though the basal grade of lamiids (Aquifoliales, Garryales, Icacinales) has predominately drupaceous fruits, recent ancestral character state reconstruction of the fruit type shows that the backbone of the remaining lamiid phylogeny have capsular fruits (Zhang & al., 2020). The evolution of placentation type, is also straight forward, with all asterids having primarily axile placentation, from which, not that frequently, other types evolved (parietal at least 22 times, apical at least 14 times, basal at least 7 times, free central at least 4 times; Shivaprakash & Bawa, 2022). Consequently, capsules and axile placentation seem to be two important starting points for fruit diversification not only in lamiids, but also in other lineages of asterids such as Ericales.

Considering these important insights into fruit evolution of asterids, one of the most successful angiosperm lineages, and having a fairly complete and well resolved phylogeny for the Boraginales (chapter 6), a fertile ground emerges for the continuation of the discussion of section 7.2. Convergent fruit evolution is apparent in the two clades of Boraginales (Figs. 7.1 & 7.2). The hypothetical evolutionary series provided here, is focused mainly on those character states which were identified as crucial for fruit evolution by this thesis. These are mainly the placentation type, but also loss of dehiscence, carpel bulging and ovule number. To clarify, the term carpel bulging here describes the separate bulging of each locule of a carpel and it is associated with the false septa and the apical septum (see also Endress, 2011). These processes are crucial for the formation of divergent, complex internal ovary architectures, the evolution of which is illustrated for each family in a hypothetical chronological order, when possible (Figs 7.1 & 7.2). According to Shivaprakash & Bawa (2022), Boraginales originated from an ancestor with axile placentation, while our and literature data indicate that the basal lineages of both Boraginales I and II have intrusive parietal placentation (Hofmann & al., 2016; Weigend & Hilger, 2016), except for Wellstediaceae, which may have apical placentation (Luebert & al., 2016). Here, it is important to clarify that the placentation of *Codon* has been erroneously defined as axile in the literature (Jeiter & al., 2016; Luebert & al., 2016), but is clearly intrusive parietal (see anatomical sections in Jeiter & al., 2016) and similar to the placentation found in Namaceae and Hydrophyllaceae. Therefore, intrusive parietal placentation is proposed as the most probable character state for the crown node of Boraginales as it seems to be a transitional state between axile and parietal placentation, and can easily evolve towards both directions as illustrated in figures 7.1 and 7.2. Furthermore, parietal placentation is connected to pluri-ovulate ovaries (Figs. 7.1 & 7.2) and permits both an increase and decrease in ovule number (the range is four to ca. 200 ovules; Vasile & al., 2021). Axile placentation of Boraginales, on the other hand, is connected to a fixed number of four ovules completely separated into four, rarely two locules. This separation affected the dynamics of ovule competition and was accompanied by a loss of dehiscence and carpel bulging, allowing for complex pericarp subdivision (in schizocarps and eremocarps). In the absence of carpel bulging we find complex subdivisions of the endocarp (in drupes). There are cases, e.g., in Cordiaceae, where the endocarp remains entire throughout development, and does not form a separate unit with each ovule/seed. The lack of a spatial constraint can affect ovule competition and therefore, Cordiaceae often have abortive ovules, resulting in one-seeded drupes (Gottschling & al., 2016; Luebert & al., 2016; Heigl & al., 2020).

Mapping fruit evolution onto the Boraginales phylogeny in detail is an extremely complex task. The evolution of eremocarps in Boraginaceae remains a completely uncharted territory although in Boraginales I there are specific character states transitions in each clade resulting in different fruit types (Fig. 7.1). We know very little about the evolution of gynobasic styles and eremocarpic fruits (Moore, 1936; Hilger, 2014), which is likely a key innovation for the extreme diversification in this clade. Joint efforts

Fruit evolution in Boraginales I and basal grade of Boraginales II



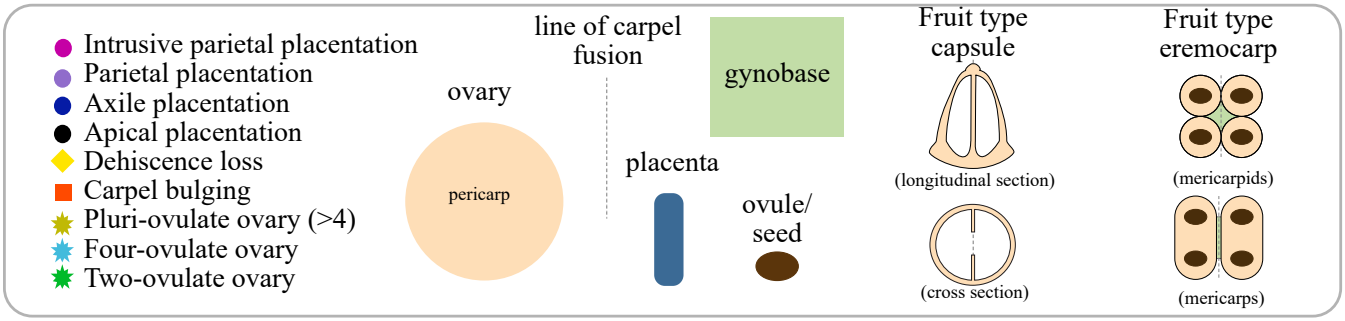


Figure 7.1 Hypothetical evolutionary series of fruit diversification in Boraginales I (Codonaceae, Wellstediaceae, Boraginaceae) and in the basal grade of Boraginales II (Namaceae, Hydrophyllaceae). Phylogenetic reconstruction is based on chapter 6 and chapter 2. Morphological characters are defined based on chapters 4 and 5 and literature data (Di Fulvio, 1991; Di Fulvio & al., 1997, 1999; Hilger & Weigend, 2016; Jeiter & al., 2016; Luebert & al., 2016; Weigend & al., 2016; Weigend & Hilger, 2016; Vasile & al., 2021). Symbols at the tips, rather than at the branches, indicate derived states that have evolved at least once within the respective clades. The schematic representation of gynoecia/fruits follows the terminology presented in Box 1 of chapter 1 and in chapter 4. Capsules are loculicidal except for Wellstediaceae. Question marks indicate uncertainty or missing data. Within Namaceae, and possibly also within Hydrophyllaceae, internal ovary architectures remain largely unexplored, and only the known types are shown here. The direction of evolution of different internal ovary architectures within the basal grade of Boriganles II is not displayed due to its complexity and due to uncertainty.

for collecting detailed ontogenetic gynoecial data on Boraginaceae, but also in gynobasic Lamiaceae, are urgently required to improve our understanding of eremocarp evolution. Fruit evolution is more nuanced in Boraginales II compared to Boraginales I, because the same fruit types arose several times independently (e.g. schizocarps; fig. 7.2). Detailed morphological studies of gynoecium and fruit development of Ehretiaceae and Heliotropiaceae, where different fruit types with a lot of different levels of seed and pericarp integration evolved, could provide new insights into the fruit evolution of the order.

7.4. Conclusion and possible directions for further studies based on this dissertation

The ever-advancing molecular techniques and computational tools in molecular evolution offer us the opportunity to get closer to the complete tree of life. A comprehensive understanding of the tree of life is required to fully understand evolutionary patterns and to chart the properties of life on Earth on a central biological roadmap (Baker & al., 2022). This dissertation may be a small, but hopefully important contribution to understanding angiosperms phylogeny and improving our knowledge of evolutionary processes and patterns of flower and fruit diversification.

Different sequencing approaches were tested, and their advantages and limitations are discussed in detail in the individual chapters 2, 3 and 6, but also in the final discussion (chapter 7). This and previous studies show that the use of standard plastid and nuclear markers can provide an overall accurate reconstruction of phylogenetic relationships of Boraginales, while plastid phylogenies are also useful for a first estimation of spatiotemporal evolution of the order or individual clades. However, there is a threshold above which phylogenetic resolution cannot be further improved regardless of the size and the coverage of the taxon sampling. Particular problematic clades of the basal grade of Boraginales II were detected and the evolutionary history of one of them (*Phacelia* sect. *Glandulosae*) was resolved using a phylogenomic approach based on double-digested RAD-seq data and an updated taxonomy is suggested. A similar approach for elucidating the evolutionary history of other clades such as *Wigandia*, *Nama* or other sections of *Phacelia* would permit the development of a much improved, stable taxonomy for the basal grade of Boraginales II. At the order level, another phylogenomic approach was tested (target enrichment sequencing), and specifically the Angiosperms353 probe set was used. This method yielded a highly resolved and supported phylogeny (chapter 6) across all different taxonomic ranks of Boraginales. Future studies should focus on individual clades, where many individual questions regarding phylogenetic relationships remain unresolved (e.g., intergeneric relationships of Ehretiaceae). This approach also opens up new possibilities for revisiting topics such as divergence time estimates and the historical biogeography of Boraginales.

Fruit evolution in Boraginales II

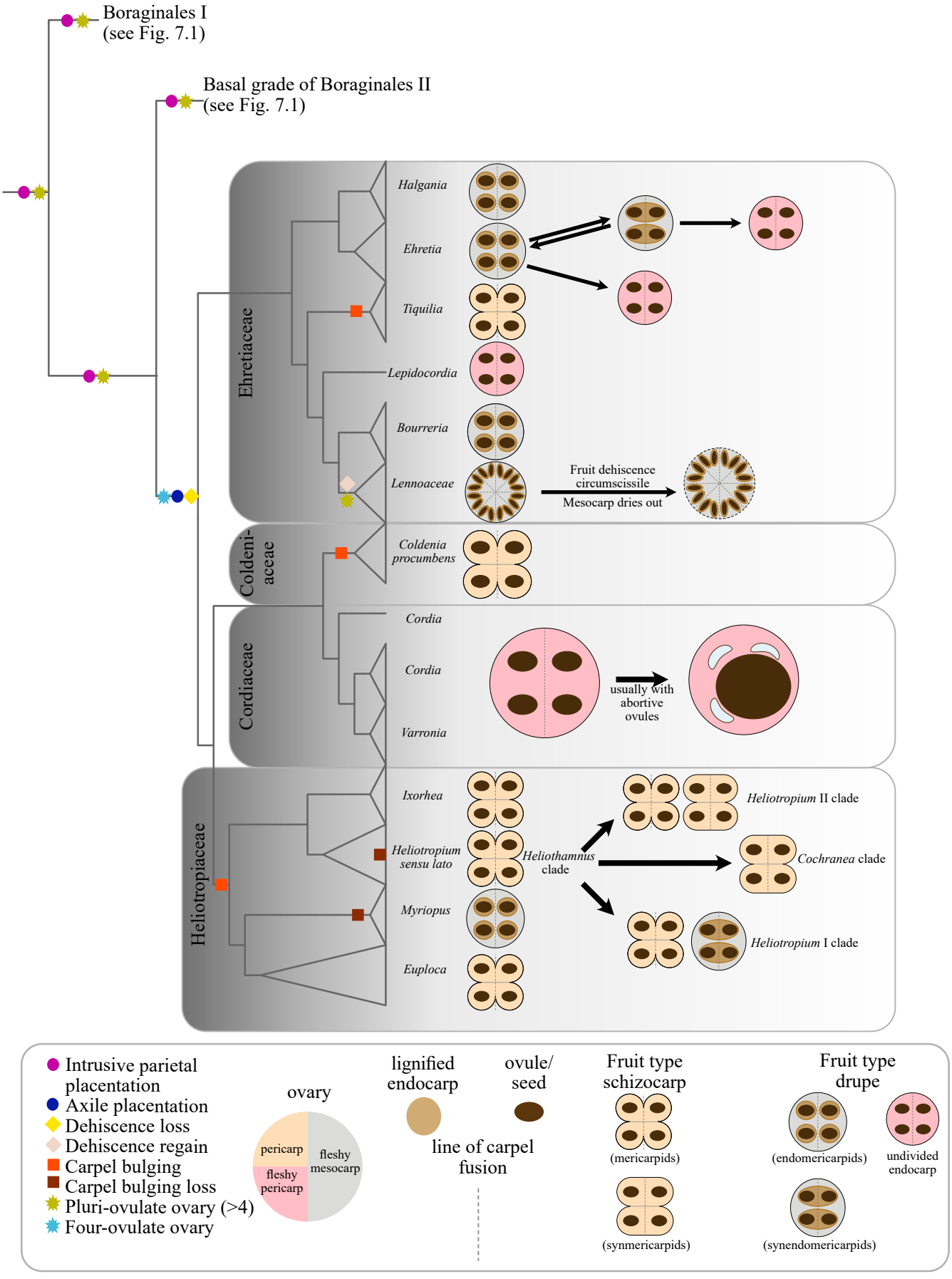


Figure 7.2 Hypothetical evolutionary series of fruit diversification in Boraginales II (Ehretiaceae, Coldeniaceae, Cordiaceae, Heliotropiaceae) excluding the basal grade of Boraginales II. Phylogenetic reconstruction is based on chapter 6. Morphological characters are defined based on chapters 4 and 5 and literature data (Diane & al., 2002, 2016; Gottschling & Hilger, 2004; Luebert & al., 2011, 2016; Gottschling & al., 2014a, 2014b; Heigl & al., 2020). Symbols at the tips, rather than at the branches, indicate derived states that have evolved at least once within the respective clades. The schematic representation of gynoecia/fruits follows the terminology presented in Box 1 of chapter 1 and in chapter 5. The direction of evolution of different internal ovary architectures is displayed whenever possible, however all families except for the monotypic Coldeniaceae remain largely unexplored in regards to fruit evolution.

Furthermore, this thesis investigates flower and fruit ontogeny and morphology of selected taxa as complex 3D units resulting from the combination of different processes and the synorganization of structures during ontogeny. The results of chapters 4 and 5 reveal that processes such as heterochrony and heterotopy have caused differentiation and dedifferentiation of structures, leading to a great diversity of fruit morphologies across the order. It was possible to detect key characters (placentation type and number of locules, ovule number and type of seed and pericarp integration) which can be used for reconstructing the fruit evolution in the order. Similar detailed ontogenetic and morphological studies are still missing for a range of groups in Boraginales (e.g., Boraginaceae, Wellstediaceae, Namaceae, Heliotropiaceae, Hoplestigmataceae, Ehretiaceae) and will certainly contribute to a more accurate and more detailed reconstruction of trait evolution.

Most importantly, this thesis underscores the bidirectional relationship of morphology and molecular phylogenetics. When the interpretation of morphology is conflicting or doubtful, DNA sequences can be enlightening for classifying taxa, or the other way around, as it is nicely demonstrated in several chapters of this thesis. A finer subdivision of Boraginales into several families has been for long a topic of dispute, but the findings of this work support it to a great extent and demonstrate that such a refined classification is necessary to adequately reflect fruit heterogeneity. However, this not always the case, as for example Lennoaceae should be treated from now on as Ehretiaceae, and the subdivision of the basal grade of Boraginales II into Namaceae and Hydrophyllaceae remains ambiguous. Similar multimethodological studies in other group of asterids, representing similar, sometimes parallel trends in fruit evolution (e.g. Lamiales, Gentianales) might help in understanding important diversification patterns in flowering plants.

7.5 Literature cited

Almeida, R.F. de, Pellegrini, M.O.O., Morais, I.L. de, Simo-Bianchini, R., Rattanakrajang, P., Cheek, M., & Simões, A.R.G. 2023. Barking up the wrong tree: the dangers of taxonomic misidentification in molecular phylogenetic studies. *Plant Ecology and Evolution* 156: 146–159.

Bacon, J. 1987. Systematics of *Nama* (Hydrophyllaceae): Seed coat morphology of *Lemmonia californica* and *Nama* species allies with *Nama demissum*. *Aliso* 11: 441–450.

Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers, T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B., Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., Mcginnie, C., Pokorny, L., Roy, S., Stone, M., Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J., & Forest, F. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* 71: 301–319.

Berg, R.Y. 1985. Gynoecium and development of embryo sac, endosperm, and seed in *Pholistoma* (Hydrophyllaceae) relative to taxonomy. *American Journal of Botany* 72: 1775–1787.

Berg, R.Y. 2009. Embryo sac, endosperm, and seed of *Nemophila* (Boraginaceae) relative to taxonomy, with a remark on embryogeny in *Pholistoma*. *American Journal of Botany* 96: 565–579.

Boluda, C.G., Christe, C., Randriarisoa, A., Gautier, L., & Naciri, Y. 2021. Species delimitation and conservation in taxonomically challenging lineages: the case of two clades of *Capurodendron* (Sapotaceae) in Madagascar. *Plants* 10: 1702.

Bonneuil, C. 2002. The manufacture of species: Kew Gardens, the empire and the standardisation of taxonomic practices in late 19th century botany. Pp. 189–215 in: M.N. Bourguet, C. Licoppe, H.O. Sibum (eds.), *Instruments, Travel and Science. Itineraries of precision from the 17th to the 20th century*. London: Routledge.

Brand, A. 1913. *Hydrophyllaceae*, Das Pflanzenreich, (Heft 59) IV. 251. Leipzig: Wilhelm Engelmann: Engler.

Britton, D.M. 1951. Cytogenetic studies on the Boraginaceae. *Brittonia* 7: 233–266.

Chacón, J., Luebert, F., Hilger, H.H., Ovchinnikova, S., Selvi, F., Cecchi, L., Guilliams, C.M., Hasenstab-Lehman, K., Sutorý, K., Simpson, M.G., & Weigend, M. 2016. The borage family (Boraginaceae s.str.): a revised infrafamilial classification based on new phylogenetic evidence, with emphasis on the placement of some enigmatic genera. *TAXON* 65: 523–546.

Chuang, T.I., & Constance, L. 1992. Seeds and systematics in Hydrophyllaceae: tribe Hydrophylleae. *American Journal of Botany* 79: 257–264.

Cohen, J. 2022. Phylogenomics, floral evolution, and biogeography of *Lithospermum* L. (Boraginaceae). *Molecular Phylogenetics and Evolution* 166: 107317.

Di Fulvio, T.E. 1991. Morfología y vascularización floral *de Lemmonia californica* y *Nama jamaicense* (Hydrophyllaceae). *Kurtziana* 21: 23–88.

Di Fulvio, T.E. 1998. Vascularización del gineceo y posición taxonómica de *Nama lobbii*y *N. rothrockii* (Hydrophyllaceae). *Kurtziana* 26: 7–12.

Di Fulvio, T.E., Cosa, M.T., & Dottori, N. 1997. Morfología y vascularización floral en *Turricula*, *Eriodictyon* y *Wigandia* (Phacelieae; Hydrophyllaceae) con relación a la taxonomía. *Kurtziana* 25: 47–66.

Di Fulvio, T.E., Cosa, M.T., & Dottori, N. 1999. Morfología y vascularización floral de *Draperia*, *Emmenanthe*, *Hesperochiron*, *Romanzoffia* y *Tricardia* (Phacelieae, Hydrophyllaceae). *Kurtziana* 27: 187–209.

Diane, N., Förther, H., & Hilger, H.H. 2002. A systematic analysis of *Heliotropium*, *Tournefortia*, and allied taxa of the Heliotropiaceae (Boraginales) based on ITS1 sequences and morphological data. *American Journal of Botany* 89: 287–295.

Diane, N., Hilger, H.H., Förther, M., Weigend, M., & Luebert, F. 2016. Heliotropiaceae. Pp. 203–211 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*, vol. XIV. Heidelberg: Springer

Endress, P.K. 2011. Evolutionary diversification of the flowers in angiosperms. *American Journal of Botany* 98: 370–396.

Ferguson, D.M. 1998a. Phylogenetic analysis and relationships in Hydrophyllaceae based on *ndhF* sequence data. *Systematic Botany* 23: 253–268.

Ferguson, D.M. 1998b. *Phylogenetic systematics of Hydrophyllaceae*. Dissertation, Harvard University, Cambridge, Massachusetts, U.S.A.

Givnish, T., Millam, K., Berry, P., & Sytsma, K. 2007. Phylogeny, adaptive radiation, and historical biogeography of Bromeliaceae inferred from *ndhF* sequence data. *Aliso: A Journal of Systematic and Floristic Botany* 23: 3–26.

Gottschling, M. 2004. Floral ontogeny in *Bourreria* (Ehretiaceae, Boraginales). *Flora - Morphology, Distribution, Functional Ecology of Plants* 199: 409–423.

Gottschling, M., & Hilger, H.H. 2004. Characterisation of a novel fruit type found in *Ehretia* (Ehretiaceae, Boraginales). *Blumea - Biodiversity, Evolution and Biogeography of Plants* 49: 145–153.

Gottschling, M., Hilger, H.H., Wolf, M., & Diane, N. 2001. Secondary structure of the ITS1 transcript and its application in a reconstruction of the phylogeny of Boraginales. *Plant Biol (Stuttg)* 3: 629–636.

Gottschling, M., Luebert, F., Hilger, H.H., & Miller, J.S. 2014a. Molecular delimitations in the Ehretiaceae (Boraginales). *Molecular Phylogenetics and Evolution* 72: 1–6.

Gottschling, M., Miller, J.S., Weigend, M., & Hilger, H.H. 2005. Congruence of a phylogeny of Cordiaceae (Boraginales) inferred from ITS1 sequence data with morphology, ecology, and biogeography. *Annals of the Missouri Botanical Garden* 92: 425–437.

Gottschling, M., Nagelmüller, S., & Hilger, H.H. 2014b. Generative ontogeny in *Tiquilia* (Ehretiaceae: Boraginales) and phylogenetic implications. *Biol J Linn Soc* 112: 520–534.

Gottschling, M., Weigend, M., & Hilger, H.H. 2016. Ehretiaceae. Pp. 165–178 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants,* vol. XIV. Heidelberg: Springer.

Hansen, D.R., Spicer, G.S., & Patterson, R. 2009. Phylogenetic relationships between and within *Phacelia* sections *Whitlavia* and *Gymnobythus* (Boraginaceae). *Systematic Botany* 34: 737–746.

Hasenstab-Lehman, K. 2017. Phylogenetics of the Borage family: delimiting Boraginales and assessing closest relatives. *Aliso* 35: 41–49.

Hasenstab-Lehman, K.E., & Simpson, M.G. 2012. Cat's eyes and popcorn flowers: phylogenetic systematics of the genus *Cryptantha* s. l. (Boraginaceae). *Systematic Botany* 37: 738–757.

Heckard, L.R. 1960. Taxonomic studies in the *Phacelia magellanica* polyploid complex: with special reference to the California members. *Univ. Calif. Publ. Bot.* 32: 1–126.

Heckard, L.R. 1963. The Hydrophyllaceae. *The Quarterly Review of Biology* 38: 117–123.

Heigl, H.M.L., Kretschmann, J., Hilger, H.H., & Gottschling, M. 2020. Flower and fruit anatomy of *Cordia nodosa* Lam. and *Varronia bonplandii* Desv. (Cordiaceae, Boraginales) with phylogenetic implications. *Org Divers Evol* 20: 9–24.

Hilger, H.H. 1987. Flower and fruit development in *Wigandia caracasana* (Hydrophyllaceae). *American Journal of Botany* 74: 250–259.

Hilger, H.H. 1992. Morphology of *Heliotropium* (Boraginaceae) dispersal units. *Botanica Acta* 105: 387–393.

Hilger, H.H. 2014. Ontogeny, morphology, and systematic significance of glochidiate and winged fruits of *Cynoglosseae* and *Eritrichieae* (Boraginaceae). *Plant Diversity and Evolution* 131: 167–214.

Hilger, H.H., & Weigend, M. 2016. Wellstediaceae. Pp. 403–406 in: in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants,* vol. XIV. Heidelberg: Springer

Hitchcock, C.L. 1933a. A taxonomic study of the genus *Nama*. I. *American Journal of Botany* 20: 415–430.

Hitchcock, C.L. 1933b. A taxonomic study of the genus *Nama*. II. *American Journal of Botany* 20: 518–534.

Hofmann, M. 1999. Flower and fruit development in the genus *Phacelia* (Phacelieae, Hydrophyllaceae): characters of systematic value. *Systematics and Geography of Plants* 68: 203–212.

Hofmann, M., Walden, G.K., Hilger, H.H., & Weigend, M. 2016. Hydrophyllaceae. Pp. 221–238 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants,* vol. XIV. Heidelberg: Springer

Holstein, N., & Gottschling, M. 2018. Waking sleeping beauties: a molecular phylogeny and nomenclator of *Halgania* (Ehretiaceae, Boraginales). *Aust. Systematic Bot.* 31: 107–119.

Hörandl, E. 2022. Novel approaches for species concepts and delimitation in polyploids and hybrids. *Plants* 11: 204.

Isaac, N.J.B., Mallet, J., & Mace, G.M. 2004. Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution* 19: 464–469.

Jeiter, J., Danisch, F., & Hilger, H.H. 2016. Polymery and nectary chambers in *Codon* (Codonaceae): Flower and fruit development in a small, capsule-bearing family of Boraginales. *Flora - Morphology, Distribution, Functional Ecology of Plants* 220: 94–102.

Jeiter, J., Staedler, Y.M., Schönenberger, J., Weigend, M., & Luebert, F. 2018. Gynoecium and fruit development in *Heliotropium* sect. *Heliothamnus* (Heliotropiaceae). *International Journal of Plant Sciences* 179: 275–286.

Joyce, E.M., Appelhans, M.S., Buerki, S., Cheek, M., de Vos, J.M., Pirani, J.R., Zuntini, A.R., Bachelier, J.B., Bayly, M.J., Callmander, M.W., Devecchi, M.F., Pell, S.K., Groppo, M., Lowry, P.P., Mitchell, J., Siniscalchi, C.M., Munzinger, J., Orel, H.K., Pannell, C.M., Nauheimer, L., Sauquet, H., Weeks, A., Muellner-Riehl, A.N., Leitch, I.J., Maurin, O., Forest, F., Nargar, K., Thiele, K.R., Baker, W.J., & Crayn, D.M. 2023. Phylogenomic analyses of Sapindales support new family relationships, rapid Mid-Cretaceous Hothouse diversification, and heterogeneous histories of gene duplication. *Frontiers in Plant Science* 14: 1063174.

Karbstein, K., Tomasello, S., Hodač, L., Wagner, N., Marinček, P., Barke, B.H., Paetzold, C., & Hörandl, E. 2022. Untying Gordian knots: unraveling reticulate polyploid plant evolution by genomic data using the large *Ranunculus auricomus* species complex. *New Phytologist* 235: 2081–2098.

Karremans, A.P., Chinchilla, I.F., Rojas-Alvarado, G., Cedeño-Fonseca, M., Damián, A., Léotard, G., Karremans, A.P., Chinchilla, I.F., Rojas-Alvarado, G., Cedeño-Fonseca, M., Damián, A., & Léotard, G. 2020. A reappraisal of neotropical *Vanilla*. With a note on taxonomic inflation and the importance of alpha taxonomy in biological studies. *Lankesteriana* 20: 395–497.

Knapp, S., Lughadha, E.N., & Paton, A. 2005. Taxonomic inflation, species concepts and global species lists. *Trends in Ecology & Evolution* 20: 7–8.

Kook, E., Vedler, E., Püssa, K., Kalamees, R., Reier, Ü., & Pihu, S. 2015. Intra-individual ITS polymorphism and hybridization in *Pulmonaria obscura* Dumort. and *Pulmonaria angustifolia* L. (Boraginaceae). *Plant Syst Evol* 301: 893–910.

Levy, F., Hou, Z.H., & Taylor-Bennetts, S. 2021. A new variety of *Phacelia dubia* (Hydrophyllaceae) and a summary of variation and biogeography of the *P. dubia* complex. *The Journal of the Torrey Botanical Society* 148: 266-284

Lindsey, A.A. 1940. Floral anatomy in the Gentianaceae. *American Journal of Botany* 27: 640–652.

Luebert, F., Brokamp, G., Wen, J., Weigend, M., & Hilger, H.H. 2011. Phylogenetic relationships and morphological diversity in Neotropical *Heliotropium* (Heliotropiaceae). *TAXON* 60: 663–680.

Luebert, F., Cecchi, L., Frohlich, M.W., Gottschling, M., Guilliams, C.M., Hasenstab-Lehman, K.E., Hilger, H.H., Miller, J.S., Mittelbach, M., Nazaire, M., Nepi, M., Nocentini, D., Ober, D., Olmstead, R.G., Selvi, F., Simpson, M.G., Sutorý, K., Valdés, B., Walden, G.K., & Weigend, M. 2016. Familial classification of the Boraginales. *TAXON* 65: 502–522.

Luebert, F., Couvreur, T.L.P., Gottschling, M., Hilger, H.H., Miller, J.S., & Weigend, M. 2017. Historical biogeography of Boraginales: West Gondwanan vicariance followed by long-distance dispersal? *Journal of Biogeography* 44: 158–169.

Luebert, F., & Wen, J. 2008. Phylogenetic analysis and evolutionary diversification of *Heliotropium* sect. *Cochranea* (Heliotropiaceae) in the Atacama Desert. *Systematic Botany*. 32: 390-–402

Mitsui, Y., Chen, S.-T., Zhou, Z.-K., Peng, C.-I., Deng, Y.-F., & Setoguchi, H. 2008. Phylogeny and biogeography of the genus *Ainsliaea* (Asteraceae) in the Sino-Japanese region based on nuclear rDNA and plastid DNA sequence data. *Annals of Botany* 101: 111–124.

Moonlight, P.W., & Cardoso, D.B.O.S. 2023. A taxonomic revision of *Keraunea*, including three new species and its phylogenetic realignment with Ehretiaceae (Boraginales). *PhytoKeys* 219: 145–170.

Moore, J.A. 1936. Morphology of the Gynobase in *Mertensia*. *The American Midland Naturalist* 17: 749–752.

Moore, M.J., & Jansen, R.K. 2006. Molecular evidence for the age, origin, and evolutionary history of the American desert plant genus *Tiquilia* (Boraginaceae). *Molecular Phylogenetics and Evolution* 39: 668–687.

Naciri, Y., & Linder, H.P. 2015. Species delimitation and relationships: The dance of the seven veils. *TAXON* 64: 3–16.

Nazaire, M., & Hufford, L. 2012. A broad phylogenetic analysis of Boraginaceae: implications for the relationships of *Mertensia*. *Systematic Botany* 37: 758–783.

Nelson, D.C. 2021. The mechanism of host-induced germination in root parasitic plants. *Plant Physiology* 185: 1353–1373.

Oberprieler, C., Ott, T., & Vogt, R. 2023. Picks in the fabric of a polyploidy complex: integrative species delimitation in the tetraploid *Leucanthemum* Mill. (Compositae, Anthemideae) representatives. *Biology* 12: 288.

Refulio-Rodriguez, N.F., & Olmstead, R.G. 2014. Phylogeny of Lamiidae. *American Journal of Botany* 101: 287–299.

Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., Ma, H., & Qi, J. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Molecular Plant* 11: 414–428.

Shivaprakash, K.N., & Bawa, K.S. 2022. The evolution of placentation in flowering plants: a possible role for kin selection. *Frontiers in Ecology and Evolution* 10: 784077.

Soares, A.E.R., & Schrago, C.G. 2015. The influence of taxon sampling on Bayesian divergence time inference under scenarios of rate heterogeneity among lineages. *Journal of Theoretical Biology* 364: 31–39.

Spangler, R., Zaitchik, B., Russo, E., & Kellogg, E. 1999. Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using *ndhF* sequences. *Systematic Botany* 24: 267–281.

Taylor, S.E. 2012. *Molecular systematics and the origins of gypsophily in* Nama *L. (Boraginaceae)*. Dissertation, University of Texas, Austin, Texas, U.S.A.

Vasile, M.-A., Jeiter, J., Weigend, M., & Luebert, F. 2020. Phylogeny and historical biogeography of Hydrophyllaceae and Namaceae, with a special reference to *Phacelia* and *Wigandia*. *Systematics and Biodiversity* 18: 757–770.

Vasile, M.-A., Luebert, F., Jeiter, J., & Weigend, M. 2021. Fruit evolution in Hydrophyllaceae. *American Journal of Botany* 108: 925–945.

Villaverde, T., Pokorny, L., Olsson, S., Rincón-Barrado, M., Johnson, M.G., Gardner, E.M., Wickett, N.J., Molero, J., Riina, R., & Sanmartín, I. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.

Walden, G.K., Garrison, L.M., Spicer, G.S., Cipriano, F.W., & Patterson, R. 2014. Phylogenies and chromosome evolution of *Phacelia* (Boraginaceae: Hydrophylloideae) inferred from nuclear ribosomal and chloroplast sequence data. *Madr* 61: 16–47.

Weigend, M., & Hilger, H.H. 2010. Codonaceae-a newly required family name in Boraginales. *Phytotaxa* 10: 26–30.

Weigend, M., & Hilger, H.H. 2016. Codonaceae. Pp. 137–140 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*. Heidelberg: Springer.

Weigend, M., Luebert, F., Gottschling, M., Couvreur, T.L.P., Hilger, H.H., & Miller, J.S. 2014. From capsules to nutlets—phylogenetic relationships in the Boraginales. *Cladistics* 30: 508–518. https://doi.org/10.1111/cla.12061

Weigend, M., Luebert, F., Selvi, F., Brokamp, G., & Hilger, H.H. 2013. Multiple origins for Hound's tongues (*Cynoglossum* L.) and Navel seeds (*Omphalodes* Mill.) – The phylogeny of the borage family (Boraginaceae s.str.). *Molecular Phylogenetics and Evolution* 68: 604–618.

Weigend, M., Selvi, F., Thomas, D.C., & Hilger, H.H. 2016. Boraginaceae. Pp. 41–102 in: J.W. Kadereit & V. Bittrich (eds.), *The Families and Genera of Vascular Plants*. Heidelberg: Springer.

Wen, J., Berggren, S.T., Lee, C.-H., Ickert-Bond, S., Yi, T.-S., Yoo, K.-O., Xie, L., Shaw, J., & Potter, D. 2008. Phylogenetic inferences in *Prunus* (Rosaceae) using chloroplast *ndhF* and nuclear ribosomal ITS sequences. *Journal of Systematics and Evolution* 46: 322–332.

Wilkinson, A.M. 1944. Floral anatomy of some species of *Cornus*. *Bulletin of the Torrey Botanical Club* 71: 276–301.

Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., Rees, M., Frohlich, M.W., Qi, J., Weigend, M., & Ma, H. 2020. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and sup-port phylogenetic placement for numerous whole-genome duplications. *Molecular Biology and Evolution* 37: 3188–3210.

Appendix A

—Supplementary material of chapter 2—

Table S1. List of taxa, voucher information and GenBank accessions.

Figure S1. Complete *ndhF* tree. A, phylogenetic relationships of Hydrophyllaceae. B, phylogenetic relationships of the remaining Boraginales lineages. Numbers above branches are ML bootstrap values (≥ 50) and below branches Bayesian posterior probabilities (BPP).

Figure S2. Maximum likelihood ITS tree with bootstrap values and without collapsed nodes.

Figure S3. Bayesian inference ITS consensus tree with posterior probabilities.

Figure S4. Divergence times estimation of Boraginales based on *ndhF* dataset with their 95% confidence intervals.

Biogeobears R Script. This R script was used for performing ancestral area reconstruction of Namaceae. The same script with different input data was used for Hydrophyllaceae.

Table S1. List of taxa, voucher information and GenBank accessions.

Order	Family	Species	Voucher	Reference	DNA Nr.	*ndhF*	ITS
Boraginales	Boraginaceae s.str.	*Borago officinalis* L.		Ferguson, 1998		L36393	
Boraginales	Boraginaceae s.str.	*Cryptantha flavoculata* (Nelson) Payson	Ferguson 16 (GH)	Ferguson, 1998		AF047803	
Boraginales	Boraginaceae s.str.	*Amsinckia tessellata* A.Gray	Hilger 132/1997 (B)	Weigend et al. 2014	249	KF158024	
Boraginales	Boraginaceae s.str.	*Anchusa formosa* Selvi, Bigazzi & Bacch.	Bigazzi & Selvi 97.006 (FI)	Weigend et al. 2010, 2013	781	KF158060	
Boraginales	Boraginaceae s.str.	*Antiphytum floribundum* A.Gray	Lott et al. 5574 (TEX)	Weigend et al. 2013	2759	KF158030	
Boraginales	Boraginaceae s.str.	*Cryptantha peruviana* I.M.Johnst.	Weigend & Schwarzer 7814 (B)	Weigend et al. 2014	1681	KF158014	
Boraginales	Boraginaceae s.str.	*Cynoglossum officinale* L.	Hilger s.n. (B)	Weigend et al. (2010, 2013)	646	KF158054	
Boraginales	Boraginaceae s.str.	*Echiochilon fruticosum* Desf.	Kagiampaki s.n. (B)	Weigend et al. (2009, 2013)	2542	KF158025	
Boraginales	Boraginaceae s.str.	*Echium vulgare* L.	Mohr 597 (B)	Weigend et al. (2009, 2013)	2063	KF158018	
Boraginales	Boraginaceae s.str.	*Eritrichium aretioides* DC.	Weigend 9126 (B)	Weigend et al. (2013)	2693	KF158028	
Boraginales	Boraginaceae s.str.	*Lappula squarrosa* (Retz.) Dumort.	Mohr 591 (B)	Weigend et al. (2010, 2013)	2059	KF158017	
Boraginales	Boraginaceae s.str.	*Lithospermum officinale* L.	Werres & Ristow s.n. (B)	Weigend et al. (2009, 2013)	409	KF158077	
Boraginales	Boraginaceae s.str.	*Mertensia maritima* (L.) Gray	Hilger s.n. (B)	Weigend et al. (2010, 2013)	1230	KF158012	
Boraginales	Boraginaceae s.str.	*Moritzia lindenii* (A.DC) Benth. ex Gürke	Eriksen 59018 (MO)	Weigend et al. (2010, 2013)	1977	KF158016	
Boraginales	Boraginaceae s.str.	*Myosotis incrassata* Guss.	Hilger Kreta1998/5 (B)	Weigend et al. (2010, 2013)	378	KF158042	
Boraginales	Boraginaceae s.str.	*Onosma rhodopea* Bul.	Hilger 97/24 (B)	Weigend et al. (2009, 2013)	582	KF158052	
Boraginales	Boraginaceae s.str.	*Pentaglottis sempervirens* (L.) L.H.Bailey	Weigend 9065 (B)	Weigend et al. 2014	2626	KF158026	
Boraginales	Boraginaceae s.str.	*Plagiobothrys humilis* (Ruiz & Pav.) I.M. Johnst.	Weigend & Schwarzer 8075 (B)	Weigend et al. (2010, 2013)	2178	KF158019	
Boraginales	Boraginaceae s.str.	*Trachystemon orientalis* (L.) G.Don	Hilger s.n. (B)	Weigend et al. (2010, 2013)	666	KF158055	
Boraginales	Codonaceae	*Codon royenii* L.	Greuter 21551 (B)	Weigend et al. 2013	2373	KF158021	
Boraginales	Codonaceae	*Codon schenkii* Schinz	Walter & Walter 118 (B)	Weigend et al. 2013	2375	KF158022	
Boraginales	Coldeniaceae	*Coldenia procumbens* L.	Jongkind & Nieuwenhuis 1973 (MO)	Moore & Jansen 2006	801	DQ197255	
Boraginales	Cordiaceae	*Cordia africana* Lam.	Pope 284 (B)	Weigend et al. 2014	930	KF158068	
Boraginales	Cordiaceae	*Cordia americana* (L.) Gottschling & J.S.Mill.	Cult. HB Munich (B)	Weigend et al. 2014	916	KF158067	
Boraginales	Cordiaceae	*Cordia bonplandii* (Desv.) Roem. & Schult.	Gerlach s.n. (M)	Weigend et al. 2014	56	KF158051	
Boraginales	Cordiaceae	*Cordia bullata* (L.) Roem. & Schult.	Gottschling CUB49 (B)	Weigend et al. 2014	915	KF158066	
Boraginales	Cordiaceae	*Cordia collococca* L.	Miller & Sherman 6370 (B)	Weigend et al. 2014	1072	KF158010	
Boraginales	Cordiaceae	*Cordia dencandra* Hook. & Arn	Luebert & Kritzner 1873 (SGO)	Luebert & Wen 2008	FL74	EF688954	
Boraginales	Cordiaceae	*Cordia guineensis* Thonn.	Cult. HB Berlin-Dahlem (B)	Weigend et al. 2014	397	KF158044	
Boraginales	Cordiaceae	*Cordia lutea* Lam.	Weigend et al. 2000/707 (B)	Weigend et al. 2014	988	KF158072	
Boraginales	Cordiaceae	*Cordia monoica* Roxb.	Cubr 38073 (B)	Weigend et al. 2014	396	KF158043	

Boraginales	Cordiaceae	*Cordia myxa* L.	Cult. HB Berlin-Dahlem (B)	Luebert et al. (2011b); Mansion et al. (2009)	377	KF158041	
Boraginales	Cordiaceae	*Cordia nevillii* Alston		Refulio-Rodriguez & Olmstead 2014		HQ384860	
Boraginales	Cordiaceae	*Cordia nodosa* Lam.	D. Yu (no voucher)	Ferguson, 1998		AF047808	
Boraginales	Cordiaceae	*Cordia oncocalyx* Allemão	Quintana et al. 604 (MO)	Weigend et al. 2014	1260	KF158013	
Boraginales	Cordiaceae	*Cordia saccellia* Gottschling & J.S.Mil.	Taylor et al. 11353 (MO)	Weigend et al. 2014	815	KF158062	
Boraginales	Cordiaceae	*Cordia sinensis* Lam.	Breckle 3999 (B)	Mansion et al. 2009	931	KF158069	
Boraginales	Cordiaceae	*Cordia sonorae* N.E.Rose	Miller & Campos Ríos 2956 (B)	Weigend et al. 2014	979	KF158071	
Boraginales	Cordiaceae	*Cordia trichotoma* (Vell.) Arráb. ex Steud.	Hilger et al. ARG 95/58 (B)	Mansion et al. 2009	118	EU599739	
Boraginales	Cordiaceae	*Cordia weddellii* I.M.Johnst.	Mereles & Degen 5513 (MO	Weigend et al. 2014	1206	KF158010	
Boraginales	Ehretiaceae	*Bourreria costaricensis* (Standl.) Gentry	Alverson & Ferguson 2526 (GH)	Ferguson, 1998		AF047797	
Boraginales	Ehretiaceae	*Bourreria moaensis* Britton	Beurton & Köhler s.n. (B)	Weigend et al. 2014	450	KF158048	
Boraginales	Ehretiaceae	*Bourreria succulenta* Jacq.	Olmstead 96-114 (WTU)	Moore & Jansen 2006		DQ197257	
Boraginales	Ehretiaceae	*Coldenia procumbens* L.	Jongkind and Nieuwenhuis 1973 (MO)	Moore & Jansen 2006		DQ197255	
Boraginales	Ehretiaceae	*Ehretia anacua* I.M. Johnst	Cult. at Univ. of Texas	Moore & Jansen 2006		DQ197256	
Boraginales	Ehretiaceae	*Ehretia asperula* Zoll. & Moritzi		Pham et al. 2017 (unpub.)			KY320205
Boraginales	Ehretiaceae	*Ehretia aquatica* (Lour.) Gottschling & Hilger	Jongkind 2517 (MO)	Luebert et al. (2011b); Mansion et al. (2009)	792	KF158061	
Boraginales	Ehretiaceae	*Ehretia cymosa* Thonn.	de Wilde 4230 (B)	Luebert et al. (2011b); Mansion et al. (2009)	416	KF158046	
Boraginales	Ehretiaceae	*Ehretia longiflora* Champ. ex Benth.	King s.n. (B)	Weigend et al. 2014	863	KF158063	
Boraginales	Ehretiaceae	*Ehretia microphylla* Lam.	Cult. HB Singapore	Weigend et al. 2014	406	KF158045	
Boraginales	Ehretiaceae	*Ehretia ovalifolia* Hassk.		Ferguson, 1998		AF047800	
Boraginales	Ehretiaceae	*Ehretia tinifolia* L.	Gottschling CUB52 (B)	Luebert et al. (2011b)	439	KF158047	
Boraginales	Ehretiaceae	*Halgania cyanea* Lindl.	Greuter 18801 (B)	Weigend et al. 2014	880	KF158064	
Boraginales	Ehretiaceae	*Rochefortia lundellii* Camp ex Lundell	Wendt 2828 (B)	Weigend et al. 2014	913	KF158065	
Boraginales	Ehretiaceae	*Tiquilia canescens* (DC.) A.T. Richardson	Moore 239	Moore & Jansen 2006		DQ197258	
Boraginales	Ehretiaceae	*Tiquilia conspicua* (I.M. Johnst.) A.T. Richardson	Moore 294	Moore & Jansen 2006		DQ197278	
Boraginales	Ehretiaceae	*Tiquilia cuspidata* (I.M. Johnst.) A.T. Richardson	Moore 223	Moore & Jansen 2006		DQ197277	
Boraginales	Ehretiaceae	*Tiquilia darwinii* (Hook.f.) A.T. Richardson	Tye 573	Moore & Jansen 2006		DQ197276	
Boraginales	Ehretiaceae	*Tiquilia elongata* (Rusby) A.T. Richardson	Moore 289	Moore & Jansen 2006		DQ197279	
Boraginales	Ehretiaceae	*Tiquilia gossypina* (Wooton and Standl.) A.T. Richardson	Moore 134	Moore & Jansen 2006		DQ197267	

Boraginales	Ehretiaceae	*Tiquilia greggii* (Torr. and A. Gray) A.T. Richardson	Moore 133	Moore & Jansen 2006		DQ197259	
Boraginales	Ehretiaceae	*Tiquilia hispidissima* (Torr. and A. Gray) A.T. Richardson	Moore 131	Moore & Jansen 2006		DQ197268	
Boraginales	Ehretiaceae	*Tiquilia latior* (I.M. Johnst.) A.T. Richardson	Moore 216	Moore & Jansen 2006		DQ197272	
Boraginales	Ehretiaceae	*Tiquilia mexicana* (S. Watson) A.T. Richardson	Moore 117	Moore & Jansen 2006		DQ197261	
Boraginales	Ehretiaceae	*Tiquilia nuttallii* (Benth.) A.T. Richardson	Moore 218	Moore & Jansen 2006		DQ197282	
Boraginales	Ehretiaceae	*Tiquilia palmeri* (A. Gray) A.T. Richardson	Moore 202	Moore & Jansen 2006		DQ197281	
Boraginales	Ehretiaceae	*Tiquilia paronychioides* (Phil.) A.T. Richardson	Moore 300	Moore & Jansen 2006		DQ197275	
Boraginales	Ehretiaceae	*Tiquilia plicata* (Torr.) A.T. Richardson	Moore 196	Moore & Jansen 2006		DQ197274	
Boraginales	Ehretiaceae	*Tiquilia purpusii* (Brandegee) A.T. Richardson	Moore 109	Moore & Jansen 2006		DQ197273	
Boraginales	Ehretiaceae	*Tiquilia turneri* A.T. Richardson	Moore 89	Moore & Jansen 2006		DQ197262	
Boraginales	Ehretiaceae	*Tiquilia*"Durango" (undescribed species)	Moore 260	Moore & Jansen 2006		DQ197260	
Boraginales	Heliotropiaceae	*Ceballosia fruticosa* (L.f.) Kunkel ex Förther	Weigend & Weigend 8703 (B)	Luebert et al. 2011a	FL96	KF158075	
Boraginales	Heliotropiaceae	*Euploca campestris* (Griseb.) Diane & Hilger	Nee & Wen 53873 (US)	Walden et al. 2014		EF688908	
Boraginales	Heliotropiaceae	*Euploca pilosa* (Ruiz & Pav.) Luebert	Dillon 8819 (F)	Luebert & Wen 2008	FL73	EF688953	
Boraginales	Heliotropiaceae	*Euploca procumbens* (Mill.) Diane & Hilger	Nee & Wen 53873 (US)	Luebert & Wen 2008	FL5	EF688908	
Boraginales	Heliotropiaceae	*Heliotropium adenogynum* I.M.Johnst.	Cano 10058 (USM)	Luebert et al. 2011a	726	KF158057	
Boraginales	Heliotropiaceae	*Heliotropium arborescens* L.		Ferguson, 1998		AF014000	
Boraginales	Heliotropiaceae	*Heliotropium curassavicum* L.	Luebert & García 2521 (SGO)	Luebert & Wen 2008	FL61	EF688949	
Boraginales	Heliotropiaceae	*Heliotropium erosum* Lehm.	Zippel 00/69 (BSB)	Luebert et al. 2011a	677	KF158056	
Boraginales	Heliotropiaceae	*Heliotropium europaeum* L.	Hilger 97/6 (BSB)	Luebert et al. 2011a	496	KF158050	
Boraginales	Heliotropiaceae	*Heliotropium incanum* Ruiz & Pav.	Weigend 2000/162 (NY)	Luebert et al. 2011a	733	KF158058	
Boraginales	Heliotropiaceae	*Heliotropium linariifolium* Phil.		Luebert & Wen 2008			EF688870
Boraginales	Heliotropiaceae	*Heliotropium nicotianifolium* Poir.	Nee & Wen 53843 (US)	Luebert & Wen 2008	FL3	EF688906	
Boraginales	Heliotropiaceae	*Heliotropium paronychioides A*.DC.	Luebert & Teillier 2241 (SGO)	Luebert & Wen 2008	FL60	EF688948	
Boraginales	Heliotropiaceae	*Heliotropium pycnophyllum* Phil.	Luebert & García 2813 (SGO)	Luebert & Wen 2008	FL31	EF688920	
Boraginales	Heliotropiaceae	*Heliotropium stenophyllum* Hook. & Arn.	Luebert & Becker 1990 (SGO)	Luebert & Wen 2008	FL69	EF688950	
Boraginales	Heliotropiaceae	*Heliotropium zeylanicum* (Burm.f.) Lam.	Hilger 94/04 (BSB)	Luebert et al. 2011a	612	KF158053	
Boraginales	Heliotropiaceae	*Ixorhea tschudiana* Fenzl	Hilger et al. ARG 95/65 (BSB)	Luebert et al. 2011a	176	KF158015	

Boraginales	Heliotropiaceae	*Myriopus salzmannii* (DC.) Diane & Hilger	Nee & Wen 53848 (US)	Luebert & Wen 2008	FL2	EF688905	
Boraginales	Heliotropiaceae	*Tournefortia acutiflora* M. Martens & Galeotti	Ferguson 60 (GH)	Ferguson, 1998		AF047813	
Boraginales	Heliotropiaceae	*Tournefortia cf. buchtienii* Killip	Nee & Wen 53944 (US)	Luebert & Wen 2008	FL7	EF688910	
Boraginales	Heliotropiaceae	*Tournefortia hirsutissima* L.	Wen 8706 (US)	Luebert & Wen 2008	FL67	KF158073	
Boraginales	Hoplestigmataceae	*Hoplestigma klaineanum* Pierre	Carvalho 5089 (NY)	Weigend et al. 2014	TC-NY002	KF158076	
Boraginales	Hydrophyllaceae	Draperia *systyla* (A. Gray) Torr.	1752 (BONN); Thorne 53719 (RSA)	Ferguson, 1998	W5140	MT176345	AF091155
Boraginales	Hydrophyllaceae	*Ellisia nyctelea* L.	Olmstead 95-16 (COLO)	Ferguson, 1998		AF047796	AF091157
Boraginales	Hydrophyllaceae	*Emmenanthe penduliflora* Benth.	Steve Boyd 11494	Walden 2015 (unpub.)	W5621	MT176365	KT693155
Boraginales	Hydrophyllaceae	*Emmenanthe penduliflora* Benth. var. *penduliflora*	Ferguson 47 (GH)	Ferguson, 1998; Thornhill et al. 2017		AF047815	MF964104
Boraginales	Hydrophyllaceae	*Eucrypta chrysanthemifolia* (Benth.) Greene	Ferguson 48 (GH)	Ferguson, 1998		AF047814	AF091165
Boraginales	Hydrophyllaceae	*Eucrypta micrantha* (Torr.) A. Heller	Chambers 5832 (GH)	Ferguson, 1998		AF047773	AF091166
Boraginales	Hydrophyllaceae	*Hesperochiron pumilis* (Griseb.) Porter	Olmstead & Ferguson 92 (GH)	Ferguson, 1998		AF047783	AF091167
Boraginales	Hydrophyllaceae	*Howellanthus dalesianus* (J. T. Howell) Walden & R. Patt.	Patterson & Santos 1982 (SFSU)	Walden et al. 2014; Thornhill et al. 2017		JQ249933	MF964147
Boraginales	Hydrophyllaceae	*Hydrophyllum canadense* L.	Hilger s.n. 1998 (B)	Weigend et al. 2014	324	KF158040	
Boraginales	Hydrophyllaceae	*Hydrophyllum capitatum* Douglas ex Benth. var. *capitatum*	Ferguson 116 (GH)	Ferguson, 1998		AF047785	AF091169
Boraginales	Hydrophyllaceae	*Hydrophyllum fendleri* (A. Gray) A. Heller	s.col. s.n. (BONN)		W5142	MT176347	
Boraginales	Hydrophyllaceae	*Hydrophyllum tenuipes* A. Heller	1556 (BONN); Ferguson 116 (GH)	Ferguson, 1998	W5138	MT176344	AF091170
Boraginales	Hydrophyllaceae	*Hydrophyllum virginianum* L.	1820 (BONN)		W5141	MT176346	
Boraginales	Hydrophyllaceae	*Nemophila menziesii* Hook. & Arn.	Weigend 9341 (B); - ; US:Ingram92	Walden et al. 2014; Barr, 2015; Zuniga, 2019 (unpub.)	3067	KF158039	KP208583; MH766931
Boraginales	Hydrophyllaceae	*Nemophila parviflora* Douglas ex Benth.	Ferguson 96 (GH)	Ferguson, 1998		AF047782	AF091184
Boraginales	Hydrophyllaceae	*Phacelia adenophora* J. T. Howell	Tiehm 15643 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ249940	AY630260
Boraginales	Hydrophyllaceae	*Phacelia affinis* A. Gray	Dempcy 137-1 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ249941	AY630264
Boraginales	Hydrophyllaceae	*Phacelia arizonica* A. Gray	L. M. Garrison 32 (SFSU)	Walden et al. 2014			JX233426
Boraginales	Hydrophyllaceae	*Phacelia argentea* A. Nelson & J. F. Macbr.	Ferguson 82 (GH)	Ferguson, 1998		AF047810	AF091185
Boraginales	Hydrophyllaceae	*Phacelia artemisioides* Griseb.	Sancho Gisela et.al 382 (BONN, LP)		W5608	MT176356	MT137551
Boraginales	Hydrophyllaceae	*Phacelia bicolor* Torr.	A. Tiehm 15523 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ249945	AY630266
Boraginales	Hydrophyllaceae	*Phacelia bolanderi* A. Gray	W. Roderick s.n. (UCB61.0093); -	Walden et al. 2014; Ferguson, 2000 (unpub.)		JQ249946	AF091187
Boraginales	Hydrophyllaceae	*Phacelia bombycina* Wooton & Standl.	Garrison 43 (SFSU); Garrison 15 (SFSU)	Walden et al. 2014		JQ249947	JX233427
Boraginales	Hydrophyllaceae	*Phacelia brachyantha* Benth.	F.Luebert 3182 (BONN)		W5600	MT176353	MT137550

Boraginales	Hydrophyllaceae	*Phacelia brachyloba* (Benth.) A. Gray	Walden 68 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ249948	AY630271
Boraginales	Hydrophyllaceae	*Phacelia breweri* A.Gray		Gilbert et al. 2005			AY630274
Boraginales	Hydrophyllaceae	*Phacelia californica* Cham.	Roderick s.n. (UCB)	Walden et al. 2014		JQ249950	AY630275
Boraginales	Hydrophyllaceae	*Phacelia calthifolia* Brand.		Gilbert et al. 2005			AY630276
Boraginales	Hydrophyllaceae	*Phacelia campanularia* A. Gray	M. Weigend s.n. (BONN); -	Ferguson, 2000 (unpub.)	W5615	MT176362	AF091188
Boraginales	Hydrophyllaceae	*Phacelia campanularia* A. Gray var. *vasiformis* Walden & R. Patt.	Walden 20 (SFSU); DH49	Walden et al. 2014; Hansen et al. 2010		JQ249953	FJ814647
Boraginales	Hydrophyllaceae	*Phacelia campanularia* A. Gray var. *campanularia*	Hansen 22 (SFSU); DH16	Walden et al. 2014; Hansen et al. 2010		JQ249952	FJ814643
Boraginales	Hydrophyllaceae	*Phacelia capitata* Kruckeb.	Halse 7466 (SFSU)	Walden et al. 2014		JQ249954	
Boraginales	Hydrophyllaceae	*Phacelia cephalotes* A. Gray,	Higgins 25490 (DAV)	Walden et al. 2014		JQ249935	
Boraginales	Hydrophyllaceae	*Phacelia cicutaria* Greene var. *cicutaria*	Garrison 02 (SFSU)	Walden et al. 2014		JQ249955	JX233431
Boraginales	Hydrophyllaceae	*Phacelia cicutaria* Greene var. *hispida* (A. Gray) J. T. Howell	Rebman 14303 (SDNHM); L. M. Garrison 24 (SFSU)	Walden et al. 2014		JQ249936	JX233433
Boraginales	Hydrophyllaceae	*Phacelia coerulea* Greene	Garrison 30 (SFSU)	Walden et al. 2014		JQ249956	JX233435
Boraginales	Hydrophyllaceae	*Phacelia congesta* Hook.	Halse 4436 (GH)	Ferguson, 1998		AF047780	AF091189
Boraginales	Hydrophyllaceae	*Phacelia cookei* Constance & Heckard,	Schierenbeck s.n. (SFSU)	Walden et al. 2014		JQ249957	
Boraginales	Hydrophyllaceae	*Phacelia corrugata* A.Nelson	L. M. Garrison 53 (SFSU)	Walden et al. 2014			JX233437
Boraginales	Hydrophyllaceae	*Phacelia cottamii* N. D. Atwood	Atwood 24858 (DAV)	Walden et al. 2014		JQ249958	
Boraginales	Hydrophyllaceae	*Phacelia crenulata* Torr. var. *crenulata*	Garrison 44 (SFSU); Garrison 05 (SFSU)	Walden et al. 2014		JQ249959	JX233447
Boraginales	Hydrophyllaceae	*Phacelia cryptantha* Greene	Garrison 20 (SFSU)	Walden et al. 2014			JX233453
Boraginales	Hydrophyllaceae	*Phacelia cumingii* (Benth.) A. Gray	M. Weigend et al. 5995 (B)		W5619	MT176364	
Boraginales	Hydrophyllaceae	*Phacelia cumingii* (Benth.) A. Gray	M. Weigend et al. 5905 (B)		W5618	MT176363	
Boraginales	Hydrophyllaceae	*Phacelia cumingii* (Benth.) A. Gray	F.Luebert,T.Böhnert 3546 (BONN); -	Gilbert et al. 2005	W5144	MT176348	AY630282
Boraginales	Hydrophyllaceae	*Phacelia cumingii* (Benth.) A. Gray	F.Luebert,T.Böhnert 3823 (BONN)		W5609	MT176357	MT137552
Boraginales	Hydrophyllaceae	*Phacelia davidsonii* A. Gray	Walden 72 (SFSU)	Walden et al. 2014		JQ249962	
Boraginales	Hydrophyllaceae	*Phacelia demissa* A. Gray,	Atwood 30342 (DAV)	Walden et al. 2014		JQ249963	
Boraginales	Hydrophyllaceae	*Phacelia distans* Benth. sensu lato*	Garrison 21 (SFSU); Garrison 14 (SFSU)	Walden et al. 2014		JQ250026	JX233454
Boraginales	Hydrophyllaceae	*Phacelia divaricata* A. Gray	Walden 35 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ249964	AY630285
Boraginales	Hydrophyllaceae	*Phacelia duglasii* (Benth.) Torr.	-	Gilbert et al. 2005			AY630286
Boraginales	Hydrophyllaceae	*Phacelia dubia* (L.) Trel. & Small	-	Glass & Levy 2011			JN619425
Boraginales	Hydrophyllaceae	*Phacelia egena* (Brand) J.T.Howell	Felix Trabert s.n. (BONN); -	Gilbert et al. 2005	W5136	MT176343	AY630287
Boraginales	Hydrophyllaceae	*Phacelia eisenii* Brandegee	Kruse-Pickler 21 (SFSU)	Walden et al. 2014		JQ249965	
Boraginales	Hydrophyllaceae	*Phacelia formosula* Osterh.	Atwood 5830 (DAV)	Walden et al. 2014		JQ249966	

Boraginales	Hydrophyllaceae	*Phacelia franklinii* (R. Br.) A. Gray,	Harms 20025 (GH); -	Ferguson, 1998; Kuzmina, 2018		AF047772	MG218617
Boraginales	Hydrophyllaceae	*Phacelia fremontii* Torr.	-	Gilbert et al. 2005			AY630289
Boraginales	Hydrophyllaceae	*Phacelia glaberrima* (S.Watson) J.T.Howell	-	Gilbert et al. 2005		AY630291	
Boraginales	Hydrophyllaceae	*Phacelia glandulifera* Piper	Dempcy 119-2 (SFSU)	Walden et al. 2014; Gilbert et al. 2005		JQ249968	AY630293
Boraginales	Hydrophyllaceae	*Phacelia glandulosa*	Weber & Randolph 17444 (TEX)	Walden et al. 2014			JX233456
Boraginales	Hydrophyllaceae	*Phacelia glechomifolia* A. Gray,	Atwood 30196 (DAV)	Walden et al. 2014		JQ249969	
Boraginales	Hydrophyllaceae	*Phacelia grandiflora* (Benth.) A. Gray	T.Joßberger s.n. (BONN); DH2	Hansen et al. 2010	W5613	MT176361	FJ814648
Boraginales	Hydrophyllaceae	*Phacelia greenei* J. T. Howell	Patterson 1981 (SFSU)	Walden et al. 2014		JQ249973	
Boraginales	Hydrophyllaceae	*Phacelia grisea* A. Gray	Walden 151 (SFSU)	Walden et al. 2014		JQ249974	
Boraginales	Hydrophyllaceae	*Phacelia gymnoclada* Torr. ex S. Watson	Zika 12351 (GH)	Ferguson, 1998; Gilbert et al. 2005		AF047793	AY630295
Boraginales	Hydrophyllaceae	*Phacelia gypsogenia* I.M.Johnst.	Aramberri 00248757 (TEX)	Walden et al. 2014			JX233457
Boraginales	Hydrophyllaceae	*Phacelia hastata* Douglas ex Lehm. var. *compacta* (Brand) *cronquist*	Tiehm 15577 (SFSU)	Walden et al. 2014; Gilbert et al. 2005		JQ249975	AY630299
Boraginales	Hydrophyllaceae	*Phacelia heterophylla* Pursh var. *virgata* (Greene) R. D. Dorn	Gauna s.n. (SFSU)	Walden et al. 2014		JQ249976	
Boraginales	Hydrophyllaceae	*Phacelia heterophylla* Pursh.	R.Halse 5088 (MO)	Ferguson, 1998	W5605	MT176355	AF091192
Boraginales	Hydrophyllaceae	*Phacelia hirsuta* Nutt.	Bufford et al. 25539 (A)	Ferguson, 1998		AF047777	AF091193
Boraginales	Hydrophyllaceae	*Phacelia howelliana* N. D. Atwood	Garrison 59 (SFSU)	Walden et al. 2014		JQ249979	JX233458
Boraginales	Hydrophyllaceae	*Phacelia hubbyi* (J. F. Macbr.) L. M. Garrison	Garrison 23 (SFSU)	Walden et al. 2014		JQ249980	
Boraginales	Hydrophyllaceae	*Phacelia humilis* Torr. & A. Gray	Ferguson 21 (GH)	Ferguson, 1998		AF047817	AF091194
Boraginales	Hydrophyllaceae	*Phacelia hydrophylloides* Torr. ex A. Gray	Hankamp s.n. (SFSU)	Walden et al. 2014		JQ249982	AY630304
Boraginales	Hydrophyllaceae	*Phacelia imbricata* Greene		Gilbert et al. 2005			AY630305
Boraginales	Hydrophyllaceae	*Phacelia insularis* Munz var. *insularis*	Hansen 55 (SFSU)	Walden et al. 2014; Hansen et al. 2010		JQ249937	FJ814627
Boraginales	Hydrophyllaceae	*Phacelia integrifolia* var. *integrifolia*	Garrison 37 (SFSU)	Walden et al. 2014			JX233460
Boraginales	Hydrophyllaceae	*Phacelia inundata* J. T. Howell	Tiehm 15553 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ249983	AY630306
Boraginales	Hydrophyllaceae	*Phacelia inyoensis* (J.F.Macbr.) J.T.Howell		Gilbert et al. 2005			AY630307
Boraginales	Hydrophyllaceae	*Phacelia ivesiana* Torr. var. *pediculoides* J. T. Howell	Walden 22 (SFSU)	Walden et al. 2014; Hansen et al. 2010		JQ249984	FJ814623
Boraginales	Hydrophyllaceae	*Phacelia keckii* Munz & I. M. Johnst.	Walden 71 (SFSU)	Walden et al. 2014		JQ249985	
Boraginales	Hydrophyllaceae	*Phacelia laxiflora* J. T. Howell	Howell 26440 (DAV19059)	Walden et al. 2014		JQ249986	
Boraginales	Hydrophyllaceae	*Phacelia leibergii* Brand	Dempcy 121 (SFSU)	Walden et al. 2014; Gilbert et al. 2005		JQ249987	AY630311
Boraginales	Hydrophyllaceae	*Phacelia linearis* (Pursh) Holz.	Gauna s.n. (SFSU)	Walden et al. 2014		JQ249988	AF091195
Boraginales	Hydrophyllaceae	*Phacelia longipes* Torr. ex A.Gray	DH57	Hansen et al. 2010			FJ814645
Boraginales	Hydrophyllaceae	*Phacelia lutea* (Hook. & Arn.) J. T. Howell	Arnold Thiem 12085; -	Gilbert et al. 2005	W5622	MT176366	AY630314
Boraginales	Hydrophyllaceae	*Phacelia malvifolia* Cham.	T.Joßberger 1488 (BONN)		W5597	MT176352	MT137549

Boraginales	Hydrophyllaceae	*Phacelia mammillarensis* N.D. Atwood	Garrison 63 (SFSU)	Walden et al. 2014			JX233464
Boraginales	Hydrophyllaceae	*Phacelia marshall-johnstonii* N.D.Atwood & Pinkava		Walden et al. 2014			JX233465
Boraginales	Hydrophyllaceae	*Phacelia minor* (Harvey) Thell. ex F. Zimm.	Hansen 14 (SFSU); -	Walden et al. 2014; Ferguson, 1998		JQ249990	AF091197
Boraginales	Hydrophyllaceae	*Phacelia minor* x *Phacelia parryi*	Hansen 30 (SFSU) ; DH30	Walden et al. 2014; Hansen et al. 2010		JQ249992	FJ814634
Boraginales	Hydrophyllaceae	*Phacelia monoensis* Halse	Kruse-Pickler 16 (SFSU)	Walden et al. 2014; Gilbert et al. 2005		JQ249993	AY630315
Boraginales	Hydrophyllaceae	*Phacelia nashiana* Jeps.	Hansen 04 (SFSU); DH04	Walden et al. 2014; Hansen et al. 2010		JQ249995	FJ814637
Boraginales	Hydrophyllaceae	*Phacelia neglecta* M.E.Jones		Gilbert et al. 2005			AY630319
Boraginales	Hydrophyllaceae	*Phacelia nemoralis* Greene var. *nemoralis*		Gilbert et al. 2005			AY630320
Boraginales	Hydrophyllaceae	*Phacelia nemoralis* Greene var. *oregonensis* (Heckard) Walden & R. Patt	Halse 7486 (SFSU)	Walden et al. 2014		JQ249998	AY630320
Boraginales	Hydrophyllaceae	*Phacelia neomexicana* Thurb. ex Torr.		Walden et al. 2014			JX233466
Boraginales	Hydrophyllaceae	*Phacelia novenmillensis* Munz	Ertter 7005 (DAV)	Walden et al. 2014		JQ249999	
Boraginales	Hydrophyllaceae	*Phacelia pachyphylla* A. Gray	Dempcy 130-3 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ250000	AY630323
Boraginales	Hydrophyllaceae	*Phacelia palmeri* Torr. ex S.Watson	Garrison 51 (SFSU)	Walden et al. 2014			JX233467
Boraginales	Hydrophyllaceae	*Phacelia parryi* Torr	Allen s.n. (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ250001	AY653742
Boraginales	Hydrophyllaceae	*Phacelia patuliflora* (Engelm. & A. Gray) A. Gray	Erten & Woodruff 5636	Ferguson, 1998	W5623	MT176367	AF091198
Boraginales	Hydrophyllaceae	*Phacelia pauciflora* S. Watson	Webster 26155 (DAV)	Walden et al. 2014		JQ250004	
Boraginales	Hydrophyllaceae	*Phacelia pedicellata* A. Gray	Walden 127 (SFSU); Garrison 07 (SFSU)	Walden et al. 2014		JQ250006	JX233468
Boraginales	Hydrophyllaceae	*Phacelia perityloides* Coville	Garrison 47 (SFSU)	Walden et al. 2014		JQ250007	JX233470
Boraginales	Hydrophyllaceae	*Phacelia petrosa* N. D. Atwood, F.J. Sm. & T. A. Knight	Atwood 30142 (DAV)	Walden et al. 2014		JQ250008	
Boraginales	Hydrophyllaceae	*Phacelia pinnatifida* Griseb. ex Wedd.	Sancho Gisela et.al 378 (BONN, LP)		W5610	MT176358	MT137553
Boraginales	Hydrophyllaceae	*Phacelia popei* Torr. & A. Gray	Garrison 36 (SFSU)	Walden et al. 2014		JQ250009	JX233471
Boraginales	Hydrophyllaceae	*Phacelia pringlei* A. Gray	Patterson 1980 (SFSU)	Walden et al. 2014		JQ250010	
Boraginales	Hydrophyllaceae	*Phacelia procera* A. Gray	Thiers 23458 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ250012	AY630325
Boraginales	Hydrophyllaceae	*Phacelia pulchellae* A. Gray	Garrison 52 (SFSU)	Walden et al. 2014		JQ250013	JX233473
Boraginales	Hydrophyllaceae	*Phacelia purpusii* Brandegee	Barron s.n. (DAV)	Walden et al. 2014		JQ250014	
Boraginales	Hydrophyllaceae	*Phacelia rafaelensis* N. D. Atwood	Garrison 54 (SFSU)	Walden et al. 2014		JQ250015	JX233474
Boraginales	Hydrophyllaceae	*Phacelia ramosissima* Douglas ex Lehm.	Ferguson 10 (GH)	Ferguson, 1998		AF047821	AF091199
Boraginales	Hydrophyllaceae	*Phacelia ramosissima* Douglas ex Lehm. var. *austrolitoralis* Munz	Walden 67 (SFSU)	Walden et al. 2014		JQ250016	
Boraginales	Hydrophyllaceae	*Phacelia ramosissima* Douglas ex Lehm. var. *eremophila* (Greene) J. F. Macbr.	Walden 83 (SFSU); -	Walden et al. 2014; Gilbert et al. 2005		JQ250024	AY630327

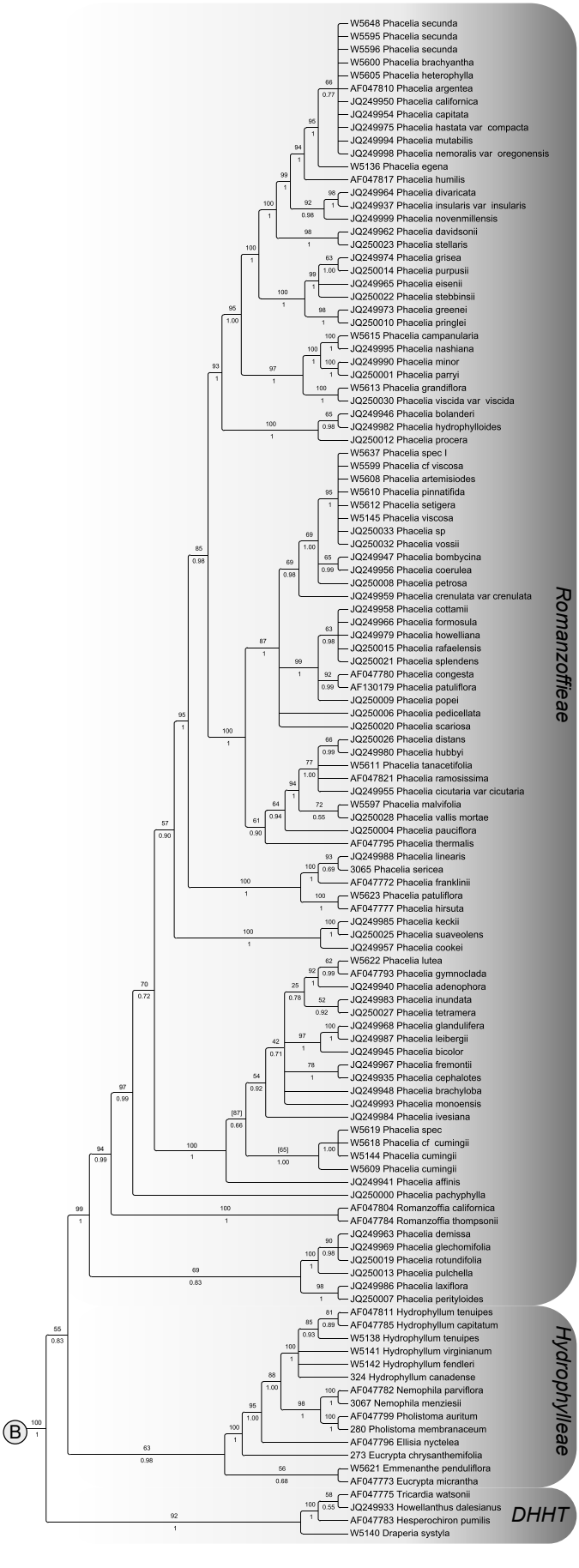
Boraginales	Hydrophyllaceae	*Phacelia ramosissima* Douglas ex Lehm. var. *subglabra* (Torr.) cronquist	Walden 200a (SFSU)	Walden et al. 2014		JQ250018	
Boraginales	Hydrophyllaceae	*Phacelia ramosissima* Douglas ex Lehm. var. *latifolia* M. Peck	Walden 76 (SFSU)	Walden et al. 2014		JQ250017	
Boraginales	Hydrophyllaceae	*Phacelia rotundifolia* Torr. ex S. Watson	Morefield & MaCarty 3274 (GH); -	Ferguson, 1998; Gilbert et al. 2005		AF047779	AY630328
Boraginales	Hydrophyllaceae	*Phacelia rupestris* Greene	Carr 16928	Walden et al. 2014			JX233475
Boraginales	Hydrophyllaceae	*Phacelia salina* (A.Nelson) J.T.Howell	-	Gilbert et al. 2005			AY630329
Boraginales	Hydrophyllaceae	*Phacelia scariosa* Brandegee	Leon 3274 (SDNHM)	Walden et al. 2014		JQ250020	JX233476
Boraginales	Hydrophyllaceae	*Phacelia scopulina* (A.Nelson) J.T.Howell	-	Gilbert et al. 2005		AY630330	
Boraginales	Hydrophyllaceae	*Phacelia secunda* J.F.Gmel.	M. Weigend et al. (B)		W5595	MT176350	
Boraginales	Hydrophyllaceae	*Phacelia secunda* J.F.Gmel.	M. Weigend et al. (B)		W5596	MT176351	
Boraginales	Hydrophyllaceae	*Phacelia secunda* J.F.Gmel.	F.Luebert 3224 (BONN)		W5648	MT176369	MT137555
Boraginales	Hydrophyllaceae	*Phacelia sericea* (Graham) A. Gray	Weigend 9346 (B)	Walden et al. 2014; Ferguson, 1998	3065	KF158037	AF091201
Boraginales	Hydrophyllaceae	*Phacelia sertigera* Phil.	T.Joßberger s.n. (BONN)		W5612	MT176360	MT137554
Boraginales	Hydrophyllaceae	*Phacelia spec. nov.*	M.Weigend 9991 (USM, HUSA, BONN)		W5637	MT176368	
Boraginales	Hydrophyllaceae	*Phacelia splendens* Eastwood,	Atwood and Welsh 22060 (TEX, n.v.)	Walden et al. 2014		JQ250021	JX233477
Boraginales	Hydrophyllaceae	*Phacelia stebbinsii* Constance & Heckard	Stebbins 7761 (DAV)	Walden et al. 2014		JQ250022	
Boraginales	Hydrophyllaceae	*Phacelia stellaris* Brand	Burrascano s.n. (SFSU)	Walden et al. 2014		JQ250023	
Boraginales	Hydrophyllaceae	*Phacelia suaveolens* Greene	Parker s.n. (SFSU)	Walden et al. 2014		JQ250025	
Boraginales	Hydrophyllaceae	*Phacelia tanacetifolia* Benth.	M.Weigend 9348 (B); Garrison 3	Hansen et al. 2010	W5611	MT176359	FJ814654
Boraginales	Hydrophyllaceae	*Phacelia tetramera* J. T. Howell	Tiehm 15641 (SFSU)	Walden et al. 2014		JQ250027	AY630333
Boraginales	Hydrophyllaceae	*Phacelia thermalis* Greene	Ferguson 125 (GH)	Ferguson, 1998		AF047795	AF091202
Boraginales	Hydrophyllaceae	*Phacelia vallis-mortae* J. W. Voss	Garrison 04 (SFSU)	Walden et al. 2014		JQ250028	
Boraginales	Hydrophyllaceae	*Phacelia viscida* (Benth.) Torr. var. *albiflora* A. Gray	Hansen 56 (SFSU); DH54	Walden et al. 2014; Hansen et al. 2010		JQ250031	FJ814650
Boraginales	Hydrophyllaceae	*Phacelia viscida* (Benth.) Torr. var. *viscida*	Hansen 54 (SFSU); DH56	Walden et al. 2014; Hansen et al. 2010		JQ250030	FJ814652
Boraginales	Hydrophyllaceae	*Phacelia viscosa* Phil.	F.Luebert,T.Böhnert 3640 (BONN)		W5145	MT176349	
Boraginales	Hydrophyllaceae	*Phacelia viscosa* Phil.	M. Weigend & Ch. Schwarzer 7828 (BSB)		W2367		MT137548
Boraginales	Hydrophyllaceae	*Phacelia vossii* N. D. Atwood		Walden et al. 2014		JQ250032	JX233480
Boraginales	Hydrophyllaceae	*Phacelia welshii* N.D.Atwood	Garrison 56 (SFSU)	Walden et al. 2014			JX233481
Boraginales	Hydrophyllaceae	Pholistoma *auritum* (Lindl.) Lilja	Ferguson 41 (GH)	Ferguson, 1998		AF047799	AF091204
Boraginales	Hydrophyllaceae	Pholistoma *membranaceum* (Benth.) Constance	Hoffmann H55/98 (B)	Walden et al. 2014	280	KF158031	
Boraginales	Hydrophyllaceae	*Romanzoffia californica* Greene	Ferguson 128 (GH)	Ferguson, 1998		AF047804	AF091205

Boraginales	Hydrophyllaceae	*Romanzoffia thompsonii* Marttala	Ferguson 134 (GH)	Ferguson, 1998		AF047784	AF091206
Boraginales	Hydrophyllaceae	*Tricardia watsonii* Torr. ex S. Watson	Morefield & MaCarty 3375 (GH)	Ferguson, 1998		AF047775	AF091209
Boraginales	Namaceae	*Eriodictyon californicum* (Hook. & Arn.) Torr.	s. col. 2663 (BONN)		W5662	MT176379	
Boraginales	Namaceae	*Eriodictyon californicum* (Hook. & Arn.) Torr.	Ferguson 59 (GH)	Ferguson, 1998		AF047820	AF091159
Boraginales	Namaceae	*Eriodictyon crassifolium* Benth	s. col. 1611 (BONN)		W5661	MT176378	
Boraginales	Namaceae	*Eriodictyon crassifolium* Benth. var. *crassifolium*	Ferguson 3 (GH); -	Ferguson, 1998		AF047789	AF091161
Boraginales	Namaceae	*Eriodictyon crassifolium* Benth. var. *nigrescens* Brand	J.Miller et al. 7978 (MO); -		W5602	MT176354	AF091162
Boraginales	Namaceae	*Eriodictyon sessilifolium* Greene	Ferguson 51 (GH)	Ferguson, 1998		AF047787	AF091163
Boraginales	Namaceae	*Eriodictyon trichocalyxA*. Heller var. *trichocalyx*	Ferguson 30 (GH)	Ferguson, 1998		AF047801	AF091164
Boraginales	Namaceae	*Nama aretioides* (Hook. & Arn.) Brand	Charlton 5415 (RSA)	Ferguson, 1998		AF047768	AF091172
Boraginales	Namaceae	*Nama californica* (A. Gray) J.D. Bacon	Thorne 53340 (RSA)	Ferguson, 1998		AF047769	AF091173
Boraginales	Namaceae	*Nama demissa* A.Gray	Hoffmann 10/98 (B)	Weigend et al. 2014	281	KF158032	AF091174
Boraginales	Namaceae	*Nama densa* Lemmon	s. col. s.n (BONN)		W5651	MT176370	
Boraginales	Namaceae	*Nama jamaicensis* L.	Halse 4439 (GH)	Ferguson, 1998		AF047791	AF091175
Boraginales	Namaceae	*Nama lobbii* A. Gray	Gustafson 3170 (RSA)	Ferguson, 1998		AF047764	AF091176
Boraginales	Namaceae	*Nama parvifolia* (Torr.) Greenm	Halse 4433 (GH)	Ferguson, 1998		AF047792	AF091177
Boraginales	Namaceae	*Nama pusilla* Lemmon ex A. Gray	Chambers 5827 (GH)	Ferguson, 1998		AF047794	AF091178
Boraginales	Namaceae	*Nama rothrockii* A.Gray	Wisura 4569 (RSA)	Ferguson, 1998			AF091179
Boraginales	Namaceae	*Nama rothrockii* A.Gray	Hilger & Hofmann USA 97/220 (B)	Weigend et al. 2014	243	KF158023	
Boraginales	Namaceae	*Nama sericea* Willd. ex Roem. & Schult.	Ferguson 53 (GH)	Ferguson, 1998		AF047798	AF091180
Boraginales	Namaceae	*Nama stenocarpa* A.Gray	Kellogg s.n. (RSA)	Ferguson, 1998		AF047766	AF091181
Boraginales	Namaceae	*Nama undulata* Kunth	van Devender 96-207 (RSA)	Ferguson, 1998		AF047771	AF091182
Boraginales	Namaceae	*Turricula parryi* (A. Gray) J.F. Macbr	Ferguson 29 (GH)	Ferguson, 1998		AF047816	AF091210
Boraginales	Namaceae	*Wigandia caracasana* Kunth.	Acuña et al. 1233 (USJ)		W5658	MT176375	
Boraginales	Namaceae	*Wigandia caracasana* Kunth.	Acuña et al. 1257 (USJ)		W5659	MT176376	
Boraginales	Namaceae	*Wigandia caracasana* Kunth.	Ferguson 62 (GH)	Ferguson, 1998		AF047812	AF091211
Boraginales	Namaceae	*Wigandia crispa* (Tafalla ex Ruiz & Pav.) Kunth	M. Weigend et al. 9533 (MJ14/034) BONN (2), HUT (1), USM (1)		W5656	MT176373	
Boraginales	Namaceae	*Wigandia ecuadorensis* Cornejo	T.Joßberger TJH-001 (BONN)		W5653	MT176371	
Boraginales	Namaceae	*Wigandia urens* (Ruiz & Pav.) Kunth	T.Joßberger COL-06 (BONN)		W5654	MT176372	
Boraginales	Namaceae	*Wigandia cf.* urens (Ruiz & Pav.) Kunth	M.Weigend et al. 9650 (MJ14/150)		W5660	MT176377	

			(BONN)				
Boraginales	Namaceae	*Wigandia cf. urens* (Ruiz & Pav.) Kunth	M.Weigend et al. 9650 (MJ14/150) BONN (2), HUT (1), USM (1)		W5657	MT176374	
Boraginales	Namaceae	*Wigandia urens* (Ruiz & Pav.) Kunth	Wen 8671 (US)	Ferguson (1999); Luebert & Wen (2008)	FL68	KF158074	AF091212
Boraginales	Wellstediaceae	*Wellstedia dinteri* Pilg.	Hilger s.n. 1994 (B)	Weigend et al. 2013	2670	KF158027	
Boraginales	Wellstediaceae	*Wellstedia socotrana* Balf.f.	Thulin et al. 10155 (UPS)	Weigend et al. 2013	3003	KF158035	
Boraginales	Wellstediaceae	*Wellstedia somalensis* Thulin & A.Johanss.	Thulin et al. 10084 (UPS)	Weigend et al. 2014	3004	KF158036	
Gentianales	Apocynaceae	*Nerium oleander* L.		Olmstead et al. (2000)		AF130168	
Gentianales	Loganiaceae	*Logania vaginalis* F.Muell.	Bremer 3013 (UPS)	Backlund et al. (2000)		AJ235837	
Gentianales	Rubiaceae	*Coffea arabica* L.		Nalapalli et al. (2006; unpub.)		NC_008535	
Gentianales	Gelsemiaceae	*Pteleocarpa lamponga* Bakh. ex K.Heyne	N. S. Mat Asri, FRI 26732 (KEP)	Refulio-Rodriguez & Olmstead 2014		HQ384849	

Figure S1

A. Boraginales phylogeny (Hydrophyllaceae) - maximum likelihood *ndhF* tree.



B. Boraginales phylogeny (remaining lineages) - maximum likelihood *ndhF* tree

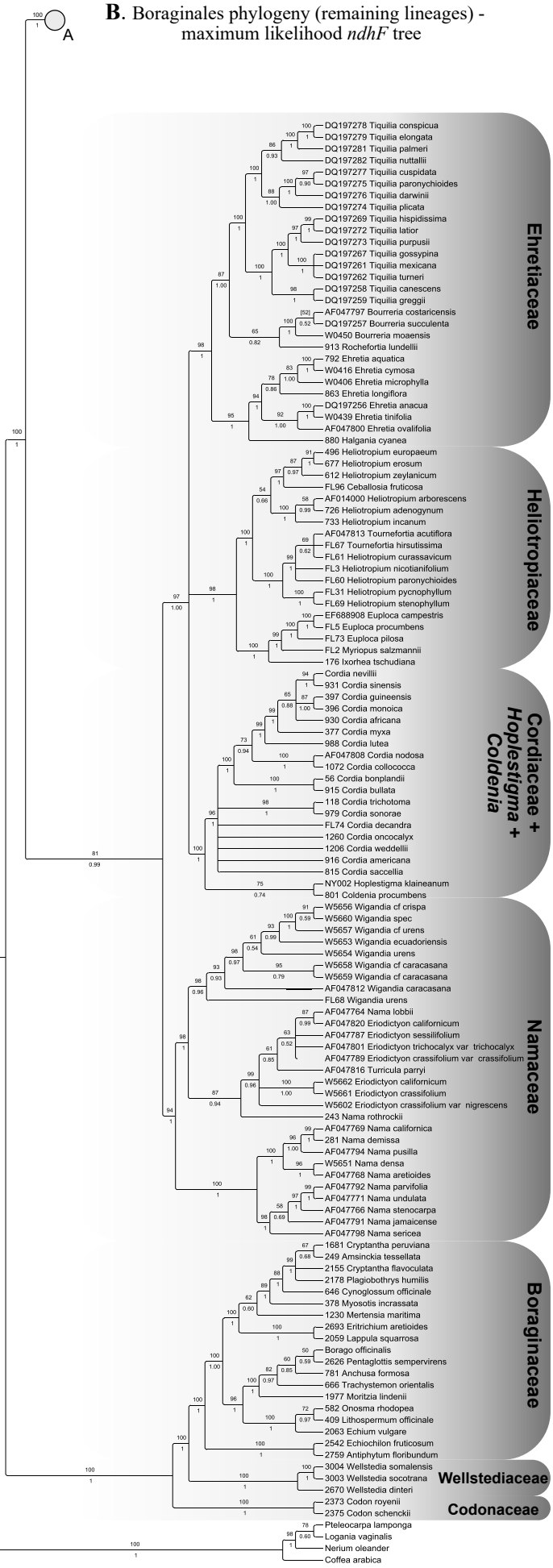


Figure S2. Maximum likelihood ITS tree with bootstrap values and without collapsed nodes.

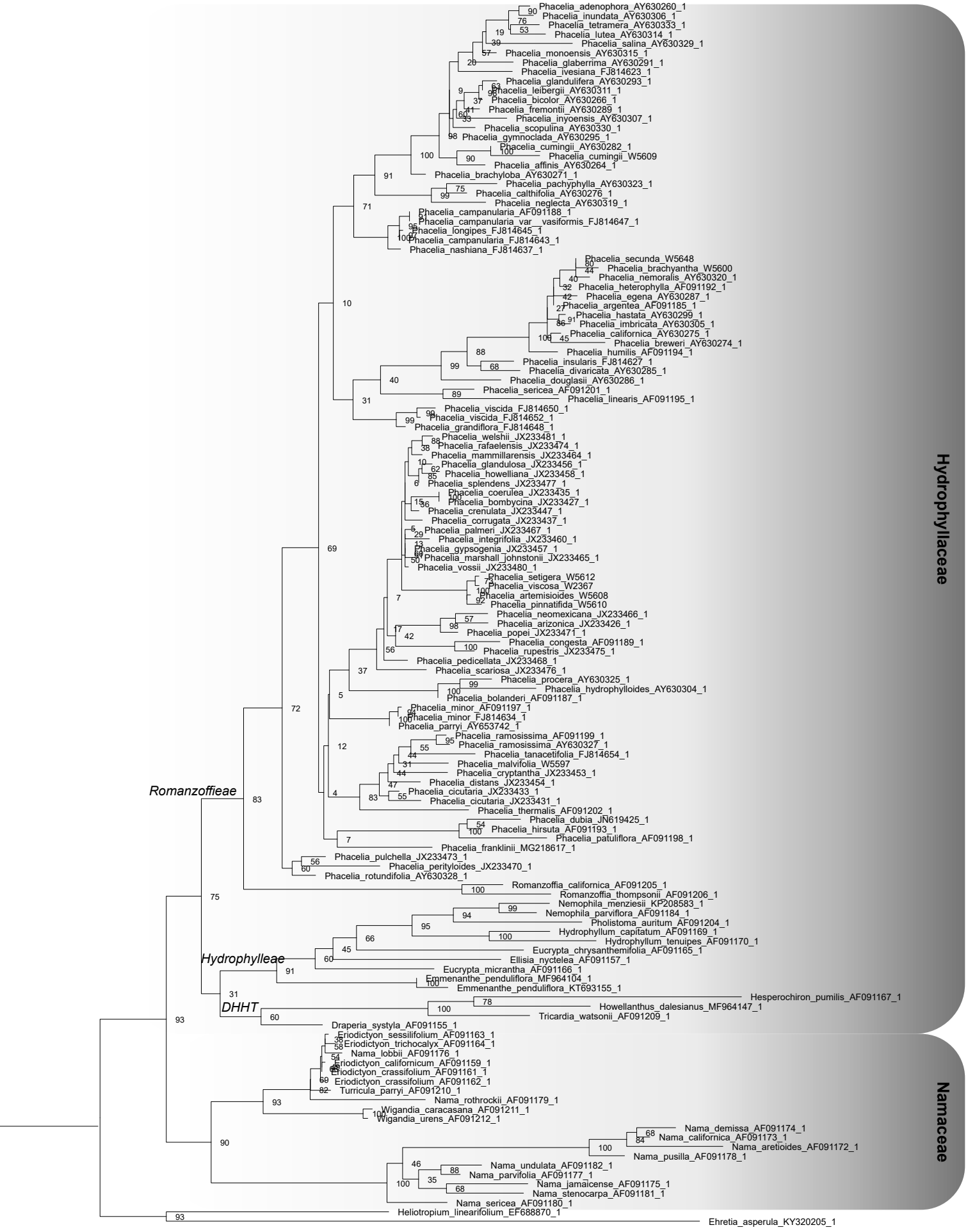
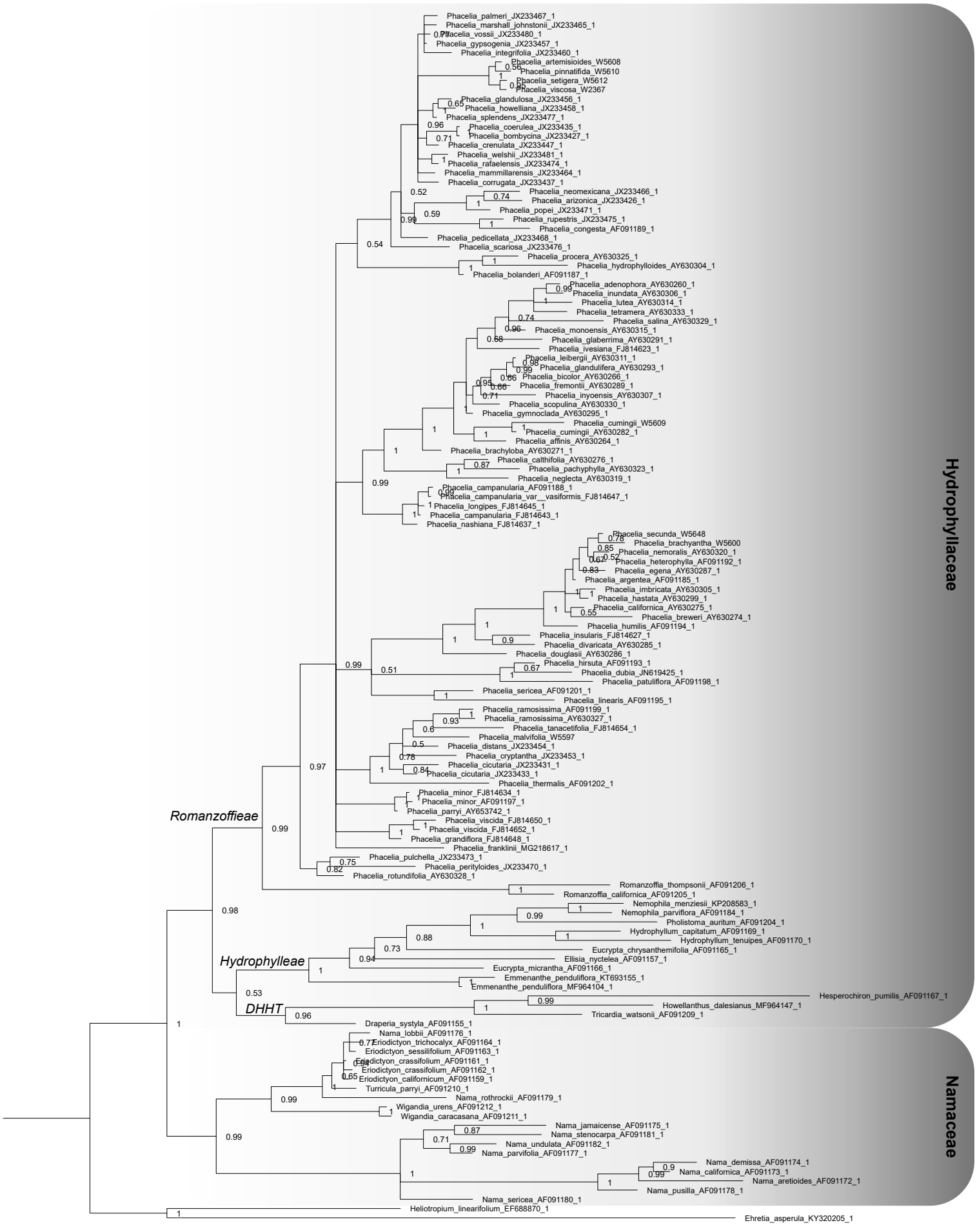


Figure S3. Bayesian inference ITS consensus tree with posterior probabilities.



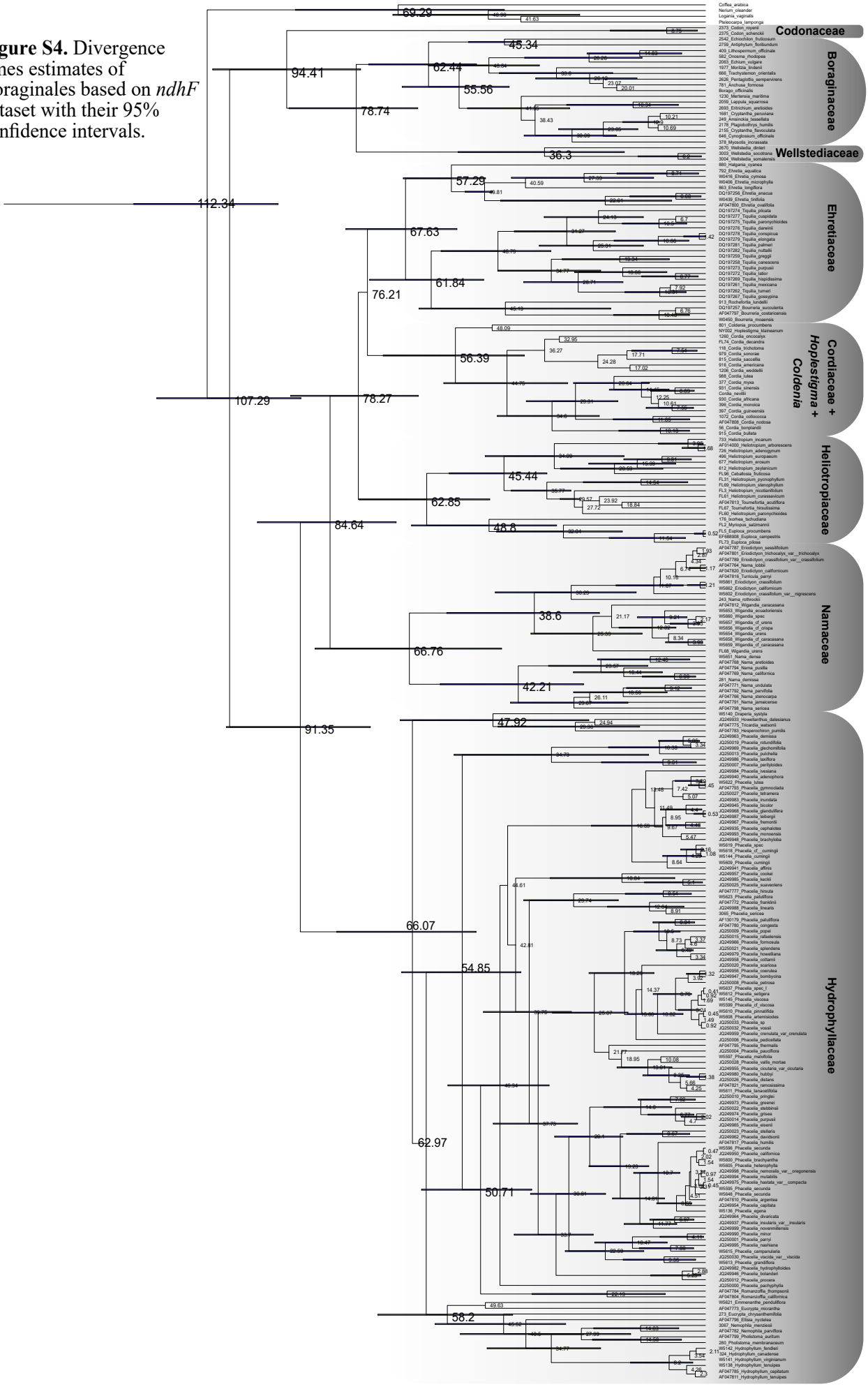


Figure S4. Divergence times estimates of Boraginales based on *ndhF* dataset with their 95% confidence intervals.

Biogeobears R Script. This R script was used for performing ancestral area reconstruction of Namaceae. The same script with different input data was used for Hydrophyllaceae.

```
setwd("C:/Users/mvasile/sciebo/Biogeography/Namaceae")

library(ape)
tree <- read.tree("Namaceae.tree")

if(!require(installr)) {
  install.packages("installr"); require(installr)} #load / install+load
installr

# using the package:
updateR()

install.packages("optimx", dependencies=TRUE,
repos="http://cran.rstudio.com")
install.packages("snow")

install.packages("rexpokit")
install.packages("ape")
install.packages("magik")

library(snow)
install.packages("phylobase", dependencies=TRUE,
repos="http://cran.rstudio.com")
install.packages("devtools")
library(devtools)
devtools::install_github(repo="nmatzke/BioGeoBEARS",INSTALL_opts=
"--byte-compile", dependencies=T)

library(BioGeoBEARS)

library(parallel)
library(optimx)
install.packages("ape")
library(ape)
install.packages("cladoRcpp")
library(cladoRcpp)
install.packages("geoscale")
library(geoscale)
install.packages("strap")
library(strap)#for ploting geologic time scale
library(phyloch) #for ploting height 95% bars
load("Namaceae_Biogeobears.rda")

#with package phyloch:
p <- read.nexus("Boraginales_Gentianales_treeannotator.tree")
pB<-read.beast("/Users/mvasile/sciebo/Biogeography/Namaceae/
Boraginales_Gentianales_treeannotator.tree")#read beast tree again with beast stats
#pBC<-drop.tip(ladderize(pB,right = F),"W5660_Wigandia_spec")
#plot(pBC, cex=0.2) ##first ladderize and extract clade

pB1<-ladderize(extract.clade(pB,node=getMRCA(pB,c(
"AF047787_Eriodictyon_sessilifolium","AF047798_Nama_sericea"))), right = F)

pB1
pB0<-extract.clade2(pB,node=getMRCA(pB,tip=c("AF047798_Nama_sericea",
"AF047787_Eriodictyon_sessilifolium")))
plot(pB1)
tiplabels()
pB1$tip.label
nodelabels()
write.tree(pB1, file = "Namaceae_new2.tree")

getwd()

list.files()
BioGeoBEARS_model_object =define_BioGeoBEARS_model_object(minval_anagenesis = 1e-15,
minval_cladogenesis = 1e-03, maxval = 3)
BioGeoBEARS_run_object = define_BioGeoBEARS_run(BioGeoBEARS_model_object=
  BioGeoBEARS_model_object,trfn="Namaceae.tree",geogfn="Distr_Nam.txt",num_cores_to_use=4,
  max_range_size=3)
```

```
BioGeoBEARS_run_object$geogfn = "Distr_Nam.txt"

Namaceae = bears_optim_run(BioGeoBEARS_run_object)

Namaceae
states_list_0based_index = rcpp_areas_list_to_states_list(areas=c("A","B","C"),maxareas=3)


colors_matrix = get_colors_for_numareas(3)
colors_list_for_states = mix_colors_for_states(colors_matrix, states_list_0based_index)

colors_list_for_states
colors_list_for_states

nodestates<-get_ML_states_from_relprobs(Namaceae$
ML_marginal_prob_each_state_at_branch_top_AT_node[,2:7],statenames=c("A","B",
"C", "AB", "AC","BC", "ABC" ))

phy<-read.tree("Namaceae.tree")
phy$tip.label
phy$root.time<-branching.times(phy)[1]

coltips<-read.table("Colours_Nam.txt",row.names=1)
coltips<-coltips[phy$tip.label,]
phy$tip.label

coltips

#pB1
pB1$height
#pB1$root.time <- branching.times(pB1)[1]
pB1$"height_95%_HPD_MAX"
pB1$"height_95%_HPD_MIN"
pB1$root.time <- pB0$height[1]
pB1$"height_95%_HPD_MIN"<-pB0$"height_95%_HPD_MIN"
pB1$"height_95%_HPD_MAX"<-pB0$"height_95%_HPD_MAX"
length(branching.times(pB1))

# With only AXIS:

pdf("Namaceae.pdf", width=10, height=10)


par(xpd=NA)
phy$root.time<-branching.times(phy)[1]


plot.phylo(phy,type = "p", cex=0.6,use.edge.length = T,label.offset = 4.0, font=,edge.width
= 1.3,show.node.label = FALSE)

plot_BioGeoBEARS_results(Namaceae,plotwhat="pie",tipcex=0.3,statecex=0.3,plotsplits=F,
    plotlegend=F,legend_cex=0.5,titlecex=0,label.offset=0.8, cornercoords_loc="C:/Users/
mvasile/Documents/R/R-3.6.0/library/BioGeoBEARS/extdata/a_scripts",skiptree=T,tipboxes_TF=F)
nodelabels(nodestates[30:57], adj=c(1.5,-1), frame="none",cex=0.5)

axisPhylo(x=12.75, y = 72, cex.axis=0.75, line=4.2)

#table_colors<-read.table("Colours_Nam.txt", header = F, sep=" ")
tiplabels(pch=22, bg=as.character(coltips$V2), cex = 0.6, adj = 0.900, lwd=0.8)
tiplabels(pch=22, bg=as.character(coltips$V3), cex = 0.6, adj = 2.000, lwd=0.8)
tiplabels(pch=22, bg=as.character(coltips$V4), cex = 0.6, adj = 3.100, lwd=0.8)


color1<- c("blue","green", "red")
legend(x= 12.75, y =72, title="Biogeographic units",legend=c("North America (A)","Central
America (B)", "South America (C)","North and Central America (AB)", "North and South
America (AC)", "Central and South America (CB)", "North, Central and South America (ABC)"),
cex=0.9, pch=22, pt.bg = color1, pt.lwd=0.2,bty="n")
text(x=50.72, y = 100.7,"A B C",cex=0.6)

dev.off()
```

```
# With GEOSCALE:

pdf("Namaceae_18_02_2020b.pdf", width=12, height=10)
par(xpd=NA)
phy$root.time<-branching.times(phy)[1]

geoscalePhylo(tree=pB1,boxes="Epoch",units=c("Period", "Epoch"),cex=0.7,
use.edge.length = T,cex.ts=0.55,cex.tip=0.6,x.lim=c(-20,90),y.lim=c(2,
30),
width=1,quat.rm=T,tick.scale="no",erotate=0,label.offset=4)

HPDbars(pB1, label = c("height_95%_HPD"), col = rgb(153, 255, 255, 130,
 names = NULL, maxColorValue = 255), lwd = 5)

plot_BioGeoBEARS_results(Namaceae,plotwhat="pie",tipcex=0.5,
statecex=0.45,
plotsplits=F,plotlegend=F,legend_cex=0.5,titlecex=0,label.offset=0.8,
cornercoords_loc="C:/Users/mvasile/Documents/R/R-
3.6.0/library/BioGeoBEARS/
extdata/a_scripts",skiptree=T,tipboxes_TF=F)

#nodelabels(nodestates[29:57], adj=c(1.5,-1), frame="none",cex=0.7)
#nodestates[29 means tips number (number of species + 1)
#and 57 means number of tips (28) + internal nodes]


axisPhylo(cex.axis=0.75, line=6)
#
table_colors<-read.table("Colours_Nam.txt", header = F, sep=" ")
tiplabels(pch=22, bg=as.character(coltips$V2), cex = 0.8, adj = 0.900,
lwd=0.5)
tiplabels(pch=22, bg=as.character(coltips$V3), cex = 0.8, adj = 1.800,
lwd=0.5)
tiplabels(pch=22, bg=as.character(coltips$V4), cex = 0.8, adj = 2.700,
lwd=0.5)
color1<- c("blue","green","red", "Cyan", "darkmagenta", "olivedrab",
"white")

legend(x= -5.5, y = 31, title="Biogeographic units",legend=c("North
America (A)"
,"Central America (B)", "South America (C)","North and Central America
(AB)",
 "North and South America (AC)", "Central and South America (BC)",
  "North, Central and South America (ABC)"), cex=1, pch=21, pt.bg =
color1,
   pt.lwd=0.2,bty="n")

text(x=68.2, y = 29,"A  B  C",cex=0.55)
dev.off()

Namaceae_data<-Namaceae$ML_marginal_prob_each_state_at_branch_top_AT_node
write.table(Namaceae_data, file='Namaceae_prob.csv', sep=';', dec=',',
 row.names=FALSE)

phy

plot(phy)
nodelabels()

save.image("Namaceae_Biogeobears.rda")
branching.times(tree)
plot.
```

Appendix B

—Supplementary material of chapter 3—

Table S1. Species included in this study and voucher information.

Supplementary material S1. RAxML trees of the genus *Phacelia* based on ddRAD sequences for the 1st and 2nd sets of assemblies. Trees are presented in the following order: 1st set of assembly (85%, 90%, 95% clustering thresholds) and 2nd set of assembly (85%, 90%, 95% clustering thresholds).

Supplementary material S2. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 3rd set of assemblies.

Supplementary material S3. TetRAD trees of the genus *Phacelia* based on ddRAD sequences using the 3rd set of assemblies.

Supplementary material S4. TreeMix iterations at different SNPs subsamples.

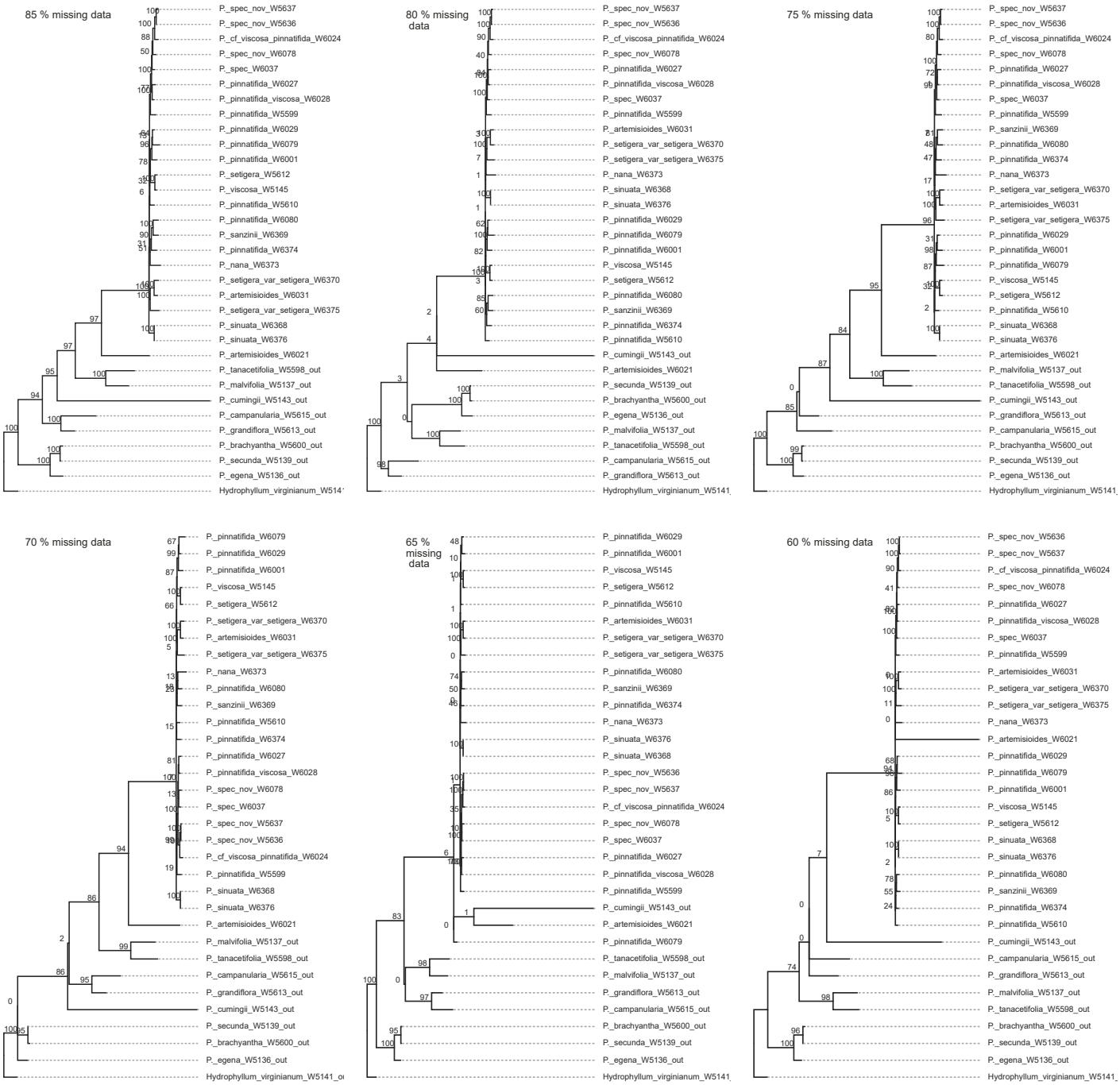
Jupyter-Notebooks. Python scripts documenting the following parts of the downstream analyses: assembly of *Phacelia* ddRAD data with assembly statistics; phylogenetic analyses (RAxML and Tetrad); Structure Analysis; PCA Analysis; Treemix Analysis.

Table S1. Species included in this study and voucher information.

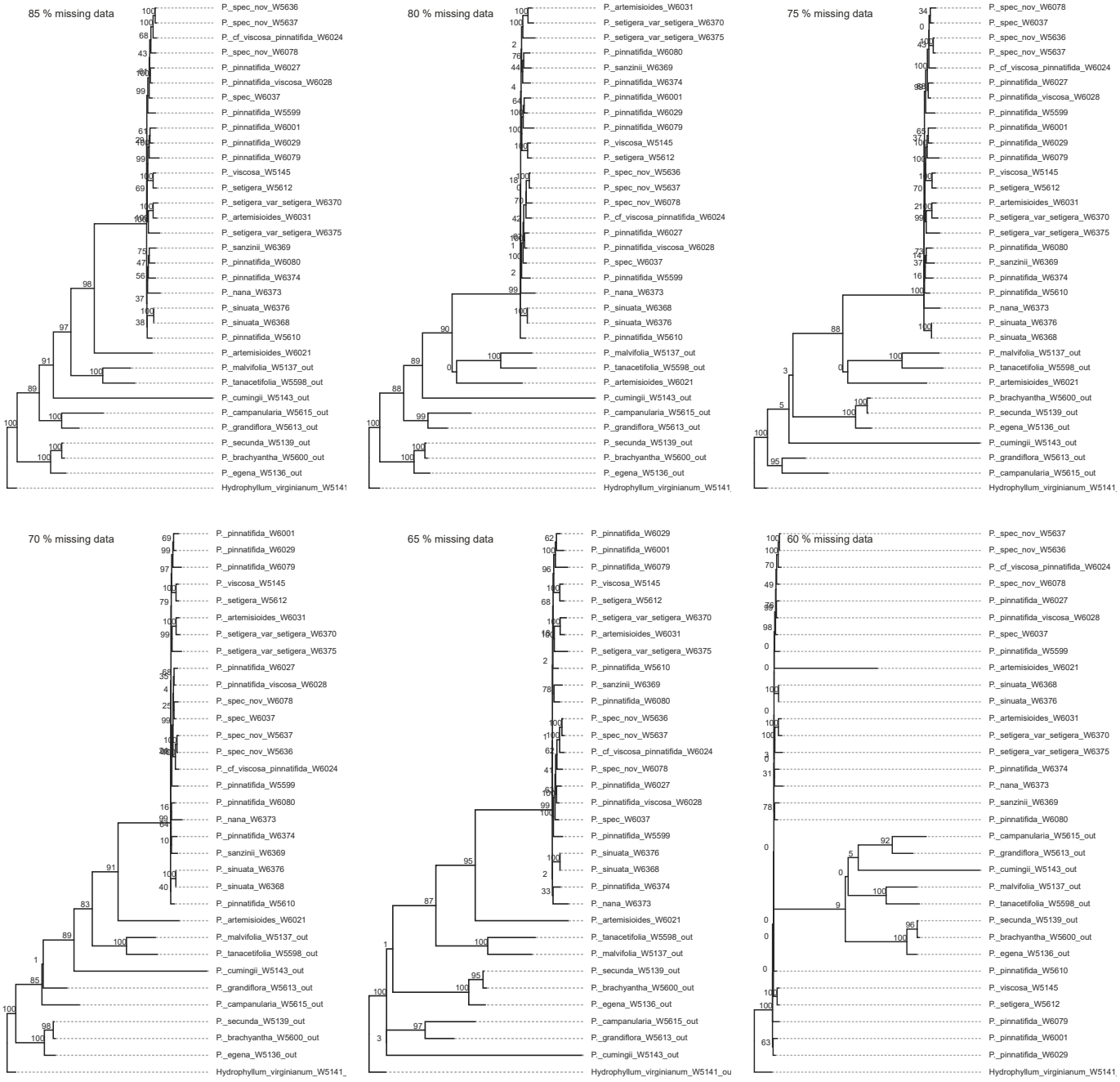
Voucher information (geographic origin, collector, collection number, herbarium, IPEN if applicable) and DNA numbers of all species included in this study. A dash (–) indicates missing information. All sequences generated for this study will be deposited at the European Nucleotide Archive (ENA)
Hydrophyllum virginianum L., –, –, –, XX-0-BONN-36861, W5141; ***Phacelia artemisioides*** Griseb. **1**, Argentina, S. Gisela et al., 382 (LP), W6031; ***Phacelia artemisioides* 2**, Argentina, Leuenberger & Arroyo, 3708 (B), W6021; ***Phacelia brachyantha*** Benth., –, F. Luebert & A. Moreira, 3041 (B), W5600; ***Phacelia campanularia*** A. Gray, –, M. Weigend, s.n. (BONN), XX-0-BONN-34716, W5615; ***Phacelia cumingii*** (Benth.) A. Gray, Chile, Luebert & Böhnert 3696 (BONN), W5143; ***Phacelia egena*** (Brand) Greene ex C. F. Baker, –, Felix Trabert, s.n. (BONN), XX-0-BONN-37009, W5136; ***Phacelia grandiflora*** (Benth.) A. Gray, –, T. Joßberger, s.n. (BONN), XX-0-BONN-34686, W5613; ***Phacelia malvifolia*** Cham. & Schltdl., –, T. Joßberger, 1486 (BONN), XX-0-BONN-37883, W5137; ***Phacelia nana*** Wedd., Argentina, J.M. Acosta 1159 (SI), W6373; ***Phacelia pinnatifida*** Griseb. ex Wedd. **1**, Argentina, J.M. Acosta, 877 (SI), W6374; ***Phacelia pinnatifida* 2**, Argentina, Novara, 6098 (B), W6080; ***Phacelia pinnatifida* 3.**, Argentina, S. Gisela et al., 378 (LP), W5610; ***Phacelia pinnatifida* 4**, Peru, M. Weigend & K.A. Peña Ramos, MW9739 (BONN), W5599; ***Phacelia pinnatifida* 5**, Peru, Schwabe, s.n. (B), W6079; ***Phacelia pinnatifida* 6**, Peru, H.H. Hilger, 2001/9 (B), W6001; ***Phacelia pinnatifida* 7**, Peru, Cáceres, Perez, Zuniga, 7512 (HUSA), W6029; ***Phacelia pinnatifida* 8**, Peru, Cordova et al., 025 (HUSA), W6027; ***Phacelia sanzinii*** Hicken, Argentina, Paula-Souza et al., 7873 (SI)[1], W6369; ***Phacelia secunda*** J.F.Gmel., –, M. Neumann, CL73/15 (BONN), XX-0-BONN-38409, W5139; ***Phacelia setigera*** Phil. **1**, Argentina**,** J.M. Acosta 1046 (SI), W6375; ***Phacelia setigera* 2**, Argentina, L. Aagesen et al. 12082 (SI), W6370; ***Phacelia setigera* 3**, Chile, M. Belov (BONN), CL-0-BONN-34622, W5612; ***Phacelia setigera* var. *humahuaquense*** Deginani, Argentina, M.V. Nicola et al. 13528 (SI), W6371; ***P. sinuata*** Phil., Argentina, N. Brignone 31 (SI), W6368, W6376; ***Phacelia spec. nov.***, Peru, M. Weigend, MW9974 (BONN), W5636; ***Phacelia spec. nov.,*** Peru, M. Weigend, MW9991 (BONN), W5637; ***Phacelia spec. nov.***, Peru, D. B. Montesinos, 7370 (BONN), W6078; ***Phacelia spec.***, Peru, Caceres & Fernadez, 6184 (B), W6037; ***Phacelia tanacetifolia*** Benth., –, –, –, XX-0-BONN-31966, W5598; ***Phacelia viscosa*** Phil. **1**, Chile, Luebert & Böhnert 3640 (BONN) W5145; ***Phacelia* cf. *viscosa* 2**, Peru, Weigend, 9192-2 (B), W6024; ***Phacelia viscosa* 3.**, Peru, Caceres & Baldarrago, 3282 (HUSA) W6028;

[1] *The macromorphology and locality of the P. sanzinii specimen (Paula-Souza et al., 7873 SI) coincides that of previous collections and descriptions but due to slightly divergent micromorphology (developed scales) the sample is treated as P. sanzinii but mentioned as P. sanzinii (cf. pinnatifida) in the figures.*

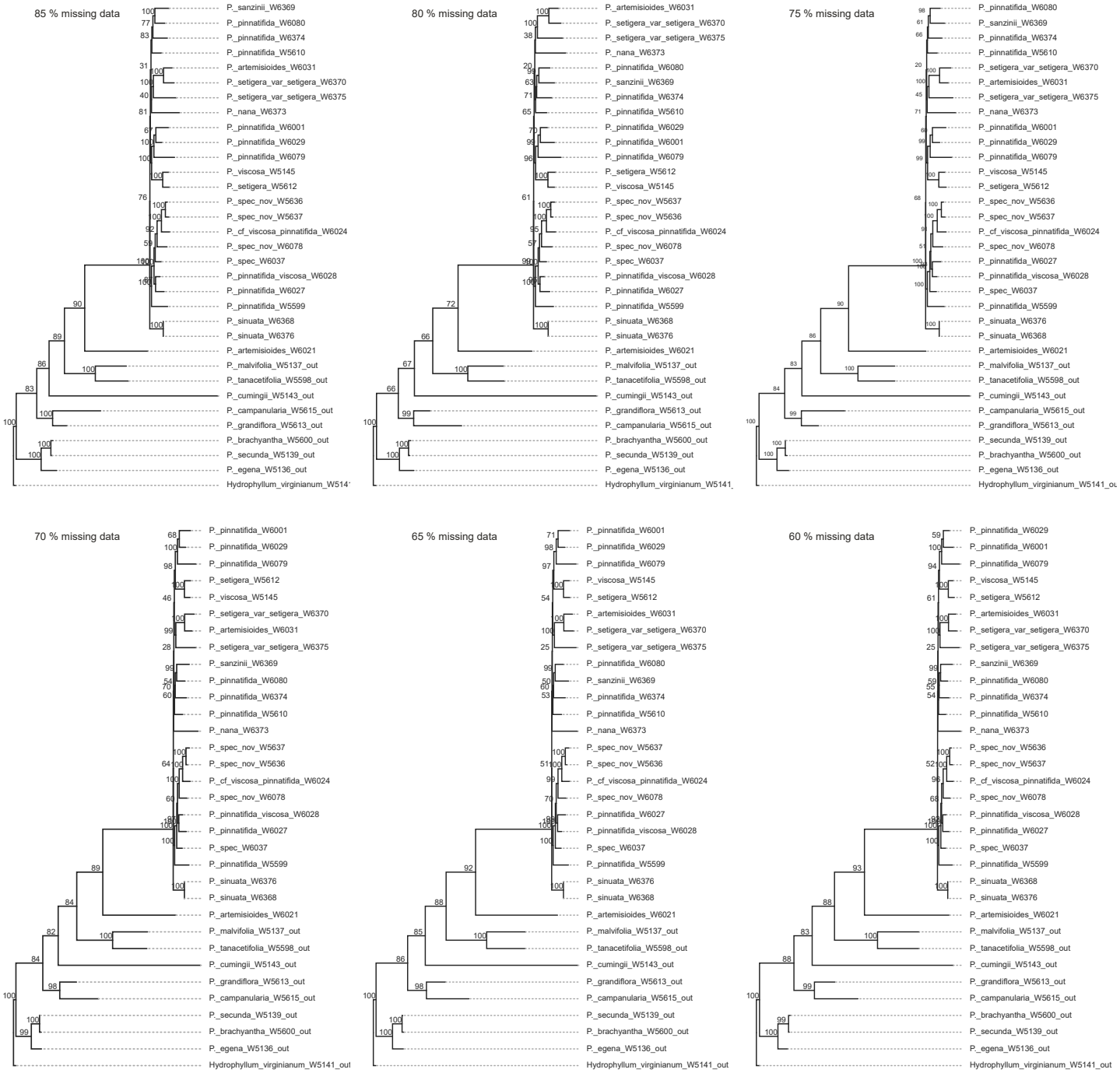
Supplementary material S1. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 1st set of assemblies. Clustering threshold 85% with six different percentages of missing data.



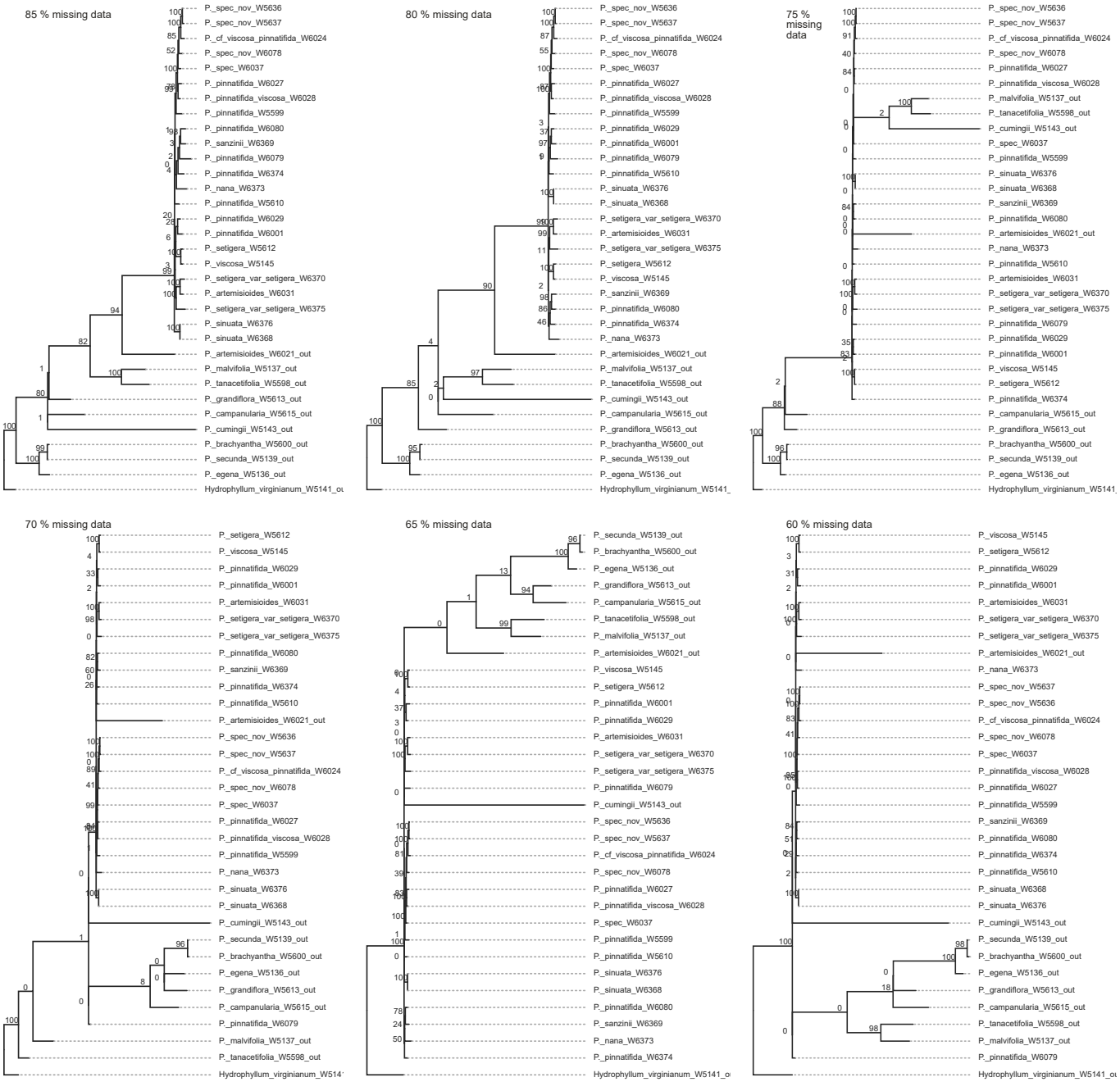
Supplementary material S1. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 1st set of assemblies. Clustering threshold 90% with six different percentages of missing data.



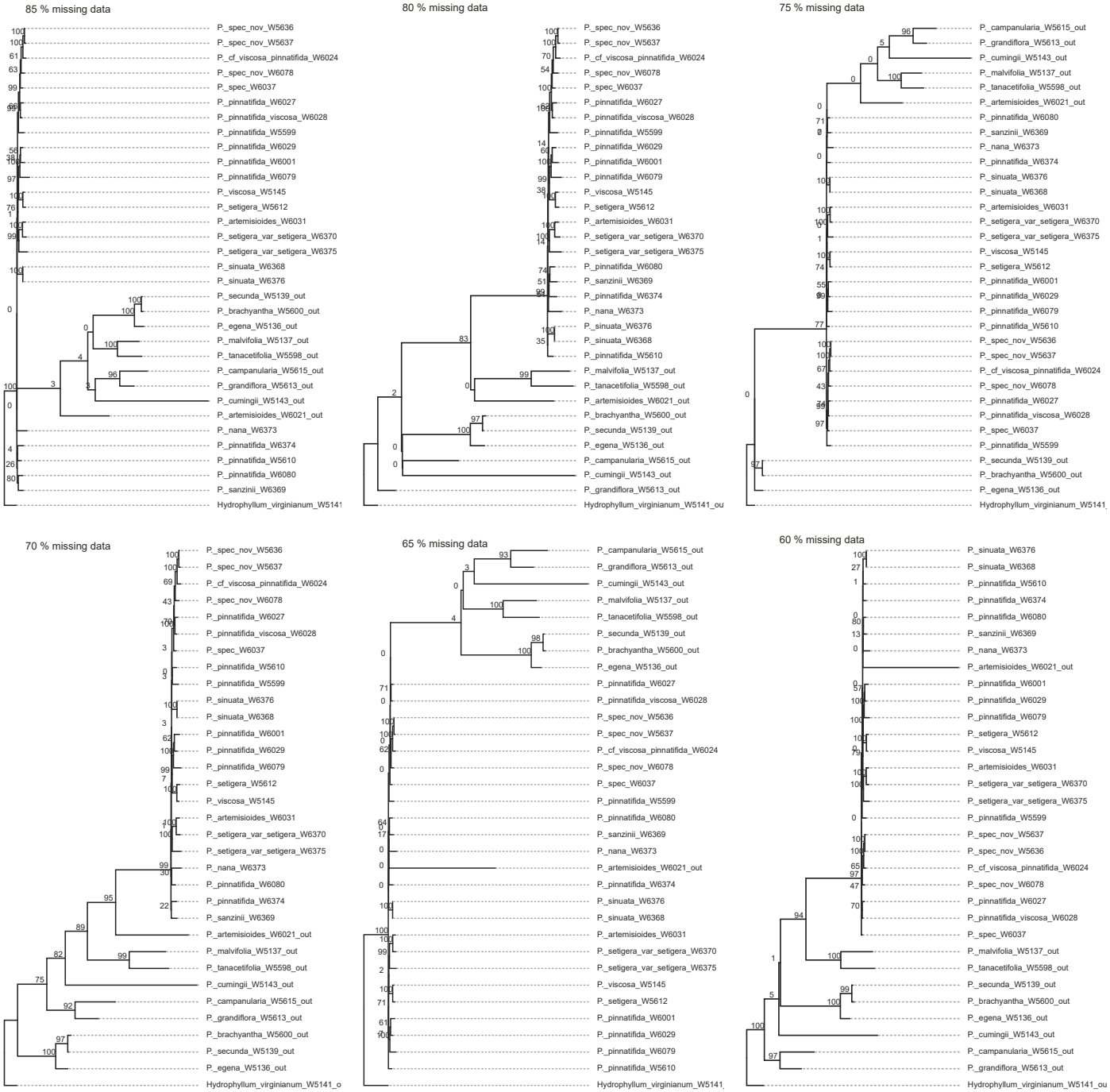
Supplementary material S1. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 1st set of assemblies. Clustering threshold 95% with six different percentages of missing data.



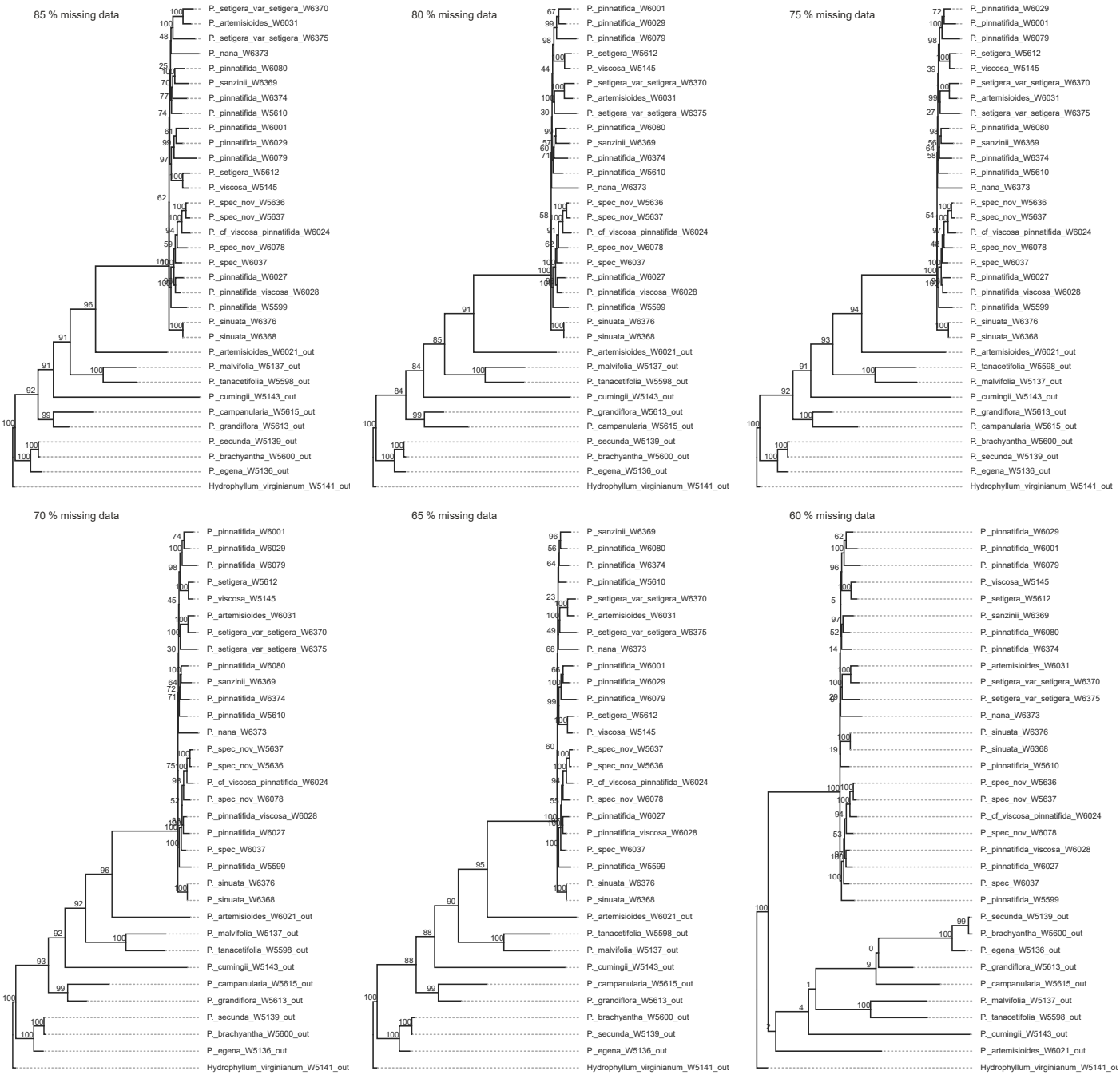
Supplementary material S1. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 2nd set of assemblies. Clustering threshold 85% with six different percentages of missing data.



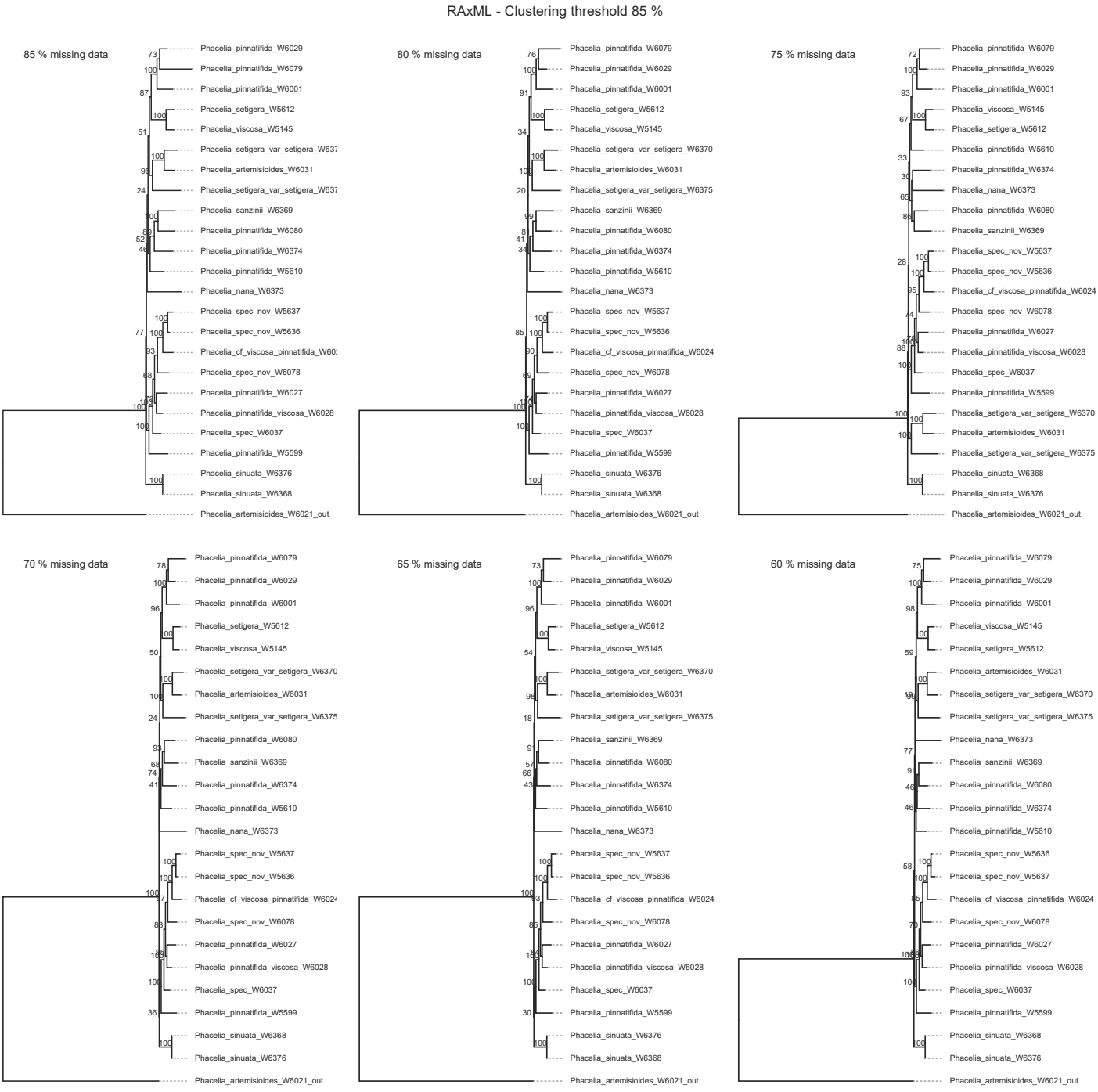
Supplementary material S1. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 2nd set of assemblies. Clustering threshold 90% with six different percentages of missing data.



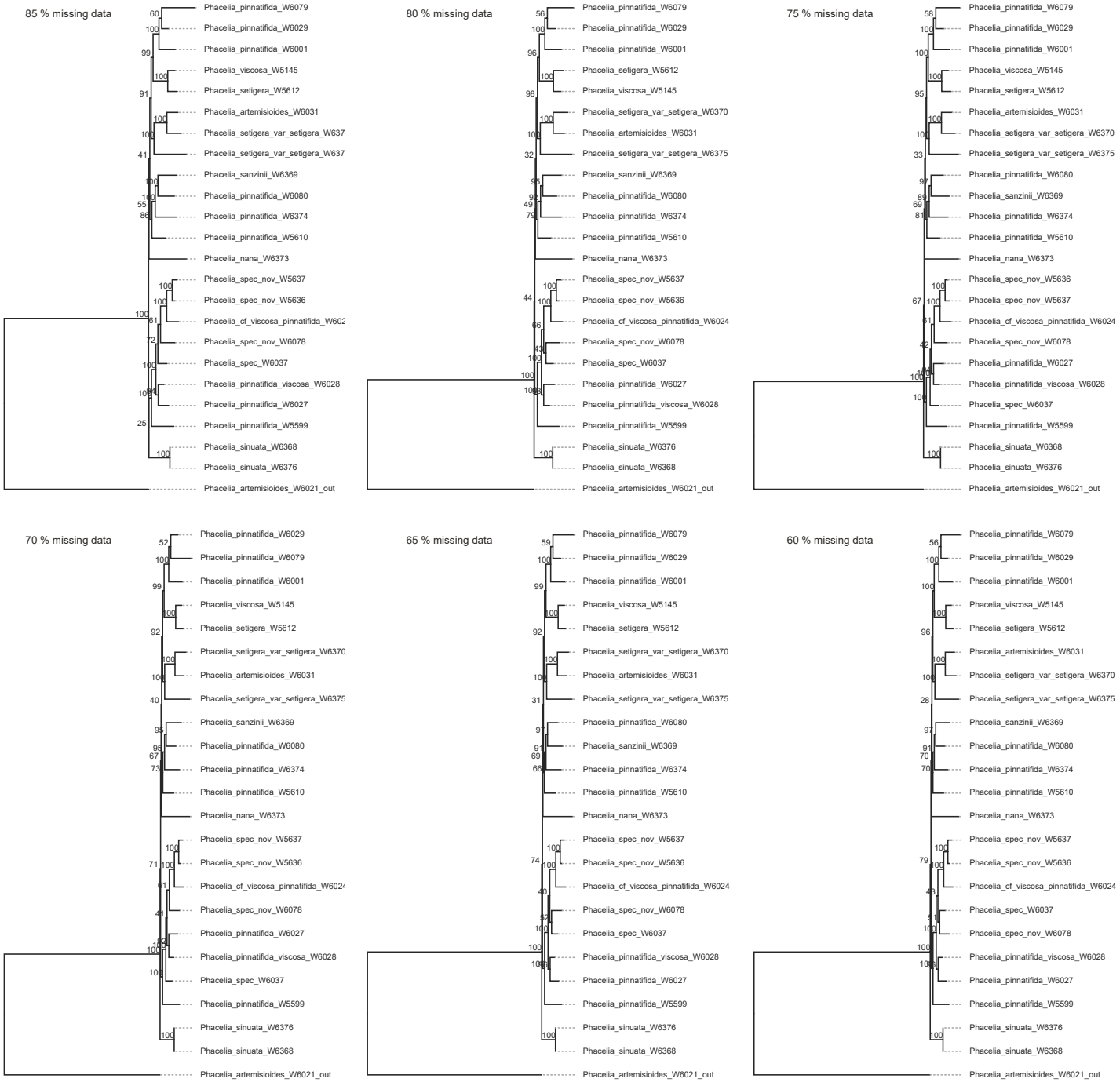
Supplementary material S1. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 2nd set of assemblies. Clustering threshold 95% with six different percentages of missing data.



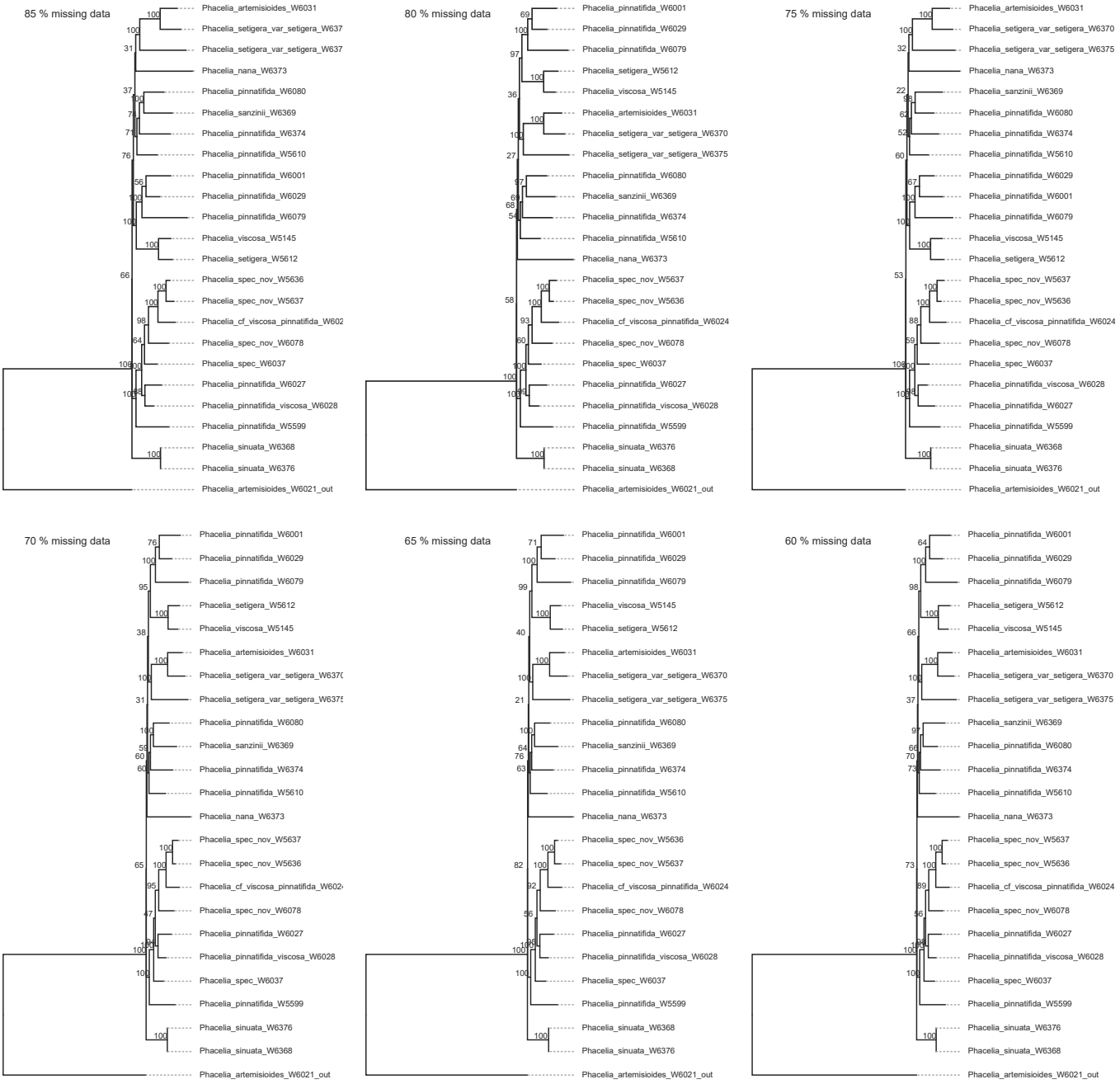
Supplementary material S2. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 3rd set of assemblies. Clustering threshold 85% with six different percentages of missing data.



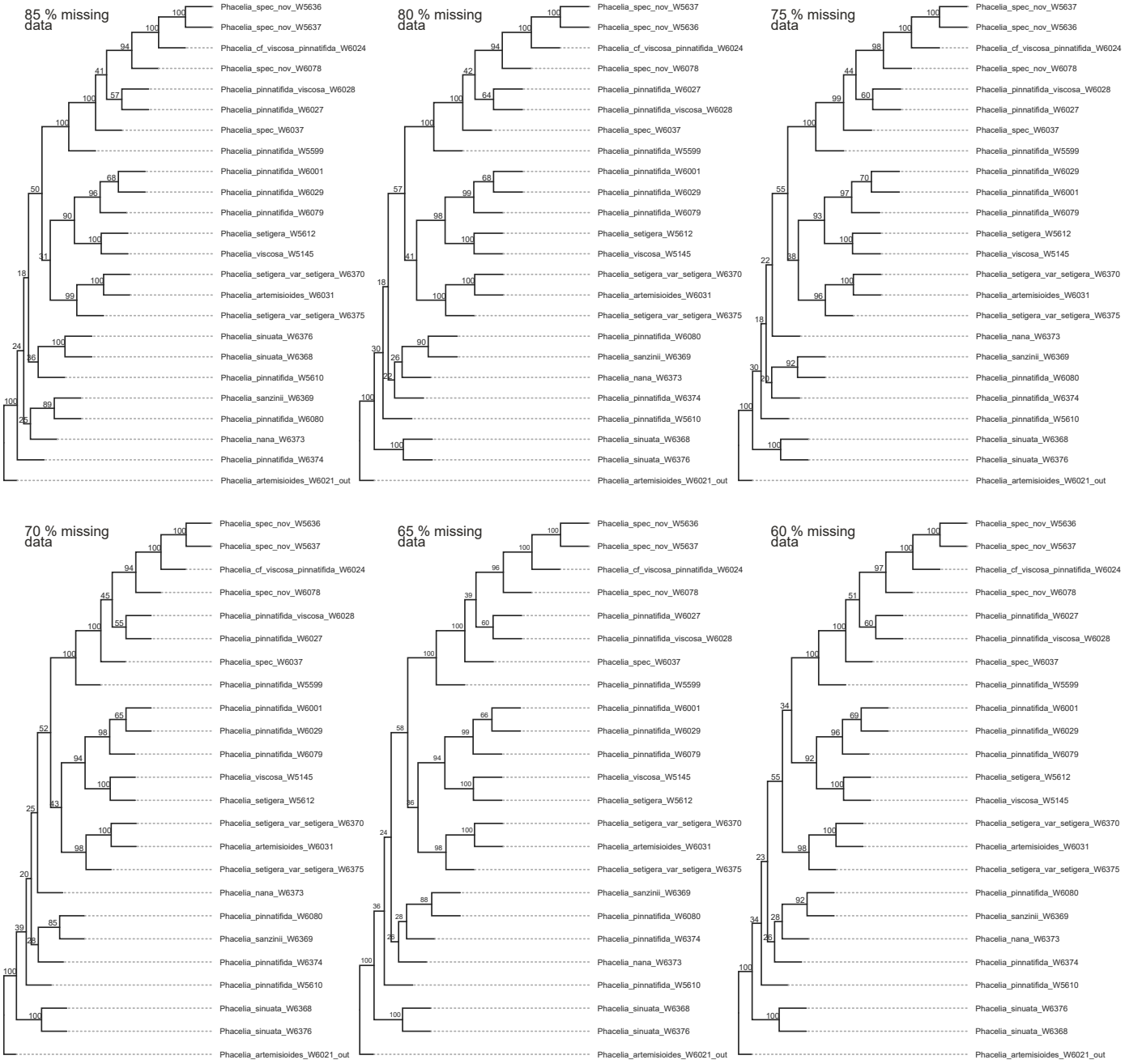
Supplementary material S2. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 3rd set of assemblies. Clustering threshold 90% with six different percentages of missing data.



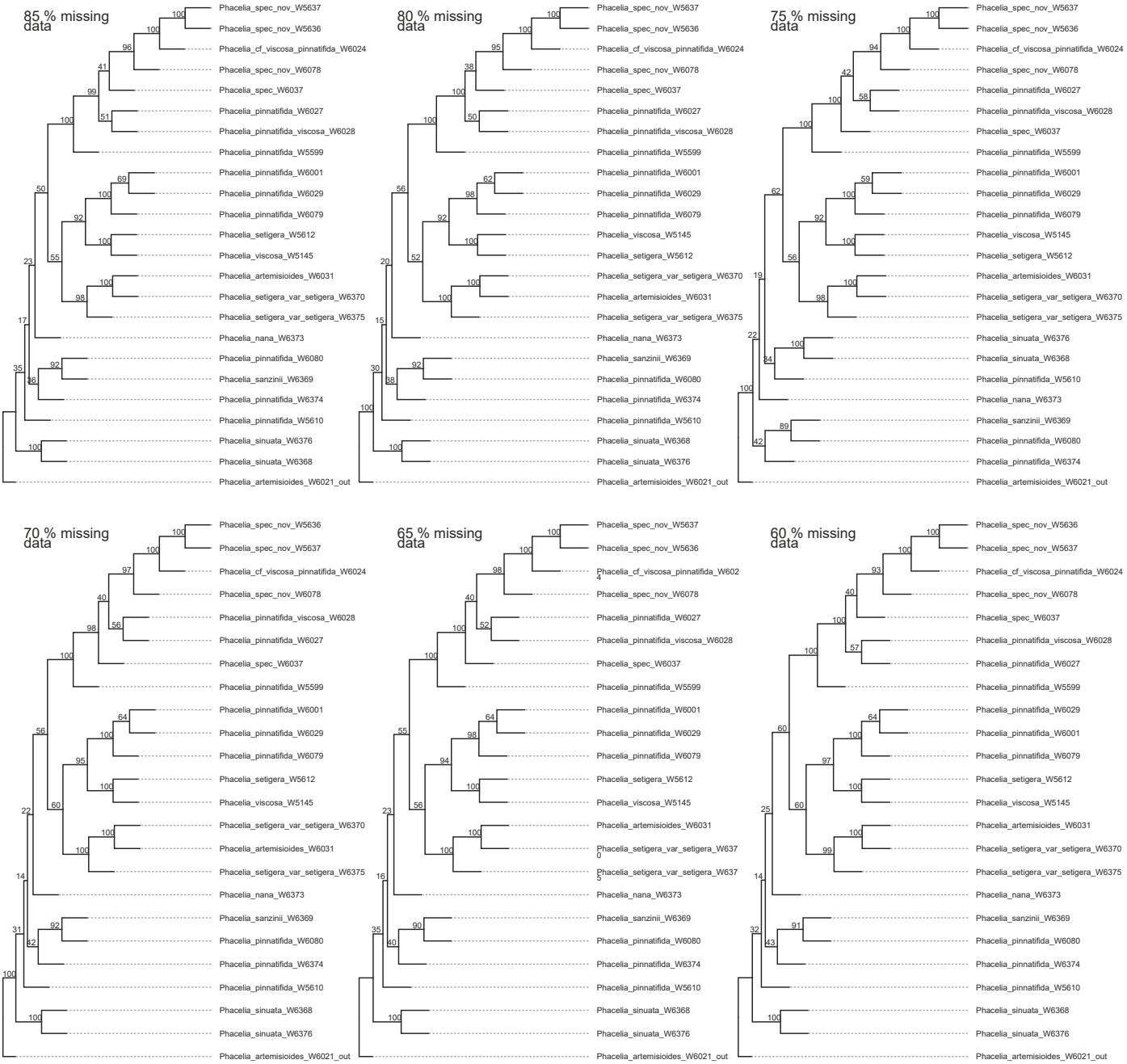
Supplementary material S2. RAxML trees of the genus *Phacelia* based on ddRAD sequences using the 3rd set of assemblies. Clustering threshold 95% with six different percentages of missing data.



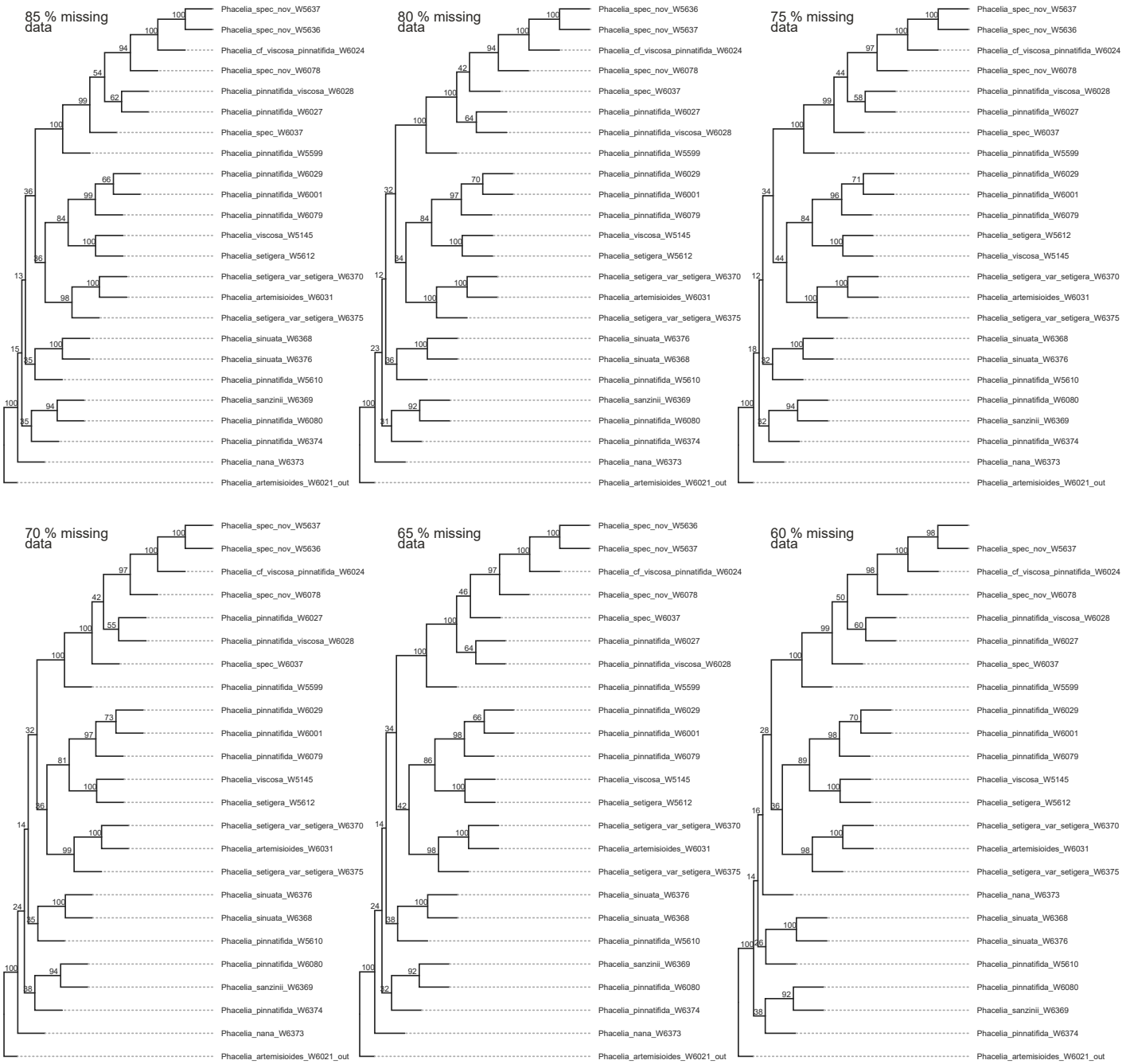
Supplementary material S3. TetRAD trees of the genus *Phacelia* based on ddRAD sequences using the 3rd set of assemblies. Clustering threshold 85 % with six different percentages of missing data.



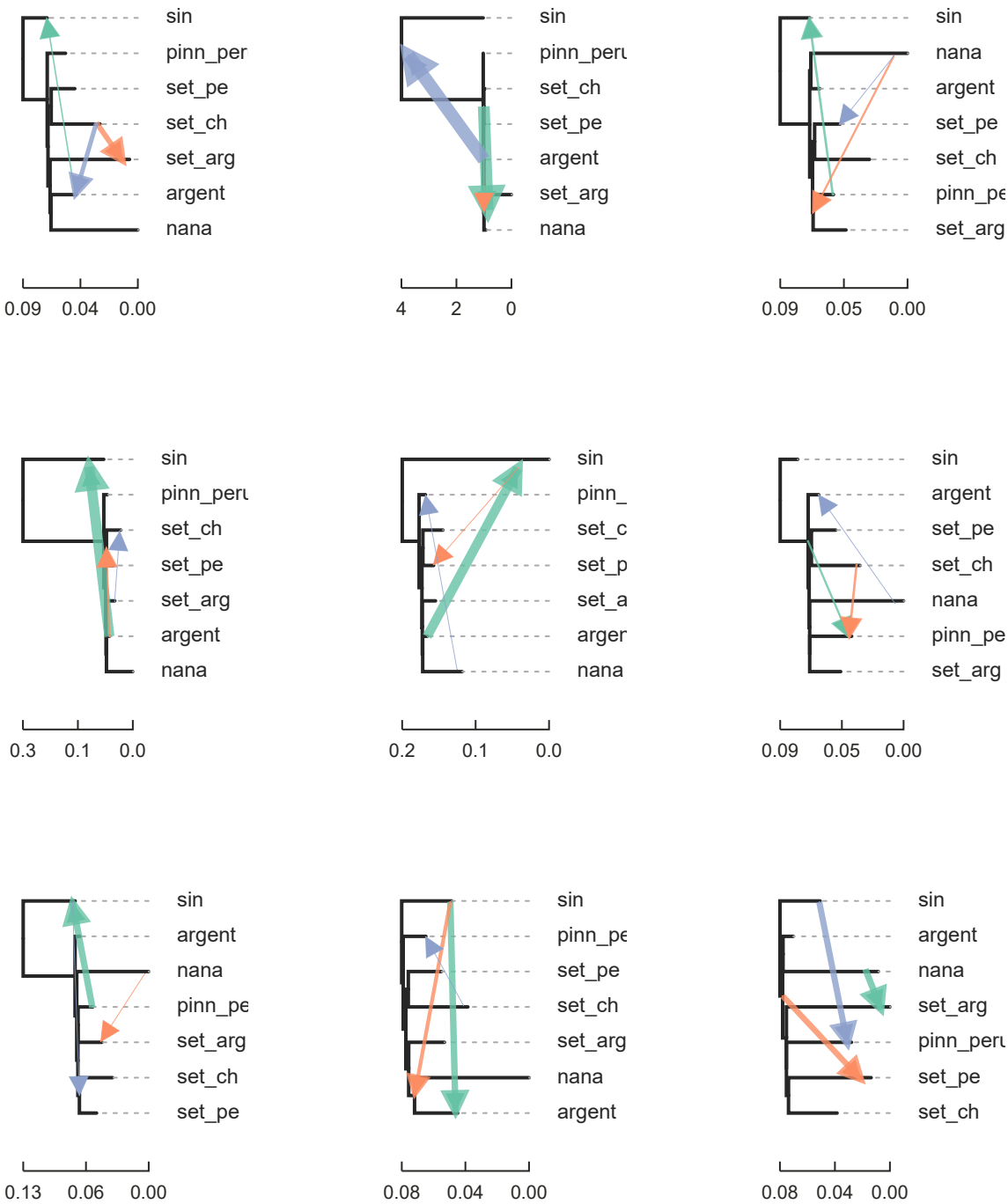
Supplementary material S3. TetRAD trees of the genus *Phacelia* based on ddRAD sequences using the 3rd set of assemblies. Clustering threshold 90 % with six different percentages of missing data.



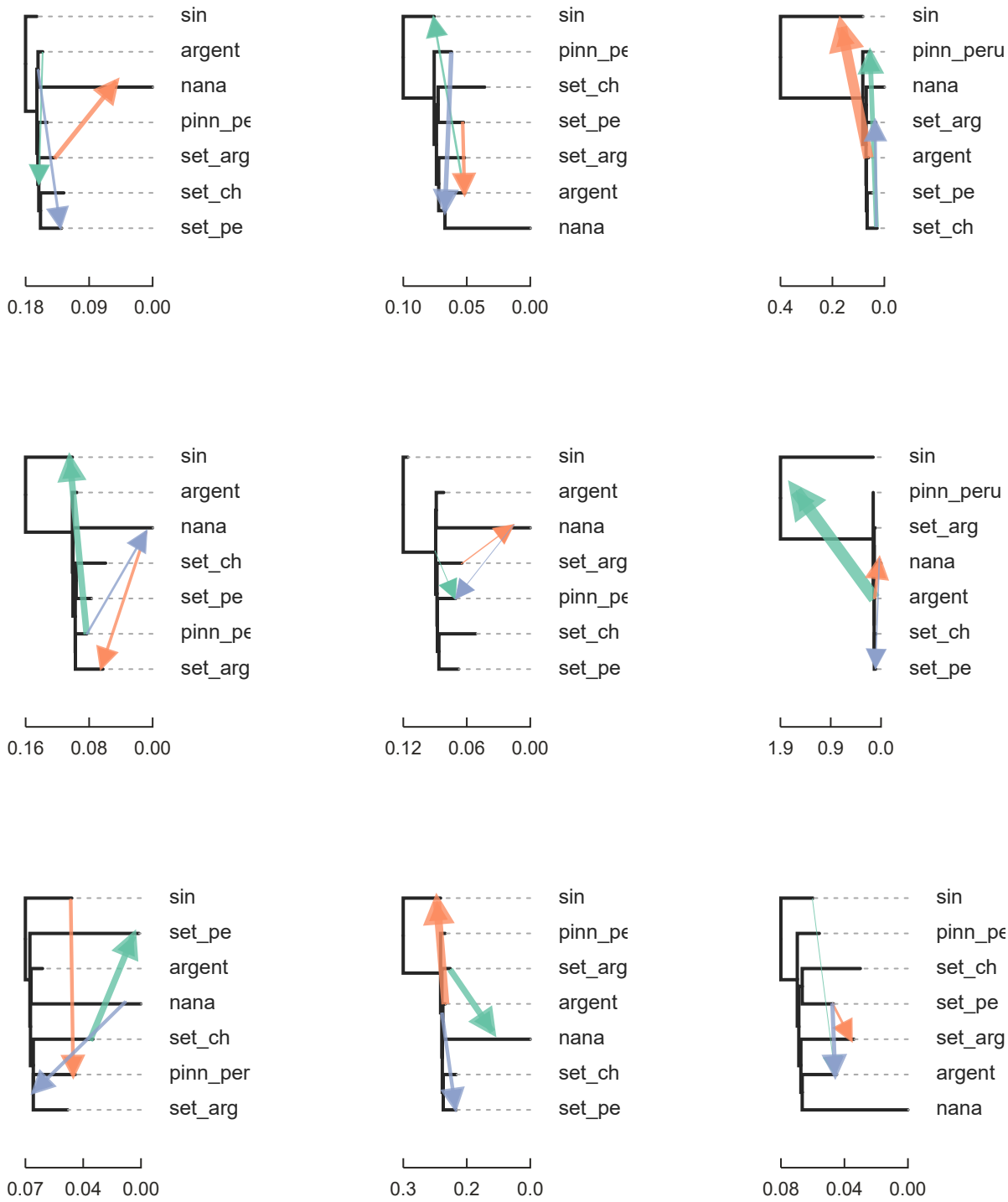
Supplementary material S3. TetRAD trees of the genus *Phacelia* based on ddRAD sequences using the 3rd set of assemblies. Clustering threshold 95 % with six different percentages of missing data.



Supplementary material S4. TreeMix iterations at different SNPs subsamples.



Supplementary material S4. TreeMix iterations at different SNPs subsamples.



Jupyter-Notebooks

Assembly and Phylogenetic Analyses (RAxML & Tetrad)

Phylogeny of *Phacelia* section *Glandulosae* in South America

[5]:
```
#import programs you need
import ipyrad as ip
import ipyrad.analysis as ipa
import ipyparallel as ipp
import pandas as pd
import toytree
import toyplot

##print Version of ipyrad and toytree
print("ipyrad v. {}".format(ip.__version__))
print("toytree v. {}".format(toytree.__version__))

##print version of Python
from platform import python_version
print("Python v.", python_version())
```

```
ipyrad v. 0.9.84
toytree v. 2.0.5
Python v. 3.10.5
```

1.0.1 Parallel processes on independent Python kernels

To start a parallel client you must run the command-line program 'ipcluster'. This will essentially start a number of independent Python processes(kernels) which we can then send bits of work to do. The cluster can be stopped and restarted independently of this notebook, which is convenient for working on a cluster where connecting to many cores is not always immediately available.

Open a terminal and type the following command to start an ipcluster instance with N engines.

[ ]:
```
## ipcluster start --n=40
```

[38]:
```
## connect to cluster
ipyclient = ipp.Client()
print(ip.cluster_info(ipyclient))
```

```
Parallel connection | Cryptantha: 40 cores
None
```

[39]:
```
## Provide a name for the assembly
data = ip.Assembly("Phacelia")
```

```
New Assembly: Phacelia
```

[40]:
```
## set parameters
data.set_params("project_dir", "./Phac_Assembly")
data.set_params("sorted_fastq_path", "./Phac_fastq/*.gz")
data.set_params("clust_threshold", 0.85)
data.set_params("max_Hs_consens", 0.05)
data.set_params("max_SNPs_locus", 0.2)
data.set_params("restriction_overhang", ('TGCAG', 'GGCC'))
data.set_params("output_formats", "*")
data.set_params("datatype", "ddrad")

## see / print all parameters
data.get_params()
```

[40]:
```
0   assembly_name               Phacelia
1   project_dir                 ./Phac_Assembly
2   raw_fastq_path
3   barcodes_path
4   sorted_fastq_path           ./Phac_fastq/*.gz
5   assembly_method             denovo
6   reference_sequence
7   datatype                    ddrad
8   restriction_overhang        ('TGCAG', 'GGCC')
9   max_low_qual_bases          5
10  phred_Qscore_offset         33
11  mindepth_statistical        6
12  mindepth_majrule            6
13  maxdepth                    10000
14  clust_threshold             0.85
15  max_barcode_mismatch        0
16  filter_adapters             2
17  filter_min_trim_len         35
18  max_alleles_consens         2
19  max_Ns_consens              0.05
20  max_Hs_consens              0.05
21  min_samples_locus           4
22  max_SNPs_locus              0.2
23  max_Indels_locus            8
24  max_shared_Hs_locus         0.5
25  trim_reads                  (0, 0, 0, 0)
26  trim_loci                   (0, 0, 0, 0)
27  output_formats              ['l', 'p', 's', 'n', 'k', 'a', 'g', 'G', 'u',
'v', 't', 'm']
28  pop_assign_file
29  reference_as_filter
```

1.0.2 Assemble the data from step 1 to 6

[41]:
```
## run steps 1 & 2 of the assembly
data.run("12", force = True)
```

```
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:04 | loading reads         | s1 |
[####################] 100% 0:00:45 | processing reads      | s2 |
```

[42]:
```
##use data.branch to make branches with different cluster thresholds
## set cluster treshold to 85 && run assembly steps 3-6
data_clust85 = data.branch("data_clust85")
data_clust85.set_params("clust_threshold", 0.85)
data_clust85.run("3456", force = True)
```

```
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:11 | dereplicating         | s3 |
[####################] 100% 0:20:49 | clustering/mapping    | s3 |
[####################] 100% 0:00:04 | building clusters     | s3 |
[####################] 100% 0:00:00 | chunking clusters     | s3 |
[####################] 100% 0:04:30 | aligning clusters     | s3 |
[####################] 100% 0:00:34 | concat clusters       | s3 |
[####################] 100% 0:00:02 | calc cluster stats    | s3 |
[####################] 100% 0:00:34 | inferring [H, E]      | s4 |
[####################] 100% 0:00:02 | calculating depths    | s5 |
[####################] 100% 0:00:03 | chunking clusters     | s5 |
[####################] 100% 0:01:32 | consens calling       | s5 |
[####################] 100% 0:00:02 | indexing alleles      | s5 |
[####################] 100% 0:00:02 | concatenating inputs  | s6 |
[####################] 100% 0:01:28 | clustering across     | s6 |
[####################] 100% 0:00:02 | building clusters     | s6 |
[####################] 100% 0:00:11 | aligning clusters     | s6 |
```

[43]:
```
#show assemby stats until step 6
data_clust85.stats.sort_values(by=['hetero_est'])
data_clust85.stats
```

[43]:

	state	reads_raw	\
Hydrophyllum_virginianum_W5141_out	6	1669993	
Phacelia_artemisioides_W6021_out	6	1553643	
Phacelia_artemisioides_W6031	6	1105184	
Phacelia_brachyantha_W5600_out	6	1127974	
Phacelia_campanularia_W5615_out	6	1595084	
Phacelia_cf_viscosa_pinnatifida_W6024	6	1544846	

```
Phacelia_cumingii_W5143_out                       6    1303875
Phacelia_egena_W5136_out                          6    1593280
Phacelia_grandiflora_W5613_out                    6    1278173
Phacelia_malvifolia_W5137_out                     6    1823997
Phacelia_nana_W6373                               6    1163545
Phacelia_pinnatifida_W5599                        6    1136178
Phacelia_pinnatifida_W5610                        6    1280762
Phacelia_pinnatifida_W6001                        6    1611986
Phacelia_pinnatifida_W6027                        6    1257435
Phacelia_pinnatifida_W6029                        6    1375613
Phacelia_pinnatifida_W6079                        6    3148834
Phacelia_pinnatifida_W6080                        6    2140785
Phacelia_pinnatifida_W6374                        6    2407848
Phacelia_pinnatifida_viscosa_W6028                6    1683689
Phacelia_sanzinii_W6369                           6     304136
Phacelia_secunda_W5139_out                        6    1093797
Phacelia_setigera_W5612                           6    1299484
Phacelia_setigera_var_humahuaquense_W6371         6      26451
Phacelia_setigera_var_setigera_W6370              6     462109
Phacelia_setigera_var_setigera_W6375              6    2029087
Phacelia_sinuata_W6368                            6     451907
Phacelia_sinuata_W6376                            6    1740151
Phacelia_spec_W6037                               6    1199094
Phacelia_spec_nov_W5636                           6     899974
Phacelia_spec_nov_W5637                           6     792010
Phacelia_spec_nov_W6078                           6    1178943
Phacelia_tanacetifolia_W5598_out                  6    1439544
Phacelia_viscosa_W5145                            6    2021555

                                        reads_passed_filter   \
Hydrophyllum_virginianum_W5141_out                   1669453
Phacelia_artemisioides_W6021_out                     1552821
Phacelia_artemisioides_W6031                         1104793
Phacelia_brachyantha_W5600_out                       1127504
Phacelia_campanularia_W5615_out                      1594787
Phacelia_cf_viscosa_pinnatifida_W6024                1543628
Phacelia_cumingii_W5143_out                          1303555
Phacelia_egena_W5136_out                             1592593
Phacelia_grandiflora_W5613_out                       1277793
Phacelia_malvifolia_W5137_out                        1823706
Phacelia_nana_W6373                                  1162801
Phacelia_pinnatifida_W5599                           1135732
Phacelia_pinnatifida_W5610                           1280243
Phacelia_pinnatifida_W6001                           1611340
Phacelia_pinnatifida_W6027                           1257027
Phacelia_pinnatifida_W6029                           1375056
Phacelia_pinnatifida_W6079                           3144291
```

Phacelia_pinnatifida_W6080	2132998
Phacelia_pinnatifida_W6374	2406401
Phacelia_pinnatifida_viscosa_W6028	1682910
Phacelia_sanzinii_W6369	303948
Phacelia_secunda_W5139_out	1093426
Phacelia_setigera_W5612	1298953
Phacelia_setigera_var_humahuaquense_W6371	26430
Phacelia_setigera_var_setigera_W6370	461800
Phacelia_setigera_var_setigera_W6375	2027856
Phacelia_sinuata_W6368	451671
Phacelia_sinuata_W6376	1739216
Phacelia_spec_W6037	1198606
Phacelia_spec_nov_W5636	899681
Phacelia_spec_nov_W5637	791701
Phacelia_spec_nov_W6078	1177635
Phacelia_tanacetifolia_W5598_out	1438556
Phacelia_viscosa_W5145	2020701

	clusters_total	clusters_hidepth	\
Hydrophyllum_virginianum_W5141_out	88334	16657	
Phacelia_artemisioides_W6021_out	236570	18216	
Phacelia_artemisioides_W6031	71568	11080	
Phacelia_brachyantha_W5600_out	73475	13899	
Phacelia_campanularia_W5615_out	91859	14423	
Phacelia_cf_viscosa_pinnatifida_W6024	199755	28576	
Phacelia_cumingii_W5143_out	95070	17390	
Phacelia_egena_W5136_out	120612	18002	
Phacelia_grandiflora_W5613_out	100416	20038	
Phacelia_malvifolia_W5137_out	82891	15418	
Phacelia_nana_W6373	192946	10078	
Phacelia_pinnatifida_W5599	120337	10088	
Phacelia_pinnatifida_W5610	171586	14465	
Phacelia_pinnatifida_W6001	84415	11877	
Phacelia_pinnatifida_W6027	75042	11508	
Phacelia_pinnatifida_W6029	85960	15260	
Phacelia_pinnatifida_W6079	305007	42111	
Phacelia_pinnatifida_W6080	845427	31975	
Phacelia_pinnatifida_W6374	345683	15771	
Phacelia_pinnatifida_viscosa_W6028	101344	16864	
Phacelia_sanzinii_W6369	69403	7349	
Phacelia_secunda_W5139_out	62381	13442	
Phacelia_setigera_W5612	63731	11303	
Phacelia_setigera_var_humahuaquense_W6371	21300	105	
Phacelia_setigera_var_setigera_W6370	135583	8557	
Phacelia_setigera_var_setigera_W6375	175039	10075	
Phacelia_sinuata_W6368	62694	8477	
Phacelia_sinuata_W6376	134315	16910	

Phacelia_spec_W6037	113897	18234
Phacelia_spec_nov_W5636	53396	10781
Phacelia_spec_nov_W5637	71385	10000
Phacelia_spec_nov_W6078	98596	9958
Phacelia_tanacetifolia_W5598_out	224345	36888
Phacelia_viscosa_W5145	158947	22156

	hetero_est	error_est	\
Hydrophyllum_virginianum_W5141_out	0.011	0.003	
Phacelia_artemisioides_W6021_out	0.014	0.002	
Phacelia_artemisioides_W6031	0.004	0.002	
Phacelia_brachyantha_W5600_out	0.006	0.002	
Phacelia_campanularia_W5615_out	0.009	0.002	
Phacelia_cf_viscosa_pinnatifida_W6024	0.015	0.002	
Phacelia_cumingii_W5143_out	0.009	0.002	
Phacelia_egena_W5136_out	0.020	0.003	
Phacelia_grandiflora_W5613_out	0.010	0.002	
Phacelia_malvifolia_W5137_out	0.005	0.002	
Phacelia_nana_W6373	0.010	0.002	
Phacelia_pinnatifida_W5599	0.005	0.002	
Phacelia_pinnatifida_W5610	0.008	0.002	
Phacelia_pinnatifida_W6001	0.006	0.002	
Phacelia_pinnatifida_W6027	0.009	0.002	
Phacelia_pinnatifida_W6029	0.006	0.002	
Phacelia_pinnatifida_W6079	0.011	0.003	
Phacelia_pinnatifida_W6080	0.055	0.006	
Phacelia_pinnatifida_W6374	0.012	0.002	
Phacelia_pinnatifida_viscosa_W6028	0.009	0.002	
Phacelia_sanzinii_W6369	0.008	0.002	
Phacelia_secunda_W5139_out	0.008	0.002	
Phacelia_setigera_W5612	0.005	0.002	
Phacelia_setigera_var_humahuaquense_W6371	0.023	0.010	
Phacelia_setigera_var_setigera_W6370	0.011	0.002	
Phacelia_setigera_var_setigera_W6375	0.011	0.002	
Phacelia_sinuata_W6368	0.006	0.002	
Phacelia_sinuata_W6376	0.015	0.002	
Phacelia_spec_W6037	0.012	0.002	
Phacelia_spec_nov_W5636	0.006	0.002	
Phacelia_spec_nov_W5637	0.006	0.002	
Phacelia_spec_nov_W6078	0.013	0.002	
Phacelia_tanacetifolia_W5598_out	0.015	0.003	
Phacelia_viscosa_W5145	0.004	0.002	

	reads_consens
Hydrophyllum_virginianum_W5141_out	15007
Phacelia_artemisioides_W6021_out	16105
Phacelia_artemisioides_W6031	10620

```
Phacelia_brachyantha_W5600_out                    13150
Phacelia_campanularia_W5615_out                   13447
Phacelia_cf_viscosa_pinnatifida_W6024             25013
Phacelia_cumingii_W5143_out                       16193
Phacelia_egena_W5136_out                          14741
Phacelia_grandiflora_W5613_out                    18187
Phacelia_malvifolia_W5137_out                     14763
Phacelia_nana_W6373                                9178
Phacelia_pinnatifida_W5599                         9614
Phacelia_pinnatifida_W5610                        13507
Phacelia_pinnatifida_W6001                        11227
Phacelia_pinnatifida_W6027                        10648
Phacelia_pinnatifida_W6029                        14499
Phacelia_pinnatifida_W6079                        37883
Phacelia_pinnatifida_W6080                        18499
Phacelia_pinnatifida_W6374                        14362
Phacelia_pinnatifida_viscosa_W6028                15597
Phacelia_sanzinii_W6369                            6869
Phacelia_secunda_W5139_out                        12585
Phacelia_setigera_W5612                           10723
Phacelia_setigera_var_humahuaquense_W6371            75
Phacelia_setigera_var_setigera_W6370               7704
Phacelia_setigera_var_setigera_W6375               9229
Phacelia_sinuata_W6368                             8046
Phacelia_sinuata_W6376                            14961
Phacelia_spec_W6037                               16445
Phacelia_spec_nov_W5636                           10259
Phacelia_spec_nov_W5637                            9431
Phacelia_spec_nov_W6078                            8909
Phacelia_tanacetifolia_W5598_out                  32478
Phacelia_viscosa_W5145                            21383
```

[44]:
```
## set cluster treshold to 90 && run assembly steps 3-6
data_clust90 = data.branch("data_clust90")
data_clust90.set_params("clust_threshold", 0.90)
data_clust90.run("3456", force = True)
```

```
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:07 | dereplicating        | s3 |
[####################] 100% 0:22:21 | clustering/mapping   | s3 |
[####################] 100% 0:00:03 | building clusters    | s3 |
[####################] 100% 0:00:00 | chunking clusters    | s3 |
[####################] 100% 0:04:40 | aligning clusters    | s3 |
[####################] 100% 0:00:34 | concat clusters      | s3 |
[####################] 100% 0:00:02 | calc cluster stats   | s3 |
[####################] 100% 0:00:32 | inferring [H, E]     | s4 |
[####################] 100% 0:00:02 | calculating depths   | s5 |
```

```
[####################] 100% 0:00:03 | chunking clusters    | s5 |
[####################] 100% 0:01:37 | consens calling      | s5 |
[####################] 100% 0:00:03 | indexing alleles     | s5 |
[####################] 100% 0:00:03 | concatenating inputs | s6 |
[####################] 100% 0:01:58 | clustering across    | s6 |
[####################] 100% 0:00:02 | building clusters    | s6 |
[####################] 100% 0:00:13 | aligning clusters    | s6 |
```

[45]:
```
## show assemby stats until step 6
data_clust90.stats.sort_values(by=['reads_consens'])
data_clust90.stats
```

[45]:

	state	reads_raw	\
Hydrophyllum_virginianum_W5141_out	6	1669993	
Phacelia_artemisioides_W6021_out	6	1553643	
Phacelia_artemisioides_W6031	6	1105184	
Phacelia_brachyantha_W5600_out	6	1127974	
Phacelia_campanularia_W5615_out	6	1595084	
Phacelia_cf_viscosa_pinnatifida_W6024	6	1544846	
Phacelia_cumingii_W5143_out	6	1303875	
Phacelia_egena_W5136_out	6	1593280	
Phacelia_grandiflora_W5613_out	6	1278173	
Phacelia_malvifolia_W5137_out	6	1823997	
Phacelia_nana_W6373	6	1163545	
Phacelia_pinnatifida_W5599	6	1136178	
Phacelia_pinnatifida_W5610	6	1280762	
Phacelia_pinnatifida_W6001	6	1611986	
Phacelia_pinnatifida_W6027	6	1257435	
Phacelia_pinnatifida_W6029	6	1375613	
Phacelia_pinnatifida_W6079	6	3148834	
Phacelia_pinnatifida_W6080	6	2140785	
Phacelia_pinnatifida_W6374	6	2407848	
Phacelia_pinnatifida_viscosa_W6028	6	1683689	
Phacelia_sanzinii_W6369	6	304136	
Phacelia_secunda_W5139_out	6	1093797	
Phacelia_setigera_W5612	6	1299484	
Phacelia_setigera_var_humahuaquense_W6371	6	26451	
Phacelia_setigera_var_setigera_W6370	6	462109	
Phacelia_setigera_var_setigera_W6375	6	2029087	
Phacelia_sinuata_W6368	6	451907	
Phacelia_sinuata_W6376	6	1740151	
Phacelia_spec_W6037	6	1199094	
Phacelia_spec_nov_W5636	6	899974	
Phacelia_spec_nov_W5637	6	792010	
Phacelia_spec_nov_W6078	6	1178943	
Phacelia_tanacetifolia_W5598_out	6	1439544	
Phacelia_viscosa_W5145	6	2021555	

	reads_passed_filter	\
Hydrophyllum_virginianum_W5141_out	1669453	
Phacelia_artemisioides_W6021_out	1552821	
Phacelia_artemisioides_W6031	1104793	
Phacelia_brachyantha_W5600_out	1127504	
Phacelia_campanularia_W5615_out	1594787	
Phacelia_cf_viscosa_pinnatifida_W6024	1543628	
Phacelia_cumingii_W5143_out	1303555	
Phacelia_egena_W5136_out	1592593	
Phacelia_grandiflora_W5613_out	1277793	
Phacelia_malvifolia_W5137_out	1823706	
Phacelia_nana_W6373	1162801	
Phacelia_pinnatifida_W5599	1135732	
Phacelia_pinnatifida_W5610	1280243	
Phacelia_pinnatifida_W6001	1611340	
Phacelia_pinnatifida_W6027	1257027	
Phacelia_pinnatifida_W6029	1375056	
Phacelia_pinnatifida_W6079	3144291	
Phacelia_pinnatifida_W6080	2132998	
Phacelia_pinnatifida_W6374	2406401	
Phacelia_pinnatifida_viscosa_W6028	1682910	
Phacelia_sanzinii_W6369	303948	
Phacelia_secunda_W5139_out	1093426	
Phacelia_setigera_W5612	1298953	
Phacelia_setigera_var_humahuaquense_W6371	26430	
Phacelia_setigera_var_setigera_W6370	461800	
Phacelia_setigera_var_setigera_W6375	2027856	
Phacelia_sinuata_W6368	451671	
Phacelia_sinuata_W6376	1739216	
Phacelia_spec_W6037	1198606	
Phacelia_spec_nov_W5636	899681	
Phacelia_spec_nov_W5637	791701	
Phacelia_spec_nov_W6078	1177635	
Phacelia_tanacetifolia_W5598_out	1438556	
Phacelia_viscosa_W5145	2020701	

	clusters_total	clusters_hidepth	\
Hydrophyllum_virginianum_W5141_out	92585	17319	
Phacelia_artemisioides_W6021_out	244224	18204	
Phacelia_artemisioides_W6031	73870	11113	
Phacelia_brachyantha_W5600_out	76262	13994	
Phacelia_campanularia_W5615_out	95205	14494	
Phacelia_cf_viscosa_pinnatifida_W6024	209322	28888	
Phacelia_cumingii_W5143_out	98232	17487	
Phacelia_egena_W5136_out	126402	18626	
Phacelia_grandiflora_W5613_out	103386	20219	

Phacelia_malvifolia_W5137_out	86005	15521
Phacelia_nana_W6373	198581	10029
Phacelia_pinnatifida_W5599	123345	10153
Phacelia_pinnatifida_W5610	176933	14443
Phacelia_pinnatifida_W6001	87605	11959
Phacelia_pinnatifida_W6027	77996	11498
Phacelia_pinnatifida_W6029	89103	15329
Phacelia_pinnatifida_W6079	323655	42169
Phacelia_pinnatifida_W6080	908848	29250
Phacelia_pinnatifida_W6374	354963	15663
Phacelia_pinnatifida_viscosa_W6028	104998	16941
Phacelia_sanzinii_W6369	71863	7348
Phacelia_secunda_W5139_out	64932	13603
Phacelia_setigera_W5612	66612	11414
Phacelia_setigera_var_humahuaquense_W6371	21525	96
Phacelia_setigera_var_setigera_W6370	141251	8388
Phacelia_setigera_var_setigera_W6375	181809	9956
Phacelia_sinuata_W6368	65434	8546
Phacelia_sinuata_W6376	140544	17260
Phacelia_spec_W6037	118658	18310
Phacelia_spec_nov_W5636	55403	10889
Phacelia_spec_nov_W5637	73786	10021
Phacelia_spec_nov_W6078	104034	10028
Phacelia_tanacetifolia_W5598_out	232474	36953
Phacelia_viscosa_W5145	163995	22219

	hetero_est	error_est	\
Hydrophyllum_virginianum_W5141_out	0.008	0.002	
Phacelia_artemisioides_W6021_out	0.009	0.002	
Phacelia_artemisioides_W6031	0.003	0.002	
Phacelia_brachyantha_W5600_out	0.003	0.002	
Phacelia_campanularia_W5615_out	0.006	0.002	
Phacelia_cf_viscosa_pinnatifida_W6024	0.011	0.002	
Phacelia_cumingii_W5143_out	0.006	0.002	
Phacelia_egena_W5136_out	0.016	0.002	
Phacelia_grandiflora_W5613_out	0.009	0.002	
Phacelia_malvifolia_W5137_out	0.003	0.002	
Phacelia_nana_W6373	0.006	0.002	
Phacelia_pinnatifida_W5599	0.003	0.002	
Phacelia_pinnatifida_W5610	0.005	0.002	
Phacelia_pinnatifida_W6001	0.004	0.002	
Phacelia_pinnatifida_W6027	0.006	0.002	
Phacelia_pinnatifida_W6029	0.004	0.002	
Phacelia_pinnatifida_W6079	0.007	0.002	
Phacelia_pinnatifida_W6080	0.030	0.004	
Phacelia_pinnatifida_W6374	0.007	0.001	
Phacelia_pinnatifida_viscosa_W6028	0.007	0.002	

Phacelia_sanzinii_W6369	0.005	0.002
Phacelia_secunda_W5139_out	0.005	0.002
Phacelia_setigera_W5612	0.004	0.002
Phacelia_setigera_var_humahuaquense_W6371	0.015	0.008
Phacelia_setigera_var_setigera_W6370	0.005	0.002
Phacelia_setigera_var_setigera_W6375	0.005	0.001
Phacelia_sinuata_W6368	0.004	0.002
Phacelia_sinuata_W6376	0.009	0.002
Phacelia_spec_W6037	0.008	0.002
Phacelia_spec_nov_W5636	0.004	0.002
Phacelia_spec_nov_W5637	0.003	0.002
Phacelia_spec_nov_W6078	0.008	0.002
Phacelia_tanacetifolia_W5598_out	0.012	0.002
Phacelia_viscosa_W5145	0.003	0.002

	reads_consens
Hydrophyllum_virginianum_W5141_out	16019
Phacelia_artemisioides_W6021_out	16640
Phacelia_artemisioides_W6031	10834
Phacelia_brachyantha_W5600_out	13568
Phacelia_campanularia_W5615_out	13875
Phacelia_cf_viscosa_pinnatifida_W6024	26479
Phacelia_cumingii_W5143_out	16776
Phacelia_egena_W5136_out	16016
Phacelia_grandiflora_W5613_out	18715
Phacelia_malvifolia_W5137_out	15150
Phacelia_nana_W6373	9448
Phacelia_pinnatifida_W5599	9856
Phacelia_pinnatifida_W5610	13916
Phacelia_pinnatifida_W6001	11546
Phacelia_pinnatifida_W6027	10993
Phacelia_pinnatifida_W6029	14863
Phacelia_pinnatifida_W6079	39805
Phacelia_pinnatifida_W6080	21492
Phacelia_pinnatifida_W6374	14843
Phacelia_pinnatifida_viscosa_W6028	16073
Phacelia_sanzinii_W6369	7036
Phacelia_secunda_W5139_out	13034
Phacelia_setigera_W5612	11024
Phacelia_setigera_var_humahuaquense_W6371	78
Phacelia_setigera_var_setigera_W6370	7934
Phacelia_setigera_var_setigera_W6375	9486
Phacelia_sinuata_W6368	8267
Phacelia_sinuata_W6376	15894
Phacelia_spec_W6037	17259
Phacelia_spec_nov_W5636	10551
Phacelia_spec_nov_W5637	9681

```
Phacelia_spec_nov_W6078                          9295
Phacelia_tanacetifolia_W5598_out                34066
Phacelia_viscosa_W5145                          21679
```

[46]:
```
## set cluster treshold to 95 && run assembly steps 3-6
data_clust95 = data.branch("data_clust95")
data_clust95.set_params("clust_threshold", 0.95)
data_clust95.run("3456", force = True)
```

```
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:08 | dereplicating        | s3 |
[####################] 100% 0:26:22 | clustering/mapping   | s3 |
[####################] 100% 0:00:03 | building clusters    | s3 |
[####################] 100% 0:00:00 | chunking clusters    | s3 |
[####################] 100% 0:04:21 | aligning clusters    | s3 |
[####################] 100% 0:00:35 | concat clusters      | s3 |
[####################] 100% 0:00:02 | calc cluster stats   | s3 |
[####################] 100% 0:00:33 | inferring [H, E]     | s4 |
[####################] 100% 0:00:02 | calculating depths   | s5 |
[####################] 100% 0:00:03 | chunking clusters    | s5 |
[####################] 100% 0:01:31 | consens calling      | s5 |
[####################] 100% 0:00:02 | indexing alleles     | s5 |
[####################] 100% 0:00:03 | concatenating inputs | s6 |
[####################] 100% 0:02:11 | clustering across    | s6 |
[####################] 100% 0:00:02 | building clusters    | s6 |
[####################] 100% 0:00:11 | aligning clusters    | s6 |
```

[47]:
```
## show assemby stats until step 6
data_clust95.stats.sort_values(by=['reads_consens'])
data_clust95.stats
```

[47]:

	state	reads_raw	\
Hydrophyllum_virginianum_W5141_out	6	1669993	
Phacelia_artemisioides_W6021_out	6	1553643	
Phacelia_artemisioides_W6031	6	1105184	
Phacelia_brachyantha_W5600_out	6	1127974	
Phacelia_campanularia_W5615_out	6	1595084	
Phacelia_cf_viscosa_pinnatifida_W6024	6	1544846	
Phacelia_cumingii_W5143_out	6	1303875	
Phacelia_egena_W5136_out	6	1593280	
Phacelia_grandiflora_W5613_out	6	1278173	
Phacelia_malvifolia_W5137_out	6	1823997	
Phacelia_nana_W6373	6	1163545	
Phacelia_pinnatifida_W5599	6	1136178	
Phacelia_pinnatifida_W5610	6	1280762	
Phacelia_pinnatifida_W6001	6	1611986	
Phacelia_pinnatifida_W6027	6	1257435	

```
Phacelia_pinnatifida_W6029                          6    1375613
Phacelia_pinnatifida_W6079                          6    3148834
Phacelia_pinnatifida_W6080                          6    2140785
Phacelia_pinnatifida_W6374                          6    2407848
Phacelia_pinnatifida_viscosa_W6028                  6    1683689
Phacelia_sanzinii_W6369                             6     304136
Phacelia_secunda_W5139_out                          6    1093797
Phacelia_setigera_W5612                             6    1299484
Phacelia_setigera_var_humahuaquense_W6371           6      26451
Phacelia_setigera_var_setigera_W6370                6     462109
Phacelia_setigera_var_setigera_W6375                6    2029087
Phacelia_sinuata_W6368                              6     451907
Phacelia_sinuata_W6376                              6    1740151
Phacelia_spec_W6037                                 6    1199094
Phacelia_spec_nov_W5636                             6     899974
Phacelia_spec_nov_W5637                             6     792010
Phacelia_spec_nov_W6078                             6    1178943
Phacelia_tanacetifolia_W5598_out                    6    1439544
Phacelia_viscosa_W5145                              6    2021555

                                              reads_passed_filter   \
Hydrophyllum_virginianum_W5141_out                        1669453
Phacelia_artemisioides_W6021_out                          1552821
Phacelia_artemisioides_W6031                              1104793
Phacelia_brachyantha_W5600_out                            1127504
Phacelia_campanularia_W5615_out                           1594787
Phacelia_cf_viscosa_pinnatifida_W6024                     1543628
Phacelia_cumingii_W5143_out                               1303555
Phacelia_egena_W5136_out                                  1592593
Phacelia_grandiflora_W5613_out                            1277793
Phacelia_malvifolia_W5137_out                             1823706
Phacelia_nana_W6373                                       1162801
Phacelia_pinnatifida_W5599                                1135732
Phacelia_pinnatifida_W5610                                1280243
Phacelia_pinnatifida_W6001                                1611340
Phacelia_pinnatifida_W6027                                1257027
Phacelia_pinnatifida_W6029                                1375056
Phacelia_pinnatifida_W6079                                3144291
Phacelia_pinnatifida_W6080                                2132998
Phacelia_pinnatifida_W6374                                2406401
Phacelia_pinnatifida_viscosa_W6028                        1682910
Phacelia_sanzinii_W6369                                    303948
Phacelia_secunda_W5139_out                                1093426
Phacelia_setigera_W5612                                   1298953
Phacelia_setigera_var_humahuaquense_W6371                   26430
Phacelia_setigera_var_setigera_W6370                      461800
Phacelia_setigera_var_setigera_W6375                      2027856
```

Phacelia_sinuata_W6368	451671
Phacelia_sinuata_W6376	1739216
Phacelia_spec_W6037	1198606
Phacelia_spec_nov_W5636	899681
Phacelia_spec_nov_W5637	791701
Phacelia_spec_nov_W6078	1177635
Phacelia_tanacetifolia_W5598_out	1438556
Phacelia_viscosa_W5145	2020701

	clusters_total	clusters_hidepth	\
Hydrophyllum_virginianum_W5141_out	102210	18387	
Phacelia_artemisioides_W6021_out	260061	17980	
Phacelia_artemisioides_W6031	78336	11174	
Phacelia_brachyantha_W5600_out	82065	14027	
Phacelia_campanularia_W5615_out	102407	14643	
Phacelia_cf_viscosa_pinnatifida_W6024	225735	28610	
Phacelia_cumingii_W5143_out	104263	17428	
Phacelia_egena_W5136_out	139749	19799	
Phacelia_grandiflora_W5613_out	113308	20837	
Phacelia_malvifolia_W5137_out	93741	15633	
Phacelia_nana_W6373	206548	10008	
Phacelia_pinnatifida_W5599	128661	10194	
Phacelia_pinnatifida_W5610	185142	14463	
Phacelia_pinnatifida_W6001	94375	12034	
Phacelia_pinnatifida_W6027	83824	11509	
Phacelia_pinnatifida_W6029	96008	15406	
Phacelia_pinnatifida_W6079	354185	42046	
Phacelia_pinnatifida_W6080	979498	26416	
Phacelia_pinnatifida_W6374	368344	15706	
Phacelia_pinnatifida_viscosa_W6028	115248	16888	
Phacelia_sanzinii_W6369	75180	7368	
Phacelia_secunda_W5139_out	70433	13805	
Phacelia_setigera_W5612	72782	11484	
Phacelia_setigera_var_humahuaquense_W6371	21754	86	
Phacelia_setigera_var_setigera_W6370	148317	8328	
Phacelia_setigera_var_setigera_W6375	192400	10014	
Phacelia_sinuata_W6368	69850	8656	
Phacelia_sinuata_W6376	150843	17539	
Phacelia_spec_W6037	128089	18290	
Phacelia_spec_nov_W5636	59872	10975	
Phacelia_spec_nov_W5637	78327	10103	
Phacelia_spec_nov_W6078	113897	10118	
Phacelia_tanacetifolia_W5598_out	248071	36642	
Phacelia_viscosa_W5145	175155	22294	

	hetero_est	error_est	\
Hydrophyllum_virginianum_W5141_out	0.005	0.002	

Phacelia_artemisioides_W6021_out	0.004	0.002
Phacelia_artemisioides_W6031	0.001	0.002
Phacelia_brachyantha_W5600_out	0.001	0.002
Phacelia_campanularia_W5615_out	0.003	0.002
Phacelia_cf_viscosa_pinnatifida_W6024	0.006	0.002
Phacelia_cumingii_W5143_out	0.003	0.002
Phacelia_egena_W5136_out	0.010	0.002
Phacelia_grandiflora_W5613_out	0.005	0.002
Phacelia_malvifolia_W5137_out	0.002	0.002
Phacelia_nana_W6373	0.003	0.001
Phacelia_pinnatifida_W5599	0.001	0.002
Phacelia_pinnatifida_W5610	0.003	0.002
Phacelia_pinnatifida_W6001	0.002	0.002
Phacelia_pinnatifida_W6027	0.003	0.002
Phacelia_pinnatifida_W6029	0.002	0.002
Phacelia_pinnatifida_W6079	0.004	0.002
Phacelia_pinnatifida_W6080	0.013	0.003
Phacelia_pinnatifida_W6374	0.004	0.001
Phacelia_pinnatifida_viscosa_W6028	0.003	0.002
Phacelia_sanzinii_W6369	0.003	0.001
Phacelia_secunda_W5139_out	0.003	0.002
Phacelia_setigera_W5612	0.002	0.002
Phacelia_setigera_var_humahuaquense_W6371	0.008	0.005
Phacelia_setigera_var_setigera_W6370	0.002	0.001
Phacelia_setigera_var_setigera_W6375	0.002	0.001
Phacelia_sinuata_W6368	0.002	0.001
Phacelia_sinuata_W6376	0.005	0.002
Phacelia_spec_W6037	0.004	0.002
Phacelia_spec_nov_W5636	0.002	0.002
Phacelia_spec_nov_W5637	0.002	0.002
Phacelia_spec_nov_W6078	0.004	0.001
Phacelia_tanacetifolia_W5598_out	0.008	0.002
Phacelia_viscosa_W5145	0.002	0.002

	reads_consens
Hydrophyllum_virginianum_W5141_out	17617
Phacelia_artemisioides_W6021_out	17499
Phacelia_artemisioides_W6031	11093
Phacelia_brachyantha_W5600_out	13895
Phacelia_campanularia_W5615_out	14433
Phacelia_cf_viscosa_pinnatifida_W6024	27805
Phacelia_cumingii_W5143_out	17267
Phacelia_egena_W5136_out	18469
Phacelia_grandiflora_W5613_out	20165
Phacelia_malvifolia_W5137_out	15523
Phacelia_nana_W6373	9804
Phacelia_pinnatifida_W5599	10106

```
Phacelia_pinnatifida_W5610                          14275
Phacelia_pinnatifida_W6001                          11856
Phacelia_pinnatifida_W6027                          11339
Phacelia_pinnatifida_W6029                          15223
Phacelia_pinnatifida_W6079                          41220
Phacelia_pinnatifida_W6080                          24113
Phacelia_pinnatifida_W6374                          15390
Phacelia_pinnatifida_viscosa_W6028                  16640
Phacelia_sanzinii_W6369                              7259
Phacelia_secunda_W5139_out                          13632
Phacelia_setigera_W5612                             11350
Phacelia_setigera_var_humahuaquense_W6371              82
Phacelia_setigera_var_setigera_W6370                 8204
Phacelia_setigera_var_setigera_W6375                 9854
Phacelia_sinuata_W6368                               8576
Phacelia_sinuata_W6376                              17036
Phacelia_spec_W6037                                 18033
Phacelia_spec_nov_W5636                             10860
Phacelia_spec_nov_W5637                              9964
Phacelia_spec_nov_W6078                              9838
Phacelia_tanacetifolia_W5598_out                    35725
Phacelia_viscosa_W5145                              22030
```

1.0.3 Final assembly with different min_samples_locus settings for different analysis. At this step and the next step we produce the 2nd set of assemblies where *Hydrophyllum* and all outgroup *Phacelia* + *P. artemisioides_W6021* are defined as outfroups.

First lets exclude samples with low read number (<1000 reads after step 6), which are outside the target group or with odd placements in preliminary analysis: Samples with low read number are: >Phacelia_setigera_var_humahuaquense_W6371

[48]:
```
## exclude samples from assembly with ...
keep_list = [i for i in data.samples.keys() if i not in [
    ## exclude samples with low read number (< 5000 )
    "Phacelia_setigera_var_humahuaquense_W6371",
]]

## make a new data branch from the keep_list #excl stands for exclude
excl_clust85 = data_clust85.branch("excl_clust85", subsamples = keep_list,␣
 ↪force = True)

excl_clust90 = data_clust90.branch("excl_clust90", subsamples = keep_list,␣
 ↪force = True)

excl_clust95 = data_clust95.branch("excl_clust95", subsamples = keep_list,␣
 ↪force = True)
```

1.0.4 Calculate different sets of missing data and then run the step 7 of the assembly for each one of the three clustering thresholds

We will first write the results of the percentage loop into a dictionary which can subsequently be used in the following steps instead of making the dictionary by hand

[51]:
```
## first check number of remaining samples
ingroup = excl_clust85.stats.state.count() - 10
print("Number of ingroup taxa:", ingroup)
print("Calculate different sets of missing data:")

## for loop to calculate different values for min_sample_locus
percent = [10, 15, 20, 25, 30, 35, 40]
for i in percent:
    res = ingroup / 100 * i
    print(i,"% = ", round(res))
```

```
Number of ingroup taxa: 23
Calculate different sets of missing data:
10 % =  2
15 % =  3
20 % =  5
25 % =  6
30 % =  7
35 % =  8
40 % =  9
```

1.0.5 Step 7 of the assembly for clustering threshold 85% with different sets of missing data

[52]:
```
# Cluster Treshold 85

## Run the final assembly step 7 through for loop with different␣
 ↪min_sample_locus
## based on estimated number of remaining samples MINUS outgroup

## make a dictionary with the percentage of missing data as keys and
## the actual min_sample_locus specified as values based on the number of␣
 ↪"ingroup samples"
sample_dict = {10: 2,
               15: 3,
               20: 5,
               25: 6,
               30: 7,
               35: 8,
               40: 9}

## loop over the dictionary
```

```
for key, value in sample_dict.items():
    newname = "pops{}_clust85".format(key)
    newdata = excl_clust85.branch(newname)
    newdata.populations = {
        "ingroup":  (value, [i for i in newdata.samples if "out" not in i]),
        "outgroup": (0,     [i for i in newdata.samples if "out" in i]),
         }
    ## run final step on every interation of the loop
    newdata.run("7", force = True)
```

```
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:05 | applying filters      | s7 |
[####################] 100% 0:00:06 | building arrays       | s7 |
[####################] 100% 0:00:05 | writing conversions   | s7 |
[####################] 100% 0:00:04 | indexing vcf depths   | s7 |
[####################] 100% 0:00:14 | writing vcf output    | s7 |
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:02 | applying filters      | s7 |
[####################] 100% 0:00:05 | building arrays       | s7 |
[####################] 100% 0:00:04 | writing conversions   | s7 |
[####################] 100% 0:00:03 | indexing vcf depths   | s7 |
[####################] 100% 0:00:12 | writing vcf output    | s7 |
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:03 | applying filters      | s7 |
[####################] 100% 0:00:05 | building arrays       | s7 |
[####################] 100% 0:00:03 | writing conversions   | s7 |
[####################] 100% 0:00:02 | indexing vcf depths   | s7 |
[####################] 100% 0:00:11 | writing vcf output    | s7 |
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:02 | applying filters      | s7 |
[####################] 100% 0:00:05 | building arrays       | s7 |
[####################] 100% 0:00:03 | writing conversions   | s7 |
[####################] 100% 0:00:02 | indexing vcf depths   | s7 |
[####################] 100% 0:00:10 | writing vcf output    | s7 |
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:02 | applying filters      | s7 |
[####################] 100% 0:00:05 | building arrays       | s7 |
[####################] 100% 0:00:03 | writing conversions   | s7 |
[####################] 100% 0:00:02 | indexing vcf depths   | s7 |
[####################] 100% 0:00:10 | writing vcf output    | s7 |
Parallel connection | Cryptantha: 40 cores
[####################] 100% 0:00:02 | applying filters      | s7 |
[####################] 100% 0:00:04 | building arrays       | s7 |
[####################] 100% 0:00:02 | writing conversions   | s7 |
[####################] 100% 0:00:03 | indexing vcf depths   | s7 |
[####################] 100% 0:00:10 | writing vcf output    | s7 |
Parallel connection | Cryptantha: 40 cores
```

```
[####################] 100% 0:00:05 | applying filters      | s7 |
[####################] 100% 0:00:05 | building arrays       | s7 |
[####################] 100% 0:00:02 | writing conversions   | s7 |
[####################] 100% 0:00:02 | indexing vcf depths   | s7 |
[####################] 100% 0:00:11 | writing vcf output    | s7 |
```

1.0.6 Step 7 of the assembly is repeated for clustering threshold 90% and 95% with different sets of missing data

1.1 RAxML

1.1.1 we will first try RAxML analysis for the assembly of 29.09.2022 where only Hydrophyllum was defined as an outgroup (three different clustering thresholds- different percentages of missing data for each clustering threshold). Then we will repeat the analysis for the assembly of 30.09.2022 where Hydrophyllum + the outgroup Phacelias (incl. W6021) were defined as outgroups. Finally, RaxML will be repeated for assembly of the ingroup with only outgroup W6021 (the rest samples were excluded from the assembly).

RAxML is performed in the terminal and you don't have to run it in a specific enviroment

First create an .sh file in the terminal using the vim text editor. Then put this file in the folder with the .phy files that you want to run in parallel.

In the terminal we did the .sh executableusing the command"chmod +x RAxML_ddRAD.sh" To executethe file we used the command "./RAxML_ddRAD.sh"

The text file contains the following:

/bin/bash set -e set -u set -o pipefail

for i in *.phy do raxml -f a -s $i -m GTRGAMMA -p 54321-x 123456-N 200 -T 20 -n $i done

T stands for the cores.

1.1.2 Plotting Raxml trees

[61]:
```
## Load treesclust85_20220930
tre15 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/RAxML/
 ,→RAxML_clust859095_20220930/RAxML_bipartitions.pops15_clust85.phy")
tre20 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/RAxML/
 ,→RAxML_clust859095_20220930/RAxML_bipartitions.pops20_clust85.phy")
tre25 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/RAxML/
 ,→RAxML_clust859095_20220930/RAxML_bipartitions.pops25_clust85.phy")
tre30 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/RAxML/
 ,→RAxML_clust859095_20220930/RAxML_bipartitions.pops30_clust85.phy")
tre35 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/RAxML/
 ,→RAxML_clust859095_20220930/RAxML_bipartitions.pops35_clust85.phy")
```

```
tre40 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/RAxML/
 ↪RAxML_clust859095_20220930/RAxML_bipartitions.pops40_clust85.phy")

tre15 = tre15.root(names=["Hydrophyllum_virginianum_W5141_out"])
tre20 = tre20.root(names=["Hydrophyllum_virginianum_W5141_out"])
tre25 = tre25.root(names=["Hydrophyllum_virginianum_W5141_out"])
tre30 = tre30.root(names=["Hydrophyllum_virginianum_W5141_out"])
tre35 = tre35.root(names=["Hydrophyllum_virginianum_W5141_out"])
tre40 = tre40.root(names=["Hydrophyllum_virginianum_W5141_out"])


## set dimensions of the canvas
canvas = toyplot.Canvas(width = 2000, height = 2000)

## dissect canvas into multiple cartesian areas (x1, x2, y1, y2)
ax0 = canvas.cartesian(bounds=('2%',  '30%', '5%',  '47.5%'))
ax1 = canvas.cartesian(bounds=('33%', '63%', '5%',  '47.5%'))
ax2 = canvas.cartesian(bounds=('66%', '96%', '5%',  '47.5%'))
ax3 = canvas.cartesian(bounds=('2%',  '30%', '50%', '97.5%'))
ax4 = canvas.cartesian(bounds=('33%', '63%', '50%', '97.5%'))
ax5 = canvas.cartesian(bounds=('66%', '96%', '50%', '97.5%'))

# call draw with the 'axes' argument to pass it to a specific cartesian area
style = {
    "tip_labels_align": True,
    "tip_labels_style": {"font-size": "14px"},
    "node_labels_style":{"font-size": "12px",
                         "baseline-shift": "7px",
                         "-toyplot-anchor-shift": "-13px"},
}
tre15.ladderize(1).draw(
    axes = ax0,
    **style,
    node_sizes = 0,
    node_labels = 'support');

tre20.ladderize(1).draw(
    axes = ax1,
    **style,
    node_sizes = 0,
    node_labels = 'support');

tre25.ladderize(1).draw(
    axes = ax2,
    **style,
    node_sizes = 0,
    node_labels = 'support');
```

```
tre30.ladderize(1).draw(
    axes = ax3,
    **style,
    node_sizes = 0,
    node_labels = 'support');

tre35.ladderize(1).draw(
    axes = ax4,
    **style,
    node_sizes = 0,
    node_labels = 'support');

tre40.ladderize(1).draw(
    axes = ax5,
    **style,
    node_sizes = 0,
    node_labels = 'support');

## hide the axes (e.g, ticks and splines)
ax0.show = False; ax1.show = False; ax2.show = False;
ax3.show = False; ax4.show = False; ax5.show = False;

## add names for the single trees
canvas.text(1000, 50, 'RAxML - Clustering threshold 85 %', style = {"font-size":
 ↪"24px"})
canvas.text(150, 125, '85 % missing data', style={"font-size": "18px"})
canvas.text(800, 125, '80 % missing data', style={"font-size": "18px"})
canvas.text(1450, 125, '75 % missing data', style={"font-size": "18px"})
canvas.text(150, 1025, '70 % missing data', style={"font-size": "18px"})
canvas.text(800, 1025, '65 % missing data', style={"font-size": "18px"})
canvas.text(1450, 1025, '60 % missing data', style={"font-size": "18px"});
```

[62]:
```
import toyplot.pdf
toyplot.pdf.render(canvas, "/home/marianna/Documents/Phacelia/Figures/
 ↪RAxML__clust85_20220930.pdf");
```

Tetrad Analysis

[2]:
```
## imports
import ipyrad as ip
import ipyrad.analysis as ipa
import ipyparallel as ipp
import tetrad
import toytree
import toyplot
import numpy as np   # numerical library
```

1 Print program versions; connect to cluster

[3]:
```
#ipyrad_0.9.84 installed on the system
## print Version of ipyrad, toytree, tetrad
print("ipyrad v. {}".format(ip.__version__))
print("toytree v. {}".format(toytree.__version__))
print("tetrad v. {}".format(tetrad.__version__))
## print Version of Python
from platform import python_version
print("Python v.", python_version())
```

```
ipyrad v. 0.9.84
toytree v. 2.0.5
tetrad v. 0.9.13
Python v. 3.7.0
```

[4]:
```
# After the cluster is running attach to it with ipyparallel
ipyclient = ipp.Client()
print(ip.cluster_info(ipyclient))
```

```
Parallel connection | Cristaria: 8 cores
None
```

2 Tetrad Analysis for ddRAD assembly of the ingroup

A speciestree inference tool based on the SVDQuartets algorithm Resolvesquartet trees from SNPs for all sets of quartets in a larger tree, and then joins the quartets together into a supertree.

2.0.1 we will run a multiple tetRAD analysis in a loop for all different pops of cluster threshold 85, then 90, then 95

[5]:
```
## tetrad for clust. threshold 85
## define the path to your sequence data in HDF5 format
## The input data file should be the .snps.hdf5
## read the *.snps.hdf5 files as values and store those path in
##a dictionary with assembly names as keys
dict = {
    "pop15": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops15_clust85.snps.hdf5",
    "pop20": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops20_clust85.snps.hdf5",
    "pop25": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops25_clust85.snps.hdf5",
    "pop30": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops30_clust85.snps.hdf5",
    "pop35": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops35_clust85.snps.hdf5",
    "pop40": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops40_clust85.snps.hdf5"
}
```

[7]:
```
## Iterate through the dictionary and run a tetRAD anlysis for each assembly

for key, value in dict.items():
    tet = ipa.tetrad(
        name = "Phac_tet_clust85_" + str(key),
        data = value,
        workdir = "./Phac_Analyses/tetrad/tet_clust85",
        nquartets = 1e6,
        nboots = 200)
    ## run
    tet.run(auto = True, force = True)
```

```
loading snps array [24 taxa x 26412 snps]
max unlinked SNPs per quartet [nloci]: 8653
quartet sampler [full]: 10626 / 10626
Parallel connection | Cristaria: 8 cores
initializing quartet sets database
[####################] 100% 0:00:03 | full tree * | avg SNPs/qrt: 2367
[####################] 100% 0:00:00 | boot rep. 1 | avg SNPs/qrt: 2363
[####################] 100% 0:00:00 | boot rep. 2 | avg SNPs/qrt: 2401
[####################] 100% 0:00:00 | boot rep. 3 | avg SNPs/qrt: 2374
[####################] 100% 0:00:00 | boot rep. 4 | avg SNPs/qrt: 2324
[####################] 100% 0:00:00 | boot rep. 5 | avg SNPs/qrt: 2321
[####################] 100% 0:00:00 | boot rep. 6 | avg SNPs/qrt: 2380
                                    …
```

```
[####################] 100% 0:00:00 | boot rep. 177 | avg SNPs/qrt: 2365
[####################] 100% 0:00:00 | boot rep. 178 | avg SNPs/qrt: 2384
[####################] 100% 0:00:00 | boot rep. 179 | avg SNPs/qrt: 2423
[####################] 100% 0:00:00 | boot rep. 180 | avg SNPs/qrt: 2313
[####################] 100% 0:00:00 | boot rep. 181 | avg SNPs/qrt: 2321
[####################] 100% 0:00:00 | boot rep. 182 | avg SNPs/qrt: 2368
[####################] 100% 0:00:00 | boot rep. 183 | avg SNPs/qrt: 2421
[####################] 100% 0:00:00 | boot rep. 184 | avg SNPs/qrt: 2331
[####################] 100% 0:00:00 | boot rep. 185 | avg SNPs/qrt: 2345
[####################] 100% 0:00:00 | boot rep. 186 | avg SNPs/qrt: 2360
[####################] 100% 0:00:00 | boot rep. 187 | avg SNPs/qrt: 2348
[####################] 100% 0:00:00 | boot rep. 188 | avg SNPs/qrt: 2386
[####################] 100% 0:00:00 | boot rep. 189 | avg SNPs/qrt: 2433
[####################] 100% 0:00:00 | boot rep. 190 | avg SNPs/qrt: 2351
[####################] 100% 0:00:00 | boot rep. 191 | avg SNPs/qrt: 2422
[####################] 100% 0:00:00 | boot rep. 192 | avg SNPs/qrt: 2415
[####################] 100% 0:00:00 | boot rep. 193 | avg SNPs/qrt: 2366
[####################] 100% 0:00:00 | boot rep. 194 | avg SNPs/qrt: 2326
[####################] 100% 0:00:00 | boot rep. 195 | avg SNPs/qrt: 2362
[####################] 100% 0:00:00 | boot rep. 196 | avg SNPs/qrt: 2390
[####################] 100% 0:00:00 | boot rep. 197 | avg SNPs/qrt: 2378
[####################] 100% 0:00:00 | boot rep. 198 | avg SNPs/qrt: 2430
[####################] 100% 0:00:00 | boot rep. 199 | avg SNPs/qrt: 2334
[####################] 100% 0:00:00 | boot rep. 200 | avg SNPs/qrt: 2367
```

[6]:
```
## Plot all six clust85 tetRAD coalescent trees together
## Load trees
tet15 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/tetrad/
 ↪tet_clust85/Phac_tet_clust85_pop15.tree.cons")
tet20 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/tetrad/
 ↪tet_clust85/Phac_tet_clust85_pop20.tree.cons")
tet25 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/tetrad/
 ↪tet_clust85/Phac_tet_clust85_pop25.tree.cons")
tet30 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/tetrad/
 ↪tet_clust85/Phac_tet_clust85_pop30.tree.cons")
tet35 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/tetrad/
 ↪tet_clust85/Phac_tet_clust85_pop35.tree.cons")
tet40 = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analyses/tetrad/
 ↪tet_clust85/Phac_tet_clust85_pop40.tree.cons")

tet15 = tet15.root(wildcard = "Phacelia_artemisioides_W6021_out")
tet20 = tet20.root(wildcard = "Phacelia_artemisioides_W6021_out")
tet25 = tet25.root(wildcard = "Phacelia_artemisioides_W6021_out")
tet30 = tet30.root(wildcard = "Phacelia_artemisioides_W6021_out")
tet35 = tet35.root(wildcard = "Phacelia_artemisioides_W6021_out")
tet40 = tet40.root(wildcard = "Phacelia_artemisioides_W6021_out")
```

```
## set dimensions of the canvas
canvas = toyplot.Canvas(width = 2000, height = 2000)

## dissect canvas into multiple cartesian areas (x1, x2, y1, y2)
ax0 = canvas.cartesian(bounds=('2%',  '30%', '5%',  '47.5%'))
ax1 = canvas.cartesian(bounds=('33%', '63%', '5%',  '47.5%'))
ax2 = canvas.cartesian(bounds=('66%', '96%', '5%',  '47.5%'))
ax3 = canvas.cartesian(bounds=('2%',  '30%', '50%', '97.5%'))
ax4 = canvas.cartesian(bounds=('33%', '63%', '50%', '97.5%'))
ax5 = canvas.cartesian(bounds=('66%', '96%', '50%', '97.5%'))

## define style ones and use it for every tree
style = {
    "tip_labels_align": True,
    "tip_labels_style": {"font-size": "11px"},
    "node_labels_style":{"font-size": "12px",
                         "baseline-shift": "7px",
                         "-toyplot-anchor-shift": "-13px"},
}
tet15.ladderize(1).draw(
    axes = ax0,
    **style,
    node_sizes = 0,
    node_labels = 'support');

tet20.ladderize(1).draw(
    axes = ax1,
    **style,
    node_sizes = 0,
    node_labels = 'support');

tet25.ladderize(1).draw(
    axes = ax2,
    **style,
    node_sizes = 0,
    node_labels = 'support');

tet30.ladderize(1).draw(
    axes = ax3,
    **style,
    node_sizes = 0,
    node_labels = 'support');

tet35.ladderize(1).draw(
    axes = ax4,
    **style,
    node_sizes = 0,
```

```
    node_labels = 'support');

tet40.ladderize(1).draw(
    axes = ax5,
    **style,
    node_sizes = 0,
    node_labels = 'support');

## hide the axes (e.g, ticks and splines)
ax0.show = False; ax1.show = False; ax2.show = False;
ax3.show = False; ax4.show = False; ax5.show = False;

## add names for the single trees
canvas.text(1000, 50, 'tetRAD/SVDQuartet - Clustering threshold 85 %', style =␣
 ↪{"font-size": "24px"})
canvas.text(150, 125, '85 % missing data', style={"font-size": "18px"})
canvas.text(800, 125, '80 % missing data', style={"font-size": "18px"})
canvas.text(1450, 125, '75 % missing data', style={"font-size": "18px"})
canvas.text(150, 1025, '70 % missing data', style={"font-size": "18px"})
canvas.text(800, 1025, '65 % missing data', style={"font-size": "18px"})
canvas.text(1450, 1025, '60 % missing data', style={"font-size": "18px"});
```

[10]:
```
import toyplot.pdf
toyplot.pdf.render(canvas, "/home/marianna/Documents/Phacelia/tetrad_Figures/
 ↪tetrad_clust85_20221004.pdf");
```

[7]:
```
## tetrad for clust. threshold 90
## define the path to your sequence data in HDF5 format
## The input data file should be the .snps.hdf5
## read the *.snps.hdf5 files as values and store those path in a dictionary␣
 ↪with assembly names as keys
dict = {
    "pop15": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops15_clust90.snps.hdf5",
    "pop20": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops20_clust90.snps.hdf5",
    "pop25": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops25_clust90.snps.hdf5",
    "pop30": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops30_clust90.snps.hdf5",
    "pop35": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops35_clust90.snps.hdf5",
    "pop40": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops40_clust90.snps.hdf5"
}
```

[12]:
```
## Iterate through the dictionary and run a tetRAD anlysis for each assembly

for key, value in dict.items():
    tet = ipa.tetrad(
        name = "Phac_tet_clust90_" + str(key),
        data = value,
        workdir = "./Phac_Analyses/tetrad/tet_clust90",
        nquartets = 1e6,
        nboots = 200)
    ## run
    tet.run(auto = True, force = True)
```

```
loading snps array [24 taxa x 25597 snps]
max unlinked SNPs per quartet [nloci]: 8927
quartet sampler [full]: 10626 / 10626
Parallel connection | Cristaria: 8 cores
initializing quartet sets database
[####################] 100% 0:00:03 | full tree * | avg SNPs/qrt: 2265
[####################] 100% 0:00:00 | boot rep. 1 | avg SNPs/qrt: 2222
                                    …
```

Results for clust. threshold 90% were plotted as showed above

[9]:
```
## tetrad for clust. threshold 95
## define the path to your sequence data in HDF5 format
## The input data file should be the .snps.hdf5
## read the *.snps.hdf5 files as values and store those path in a dictionary␣
 ↪with assembly names as keys
dict = {
    "pop15": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops15_clust95.snps.hdf5",
    "pop20": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops20_clust95.snps.hdf5",
    "pop25": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops25_clust95.snps.hdf5",
    "pop30": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops30_clust95.snps.hdf5",
    "pop35": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops35_clust95.snps.hdf5",
    "pop40": "/home/marianna/Documents/Phacelia/hdf5_files_202209/
 ↪pops40_clust95.snps.hdf5"
}
```

[17]:
```
## Iterate through the dictionary and run a tetRAD anlysis for each assembly

for key, value in dict.items():
    tet = ipa.tetrad(
        name = "Phac_tet_clust95_" + str(key),
        data = value,
        workdir = "./Phac_Analyses/tetrad/tet_clust95",
        nquartets = 1e6,
        nboots = 200)
    ## run
    tet.run(auto = True, force = True)
```

```
loading snps array [24 taxa x 20779 snps]
max unlinked SNPs per quartet [nloci]: 8840
quartet sampler [full]: 10626 / 10626
Parallel connection | Cristaria: 8 cores
initializing quartet sets database
[####################] 100% 0:00:03 | full tree * | avg SNPs/qrt: 1759
[####################] 100% 0:00:00 | boot rep. 1 | avg SNPs/qrt: 1741
[####################] 100% 0:00:00 | boot rep. 2 | avg SNPs/qrt: 1728
[####################] 100% 0:00:00 | boot rep. 3 | avg SNPs/qrt: 1767
[####################] 100% 0:00:00 | boot rep. 4 | avg SNPs/qrt: 1776
[####################] 100% 0:00:00 | boot rep. 5 | avg SNPs/qrt: 1728
                                    …
```

Results for clust. threshold 95% were plotted as showed above

2.0.2 Produce a cloud tree for pops40 - clustering threshold 90%

[11]:
```
treeorder = ["Phacelia_artemisioides_W6021_out", "Phacelia_sinuata_W6376",␣
 ↪"Phacelia_sinuata_W6368", "Phacelia_pinnatifida_W5610",
            "Phacelia_pinnatifida_W6374", "Phacelia_pinnatifida_W6080",␣
 ↪"Phacelia_sanzinii_W6369", "Phacelia_nana_W6373",
             "Phacelia_setigera_var_setigera_W6375",␣
 ↪"Phacelia_setigera_var_setigera_W6370", "Phacelia_artemisioides_W6031",
             "Phacelia_viscosa_W5145","Phacelia_setigera_W5612",␣
 ↪"Phacelia_pinnatifida_W6079", "Phacelia_pinnatifida_W6001",
             "Phacelia_pinnatifida_W6029", "Phacelia_pinnatifida_W5599",␣
 ↪"Phacelia_pinnatifida_W6027",
             "Phacelia_pinnatifida_viscosa_W6028", "Phacelia_spec_W6037",
            "Phacelia_spec_nov_W6078", "Phacelia_cf_viscosa_pinnatifida_W6024",␣
 ↪"Phacelia_spec_nov_W5637",
            "Phacelia_spec_nov_W5636"]
```

[12]:
```
## Load the 200 bootstrap trees from pops35 TetRad analysis - clustering␣
 ↪threshol 90% and root it
tetcloud40 = toytree.mtree("./Phac_Analyses/tetrad/tet_clust90/
 ↪Phac_tet_clust90_pop40.tree.boots")
tetcloud40.treelist = [i.root(["Phacelia_artemisioides_W6021_out"]) for i in␣
 ↪tetcloud40.treelist]

## plot the rooted bootstrap trees as a cloud tree
canvas, axes, mark = tetcloud40.draw_cloud_tree(
    height = 600,
    width = 400,

    ## define a fix tree order to make it comparable with the cons tree
    fixed_order = treeorder,
    use_edge_lengths = False,
    edge_style = {"stroke-opacity": 0.05,
                  "stroke-width": 1}
);
```

[13]:
```
## Load consensus tree and cloud tree and root it with W6021
constree40 = toytree.tree("./Phac_Analyses/tetrad/tet_clust90/
 ↪Phac_tet_clust90_pop40.tree.cons" ).root(wildcard = "W6021")

## Load TetRad bootstrap trees and root it with W6021
cloudtree40 = toytree.mtree("./Phac_Analyses/tetrad/tet_clust90/
 ↪Phac_tet_clust90_pop40.tree.boots")
cloudtree40.treelist = [i.root(["Phacelia_artemisioides_W6021_out"]) for i in␣
 ↪cloudtree40.treelist]

## set dimensions of the canvas
canvas = toyplot.Canvas(width = 1300, height = 900)

## dissect canvas into multiple cartesian areas (x1, x2, y1, y2)
ax0 = canvas.cartesian(bounds=('5%',  '47.5%', '5%',  '95%'))
ax1 = canvas.cartesian(bounds=('52.5%', '95%', '5%',  '95%'))

# call draw with the 'axes' argument to pass it to a specific cartesian area
style = {"tip_labels_align": True,
         "tip_labels_style": {"font-size": "12px"},
         "node_labels_style":{"font-size": "12px",
                              "baseline-shift": "7px",
                              "-toyplot-anchor-shift": "-13px"},
}

cstyle = {"tip_labels_align": True,
          "layout": 'l',
          "tip_labels_style": {"font-size": "12px"},
```

```
        "node_labels_style":{"font-size": "12px",
                             "baseline-shift": "7px",
                             "-toyplot-anchor-shift": "-13px"},
}

constree40.ladderize(1).draw(
    axes = ax0,
    **style,
    node_sizes = 0,
    node_labels = 'support'); #here in your notebook is:= constree.
 ↪get_node_values("support")); but is not recognized and I modified it

## plot the rooted bootstrap trees as a cloud tree
cloudtree40.draw_cloud_tree(
    axes = ax1,
    fixed_order = treeorder,   ## define a fix tree order to make it comparable␣
 ↪with the cons tree
    **cstyle,
    use_edge_lengths = False,
    #tip_labels = False,
    edge_style = {"stroke-opacity": 0.05,
                  "stroke-width": 1}
);

# hide the axes (e.g, ticks and splines)
ax0.show = False; ax1.show = False;
```

[9]:
```
import toyplot.pdf
toyplot.pdf.render(canvas, "/home/marianna/Documents/Phacelia/tetrad_Figures/
 ↪tetrad_cloud_20221121.pdf");
```

Structure Analysis

1. Population analysis of Phacelia with outgroups removed

Structure is a standard tool for examining population genetic structure based on allele frequencies within and among populations. To perform this analysis, at step 7 of the assembly (which included 23 ingroup taxa and one outgroup, i.e. W6021) a branch was created with min_samples = 12 (50% of missing data) for clustering thresholds 85%, 90% and 95%. The assembly with clustering threshold 90% and 50% misssing data was used for Structure analysis, PCA and Treemix.

[1]:
```
## imports
import ipyrad as ip
import ipyrad.analysis as ipa
import ipyparallel as ipp
import toytree
import toyplot
import numpy as np   # numerical library
```

[2]:
```
#ipyrad_0.9.84 installed on the system
## print Version of ipyrad, toytree, tetrad
print("ipyrad v. {}".format(ip.__version__))
print("toytree v. {}".format(toytree.__version__))
#print("tetrad v. {}".format(tetrad.__version__))
## print Version of Python
from platform import python_version
print("Python v.", python_version())
```

```
ipyrad v. 0.9.84
toytree v. 2.0.5
Python v. 3.10.5
```

[3]:
```
# After the cluster is running attach to it with ipyparallel
ipyclient = ipp.Client()
print(ip.cluster_info(ipyclient))
```

```
Parallel connection | Cristaria: 8 cores
None
```

[4]:
```
## load the hdf5 data for the STRUCTURE analysis
dataclust90 = "/home/marianna/Documents/Phacelia/Phac_Assembly/min12_clust90.
 ↪snps.hdf5"
```

1.0.1 Grouping the samples into population using an 'imap' based on the ML trees

[5]:
```
# group individuals into populations, according to Karuna's analysis and the␣
 ↪new topology
imap = {
    "1": ["W6368","W6376"],
    "2": ["W5599","W6027","W6028","W6037","W6078", "W6024", "W5637", "W5636"],
    "3": ["W6373"],
    "4": ["W5610", "W6374", "W6080", "W6369"],
    "5": ["W6375", "W6031","W6370"],
    "6": ["W5145", "W5612"],
    "7": ["W6001", "W6079","W6029"],

}

## require that 50% of samples have data in each group
minmap = {i: 0.5 for i in imap}
```

[7]:
```
struct = ipa.structure(
    name = "min12", #50% missing data
    data = dataclust90,
    imap = imap,
    minmap = minmap,
    mincov = 0.5,
    workdir = "/home/marianna/Documents/Phacelia/Phac_Analyses/Structure/
 ↪min12_20221117"
)
```

```
Samples: 23
Sites before filtering: 19359
Filtered (indels): 1020
Filtered (bi-allel): 210
Filtered (mincov): 579
Filtered (minmap): 5685
Filtered (subsample invariant): 6133
Filtered (minor allele frequency): 0
Filtered (combined): 10771
Sites after filtering: 8587
Sites containing missing values: 5548 (64.61%)
Missing values in SNP matrix: 13587 (6.88%)
SNPs (total): 8587
SNPs (unlinked): 4200
```

1.0.2 Running STRUCTURE and plot results

The burnin and numreps parameters determine the length of the run.

[8]:
```
struct.mainparams.burnin  = 100000
struct.mainparams.numreps = 500000

## see all mainparams
print(struct.mainparams)

#see or ser extraparams
print(struct.extraparams)
```

```
burnin             100000
extracols          0
label              1
locdata            0
mapdistances       0
markernames        0
markovphase        0
missing            -9
notambiguous       -999
numreps            500000
onerowperind       0
phased             0
phaseinfo          0
phenotype          0
ploidy             2
popdata            0
popflag            0
recessivealleles   0

admburnin           500
alpha               1.0
alphamax            10.0
alphapriora         1.0
alphapriorb         2.0
alphapropsd         0.025
ancestdist          0
ancestpint          0.9
computeprob         1
echodata            0
fpriormean          0.01
fpriorsd            0.05
freqscorr           1
gensback            2
inferalpha          1
inferlambda         0
intermedsave        0
lambda_             1.0
linkage             0
```

```
locispop            0
locprior            0
locpriorinit        1.0
log10rmax           1.0
log10rmin           -4.0
log10rpropsd        0.1
log10rstart         -2.0
maxlocprior         20.0
metrofreq           10
migrprior           0.01
noadmix             0
numboxes            1000
onefst              0
pfrompopflagonly    0
popalphas           0
popspecificlambda   0
printlambda         1
printlikes          0
printnet            1
printqhat           0
printqsum           1
randomize           0
reporthitrate       0
seed                12345
sitebysite          0
startatpopinfo      0
unifprioralpha      1
updatefreq          10000
usepopinfo          0
```

[9]:
```
## set a range of k-values to test
kvalues = [2, 3, 4, 5, 6, 7, 8]
```

[10]:
```
## submit batches of 10 replicates jobs for each value of k
for kpop in kvalues:
    struct.run(kpop = kpop, nreps = 10, seed = 12345, ipyclient = ipyclient)#, ␣
 ↪force = True)
```

```
[####################] 100% 2:16:00 | running 10 structure jobs
[####################] 100% 2:47:39 | running 10 structure jobs
[####################] 100% 3:16:33 | running 10 structure jobs
[####################] 100% 3:49:20 | running 10 structure jobs
[####################] 100% 4:14:27 | running 10 structure jobs
[####################] 100% 4:43:27 | running 10 structure jobs
[####################] 100% 5:13:25 | running 10 structure jobs
```

2 Analyse results: check results in evanno table

Note the K value for which deltaK is maximum

[11]:
```
etable = struct.get_evanno_table(kvalues)
etable
```

[11]:

	Nreps	lnPK	lnPPK	deltaK	estLnProbMean	estLnProbStdev
2	10	0.00	0.00	0.000	-34209.15	111.621
3	10	-4338.57	10175.00	0.687	-38547.72	14800.358
4	10	5836.43	14792.91	3.698	-32711.29	4000.077
5	10	-8956.48	20984.70	0.776	-41667.77	27049.697
6	10	12028.22	75045.62	24.041	-29639.55	3121.561
7	10	-63017.40	128063.70	1.062	-92656.95	120642.293
8	10	65046.30	0.00	0.000	-27610.65	5631.731

[13]:
```
## summarize results
struct.clumppparams.m = 3                ## use largegreedy algorithm
struct.clumppparams.greedy_option = 2    ## test nrepeat possible orders
struct.clumppparams.repeats = 100000     ## number of repeats
```

[14]:
```
qtable = struct.get_clumpp_table(kvalues)#, max_var_multiple=100.)
```

```
[K2] 10/10 results permuted across replicates (max_var=0).
[K3] 10/10 results permuted across replicates (max_var=0).
[K4] 10/10 results permuted across replicates (max_var=0).
[K5] 10/10 results permuted across replicates (max_var=0).
[K6] 10/10 results permuted across replicates (max_var=0).
[K7] 10/10 results permuted across replicates (max_var=0).
[K8] 10/10 results permuted across replicates (max_var=0).
```

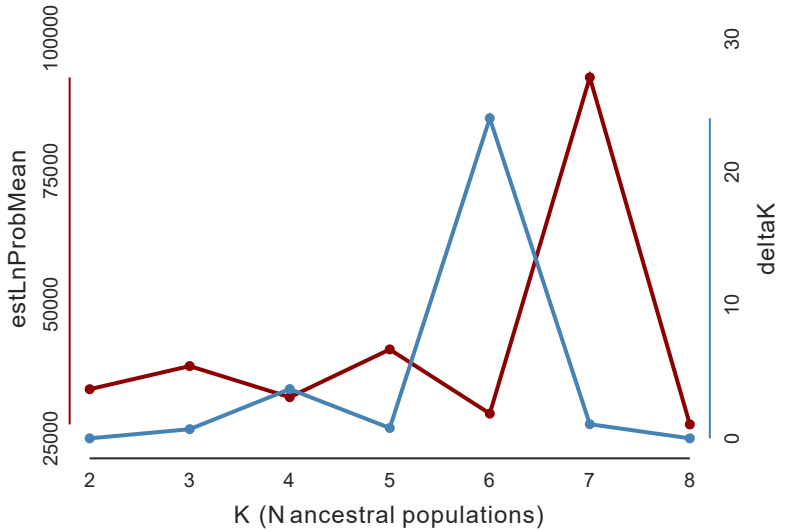
[15]:
```
# get canvas object and set size
canvas = toyplot.Canvas(width=400, height=300)

# plot the mean log probability of the models in red
axes = canvas.cartesian(ylabel="estLnProbMean")

axes.plot(etable.estLnProbMean * -1, color="darkred", marker="o")
axes.y.spine.style = {"stroke": "darkred"}

# plot delta K with its own scale bar of left side and in blue
axes = axes.share("x", ylabel="deltaK", ymax=etable.deltaK.max() + etable.
 ↪deltaK.max() * .25)
axes.plot(etable.deltaK, color="steelblue", marker="o");
axes.y.spine.style = {"stroke": "steelblue"}

# set x labels
axes.x.ticks.locator = toyplot.locator.Explicit(range(len(etable.index)),␣
 ↪etable.index)
axes.x.label.text = "K (N ancestral populations)"
```



[16]:
```
k = 6
table = struct.get_clumpp_table(k)
```

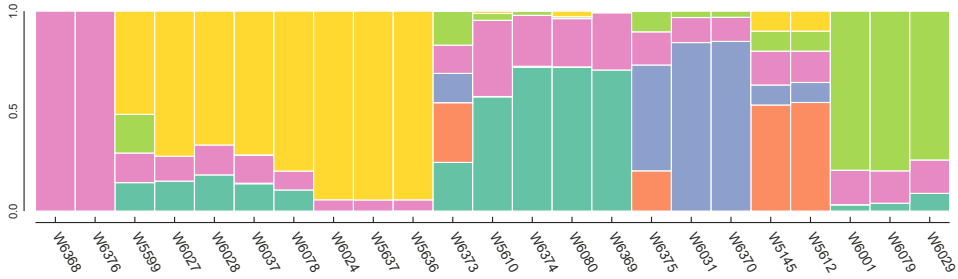
```
[K6] 10/10 results permuted across replicates (max_var=0).
```

[17]:
```
# sort list by columns
table.sort_values(by=list(range(k)), inplace=True)

# or, sort by a list of names (here taken from imap)
import itertools
onames = list(itertools.chain(*imap.values()))
table = table.loc[onames]
```

[18]:
```
# build barplot
canvas = toyplot.Canvas(width=1000, height=500)
axes = canvas.cartesian(bounds=("10%", "90%", "10%", "45%"))
axes.bars(table)

# add labels to x-axis
ticklabels = [i for i in table.index.tolist()]
axes.x.ticks.locator = toyplot.locator.Explicit(labels=ticklabels)
axes.x.ticks.labels.angle = -60
axes.x.ticks.show = True
axes.x.ticks.labels.offset = 10
axes.x.ticks.labels.style = {"font-size": "12px"}
```



[20]:
```
myorder = ["W6368", "W6376", "W5599","W6027", "W6028", "W6078", "W6037",␣
 ↪"W6024", "W5636", "W5637",
          "W6373","W5610", "W6374","W6080","W6369",  "W6375", "W6370",␣
 ↪"W6031","W5612", "W5145","W6001", "W6029","W6079"]
print("custom ordering")
print(qtable[2].loc[myorder])
```

```
custom ordering
                0          1
W6368  4.000e-01  6.000e-01
W6376  4.000e-01  6.000e-01
W5599  1.100e-03  9.989e-01
W6027  0.000e+00  1.000e+00
W6028  6.000e-04  9.994e-01
W6078  0.000e+00  1.000e+00
W6037  8.000e-04  9.992e-01
W6024  0.000e+00  1.000e+00
W5636  0.000e+00  1.000e+00
W5637  0.000e+00  1.000e+00
W6373  9.000e-01  1.000e-01
W5610  9.944e-01  5.600e-03
W6374  9.999e-01  1.000e-04
W6080  9.999e-01  1.000e-04
W6369  1.000e+00  0.000e+00
W6375  9.000e-01  1.000e-01
W6370  6.000e-01  4.000e-01
W6031  6.000e-01  4.000e-01
W5612  7.000e-01  3.000e-01
W5145  7.000e-01  3.000e-01
W6001  9.000e-01  1.000e-01
W6029  9.001e-01  9.990e-02
```

```
W6079  9.000e-01  1.000e-01
```

2.0.1 Plot all STRUCTURE results against Phylogeny

[53]:
```
# get tree from RAxML results

tre = toytree.tree("/home/marianna/Documents/Phacelia/Phac_Analysis/
 ↪RAxML_bipartitions.pops40_clust90 (copy).phy")
rtre = tre.root(wildcard = "W6021")

## further styling of plot with css
style = {"stroke":toyplot.color.black, ##near_black is giving error module␣
 ↪'toyplot.color' has no attribute 'near_black'
         "stroke-width": 0.25}

##    y1
## x1    x2
##    y2

## built & dissect canvas into multiple cartesian areas (x1, x2, y1, y2)
c = toyplot.Canvas(width = 900, height = 700)
a1 = c.cartesian(bounds=('1%', '49%', '6.25%', '94%'))       # The tree
a2 = c.cartesian(bounds=('46.5%', '55%', '5.25%', '91%'))  # K=2
a3 = c.cartesian(bounds=('55.5%', '64%', '5.25%', '91%'))  # K=3
a4 = c.cartesian(bounds=('64.5%', '73%', '5.25%', '91%'))  # K=4
a5 = c.cartesian(bounds=('73.5%', '82%', '5.25%', '91%'))  # K=5
a6 = c.cartesian(bounds=('82.5%', '91%', '5.25%', '91%'))  # K=6
a7 = c.cartesian(bounds=('91.5%', '100%', '5.25%', '91%'))  # K=7
a1.show = False
a2.show = False
a3.show = False
a4.show = False
a5.show = False
a6.show = False
a7.show = False
## draw the tree
rtre.ladderize(1).draw(
    axes = a1,
    use_edge_lengths = True,
    tip_labels_align = True,
    tip_labels_style = {"font-size": "9px"},
    node_labels = "support",
    node_sizes = 0,
    node_labels_style={"font-size": "9px",
                       "baseline-shift": "7px",
                       "-toyplot-anchor-shift": "-8px"});
```

```
## draw the STRUCTURE bar plots
## 'along' defines plot orientation; x = vertical; y = horizontal
a2.bars(qtable[2].loc[myorder], style = style, along = 'y');
a3.bars(qtable[3].loc[myorder], style = style, along = 'y');
a4.bars(qtable[4].loc[myorder], style = style, along = 'y');
a5.bars(qtable[5].loc[myorder], style = style, along = 'y');
a6.bars(qtable[6].loc[myorder], style = style, along = 'y');
a7.bars(qtable[7].loc[myorder], style = style, along = 'y');

## add header for the bar plots
c.text(460, 23, 'K = 2', style={"font-size": "13px"})
c.text(540, 23, 'K = 3', style={"font-size": "13px"})
c.text(620, 23, 'K = 4', style={"font-size": "13px"})
c.text(700, 23, 'K = 5', style={"font-size": "13px"})
c.text(780, 23, 'K = 6', style={"font-size": "13px"})
c.text(860, 23, 'K = 7', style={"font-size": "13px"})

## add deltaK values below the bar plots
c.text(460, 645, '0.0', style={"font-size": "10px"})
c.text(540, 645, '0.7', style={"font-size": "10px"})
c.text(620, 645, '3.7', style={"font-size": "10px"})
c.text(700, 645, '0.8', style={"font-size": "10px"})
c.text(780, 645, '24.0', style={"font-size": "10px"})
c.text(860, 645, '1.1', style={"font-size": "10px"})
c.text(655, 655, 'delta <b>K</b>', style={"font-size": "10px"});
```

[54]:
```
import toyplot.pdf
toyplot.pdf.render(c, "/home/marianna/Documents/Phacelia/Figures/
↪Structure_RAxML_pops40clust90_20221121.pdf");
```

PCA Analysis

1. PCA analysis

PCA stands for Principal Component Analysis. It is a dimensionality reduction method used to transform and project data points onto fewer orthogonal axes that can explain the greatest amount of variance in the data.

[1]:
```
# the path to your .snps.hdf5 database file
data = "/home/marianna/Documents/Phacelia/Phac_Assembly/min12_clust90.snps.hdf5"
```

[2]:
```
## imports
import ipyrad as ip
import ipyrad.analysis as ipa
import ipyparallel as ipp
import toyplot
import pandas as pd
import numpy as np   # numerical library
```

[3]:
```
# After the cluster is running attach to it with ipyparallel
ipyclient = ipp.Client()
print(ip.cluster_info(ipyclient))
```

```
Parallel connection | Cristaria: 8 cores
None
```

1.0.1 Grouping individuals into populations based on the topology and the 'structure analysis'

impap dictionary : assigns individual samples to populations

minimap dictionary : states the required percentage coverage in each population; here it's 50%

[4]:
```
# group individuals into populations, according to my first structure analysis.
imap ={
    "sin": ["W6368","W6376"],
    "pinn_peru": ["W5599","W6027","W6028","W6037","W6078","W6024","W5637",␣
 ↪"W5636"],
    "nana": ["W6373"],
    "argent": ["W5610","W6374", "W6080","W6369"],
    "set_arg": ["W6375", "W6031","W6370"],
```

```
    "set_ch": ["W5145", "W5612"],
    "set_pe": ["W6001", "W6079","W6029"],

}

## require that 50% of samples have data in each group
minmap = {i: 0.5 for i in imap}
```

[5]:
```
# init pca object with input data and (optional) parameter options
pca = ipa.pca(
    data=data,
    imap=imap,
    minmap=minmap,
    mincov=0.25,
    impute_method="sample",
)
```

```
Samples: 23
Sites before filtering: 19359
Filtered (indels): 1020
Filtered (bi-allel): 210
Filtered (mincov): 51
Filtered (minmap): 5685
Filtered (subsample invariant): 6133
Filtered (minor allele frequency): 0
Filtered (combined): 10771
Sites after filtering: 8587
Sites containing missing values: 5548 (64.61%)
Missing values in SNP matrix: 13587 (6.88%)
SNPs (total): 8587
SNPs (unlinked): 4200
Imputation: 'sampled'; (0, 1, 2) = 89.8%, 6.7%, 3.5%
```

[6]:
```
# run the PCA analysis
pca.run()
```

```
Subsampling SNPs: 4200/8587
```

[7]:
```
# store the PC axes as a dataframe
df = pd.DataFrame(pca.pcaxes[0], index=pca.names)

# write the PC axes to a CSV file
df.to_csv("pca_analysis20221121.csv")

# show the first ten samples and the first 10 PC axes
df.iloc[:10, :10].round(2)
```

[7]:

	0	1	2	3	4	5	6	7	8	9
W5145	-6.98	-8.80	-4.22	21.02	1.99	1.29	-1.31	-0.87	0.21	-0.08
W5599	5.78	-1.85	0.77	-1.70	-1.65	-1.08	-0.04	17.23	14.93	9.43
W5610	-4.70	-2.02	5.47	-0.69	-3.14	-6.34	-1.45	1.62	3.63	-1.35
W5612	-7.30	-8.99	-4.33	22.28	1.75	1.42	-1.51	-0.66	1.04	-0.47
W5636	17.14	1.02	-2.20	-0.17	0.49	-0.93	0.23	-8.63	4.15	1.57
W5637	17.77	1.09	-2.43	-0.40	0.83	-0.81	0.10	-9.57	4.31	1.62
W6001	-8.37	-10.48	-11.34	-9.17	1.14	1.20	0.04	-1.24	2.15	-0.78
W6024	16.02	0.59	-2.36	-0.20	0.71	-1.01	0.39	-8.49	4.32	1.54
W6027	12.08	0.57	-0.79	0.38	0.98	0.26	-0.67	10.13	-11.60	-6.05
W6028	9.01	-0.10	0.25	-0.71	0.19	0.71	-0.60	7.28	-3.15	-1.31

[8]:
```
# plot PC axes 0 and 2
pca.draw(0, 2);
# plot PC axes 0 and 1
pca.draw(0, 1);
```

[9]:
```
# plot PC axes 0 and 2 with many replicate subsamples
pca.run(nreplicates=25, seed=12345)
pca.draw(0, 2);
# plot PC axes 0 and 1 with many replicate subsamples
pca.run(nreplicates=25, seed=12345)
pca.draw(0, 1);
```

```
Subsampling SNPs: 4200/8587
Subsampling SNPs: 4200/8587
```

[11]:
```
# The outfile end in `.pdf` or `.svg`
pca.draw(0, 1, outfile="/home/marianna/Documents/Phacelia/Figures/
 ↪PCA_01_20221121.pdf")
pca.draw(0, 1, outfile="/home/marianna/Documents/Phacelia/Phac_Analyses/
 ↪Phac_PCA_01_20221121.svg")

pca.draw(0, 2, outfile="/home/marianna/Documents/Phacelia/Figures/
 ↪PCA_02_20221121.pdf")
pca.draw(0, 2, outfile="/home/marianna/Documents/Phacelia/Phac_Analyses/
 ↪Phac_PCA_02_20221121.svg")
```

[11]:
```
(<toyplot.canvas.Canvas at 0x7ff6b7dfb0d0>,
 <toyplot.coordinates.Cartesian at 0x7ff6bc1c6920>)
```

1.0.2 Grouping individuals into populations based on their geographical distribution

[12]:
```
imap2 = {
"Peru": ["W5599", "W6078", "W5636","W6037", "W6024", "W5637", "W6079", "W6001",␣
 ↪"W6028", "W6029", "W6027"],
```

```
"Argentina": ["W6368", "W6376", "W6373","W6374", "W6080", "W6369",␣
↪"W5610","W6375","W6031","W6370"],
"Chile": ["W5612","W5145"],
}
## require that 50% of samples have data in each group
minmap = {i: 0.5 for i in imap2}
```

[13]:
```
# init pca object with input data and (optional) parameter options
pca2 = ipa.pca(
    data=data,
    imap=imap2,
    minmap=minmap,
    mincov=0.25,
    impute_method="sample",
)
```

```
Samples: 23
Sites before filtering: 19359
Filtered (indels): 1020
Filtered (bi-allel): 210
Filtered (mincov): 51
Filtered (minmap): 2298
Filtered (subsample invariant): 6133
Filtered (minor allele frequency): 0
Filtered (combined): 8381
Sites after filtering: 10977
Sites containing missing values: 7938 (72.31%)
Missing values in SNP matrix: 26176 (10.37%)
SNPs (total): 10977
SNPs (unlinked): 5189
Imputation: 'sampled'; (0, 1, 2) = 87.7%, 10.4%, 1.9%
```

[14]:
```
# run the PCA analysis
pca2.run()
```

```
Subsampling SNPs: 5189/10977
```

[15]:
```
# store the PC axes as a dataframe
df = pd.DataFrame(pca2.pcaxes[0], index=pca2.names)

# write the PC axes to a CSV file
df.to_csv("pca2_analysis_20221121.csv")

# show the first ten samples and the first 10 PC axes
df.iloc[:10, :10].round(2)
```

[15]:

	0	1	2	3	4	5	6	7	8	9
W5145	-7.62	25.71	6.18	-7.34	-3.06	-1.47	-1.00	-0.23	0.70	0.31
W5599	5.49	-0.05	-3.26	4.38	0.06	-0.71	9.44	15.75	20.27	-0.91
W5610	-6.25	-1.96	-3.45	0.35	-2.26	1.55	11.10	-1.92	3.91	-1.47
W5612	-8.01	26.14	6.44	-7.70	-3.22	-1.20	-0.95	-0.00	1.29	-0.56
W5636	18.41	-2.32	3.38	-2.71	-1.62	-0.10	-2.75	-10.15	5.09	-0.47
W5637	18.68	-2.65	3.36	-2.96	-1.85	-0.11	-3.12	-10.27	4.65	-0.48
W6001	-2.61	7.34	-6.16	20.59	9.93	-2.30	-5.41	-3.88	-3.33	-9.29
W6024	17.55	-2.06	3.18	-2.87	-1.73	-0.21	-2.25	-10.01	4.94	0.34
W6027	11.33	-1.09	0.74	-2.41	-0.62	0.00	1.12	11.10	-13.79	-0.67
W6028	10.34	-1.39	0.17	-1.13	-0.19	-0.23	2.14	10.58	-4.55	0.34

[16]:
```
# plot PC axes 0 and 2
pca2.draw(0, 2);
```

[17]:
```
# plot PC axes 0 and 1
pca2.draw(0, 1);
```

[18]:
```
# plot PC axes 0 and 2 with many replicate subsamples
pca2.run(nreplicates=25, seed=12345)
pca2.draw(0, 2);
# plot PC axes 0 and 1 with many replicate subsamples
pca2.run(nreplicates=25, seed=12345)
pca2.draw(0, 1);
```

```
Subsampling SNPs: 5189/10977
Subsampling SNPs: 5189/10977
```

[19]:
```
# The outfile end in `.pdf` or `.svg`
pca2.draw(0, 1, outfile="/home/marianna/Documents/Phacelia/Figures/
 ↪PCAgeodis_01_20221121.pdf")
pca2.draw(0, 1, outfile="/home/marianna/Documents/Phacelia/Phac_Analyses/
 ↪Phacgeodis_PCA_01_20221121.svg")

pca2.draw(0, 2, outfile="/home/marianna/Documents/Phacelia/Figures/
 ↪PCAgeodis_02_20221121.pdf")
pca2.draw(0, 2, outfile="/home/marianna/Documents/Phacelia/Phac_Analyses/
 ↪Phacgeodis_PCA_02_20221121.svg")
```

[19]:
```
(<toyplot.canvas.Canvas at 0x7ff6b7f5e2f0>,
 <toyplot.coordinates.Cartesian at 0x7ff6b7f880d0>)
```

[20]:
```
# plotting all the PCA plots together
# get canvas object and set size
c = toyplot.Canvas(width=1000, height=900)

## built & dissect canvas into multiple cartesian areas
```

```
##    y1
## x1    x2
##    y2
a0 = c.cartesian(bounds=('11%','38%', '6%',  '35%'))## x1, x2, y1, y2
a1 = c.cartesian(bounds=('59%', '85%', '6%', '35%'))
a2 = c.cartesian(bounds=('11%','38%', '50%', '80%'))
a3 = c.cartesian(bounds=('59%', '85%', '50%', '80%'))


pca.draw(0, 1, axes= a0);
pca.draw(0, 2, axes= a1);
pca2.draw(0, 1, axes= a2);
pca2.draw(0, 2, axes= a3);
```

[21]:
```
# The outfile end in `.pdf` or `.svg`
import toyplot.pdf
toyplot.pdf.render(c,"/home/marianna/Documents/Phacelia/Figures/
 ↪PCA_all_20221121.pdf")

import toyplot.svg
toyplot.pdf.render(c, "/home/marianna/Documents/Phacelia/Phac_Analyses/
 ↪PCA_all_20221121.svg")
```

Treemix analysis

1. Treemix analysis

[1]:
```
## imports
import ipyrad.analysis as ipa
import toytree
import toyplot
import ipyrad as ip
import ipyparallel as ipp
import numpy as np
```

[2]:
```
##print versions
print('ipyrad', ipa.__version__)
print('toytree', toytree.__version__)
! treemix --version | grep 'TreeMix v. '
```

```
ipyrad 0.9.84
toytree 2.0.5
TreeMix v. 1.13
```

[3]:
```
# After the cluster is running attach to it with ipyparallel
ipyclient = ipp.Client()
print(ip.cluster_info(ipyclient))
```

```
Parallel connection | Cristaria: 8 cores
None
```

[4]:
```
## load the hdf5 data for the STRUCTURE analysis
data = "/home/marianna/Documents/Phacelia/Phac_Assembly/min12_clust90.snps.hdf5"
```

1.0.1 group individuals into populations according to structure analysis

[6]:
```
# group individuals into populations, according to my first structure analysis.
imap ={
    "sin": ["W6368","W6376"],
    "pinn_peru": ["W5599","W6027","W6028","W6037","W6078","W6024","W5637",␣
 ↪"W5636"],
    "nana": ["W6373"],
```

```
    "argent": ["W5610","W6374", "W6080","W6369"],
    "set_arg": ["W6375", "W6031","W6370"],
    "set_ch": ["W5145", "W5612"],
    "set_pe": ["W6001", "W6079","W6029"],

}

## require that 50% of samples have data in each group
minmap = {i: 0.5 for i in imap}
```

1.0.2 Finding the best value for m

initiate a treemix analysis object with some param arguments

[7]:
```
tmx = ipa.treemix(
    data=data,
    imap=imap,
    minmap=minmap,
    seed=123456,
    root="sin",
    m=3,
)
```

```
Samples: 23
Sites before filtering: 19359
Filtered (indels): 1020
Filtered (bi-allel): 210
Filtered (mincov): 0
Filtered (minmap): 5685
Filtered (subsample invariant): 6133
Filtered (minor allele frequency): 0
Filtered (combined): 10771
Sites after filtering: 8587
Sites containing missing values: 5548 (64.61%)
Missing values in SNP matrix: 13587 (6.88%)
SNPs (total): 8587
SNPs (unlinked): 4200
subsampled 4200 unlinked SNPs
```

[8]:
```
# print the command string that will be called and run it
print(tmx.command)
tmx.run()
```

```
/home/marianna/anaconda3/envs/treemix/bin/treemix -i
/home/marianna/Documents/Phacelia/analysis-treemix/test.treemix.in.gz -o
/home/marianna/Documents/Phacelia/analysis-treemix/test -m 3 -seed 123456 -root
sin
```

[9]:
```
# draw the resulting tree
tmx.draw_tree();
```

[10]:
```
# a gridded canvas to plot the tree
canvas = toyplot.Canvas(width=600, height=700)
## dissect canvas into multiple cartesian areas (x1, x2, y1, y2)
ax0 = canvas.cartesian(bounds=('2%',  '97%', '5%',  '97.5%'))

# select a plot grid axis and add tree to axes

tmx.draw_tree(axes=ax0);
```

[11]:
```
  def draw_tree(self, axes=None):
        """
        Returns a treemix plot on a toyplot.axes object.
        """
        # create a toytree object from the treemix tree result
        tre = toytree.tree(newick=self.results.tree)

        # draw on axes or create new ones
        if axes:
            canvas = None
            tre.draw(
                axes=axes,
                use_edge_lengths=True,
                tip_labels_align=True,
                edge_align_style={"stroke-width": 1},
                scalebar=True,
            )
        else:
            canvas, axes, mark = tre.draw(
                axes=axes,
                use_edge_lengths=True,
                tip_labels_align=True,
                edge_align_style={"stroke-width": 1},
                scalebar=True,
            )
```

[12]:
```
import toyplot.pdf
toyplot.pdf.render(canvas, "/home/marianna/Documents/Phacelia/Figures/
 ↪Treemix_m3_noiteration__20221130.pdf");
```

[13]:
```
# draw the covariance matrix
tmx.draw_cov();
```

[15]:
```
import toyplot.pdf
```

```
toyplot.pdf.render(canvas, "/home/marianna/Documents/Phacelia/Figures/
↪Treemix_covariancematrix__20221130.pdf");
```

[16]:
```
tests = {}
nadmix = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

# iterate over n admixture edges and store results in a dictionary
for adm in nadmix:
    tmx.params.m = adm
    tmx.run()
    tests[adm] = tmx.results.llik
```

[17]:
```
# plot the likelihood for different values of m
toyplot.plot(
    nadmix,
    [tests[i] for i in nadmix],
    width=350,
    height=275,
    stroke_width=3,
    xlabel="n admixture edges",
    ylabel="ln(likelihood)",
);
```

1.0.3 Iterate over different subsamples of SNPs

[19]:
```
# a gridded canvas to plot trees on
canvas = toyplot.Canvas(width=600, height=700)

# iterate over multiple set of SNPs
for i in range(9):

    # init a treemix analysis object with a random (no) seed
    tmx = ipa.treemix(
        data=data,
        imap=imap,
        minmap=minmap,
        root="sin",
        global_=True,
        m=2,
        quiet=True
    )

    # run model fit
    tmx.run()

    # select a plot grid axis and add tree to axes
    axes = canvas.cartesian(grid=(3, 3, i))
```

```
    tmx.draw_tree(axes);
```

[21]:
```
import toyplot.pdf
toyplot.pdf.render(canvas, "/home/marianna/Documents/Phacelia/Figures/
 ↪Treemix_m2__20221121.pdf");
```

[20]:
```
# a gridded canvas to plot trees on
canvas = toyplot.Canvas(width=600, height=700)

# iterate over multiple set of SNPs
for i in range(9):

    # init a treemix analysis object with a random (no) seed
    tmx = ipa.treemix(
        data=data,
        imap=imap,
        minmap=minmap,
        root="sin",
        global_=True,
        m=3,
        quiet=True
    )

    # run model fit
    tmx.run()

    # create a grid axis and add tree to axes
    axes = canvas.cartesian(grid=(3, 3, i))
    tmx.draw_tree(axes)
```

[22]:
```
import toyplot.pdf
toyplot.pdf.render(canvas, "/home/marianna/Documents/Phacelia/Figures/
 ↪Treemix_m3__20221130.pdf");
```

Appendix C

—Supplementary material of chapter 5—

Table S1. µCT-scanner settings

Table S1. µCT-scanner settings

Taxon, developmental stage	Reference no.	Source voltage (kV)	Source current (µA)	Exposure time (ms) [Frame averaging]	Rotation step	Pictures per sample	Camera binning	Filter	Rotation	Isotropic voxel size (µm)
Bourreria succulenta:										
Anthetic flower	A0023	50	200	305 [4]	0.200°	1800	2 × 2	n.f.[2]	360°	6.5
Mature fruit	A0223	60	166	1500 [5]	0.390°	923	2 × 2	Al 0.25 mm	360°	7.0
Young fruit	A0224	45	165	1000 [6]	0.190°	1894	1 × 1	n.f.	360°	2.5
Anthetic flower	A0225	45	165	1700 [5]	0.190°	1894	1 × 1	n.f.	360°	4.8
Pre-anthetic flower	A0226	45	165	1700 [5]	0.190°	1894	1 × 1	n.f.	360°	2.2
Pre-anthetic flower	A0227	45	165	3500 [5]	0.190°	989	1 × 1	n.f.	180°	1.5
Inflorescence tip	A0228	45	165	2400 [4]	0.190°	1894	1 × 1	n.f.	360°	1.5
Ehretia dicksonii:										
Mature fruit	A0216	65	153	1400 [6]	0.190°	1894	1 × 1	Al 0.25 mm	360°	4.5
Young fruit	A0217	30	200	1500 [5]	0.600°	600	1 × 1	n.f.	360°	3.0
Anthetic flower	A0218	30	200	1600 [5]	0.400°	900	1 × 1	n.f.	360°	3.0
Pre-anthetic flower	A0219	30	200	1800 [6]	0.400°	900	1 × 1	n.f.	360°	2.8
Pre-anthetic flower	A0220	30	200	4800 [4]	0.500°	720	1 × 1	n.f.	360°	1.4
Inflorescence tip	A0221	30	200	5000 [4]	0.500°	720	1 × 1	n.f.	360°	1.4
Pholisma arenarium:										
Anthetic flower	A0211	45	165	1400 [6]	0.190°	1894	1 × 1	n.f.	360°	3.0
Inflorescence tip	A0212	60	166	2000 [4]	0.190°	1894	1 × 1	Al 0.25 mm	360°	1.8
Pre-anthetic flower	A0214	60	166	2000 [4]	0.190°	1894	1 × 1	Al 0.25 mm	360°	1.8
Inflorescence tip	A0215	45	165	3600 [5]	0.190°	989	1 × 1	n.f.	180°	1.0
Mature fruit	A0222	45	165	1300 [5]	0.190°	1894	2 × 2	n.f.	360°	4.4
Young fruit	A0237	45	165	1200 [5]	0.190°	1894	1 × 1	n.f.	180°	1.8
Young fruit	A0238	60	166	700 [5]	0.190°	1894	1 × 1	Al 0.25 mm	360°	3.0
Mature fruit	A0239	60	166	600 [6]	0.190°	1894	1 × 1	Al 0.25 mm	360°	3.0
Entire plant[1]	s.n.	52	153	500 [8]	0.190°	1894	1 × 1	n.f.	360°	16.0

[1] scanned in a Skyscan 1173 system, [2] n.f. = no filter

Appendix D

—Supplementary material of chapter 6—

Table S1. Species included in this study and voucher information.

Supplementary material S1. A. Genes recovery statistics produced with the hybpiper_stats.py script available with HybPiper; B. Heatmap visualizing percentage length recovery for each gene, relative to the mean of the target-file references. The heatmap was produced with the gene_recovery_heatmap.py script available with HybPiper.

Supplementary material S2. Heatmap visualizing the paralog distribution according to the paralog report table, produced with the *paralog_retriever.py* script available with HybPiper.

Supplementary material S3. A. TOAST R Script for generating a partition block and assessing overall missing data patterns across the alignments. B. Bar plot produced with the TOAST R Script showing the missing data patterns in samples with more than 50% missing data.

Figure S1. Phylogenetic relationships among the Boraginales. Concatenation analysis with IQTREE using the retrieved exonic regions of 353 nuclear loci from the Angiosperms353 probe set. Numerical values are ultrafast bootstrap approximation (UFBoot).

Figure S2. Phylogenetic relationships among the Boraginales. ASTRAL-III species tree generated using the retrieved supercontigs of 353 nuclear loci from the Angiosperms353 probe set. Numerical values are local posterior probabilities (LPP).

Figure S3. Phylogenetic relationships among the Boraginales. Concatenation analysis with IQTREE using the retrieved supercontigs of 353 nuclear loci from the Angiosperms353 probe set. Numerical values are ultrafast bootstrap approximation (UFBoot).

Figure S4. Phylogenetic relationships among the Boraginales. ASTRAL-III species tree generated using the retrieved exonic regions of 353 nuclear loci from the Angiosperms353 probe set, but with 26 potentially paralogous genes removed. Numerical values are local posterior probabilities (LPP).

Figure S5. Phylogenetic relationships among the Boraginales. Concatenation analysis with IQTREE using the retrieved exonic regions of 353 nuclear loci from the Angiosperms353 probe set, but with 26 potentially paralogous genes removed. Numerical values are ultrafast bootstrap approximation (UFBoot).

Table S1. Species included in this study and voucher information.

Order	Family	Species	Voucher	DNA Nr.	Accession Nr.	Project / Reference
Boraginales	Boraginaceae	*Aegonychon purpurocaeruleum* (L.) Holub	Chase M.W. 6055 (K)		ERR7621358	PAFTOL
Boraginales	Boraginaceae	*Alkanna orientalis* (L.) Boiss.	Davis 42140 (K)		ERR7621508	PAFTOL
Boraginales	Boraginaceae	*Amsinckia grandiflora* (A.Gray) Kleeb ex Greene	Hong Ma HM1610 (FUS)		SRR11994239	Zhang et al., 2020
Boraginales	Boraginaceae	*Anchusa capensis* Thunb.	N/A		SRR7027846	PRJNA415507 Anchusa capensis
Boraginales	Boraginaceae	*Anchusa officinalis* L.	Hong Ma L18 (FUS)		SRR11922382	Zhang et al., 2020
Boraginales	Boraginaceae	*Antiotrema dunnianum* (Diels) Hand.-Mazz.	s.n.; 36031 (K)		ERR7621507	PAFTOL
Boraginales	Boraginaceae	*Arnebia benthamii* (Wall. ex G.Don) I.M.Johnst.	Chase M.W. 34887 (K)		ERR7621355	PAFTOL
Boraginales	Boraginaceae	*Asperugo procumbens* L.	Fayvush G., Vitek E. 09-0644 (K); K001393222		ERR7621510	PAFTOL
Boraginales	Boraginaceae	*Borago officinalis* L.	Chase M.W.. 2746 (K)		ERR7621593	PAFTOL
Boraginales	Boraginaceae	*Bothriospermum chinense* Bunge	N/A		SRR6374694	Bioproject PRJNA421868
Boraginales	Boraginaceae	*Brachybotrys paridiformis* Maxim. ex Oliv.	Pimenova E.A. s.n.; 88112 (K)		ERR7621356	PAFTOL
Boraginales	Boraginaceae	*Brunnera macrophylla* (Adams) I.M.Johnst.	s.n.; 1948-24801 (K)		ERR7621357	PAFTOL
Boraginales	Boraginaceae	*Caccinia strigosa* Boiss	Chase M.W.38129.19692 (K)		ERR7621359	PAFTOL
Boraginales	Boraginaceae	*Cerinthe major* L.	Chase M.W. 6056 (K)		ERR7621360	PAFTOL
Boraginales	Boraginaceae	*Chionocharis hookeri* (C.B.Clarke) I.M.Johnst.	Crawford et al. 571 (K)		ERR7621594	PAFTOL
Boraginales	Boraginaceae	*Craniospermum subvillosum* Lehm.	Kovtonyuk N. 741 (NSK)	W6352		
Boraginales	Boraginaceae	*Craniospermum subvillosum* Lehm.	Popov s.n.; 36034 (K)		ERR7621595	PAFTOL
Boraginales	Boraginaceae	*Cryptantha decipiens* (M.E.Jones) A.Heller	SDSU20014		SRR5713365	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Cynoglossum amplifolium* Hochst. ex A.DC.	Bidgood S., Keeley P. 288 (K); K001275673		ERR7621520	PAFTOL
Boraginales	Boraginaceae	*Cynoglottis barrelieri* (All.) Vural & Kit Tan	Mayer E., Wraber T. 39308 (K); K001393221		ERR7621509	PAFTOL
Boraginales	Boraginaceae	*Cystostemon heliocharis* (S.Moore) A.G.Mill. & Riedl	Wood J.R.I. 1606 (K); K001275666		ERR7621514	PAFTOL
Boraginales	Boraginaceae	*Decalepidanthus echioides* (Royle ex Benth.) Dickoré & Hilger	Lyon S.B. 8014 (K); K001275678		ERR7621522	PAFTOL
Boraginales	Boraginaceae	*Echium plantagineum* L.	Capelo J. et al., 96329 (LISE)		SRR16690175	NCBI Organelle RefSeqs
Boraginales	Boraginaceae	*Echium wildpretii* H.Pearson ex Hook.f.	N/A		SRR7866831	PRJNA491458
Boraginales	Boraginaceae	*Embadium johnstonii* Ising	AD 251469		ERR7599752	Genomics for Australian Plants (GAP)
Boraginales	Boraginaceae	*Eremocarya lepida* (A.Gray) Greene	SDSU19533		SRR5713420	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Gastrocotyle hispida* (Forssk.) Bunge	Abulaila K., Tehabshem Z., van Slageren M. 2009JOR16-2 (K); K001393223		ERR7621511	PAFTOL
Boraginales	Boraginaceae	*Greeneocharis circumscissa* (Hook. & Arn.) Rydb.	SDSU20663		SRR5713374	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Harpagonella arizonica* (I.M.Johnst.) Guilliams & B.G.Baldwin	ARIZ388168		SRR5713372	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Ivanjohnstonia jaunsariensis* Kazmi	Gamble J.S. 27383 (K); K000998422		ERR7621592	PAFTOL
Boraginales	Boraginaceae	*Johnstonella angustifolia* (Torr.) Hasenstab & M.G.Simpson	RSA731212		SRR5713371	Biorproject PRJNA388354 *Amsinckiinae*

Order	Family	Species	Voucher	DNA Nr.	Accession Nr.	Project / Reference
Boraginales	Boraginaceae	*Johnstonella holoptera* (A.Gray) Hasenstab & M.G.Simpson	SDSU13036		SRR5713342	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Lappula squarrosa* (Retz.) Dumort	Chase M.W. 6549 (K)		ERR7621600	PAFTOL
Boraginales	Boraginaceae	*Lasiocaryum munroi* (C.B.Clarke) I.M.Johnst.	Cutting & Vernay 81 (K)		ERR7621601	PAFTOL
Boraginales	Boraginaceae	*Lepechiniella sarawschanica* (Lipsky) Popov	Landrein S. 179 (K); K000493434		ERR7621516	PAFTOL
Boraginales	Boraginaceae	*Lindelofia anchusoides (Lindl.) Lehm.*	Chase M.W.38131.19907 (K)		ERR7621362	PAFTOL
Boraginales	Boraginaceae	*Lobostemon fruticosus* (L.) H.Buek	s.n.; 2002-438 (K)		ERR7621363	PAFTOL
Boraginales	Boraginaceae	*Mattiastrum formosum* Rech.f. & Riedl	Podlech D. 18096 (K); K001275695		ERR7621534	PAFTOL
Boraginales	Boraginaceae	*Melanortocarya obtusifolia* (Willd.) Selvi, Bigazzi, Hilger & Papini	(K); K001275692		ERR7621531	PAFTOL
Boraginales	Boraginaceae	*Mertensia paniculata* (Aiton) G.Don	132747 (ALTA)		DKFZ; ERS1829588	OneKP
Boraginales	Boraginaceae	*Microcaryum pygmaeum* I.M.Johnst.	Smith & Cave 1837 (K)		ERR7621602	PAFTOL
Boraginales	Boraginaceae	*Moritzia ciliata* (Cham.) DC. ex Meisn.	Rambo B. 44040 (K); K001275687		ERR7621528	PAFTOL
Boraginales	Boraginaceae	*Myosotis arvensis* (L.) Hill	Chase M.W. 6057 (K)		ERR7621603	PAFTOL
Boraginales	Boraginaceae	*Myosotis rehsteineri* (Hausm.) Wartm. ex Reut.	BONN577	W6354		
Boraginales	Boraginaceae	*Myosotis scorpioides* L.	BONN2988	W6353		
Boraginales	Boraginaceae	*Neatostema apulum* (L.) I.M.Johnst.	Tengwall T.A. 491 (K); K001275688		ERR7621529	PAFTOL
Boraginales	Boraginaceae	*Nonea persica* Boiss.	Hamid F.K.H., Jasi M. 41061 (K); K001275669		ERR7621518	PAFTOL
Boraginales	Boraginaceae	*Omphalodes nitida* (Willd.) Hoffmanns. & Link	BONN3768	W6357		
Boraginales	Boraginaceae	*Omphalolappula concava* (F.Muell.) Brand	AD 251455		ERR7599751	Genomics for Australian Plants (GAP)
Boraginales	Boraginaceae	*Oncaglossum pringlei* (Greenm.) Sutorý	(K); K001275690		ERR7621530	PAFTOL
Boraginales	Boraginaceae	*Onosma popoviana* (Riedl) L.Cecchi & Hilger	Rawi A., Serhang I. 23818 (K); K000450574		ERR7621591	PAFTOL
Boraginales	Boraginaceae	*Oreocarya flavoculata* A.Nelson	SDSU20030		SRR5713340	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Oreocarya paradoxa* A.Nelson	Broome C.R. 8497 (K); K001275683		ERR7621527	PAFTOL
Boraginales	Boraginaceae	*Oreocarya virgata* (Porter) Greene	Chase M.W. 6092 (K)		ERR7621361	PAFTOL
Boraginales	Boraginaceae	*Paracaryum platycalyx* Riedl	Rechinger K.H. 57545 (K); K001275663		ERR7621512	PAFTOL
Boraginales	Boraginaceae	*Paramoltkia doerfleri* (Wettst.) Greuter & Burdet	Chase M.W. 38130.19603 (K)		ERR7621364	PAFTOL
Boraginales	Boraginaceae	*Pardoglossum cheirifolium* (L.) E.Barbier & Mathez	Chase M.W. 6065 (K)		ERR7621605	PAFTOL
Boraginales	Boraginaceae	*Pectocarya linearis* (Ruiz & Pav.) DC.	SBBG132392		SRR5713356	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Pectocarya pusilla* (A.DC.) A.Gray	SBBG132394		SRR5713354	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Pentaglottis sempervirens* (L.) Tausch ex L.H.Bailey	Chase M.W. 6058 (K)		ERR7621606	PAFTOL
Boraginales	Boraginaceae	*Plagiobothrys austinae* (Greene) I.M.Johnst.	SBBG132398		SRR5713348	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Plagiobothrys myosotoides* (Lehm.) Brand	SBBG132405		SRR5713399	Biorproject PRJNA388354 *Amsinckiinae*

Order	Family	Species	Voucher	DNA Nr.	Accession Nr.	Project / Reference
Boraginales	Boraginaceae	*Plagiobothrys plurisepaleus (F.Muell.) I.M.Johnst.*	AD 99827150		ERR7599710	Genomics for Australian Plants (GAP)
Boraginales	Boraginaceae	*Plagiobothrys undulatus* (Piper) I.M.Johnst.	SBBG132409		SRR5713408	Biorproject PRJNA388354 *Amsinckiinae*
Boraginales	Boraginaceae	*Pulmonaria rubra* Schott	s.n.; 1994-1200 (K)		ERR7621365	PAFTOL
Boraginales	Boraginaceae	*Rochelia sessiliflora* (Boiss.) Khoshsokhan & Kaz.Osaloo	Hong Ma L55 (FUS)		SRR12034789	Zhang et al., 2020
Boraginales	Boraginaceae	*Selkirkia berteroi* (Colla) Hemsl.	Solrig O.T. 3807 (K)		ERR7621607	PAFTOL
Boraginales	Boraginaceae	*Solenanthus circinnatus* Ledeb.	Rechinger K.H. 53423 (K); K001275671		ERR7621519	PAFTOL
Boraginales	Boraginaceae	*Stephanocaryum dschagastanicum* (Regel) O.D.Nikif.	Botschantzev V., Vvedensky A. 6422 (K); K001275694		ERR7621533	PAFTOL
Boraginales	Boraginaceae	*Trichodesma hildebrandtii* Gürke	Hong Ma L57 (FUS)		SRR12034787	Zhang et al., 2020
Boraginales	Boraginaceae	*Trichodesma indicum* (L.) Sm.	Neha Ravindra Gupte 20	W6362		
Boraginales	Boraginaceae	*Trichodesma scottii* Balf.f.	Chase M.W. 2912 (K)		ERR7621609	PAFTOL
Boraginales	Boraginaceae	*Trigonotis peduncularis* (Trevir.) Benth. ex F.B.Forbes & Hemsl.	N/A		SRR6374711	Bioproject PRJNA421868
Boraginales	Boraginaceae	*Adelinia grandis* (Douglas ex Lehm.) J.I.Cohen	M. Schwarzländer KW_891	W2834		
Boraginales	Boraginaceae	*Adelocaryum coelestinum* (Lindl.) Brand	Neha Ravindra Gupte 18	W4990		
Boraginales	Boraginaceae	*Afrotysonia glochidiata* (R.R.Mill) R.Mill	D.G.A.Styles 5700	W5448		
Boraginales	Boraginaceae	*Alkanna matthioli* Tausch synonym of *Alkanna tinctoria* Tausch subsp. *tinctoria*	BONN2664	W6197		
Boraginales	Boraginaceae	*Amblynotus rupestris* (Georgi) Popov	Hilger H.H. 1597 (BSB)	W2294		
Boraginales	Boraginaceae	*Amsinckia spectabilis* Fisch. & C.A.Mey.	BONN1266	W6195		
Boraginales	Boraginaceae	*Anoplocaryum compressum* Ledeb.	Hilger H.H. 1607 (BSB)	W2295		
Boraginales	Boraginaceae	*Arnebia decumbens* (Vent.) Coss. & Kralik	Korolyuk A.Yu. & Korolyuk E. A. (NSK)	W4056		
Boraginales	Boraginaceae	*Bothriospermum zeylanicum* (J.Jacq.) Druce	Joßberger s.n. (BONN)	W4307		
Boraginales	Boraginaceae	*Buglossoides arvensis* subsp. *arvensis* (L.) I.M.Johnst.	Weigend M. & K. 2000 (B, NY, M)	W1602		
Boraginales	Boraginaceae	*Cerinthe major* L.	BONN1485	W4981		
Boraginales	Boraginaceae	*Craniospermum echioides* (Schrenk) Bunge synonym of *C. subvillosum Lehm.*	A.Yu. Korolyuk s.n. (NSK)	W4043		
Boraginales	Boraginaceae	*Cryptantha clevelandii* Greene	BONN3205	W6209		
Boraginales	Boraginaceae	*Cryptantha micrantha* (Torr.) I.M.Johnst. synonym of *Eremocarya micrantha* (Torr.) Greene	Weigend M. 9058 (BSB)	W2690		
Boraginales	Boraginaceae	*Cryptantha micromeres* (A.Gray) Greene synonym of *Johnstonella micromeres* (A.Gray) Hasenstab & M.G.Simpson	BONN3207	W6210		
Boraginales	Boraginaceae	*Cynoglossum australe* R.Br.	Joßberger s.n. (BONN)	W4272		
Boraginales	Boraginaceae	*Cynoglossum viridiflorum* Lehm.	Rubtsov N. J. (AA)	W6588		
Boraginales	Boraginaceae	*Dasynotus daubenmirei I.M.Johnst.*	M. Schwarzländer KW_889	W2833		
Boraginales	Boraginaceae	*Decalepidanthus primuloides* (Decne.) Dickoré & Hilger	19751894 (E)	W2702		
Boraginales	Boraginaceae	*Echiochilon fruticosum* Desf.	Kagiampaki s.n. (BSB)	W2542		
Boraginales	Boraginaceae	*Echiochilon jugatum* I.M.Johnst.	Gallagher 76965 (K)		ERR7621596	PAFTOL

Order	Family	Species	Voucher	DNA Nr.	Accession Nr.	Project / Reference
Boraginales	Boraginaceae	*Echium vulgare* L.	Mohr O. 597 (BSB)	W2063		
Boraginales	Boraginaceae	*Eritrichium nanum* (L.) Gaudin	Zippel et al. s.n. (B)	W0588		
Boraginales	Boraginaceae	*Glandora oleifolia* (Lapeyr.) D.C.Thomas	Hilger s.n. (BSB)	W1686		
Boraginales	Boraginaceae	*Hackelia deflexa* (Wahlenb.) Opiz	Frey W. s.n. (B)	W0480		
Boraginales	Boraginaceae	*Halacsya sendtneri* (Boiss.) Dörfl.	Cecchi L., Coppi A. & Selvi F. s.n. (BONN)	W4983		
Boraginales	Boraginaceae	*Harpagonella palmeri* A.Gray	Weigend M. 9056 (BSB)	W2691		
Boraginales	Boraginaceae	*Heterocaryum szovitsianum* (Fisch. & C.A.Mey.) A.DC.	Selvi & Bigazzi 28.5.00 (FI)	W0675		
Boraginales	Boraginaceae	*Hormuzakia negevensis* (Danin) Danin & Hilger	Tomer Faraj s.n. BONN3202	W6208		
Boraginales	Boraginaceae	*Huynhia pulchra* (Willd. ex Roem. & Schult.) Greuter & Burdet	Mohr O. 587 (BSB)	W2064		
Boraginales	Boraginaceae	*Lappula consanguinea* (Fisch. & C.A.Mey.) Gürke synonym of *L. squarrosa* (Retz.) Dumort.	Joßberger s.n. (BONN)	W6185		
Boraginales	Boraginaceae	*Lasiocaryum trichocarpum* (Hand.-Mazz.) I.M.Johnst.	Sinclair & Long 5405 (E)	W2603		
Boraginales	Boraginaceae	*Lindelofia longiflora* (DC.) Baill.	BONN2379	W5355		
Boraginales	Boraginaceae	*Lithospermum officinale* L.	Werres A. & Ristow M. s.n. (BSB)	W0409		
Boraginales	Boraginaceae	*Mattiastrum lithospermifolium* (Lam.) Brand	Hilger H. H. s.n. (BSB)	W1255		
Boraginales	Boraginaceae	*Memoremea scorpioides* (Haenke) A.Otero, Jim.Mejías, Valcárcel & P.Vargas	Weigend M. 9039 (BSB)	W2620		
Boraginales	Boraginaceae	*Mertensia ciliata* (E.James ex Torr.) G.Don	Ackermann M. 709 (BSB)	W2728		
Boraginales	Boraginaceae	*Mertensia lanceolata* (Pursh) DC.	Weigend M. 9171 (BSB)	W2905		
Boraginales	Boraginaceae	*Microcaryum pygmaeum* I.M.Johnst.	Boufford et al. 36208 (B, HUH)	W2970		
Boraginales	Boraginaceae	*Microparacaryum intermedium* (Fresen.) Hilger & Podlech	Kürschner H. 10-52	W2981		
Boraginales	Boraginaceae	*Moltkia aurea* Boiss.	Bigazzi M. & Selvi F. 02.14 (BSB)	W2089		
Boraginales	Boraginaceae	*Moltkiopsis ciliata* (Forssk.) I.M.Johnst.	Joßberger TJ-321 (BONN)	W4959		
Boraginales	Boraginaceae	*Moritzia dasyantha* Fresen. synonym of *M. ciliata (Cham.) DC. ex Meisn.*	Trevisan R. s.n. (BONN)	W4759		
Boraginales	Boraginaceae	*Moritzia lindenii* (A. DC) Benth. Ex Guerke	Weigend M.&K., Hennung T. & Schneider Ch. 01/436 (USM)	W0856		
Boraginales	Boraginaceae	*Myosotidium hortensia* (Decne.) Baill.	Weigend M. 9068 (BSB)	W2623		
Boraginales	Boraginaceae	*Nonea lutea* (Desr.) DC.	Ackermann M. 702 (BSB)	W2729		
Boraginales	Boraginaceae	*Ogastemma pusillum* (Coss. & Durieu ex Bonnet & Barratte) Brummitt	Joßberger TJH-300 (BONN)	W4634.2		
Boraginales	Boraginaceae	*Ogastemma pusillum* (Coss. & Durieu ex Bonnet & Barratte) Brummitt	K001275668		ERR7621517	PAFTOL
Boraginales	Boraginaceae	*Onosma alborosea* Fisch. & C.A.Mey.	Bot. Garten 19(7)4-16-74-80 (B)	W2664		
Boraginales	Boraginaceae	*Oreocarya humilis* (A.Gray) Greene	45916 BG Bonn	W6211		
Boraginales	Boraginaceae	*Paracaryum rugulosum* (DC.) Boiss.	N/A	W1320		
Boraginales	Boraginaceae	*Paracynoglossum asperrimum* (Nakai) Popov	Mohr O. 615 (BSB)	W2115		
Boraginales	Boraginaceae	*Pentaglottis sempervirens* (L.) Tausch ex L.H.Bailey	Weigend M. 9065 (BSB)	W2626		

Order	Family	Species	Voucher	DNA Nr.	Accession Nr.	Project / Reference
Boraginales	Boraginaceae	*Plagiobothrys chorisianus* var. *hickmanii* (Greene) I.M.Johnst.	BONN3219	W6206		
Boraginales	Boraginaceae	*Pontechium maculatum* (L.) Böhle & Hilger	Joßberger s.n. (BONN)	W4950		
Boraginales	Boraginaceae	*Pulmonaria angustifolia* L.	Ackermann M. 1031 (BONN)	W4082		
Boraginales	Boraginaceae	*Rindera lanata* (Lam.) Bunge	Salmaki Y. & Zarre Sh. 47363 (TUH)	W5312		
Boraginales	Boraginaceae	*Rochelia cancellata* Boiss. & Balansa	Bigazzi & Selvi 02.53 (FI)	W1301		
Boraginales	Boraginaceae	*Rochelia disperma* (L.f.) K.Koch	Yair Ur, BONN3209	W6207		
Boraginales	Boraginaceae	*Solenanthus apenninus* (L.) Fisch. & C.A.Mey.	W. Frey s.n. (BSB)	W0654		
Boraginales	Boraginaceae	*Symphytum tuberosum* L.	Hilger H.H. s.n.	W0629		
Boraginales	Boraginaceae	*Thaumatocaryon tetraquetrum* (Cham.) I.M.Johnst.	Trevisan R. s.n. (BONN)	W4757		
Boraginales	Boraginaceae	*Trachelanthus cerinthoides* (Boiss.) Kunze	Parishani M. R. 14275 (M)	W2335		
Boraginales	Boraginaceae	*Trachystemon orientalis* (L.) D.Don	Hilger H.H. s.n. (BSB)	W0666		
Boraginales	Boraginaceae	*Trigonotis formosana var. elevatovenosa* (Hayata) S.D.Shen & J.C.Wang	Weigend M. 8128 (BONN)	W4273		
Boraginales	Codonaceae	*Codon royenii* L.	Hong Ma L52 (FUS)		SRR12034792	Zhang et al., 2020
Boraginales	Codonaceae	*Codon schenckii* Schinz	Seydel R., 2968 (K)		ERR4180157	PAFTOL
Boraginales	Codonaceae	*Codon schenckii* Schinz	Lintermann P. s.n.; BONN653	W6364		
Boraginales	Coldeniaceae	*Coldenia procumbens* L.	Neha Ravindra Gupte NRG - 12	W6188		
Boraginales	Coldeniaceae	*Coldenia procumbens* L.	Cai F. Zhang 4466 (FUS)		SRR12009669	Zhang et al., 2020
Boraginales	Coldeniaceae	*Coldenia procumbens* L.	MEL 2441170A		ERR7599654	Genomics for Australian Plants (GAP)
Boraginales	Cordiaceae	*Cordia macleodii* (Griff.) Hook.f. & Thomson	Neha Ravindra Gupte NRG - 13	W6189		
Boraginales	Cordiaceae	*Cordia myxa* L.	Leeuwenberg, A.J.M. 7592 (K)		ERR5970511	PAFTOL
Boraginales	Cordiaceae	*Cordia sonorae* Rose	Miller & Campos 2956 (BSB)	W0979		
Boraginales	Cordiaceae	*Varronia cylindristachya* Ruiz & Pav.	Weigend M. et al. MJ14/147; BONN1937	W6216		
Boraginales	Cordiaceae	*Varronia rupicola* (Urb.) Britton	N/A		SRR11934229	Zhang et al., 2020
Boraginales	Ehretiaceae	*Bourreria ovata* Miers synonym of *B. succulenta* Jacq	BONN1298	W6212		
Boraginales	Ehretiaceae	*Bourreria succulenta* Jacq	Hong Ma L71 (FUS)		SRR12034773	Zhang et al., 2020
Boraginales	Ehretiaceae	*Ehretia acuminata* R.Br	Soltis and Miles 2838		EMAL; ERS1829587	OneKP
Boraginales	Ehretiaceae	*Ehretia dicksonii* Hance	s.n. (BONN)	W6367		
Boraginales	Ehretiaceae	*Ehretia microphylla Lam.*	Pullen R. 6809 (K); K000062263		ERR5970513	PAFTOL
Boraginales	Ehretiaceae	*Halgania anagalloides* Endl.	Smith R.J. & Shade A. 2 (K)		ERR7621598	PAFTOL
Boraginales	Ehretiaceae	*Halgania cyanea* Lindl.	Byrne G. 6051; PERTH 08936749	W6567		
Boraginales	Ehretiaceae	*Lepidocordia punctata* Ducke	Jangoux J., Ribeiro B.G.S. 1479 (K); K001275665		ERR7621513	PAFTOL
Boraginales	Ehretiaceae	*Rotula aquatica* Lour. synonym of *Ehretia aquatica* (Lour.) Gottschling & Hilger	Momoh J. 130 (K); K001243255		ERR7621366	PAFTOL
Boraginales	Ehretiaceae	*Tiquilia dichotoma* (Ruiz & Pav.) Pers.	Weigend & Schwarzer 8331 (BONN)	W2280		
Boraginales	Ehretiaceae	*Tiquilia litoralis* (Phil.) A.T.Richardson	Luebert F., Böhnert T., Stoll A. 3272 (BONN)	W5517.2		
Boraginales	Heliotropiaceae	*Euploca calcicola* (Fernald) J.I.M.Melo synonym of *Euploca mexicana* (A.DC.) M.W.Frohl. & M.W.Chase	N/A		XVRU; ERR2040513	OneKP
Boraginales	Heliotropiaceae	*Euploca convolvulacea* Nutt.	N/A		OUER; ERS1829577	OneKP

Order	Family	Species	Voucher	DNA Nr.	Accession Nr.	Project / Reference
Boraginales	Heliotropiaceae	*Euploca filiformis* (Lehm.) J.I.M.Melo & Semir	N/A		OEKO; ERS1829579 & ERS1829578	OneKP
Boraginales	Heliotropiaceae	*Euploca greggii* (Torr.) Halse & Feuillet	N/A		ABEH; ERS1829580	OneKP
Boraginales	Heliotropiaceae	*Euploca humifusa* (Kunth) Diane & Hilger	Gonzalez P. 1306-6 (HAJB)	W4135.1		
Boraginales	Heliotropiaceae	*Euploca karwinskyi* (I.M.Johnst.) J.I.M.Melo	N/A		NIGS; ERS1829581	OneKP
Boraginales	Heliotropiaceae	*Euploca mendocina* (Phil.) Diane & Hilger	N/A		MZOB; ERS1829582	OneKP
Boraginales	Heliotropiaceae	*Euploca racemosa* Rose & Standl.	N/A		IDGE; ERS1829583	OneKP
Boraginales	Heliotropiaceae	*Euploca tenella* (Nutt.) Feuillet & Halse	N/A		DIHD; ERS1829584	OneKP
Boraginales	Heliotropiaceae	*Euploca tenuifolia* (R.Br.) Diane & Hilger	N/A		JWEY; ERS1829574	OneKP
Boraginales	Heliotropiaceae	*Euploca texana* (I.M.Johnst.) M.W.Frohl. & M.W.Chase	N/A		MDJK; ERS1829586 & ERS1829585	OneKP
Boraginales	Heliotropiaceae	*Heliotropium arborescens* L.	Weigend M. s.n. (BONN); BONN822	W6214		
Boraginales	Heliotropiaceae	*Heliotropium curassavicum* var. *curassavicum* L.	Weigend M. 2000/678 (BSB)	W2002		
Boraginales	Heliotropiaceae	*Heliotropium glutinosum* Phil.	Luebert F. 1970 (SGO)	W1969		
Boraginales	Heliotropiaceae	*Heliotropium jaffiuelii I.M.Johnst.*	N/A	W4928		
Boraginales	Heliotropiaceae	*Heliotropium messerschmidioides* Kuntze	Chase M.W. 6025 (K)		ERR7621608	PAFTOL
Boraginales	Heliotropiaceae	*Ixorhea tschudiana* Fenzl	Renvoize S.A., Wilmot-Dear M., Kiesling R. 3407 (K); K001275679		ERR7621523	PAFTOL
Boraginales	Heliotropiaceae	*Ixorhea tschudiana* Fenzl	Schlumpberger B. BOS753 (BONN)	W6215		
Boraginales	Heliotropiaceae	*Myriopus poliochros* (Spreng.) Small	s.n. (BONN)	W4931.1		
Boraginales	Heliotropiaceae	*Myriopus volubilis* (L.) Small	Sanders A.C., Charlton D., Phillips B., Vazquez L. 10813 (K); K001275693		ERR7621532	PAFTOL
Boraginales	Heliotropiaceae	*Nogalia drepanophylla* (Baker) Verdc.	Fison T. 3 (K); K001275667		ERR7621515	PAFTOL
Boraginales	Heliotropiaceae	*Tournefortia bicolor* Sw. synonym of *Heliotropium verdcourtii* Craven	Weigend M. 9916 (BONN)	W6578		
Boraginales	Hydrophyllaceae	*Draperia systyla* (A.Gray) Torr.	BONN1752	W5140.1		
Boraginales	Hydrophyllaceae	*Ellisia nyctelea* (L.) L.	Jones G.N. 24217 (K); K001275681		ERR7621525	PAFTOL
Boraginales	Hydrophyllaceae	*Emmenanthe penduliflora* Benth.	Chase M.W. 6093 (K)		ERR7621597	PAFTOL
Boraginales	Hydrophyllaceae	*Emmenanthe penduliflora* Benth.	BONN2751	W6199		
Boraginales	Hydrophyllaceae	*Eucrypta chrysanthemifolia* (Benth.) Greene	Piehl M.A. 62131 (K); K001275680		ERR7621524	PAFTOL
Boraginales	Hydrophyllaceae	*Eucrypta chrysanthemifolia* (Benth.) Greene	BONN990	W6193		
Boraginales	Hydrophyllaceae	*Hydrophyllum canadense* L.	Chase M.W. 2548 (K)		ERR7621599	PAFTOL
Boraginales	Hydrophyllaceae	*Hydrophyllum virginianum* L.	BONN1820	W5141		
Boraginales	Hydrophyllaceae	*Nemophila menziesii* Hook. & Arn.	Chase M.W. 6550 (K)		ERR7621604	PAFTOL
Boraginales	Hydrophyllaceae	*Phacelia bolanderi* A.Gray	BONN3006	W6203		
Boraginales	Hydrophyllaceae	*Phacelia campanularia* A.Gray	M.K. Deyholos 2016-111		YQIJ; ERS1829589	OneKP
Boraginales	Hydrophyllaceae	*Pholistoma auritum* (Lindl.) Lilja	BONN2734	W6198		
Boraginales	Hydrophyllaceae	*Pholistoma membranaceum* (Benth.) Constance	Halse R.R. 6150 (K); K001275682		ERR7621526	PAFTOL
Boraginales	Hydrophyllaceae	*Romanzoffia californica* Greene	BONN3107	W6202		
Boraginales	Lennoaceae	*Lennoa madreporoides* Lex.	N/A		SMUR; ERS1829590	OneKP
Boraginales	Lennoaceae	*Pholisma arenarium* Nutt.	N/A		HANM; ERS1829591	OneKP
Boraginales	Lennoaceae	*Pholisma arenarium* Nutt.	Lee B. & Grossenbacher L. 1 (BONN)	W6182		

Order	Family	Species	Voucher	DNA Nr.	Accession Nr.	Project / Reference
Boraginales	Namaceae	*Eriodictyon californicum* (Hook. & Arn.) Decne.	2663 (BONN)	W5662.2		
Boraginales	Namaceae	*Eriodictyon crassifolium* Benth	Hong Ma L53 (FUS)		SRR12034791	Zhang et al., 2020
Boraginales	Namaceae	*Nama dichotoma* (Ruiz & Pav.) Choisy	Wood J.R.I. 11873 (K); K000543033		ERR7621521	PAFTOL
Boraginales	Namaceae	*Nama rothrockii* A.Gray	Hilger H.H. & Hofmann M. 1201 (B)	W0243		
Boraginales	Namaceae	*Turricula parryi* (A.Gray) J.F.Macbr.	Spencer M.F. 1577 (K); K001275696		ERR7621535	PAFTOL
Boraginales	Namaceae	*Wigandia ecuadorensis* Cornejo	Joßberger TJH-001 (BONN)	W5653.1		
Boraginales	Namaceae	*Wigandia urens* (Ruiz & Pav.) Kunth	30082 (K)		SRR7451086	PAFTOL
Boraginales	Wellstediaceae	*Wellstedia dinteri* Pilg.	Hilger s.n. (BSB)	W2670		
Boraginales	Wellstediaceae	*Wellstedia somalensis* Thulin & A.Johanss.	Thulin M. et al. 10084 (UPPS)	W3004		
Gentianales	Gentianaceae	*Gentiana cruciata* L.	2001-3694 (K)		ERR5033622	PAFTOL
Gentianales	Rubiaceae	*Alibertia latifolia* (Benth.) K.Schum.	Persson et al. 635 (GB)		ERR7622025	PAFTOL
Solanales	Convolvulaceae	*Ipomoea nil* (L.) Roth	Leon C.J., Lin Y-L, Zhou K -Y 498 (K)		ERR7619226	PAFTOL
Solanales	Solanaceae	*Nicotiana rosulata* (S.Moore) Domin	Chase M.W., 68280 (K)		ERR4180152	PAFTOL

Supplementary material S1. A. Genes recovery statistics produced with the *hybpiper_stats.py* script available with HybPiper;

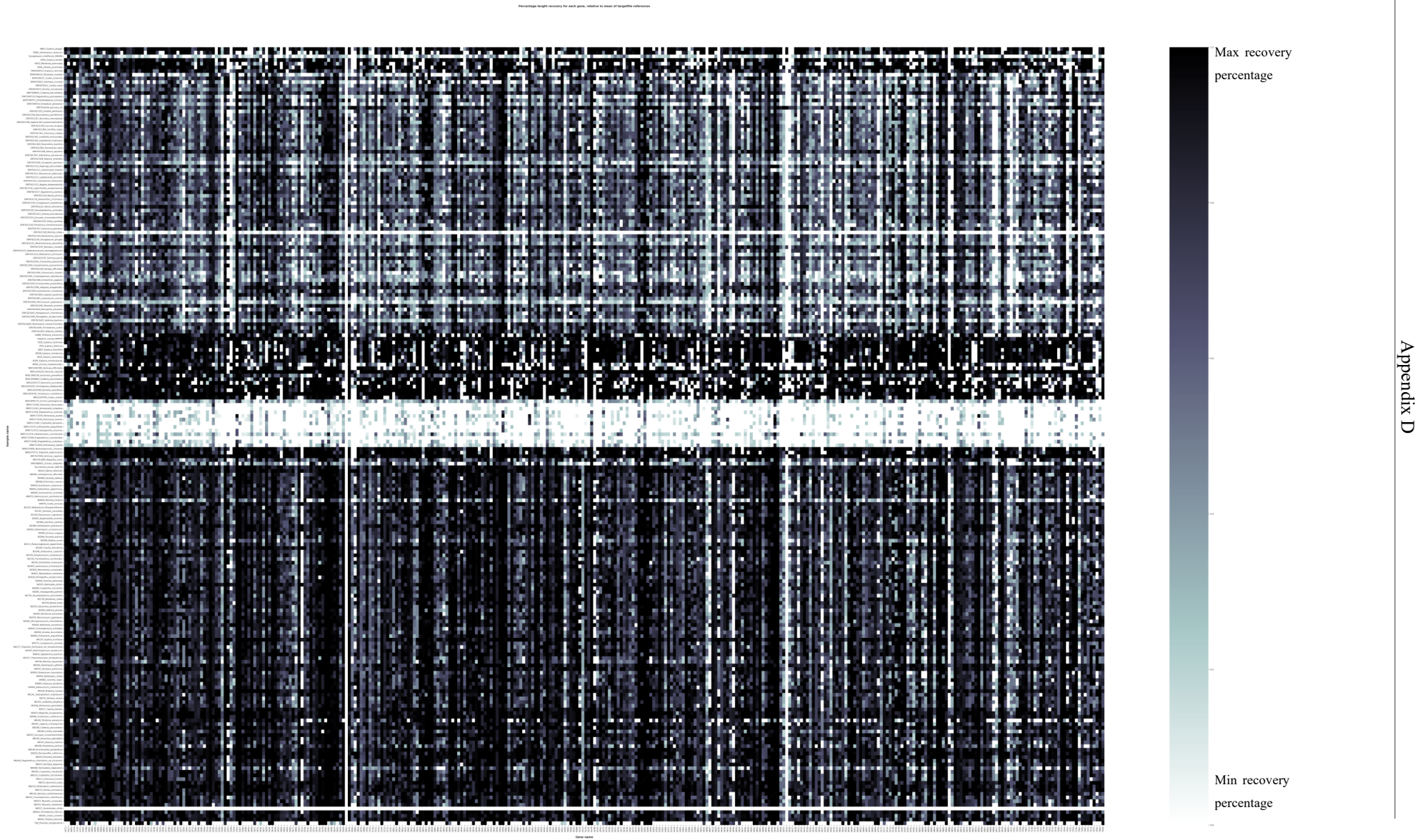
Name	NumReads	ReadsMapped	PctOnTarget	Genes Mapped	GenesWith Contigs	GenesWith Seqs	Genes At25pct	Genes At50pct	Genes At75pct	Genes At150pct	Paralog Warnings Long	Paralog Warnings Depth	GenesWithout StitchedContigs	GenesWith StitchedContigs	GenesWith StitchedContigs Skipped	GenesWith ChimeraWarning
ABEH_Euploca_greggii	24359762	188546	0.8	350	308	308	306	294	275	0	1	1	280	28	0	0
CSWE_Heliotropium_texanum	19741170	141778	0.7	350	263	263	256	239	219	0	16	22	234	29	0	0
DIHD_Euploca_tenella	23205410	213085	0.9	348	313	313	308	298	280	0	21	36	280	33	0	0
DKFZ_Mertensia_paniculata	23157920	78903	0.3	339	260	260	249	231	201	0	2	5	226	34	0	0
EMAL_Ehretia_acuminata	19491394	154362	0.8	347	304	303	301	294	272	0	7	10	264	39	0	0
ERR2040513_Euploca_calcicola	22572978	147505	0.7	349	337	337	330	302	242	1	10	19	212	125	0	0
ERR4180152_Nicotiana_rosulata	111556	37392	33.5	344	223	221	183	125	72	0	18	27	151	70	0	0
ERR4180157_Codon_schenckii	1930516	185980	9.6	349	337	332	315	270	186	0	5	11	133	199	0	0
ERR5033622_Gentiana_cruciata	1212918	61662	5.1	335	225	224	198	136	88	0	4	12	162	62	0	0
ERR5970511_Cordia_myxa	2685368	253587	9.4	351	329	321	299	236	145	0	17	39	156	165	0	0
ERR5970513_Ehretia_microphylla	574676	210064	36.6	349	336	331	301	226	112	0	4	7	112	219	0	0
ERR7599654_Coldenia_procumbens	8529186	2128147	25	350	345	341	340	317	233	0	1	3	105	236	0	0
ERR7599710_Plagiobothrys_plurisepalus	457764	60065	13.1	337	257	251	217	136	67	0	2	3	153	98	0	0
ERR7599751_Omphalolappula_concava	2007726	365826	18.2	349	320	311	297	243	158	0	13	25	166	145	0	0
ERR7599752_Embadium_jobnstonii	473490	70997	15	339	272	262	225	131	63	0	1	3	139	123	0	0
ERR7619226_Ipomoea_nil	3036762	1620750	53.4	351	349	347	340	310	236	0	8	17	128	219	0	0
ERR7621355_Arnebia_benthamii	3194944	164844	5.2	348	312	296	261	182	106	0	4	8	144	152	0	0
ERR7621356_Brachybotrys_paridiformis	2957768	443165	15	349	327	309	291	228	130	0	4	11	133	176	0	0
ERR7621357_Brunnera_macrophylla	2967988	343351	11.6	347	336	323	310	245	165	0	8	24	133	190	0	0
ERR7621358_Aegonychon_purpurocaeruleum	1942806	223053	11.5	348	326	312	289	218	123	0	7	13	132	180	0	0
ERR7621359_Caccinia_strigosa	2988698	533553	17.9	348	340	327	319	269	179	0	11	26	120	207	0	0
ERR7621360_Cerinthe_major	1704144	247762	14.5	347	324	314	297	232	140	0	11	18	136	178	0	0
ERR7621361_Oreocarya_virgata	1734920	280335	16.2	347	330	319	298	227	137	0	3	7	154	165	0	0
ERR7621362_Lindelofia_anchusoides	2572174	382090	14.9	350	337	328	314	244	146	0	10	18	144	184	0	0
ERR7621363_Lobostemon_fruticosus	3458958	458705	13.3	347	333	317	297	238	157	0	6	11	134	184	0	0
ERR7621364_Paramoltkia_doerfleri	2845900	352797	12.4	349	329	317	303	235	145	0	10	17	138	179	0	0
ERR7621365_Pulmonaria_rubra	2024060	254206	12.6	347	329	321	307	234	142	0	10	15	167	154	0	0
ERR7621366_Rotula_aquatica	3147136	862479	27.4	350	350	350	348	326	258	0	7	15	121	229	0	0
ERR7621507_Antiotrema_dunnianum	1849240	280857	15.2	347	329	311	287	198	93	0	7	11	108	203	0	0
ERR7621508_Alkanna_orientalis	1924568	308093	16	350	334	314	295	194	98	0	5	12	111	203	0	0
ERR7621509_Cynoglottis_barrelieri	1717854	201007	11.7	348	294	255	193	86	29	0	3	3	124	131	0	0
ERR7621510_Asperugo_procumbens	4365998	1280404	29.3	350	348	334	330	291	217	0	14	32	130	204	0	0
ERR7621511_Gastrocotyle_hispida	1215420	123649	10.2	343	294	273	237	148	66	0	4	7	128	145	0	0
ERR7621512_Paracaryum_platycalyx	3789036	625275	16.5	351	344	330	319	256	152	0	8	17	116	214	0	0
ERR7621513_Lepidocordia_punctata	2467296	773932	31.4	351	347	341	334	304	217	1	24	103	67	274	0	0
ERR7621514_Cystostemon_heliocharis	2627508	333788	12.7	348	330	315	291	224	122	0	8	17	130	185	0	0
ERR7621515_Nogalia_drepanophylla	3933866	392714	10	351	341	334	325	260	156	0	1	2	96	238	0	0
ERR7621516_Lepechiniella_sarawschanica	2478886	472868	19.1	344	330	316	301	239	135	0	4	11	121	195	0	0
ERR7621517_Ogastemma_pusillum	3809508	1036427	27.2	349	345	337	325	284	203	0	22	59	110	227	0	0
ERR7621518_Nonea_persica	2391452	380219	15.9	349	333	311	291	193	97	0	4	8	112	199	0	0
ERR7621519_Solenanthus_circinnatus	2784242	434957	15.6	351	340	326	316	238	129	0	7	12	123	203	0	0
ERR7621520_Cynoglossum_amplifolium	4541816	666098	14.7	352	339	317	296	214	108	0	5	9	107	210	0	0
ERR7621521_Nama_dichotoma	4008562	1144013	28.5	350	347	343	339	318	229	0	29	82	90	253	0	0
ERR7621522_Decalepidanthus_echioides	1660664	283354	17.1	350	331	312	278	189	91	0	4	7	124	188	0	0
ERR7621523_Ixhorea_tschudianna	1033688	118730	11.5	350	332	326	304	243	153	1	0	3	103	223	0	0
ERR7621524_Eucrypta_chrysanthemifolia	2167932	513240	23.7	349	340	335	327	291	196	0	2	7	108	227	0	0
ERR7621525_Ellisia_nyctelea	1114072	135099	12.1	351	331	325	306	256	141	0	4	6	123	202	0	0
ERR7621526_Pholistoma_membranaceum	2078932	248511	12	346	313	290	275	205	98	0	3	5	132	158	0	0
ERR7621527_Oreocarya_paradoxa	2473790	382136	15.4	350	335	325	302	232	139	0	5	7	139	186	0	0
ERR7621528_Moritzia_ciliata	1257505	6379	0.5	299	76	66	21	4	2	0	0	0	62	4	0	0
ERR7621529_Neatostema_apulum	2118800	198291	9.4	352	332	319	289	225	130	0	13	23	117	202	0	0
ERR7621530_Oncaglossum_pringlei	2962316	221425	7.5	351	318	298	259	162	78	0	3	6	146	152	0	0
ERR7621531_Melanortocarya_obtusifolia	4931006	1316478	26.7	336	304	300	237	123	45	0	3	5	113	187	0	0
ERR7621532_Myriopus_volubilis	4240324	966492	22.8	352	348	342	338	312	243	0	30	86	109	233	0	0
ERR7621533_Stephanocaryum_dschagastanicum	2382534	216626	9.1	351	336	323	299	239	154	0	5	12	108	215	0	0
ERR7621534_Mattiastrum_formosum	2227556	133715	6	351	307	279	236	145	68	0	2	4	151	128	0	0
ERR7621535_Turricula_parryi	1908532	477881	25	351	349	345	334	300	212	0	9	26	76	269	0	0
ERR7621591_Choriantha_popoviana	1413234	184263	13	345	321	305	271	186	83	0	1	5	132	173	0	0
ERR7621592_Ivanjohnstonia_jaunsariensis	2167470	440580	20.3	348	332	319	300	213	115	0	3	9	119	200	0	0
ERR7621593_Borago_officinalis	2004860	214663	10.7	344	317	311	296	234	138	0	17	29	134	177	0	0
ERR7621594_Chionocharis_hookeri	1960644	103140	5.3	351	265	242	205	131	56	0	3	4	137	105	0	0
ERR7621595_Craniospermum_subvillosum	3794537	56269	1.5	329	311	308	242	133	52	0	3	3	138	170	0	0
ERR7621596_Echiochilon_jugatum	1446430	77599	5.4	351	283	264	217	142	64	0	3	3	137	127	0	0
ERR7621597_Emmenanthe_penduliflora	2335738	656252	28.1	350	331	324	305	248	149	0	8	14	112	212	0	0
ERR7621598_Halgania_anagalloides	2569064	462792	18	351	348	347	343	311	236	0	2	12	99	248	0	0
ERR7621599_Hydrophyllum_canadense	2874988	160243	5.6	348	326	316	304	244	143	0	5	10	151	165	0	0
ERR7621600_Lappula_squarrosa	2911808	651290	22.4	345	335	325	314	273	179	1	12	33	133	192	0	0
ERR7621601_Lasiocaryum_munroi	3870579	57169	1.5	329	306	305	251	136	56	0	2	2	126	179	0	0
ERR7621602_Microcaryum_pygmaeum	3286922	553168	16.8	323	245	242	123	37	14	0	1	1	172	70	0	0

Name	NumReads	ReadsMapped	PctOnTarget	Genes Mapped	GenesWith Contigs	GenesWith Seqs	Genes At25pct	Genes At50pct	Genes At75pct	Genes At150pct	Paralog Warnings Long	Paralog Warnings Depth	GenesWithout StitchedContigs	GenesWith StitchedContigs	GenesWith StitchedContigs Skipped	GenesWith ChimeraWarning
ERR7621603_Myosotis_arvensis	3738514	750929	20.1	350	340	330	320	272	194	1	12	32	119	211	0	0
ERR7621604_Nemophila_menziesii	1762486	158624	9	347	304	289	276	199	115	0	5	7	196	93	0	0
ERR7621605_Pardoglossum_cheirifolium	3426136	384609	11.2	352	330	319	309	249	162	0	10	22	135	184	0	0
ERR7621606_Pentaglottis_sempervirens	2068008	278904	13.5	347	337	323	304	234	134	1	8	22	126	197	0	0
ERR7621607_Selkirkia_berteroi	2991853	40120	1.3	329	268	262	188	76	24	0	2	2	140	122	0	0
ERR7621608_Heliotropium_messerschmidioi	2695916	351684	13	350	341	337	332	294	212	0	2	8	121	216	0	0
ERR7621609_Trichodesma_scottii	2804616	399123	14.2	348	335	327	321	266	179	0	12	28	140	187	0	0
ERR7622025_Alibertia_latifolia	2699712	1027092	38	352	349	344	334	292	212	1	2	8	95	249	0	0
HANM_Pholisma_arenarium	14680210	72469	0.5	290	256	253	244	228	207	0	0	1	201	52	0	0
IDGE_Euploca_racemosa	24597622	208469	0.8	348	311	311	305	294	275	0	10	15	262	49	0	0
IPPG_Euploca_filiformis	26554464	215949	0.8	350	322	321	320	300	285	0	2	3	294	27	0	0
JWEY_Euploca_tenuifolia	19869388	146566	0.7	349	297	297	290	278	251	0	5	10	262	35	0	0
MZOB_Euploca_mendocina	24743260	202092	0.8	350	326	325	322	311	290	0	1	1	292	33	0	0
NIGS_Euploca_karwinskyi	22808660	194229	0.9	349	305	305	300	295	274	0	10	18	265	40	0	0
OUER_Euploca_convolvulacea	19563714	151237	0.8	349	315	315	310	297	269	0	11	13	282	33	0	0
SMUR_Lennoa_madreporoides	16464378	52465	0.3	274	227	224	219	197	155	0	1	2	178	46	0	0
SRR11922382_Anchusa_officinalis	23390766	300071	1.3	352	349	258	252	241	232	0	14	19	244	14	0	0
SRR11934229_Varronia_rupicola	81777322	1572611	1.9	352	351	346	346	346	338	1	30	48	305	41	0	0
SRR11994239_Amsinckia_grandiflora	104382914	653859	0.6	352	351	326	314	271	208	0	3	11	93	233	0	0
SRR12009669_Coldenia_procumbens	15020654	202952	1.4	352	342	286	284	273	250	0	3	3	271	15	0	0
SRR12034773_Bourreria_succulenta	21543798	342526	1.6	352	348	262	261	255	246	0	10	10	239	23	0	0
SRR12034787_Trichodesma_hildebrandtii	22829280	379145	1.7	352	352	297	295	290	279	0	35	38	285	12	0	0
SRR12034789_Rochelia_sessiliflora	23755168	369601	1.6	352	348	309	305	301	284	0	16	20	296	13	0	0
SRR12034791_Eriodictyon_crassifolium	24355446	444827	1.8	351	350	321	320	318	314	0	37	38	306	15	0	0
SRR12034792_Codon_royenii	23587888	353807	1.5	352	349	276	275	268	253	0	12	11	260	16	0	0
SRR16690175_Echium_plantagineum	6340101	4406	0.1	320	281	274	178	70	17	0	0	1	147	127	0	0
SRR5713340_Oreocarya_flavoculata	7592837	2430	0	302	63	61	19	3	1	0	0	0	55	6	0	0
SRR5713342_Johnstonella_holoptera	4136144	2007	0	297	53	52	21	4	1	0	0	0	47	5	0	0
SRR5713348_Plagiobothrys_austinae	4229076	3133	0.1	310	89	84	39	7	1	0	0	0	62	22	0	0
SRR5713354_Pectocarya_pusilla	4378288	3954	0.1	318	133	130	62	16	6	0	0	0	102	28	0	0
SRR5713356_Pectocarya_linearis	2741113	2769	0.1	310	84	81	35	8	3	0	0	0	65	16	0	0
SRR5713365_Cryptantha_decipiens	4382871	3137	0.1	312	92	89	43	8	3	0	0	0	68	21	0	0
SRR5713371_Johnstonella_angustifolia	5090873	2181	0	299	42	39	16	3	1	0	0	0	36	3	0	0
SRR5713372_Harpagonella_arizonica	2428255	2010	0.1	294	52	49	18	4	0	0	0	0	47	2	0	0
SRR5713374_Greeneocharis_circumscissa	2952446	3253	0.1	315	204	197	109	31	6	0	0	0	131	66	0	0
SRR5713399_Plagiobothrys_myosotoides	3413877	2293	0.1	303	54	51	18	2	0	0	0	0	43	8	0	0
SRR5713408_Plagiobothrys_undulatus	5443930	3960	0.1	319	118	116	58	16	4	0	0	0	92	24	0	0
SRR5713420_Eremocarya_lepida	4104083	2837	0.1	313	160	153	72	17	4	0	0	0	109	44	0	0
SRR6374694_Bothriospermum_chinense	29719448	258733	0.9	352	340	328	328	322	308	0	17	20	296	32	0	0
SRR6374711_Trigonotis_peduncularis	29031018	285311	1	352	335	325	322	312	283	0	10	14	296	29	0	0
SRR7027846_Anchusa_capensis	59949996	680801	1.1	352	350	333	333	332	323	0	38	44	310	23	0	0
SRR7451086_Wigandia_urens	527102	78970	15	350	317	315	307	255	158	0	4	16	130	185	0	0
SRR7866831_Echium_wildpretii	63125553	295455	0.5	339	265	265	242	155	96	0	1	1	197	68	0	0
W0243_Nama_rothrockii	2478148	1107616	44.7	350	350	350	350	343	317	0	20	48	80	270	0	0
W0409_Lithospermum_officinale	2336068	937344	40.1	350	340	339	333	315	248	0	12	22	159	180	0	0
W0480_Hackelia_deflexa	2236742	898646	40.2	349	340	339	334	316	250	0	17	29	155	184	0	0
W0588_Eritrichium_nanum	1491484	619811	41.6	349	335	333	332	306	238	0	24	43	151	182	0	0
W0629_Symphytum_tuberosum	1438046	583571	40.6	350	342	342	336	308	238	2	19	65	119	223	0	0
W0654_Solenanthus_appenninus	1643936	518788	31.6	351	337	334	332	298	237	0	15	28	135	199	0	0
W0666_Trachystemon_orientale	1590522	594527	37.4	348	342	342	338	310	240	0	11	36	120	222	0	0
W0675_Heterocaryum_szovitsianum	2658350	1171927	44.1	350	343	339	337	313	260	0	24	54	146	193	0	0
W0856_Moritizia_lindenii	150426	66513	44.2	345	278	276	245	169	98	0	10	13	186	90	0	0
W0979_Cordia_sonorae	1855978	696083	37.5	350	348	348	348	337	296	0	2	19	104	244	0	0
W1255_Mattiastrum_lithospermifolium	2100938	744965	35.5	351	339	337	334	311	244	0	13	28	144	193	0	0
W1301_Rochelia_cancellata	4569444	1883793	41.2	350	344	344	341	319	268	0	11	26	179	165	0	0
W1320_Paracaryum_rugulosum	1678338	585171	34.9	350	342	340	339	305	239	0	17	29	142	198	0	0
W1602_Buglossoides_arvensis	1698264	788294	46.4	350	338	338	335	308	238	0	20	57	135	203	0	0
W1686_Glandora_oleifolia	1504010	502338	33.4	349	336	336	334	304	235	1	14	31	162	174	0	0
W1969_Heliotropium_glutinosum	1859870	742703	39.9	350	349	349	348	340	296	0	7	8	112	237	0	0
W2002_Heliotropium_curassavicum	990390	372903	37.7	350	348	348	346	334	272	0	3	3	109	239	0	0
W2063_Echium_vulgare	2662198	1071610	40.3	349	345	341	337	308	251	0	11	33	146	195	0	0
W2064_Huynhia_pulchra	1290164	372769	28.9	349	336	336	331	298	224	1	16	37	127	209	0	0
W2089_Moltkia_aurea	609412	120111	19.7	346	315	313	303	252	162	0	8	18	154	159	0	0
W2115_Paracynoglossum_asperrimum	1597108	575989	36.1	350	342	339	333	301	226	0	13	29	145	194	0	0
W2280_Tiquilia_dichotoma	2664162	1253962	47.1	349	347	346	345	337	307	0	71	166	159	187	0	0
W2294_Amblynotus_rupestris	1428882	413133	28.9	349	334	333	331	303	239	0	11	26	158	175	0	0
W2295_Anoplocaryum_compressum	2234854	716471	32.1	351	344	342	341	310	256	0	15	42	141	201	0	0
W2335_Trachelanthus_cerinthoides	1164350	246771	21.2	348	330	330	325	280	197	0	12	26	140	190	0	0
W2542_Echiochilon_fruticosum	1446738	329705	22.8	349	337	335	331	291	214	0	13	29	125	210	0	0

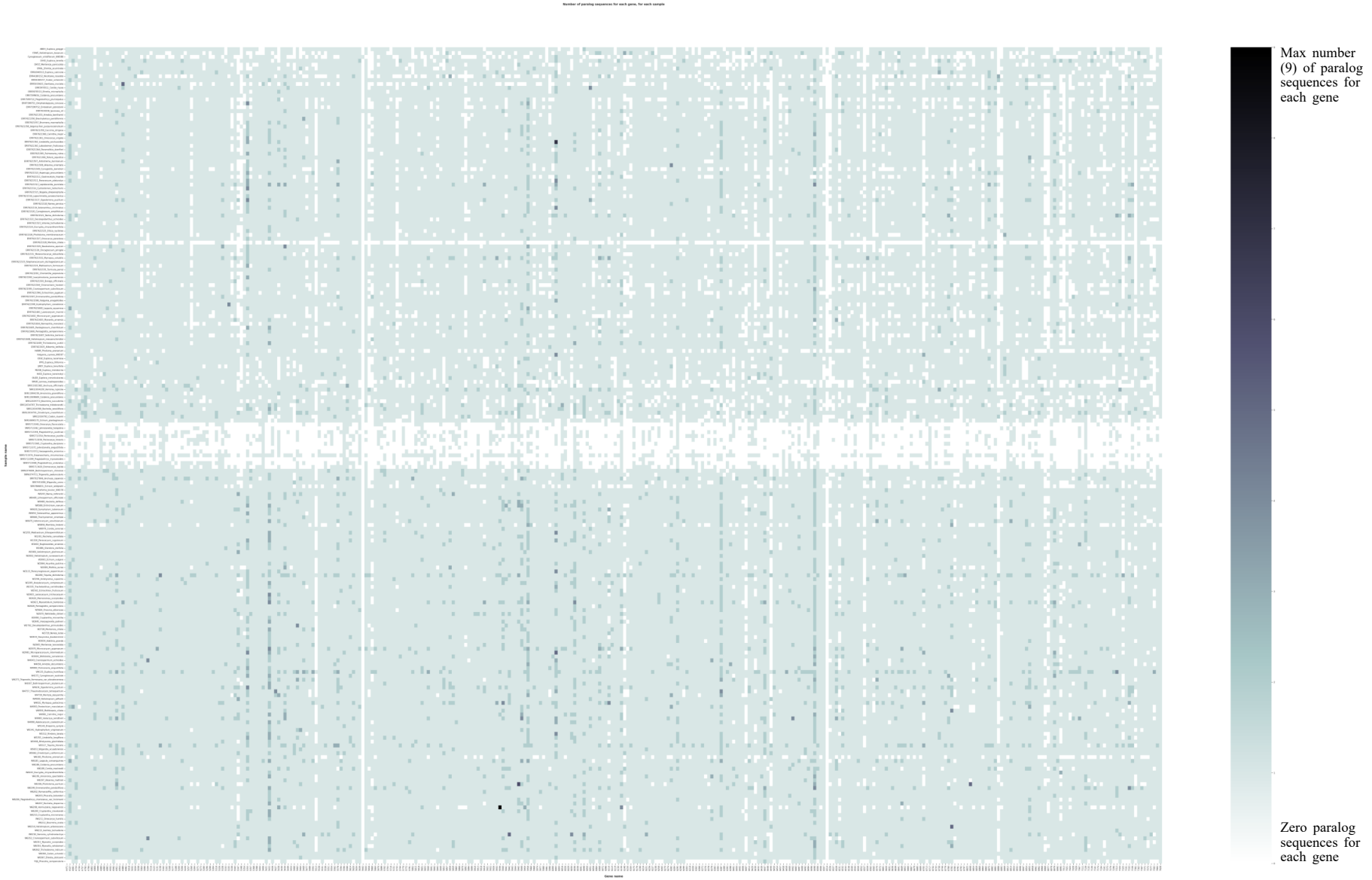
Name	NumReads	ReadsMapped	PctOnTarget	Genes Mapped	GenesWith Contigs	GenesWith Seqs	Genes At25pct	Genes At50pct	Genes At75pct	Genes At150pct	Paralog Warnings Long	Paralog Warnings Depth	GenesWithout StitchedContigs	GenesWith StitchedContigs	GenesWith StitchedContigs Skipped	GenesWith ChimeraWarning
W2603_Lasiocaryum_trichocarpum	1758284	485666	27.6	350	336	336	335	311	252	0	9	24	160	176	0	0
W2620_Memoremea_scorpioides	2515786	893654	35.5	349	341	341	339	312	254	0	10	24	161	180	0	0
W2623_Myosotidium_hortensia	2497010	880659	35.3	349	339	339	336	313	263	1	47	98	125	214	0	0
W2626_Pentaglottis_sempervirens	1220442	283190	23.2	349	338	337	332	290	226	0	11	39	121	216	0	0
W2664_Onosma_alborosea	2812286	936595	33.3	350	340	340	336	310	244	0	12	26	135	205	0	0
W2670_Wellstedia_dinteri	3381056	1075309	31.8	350	347	347	345	331	284	0	20	44	175	172	0	0
W2690_Cryptantha_micrantha	3720014	1504032	40.4	349	342	340	338	317	257	0	7	14	159	181	0	0
W2691_Harpagonella_palmeri	5536776	2660399	48	350	345	344	340	324	270	0	15	26	163	181	0	0
W2702_Decalepidanthus_primuloides	1563052	452677	29	349	339	338	333	305	233	1	11	23	147	191	0	0
W2728_Mertensia_ciliata	2487688	847242	34.1	349	342	342	341	316	260	1	15	34	147	195	0	0
W2729_Nonea_lutea	2351980	738716	31.4	349	343	342	338	319	266	0	6	15	171	171	0	0
W2833_Dasynotus_daubenmirei	2415998	849449	35.2	350	341	339	337	308	252	0	10	26	145	194	0	0
W2834_Adelinia_grande	1179628	158266	13.4	349	334	333	326	278	192	0	9	21	152	181	0	0
W2905_Mertensia_lanceolata	1860138	576837	31	350	342	342	341	306	252	0	16	35	148	194	0	0
W2970_Microcaryum_pygmaeum	2288394	804260	35.1	350	343	341	338	317	263	0	39	107	148	193	0	0
W2981_Microparacaryum_intermedium	2643344	891970	33.7	351	340	338	336	308	244	0	14	27	159	179	0	0
W3004_Wellstedia_somalensis	2463398	838570	34	350	345	344	341	317	269	1	14	30	143	201	0	0
W4043_Craniospermum_echioides	1132648	221927	19.6	350	336	335	330	287	218	0	14	35	151	184	0	0
W4056_Arnebia_decumbens	1294038	306174	23.7	350	336	334	331	294	205	1	15	37	136	198	0	0
W4082_Pulmonaria_angustifolia	2715240	648513	23.9	352	341	340	338	311	241	0	14	33	127	213	0	0
W4135_Euploca_humifusa	4766142	1787482	37.5	352	350	349	349	345	315	0	77	204	122	227	0	0
W4272_Cynoglossum_australe	2920650	778417	26.7	351	344	344	340	306	242	0	11	33	139	205	0	0
W4273_Trigonotis_formosana_var_elevatovenosa	2256208	529347	23.5	348	336	335	331	296	218	0	20	59	129	206	0	0
W4307_Bothriospermum_zeylanicum	3842740	1111462	28.9	350	343	341	337	319	257	1	14	31	144	197	0	0
W4634_Ogastemma_pusillum	5426938	2112348	38.9	352	347	346	346	330	285	1	50	155	116	230	0	0
W4757_Thaumatocaryon_tetraquetrum	1875730	245434	13.1	349	332	327	319	283	208	0	17	47	134	193	0	0
W4759_Moritzia_dasyantha	805876	134186	16.7	349	323	322	310	263	178	0	14	30	140	182	0	0
W4928_Heliotropium_jaffuelii	1779344	589947	33.2	351	350	350	349	338	288	1	6	9	123	227	0	0
W4931_Myriopus_poliochros	2445046	1091589	44.6	351	348	347	347	344	302	1	43	138	123	224	0	0
W4950_Pontechium_maculatum	1197522	264456	22.1	350	331	331	323	288	217	2	15	31	148	183	0	0
W4959_Moltkiopsis_ciliata	1409954	410517	29.1	351	338	335	331	305	233	0	8	23	154	181	0	0
W4981_Cerinthe_major	2144470	792726	37	348	339	339	336	303	250	0	14	27	176	163	0	0
W4983_Halacsya_sendtneri	1458858	519282	35.6	349	341	340	330	306	235	0	34	110	147	193	0	0
W4990_Adelocaryum_coelestinum	1775364	543329	30.6	349	341	339	335	303	232	0	14	28	145	194	0	0
W5140_Draperia_systyla	1737588	699150	40.2	350	348	348	347	336	305	0	5	25	110	238	0	0
W5141_Hydrophyllum_virginianum	1412180	235540	16.7	350	341	340	336	307	242	0	7	24	143	197	0	0
W5312_Rindera_lanata	1600106	449756	28.1	350	339	337	334	302	233	1	13	32	142	195	0	0
W5355_Lindelofia_longiflora	1927350	563268	29.2	350	343	341	337	304	234	0	19	28	141	200	0	0
W5448_Afrotysonia_glochidiata	1235928	311437	25.2	348	337	337	334	302	231	0	11	19	142	195	0	0
W5517_Tiquilia_litoralis	2309776	902328	39.1	349	345	345	345	336	295	0	63	159	155	190	0	0
W5653_Wigandia_ecuadorensis	2418412	1098774	45.4	350	350	350	350	341	309	0	10	39	83	267	0	0
W5662_Eriodictyon_californicum	3280396	1431713	43.6	350	350	350	350	344	321	0	16	63	72	278	0	0
W6182_Pholisma_arenarium	1292860	239081	18.5	349	285	278	276	256	215	0	4	9	79	199	0	0
W6185_Lappula_consanguinea	3149368	1287639	40.9	351	342	340	335	317	260	0	31	84	136	204	0	0
W6188_Coldenia_procumbens	3791974	2001881	52.8	350	347	347	347	341	311	1	2	6	149	198	0	0
W6189_Cordia_macleodii	865428	160198	18.5	351	337	334	328	293	213	0	31	72	107	227	0	0
W6193_Eucrypta_chrysanthemifolia	2338384	1122375	48	350	348	348	346	338	304	0	6	16	139	209	0	0
W6195_Amsinckia_spectabilis	4171380	1903572	45.6	349	343	340	339	316	255	0	13	62	117	223	0	0
W6197_Alkanna_mathioli	1182276	342051	28.9	348	336	335	326	301	240	0	10	32	150	185	0	0
W6198_Pholistoma_auritum	883704	160062	18.1	348	325	324	319	277	217	0	13	19	204	120	0	0
W6199_Emmenanthe_penduliflora	2654750	1278967	48.2	351	350	349	349	338	308	0	37	129	113	236	0	0
W6202_Romanzoffia_californica	3059448	1330642	43.5	350	350	349	348	343	308	1	7	21	139	210	0	0
W6203_Phacelia_bolanderi	2337630	947576	40.5	350	349	349	348	336	302	0	8	16	151	198	0	0
W6206_Plagiobothrys_chorisianus_var_hickmanii	3550380	1450190	40.8	349	339	339	335	312	265	0	9	23	157	182	0	0
W6207_Rochelia_disperma	3724950	1563232	42	350	345	345	341	320	266	0	11	25	169	176	0	0
W6208_Hormuzakia_negevensis	909512	80332	8.8	346	302	300	285	214	135	1	11	19	151	149	0	0
W6209_Cryptantha_clevelandii	4161112	1429489	34.4	349	339	339	335	316	255	0	15	28	151	188	0	0
W6210_Cryptantha_micromeres	3111594	1438630	46.2	350	341	341	337	313	253	1	8	25	155	186	0	0
W6211_Oreocarya_humilis	2174518	641384	29.5	350	339	339	336	306	252	1	8	19	136	203	0	0
W6212_Bourreria_ovata	3763236	1805070	48	350	350	350	350	343	310	0	6	20	106	244	0	0
W6214_Heliotropium_arborescens	3201378	1386300	43.3	351	350	350	350	343	308	0	6	15	109	241	0	0
W6215_Ixorhea_tschudiana	4184850	2048415	48.9	350	349	349	349	341	311	0	1	11	111	238	0	0
W6216_Varronia_cylindrostachya	1800830	612507	34	350	344	344	341	325	285	1	20	72	146	198	0	0
W6352_Craniospermum_subvillosum	1088222	297040	27.3	351	337	334	330	294	227	0	13	38	151	183	0	0
W6353_Myosotis_scorpiodes	3214508	1389818	43.2	350	342	341	337	314	250	1	17	43	152	189	0	0
W6354_Myosotis_rehsteineri	2524908	1003825	39.8	348	336	336	332	308	249	0	15	28	160	176	0	0
W6362_Trichodesma_indicum	2833678	1268074	44.8	349	342	341	339	319	278	0	21	40	139	202	0	0
W6364_Codon_schenkii	2262534	992722	43.9	350	348	347	344	332	274	0	2	12	88	259	0	0
W6367_Ehretia_dicksonii	1540310	628496	40.8	350	349	349	349	339	307	0	4	27	85	264	0	0

Name	NumReads	ReadsMapped	PctOnTarget	Genes Mapped	GenesWith Contigs	GenesWith Seqs	Genes At25pct	Genes At50pct	Genes At75pct	Genes At150pct	Paralog Warnings Long	Paralog Warnings Depth	GenesWithout StitchedContigs	GenesWith StitchedContigs	GenesWith StitchedContigs Skipped	GenesWith ChimeraWarning
YQIJ_Phacelia_campanularia	25859624	35083	0.1	346	145	143	139	133	119	0	1	2	141	2	0	0
Cynoglossum_viridiflorum_W6588	105374	24117	22.9	333	214	207	120	41	9	0	1	1	131	76	0	0
Halgania_cyanea_W6567	3595334	1825785	50.8	352	350	350	350	343	318	0	5	30	99	251	0	0
Tournefortia_bicolor_W6578	3606038	1530392	42.4	351	350	350	350	340	298	0	5	13	129	221	0	0

Supplementary material S1. B. Heatmap visualizing percentage length recovery for each gene, relative to the mean of the target-file references. The heatmap was produced with the gene_recovery_heatmap.py script available with HybPiper.



Supplementary material S2. Heatmap visualizing the paralog distribution according to the paralog report table, produced with the paralog_retriever.py script available with HybPiper.



Supplementary material S3. A. TOAST R Script for generating a partition block and assessing overall missing data patterns across the alignments.

```
library(devtools)

devtools::install_github("carolinafishes/toast")

library(toast)
library(tidyverse)

setwd("/home/marianna/Documents/Borag_353/Analyses/retrieved_dna_final/aligned/CIAlign/
Conca_Alignment_TOAST")
alignment_directory<-"/home/marianna/Documents/Borag_353/Analyses/
retrieved_dna_final/aligned/CIAlign/cleanedSampNames"

list.files(alignment_directory)
#run the Python script first to change the names, because any gap in the names will cause
a problem in the next steps ## check for missing data/sequences among the fasta files
missing <- MissingDataTable(aligned_dir = alignment_directory)
length(missing)

## write CSV tabel
write.csv(missing, file="missing_dataToastTutorial.csv")

## Mapping missing data
par(mar = c(10, 2, 4, 4) + 0.1, oma = c(7, 1, 1, .5) + 0.1)
PhyloMissmap(missing, use_tree = "FALSE", las = 2, cex.axis = 0.8)

## Alternative plotting function

## transpose the data frame "missing" to have species names as rows and the
## alignment names as columns with the length of the alignments as values
trans_missing <- as.data.frame(t(missing))
typeof(missing)

## Now add a new column with the number of NA's in the data frame and sort
## the rows in descending order listing the samples with most missing data on top
# Add a new column with the count of NA values per row
trans_missing <- trans_missing %>%
  mutate(NA_Count = rowSums(is.na(.)))%>%
  arrange(desc(NA_Count))

head(trans_missing, n = 20)

## Print the modified data frame for documentation last column is the number of missing
data
write.csv(trans_missing, file = "SequenceLength_AND_MissingData.csv")

## ::::::: Making a nicer plot than the one by Toast ::::::::

## convert data frame in tidyverse tibble & adding column name for species
trans_missing_tib <- tibble::rownames_to_column(trans_missing, var = "species")
head(trans_missing_tib)

## Convert the data frame to long format
trans_missing_long <- pivot_longer(trans_missing_tib, cols = -c(species,NA_Count),
names_to = "Alignment", values_to = "Value")

## arranges the pivot table long form in descending order of the NA_Count column
trans_missing_long <- trans_missing_long %>% filter(!grepl("^NA_Count", Alignment)) %>%
  arrange(desc(NA_Count))

head(trans_missing_long, n = 20)

## Create the plot
ggplot(trans_missing_long, aes(x = Alignment, y = species, fill = is.na(Value),
 arrange(desc(NA_Count)))) + geom_tile(color = "NA") + scale_fill_manual(values =
 c("#91bfdb", "#fc8d59"), labels = c("present", "missing")) + labs(x = "Alignments", y =
 "Samples", fill = "data") + theme_minimal() + theme(axis.text.y = element_text(size =
 5, hjust = 1), axis.text.x = element_text(angle = 90, hjust = 1, size = 3))

##THE PLOT DOES NOT LISTS THE SAMPLES IN DESCENDING ORDER BASED ON THE NA_Count
```

```
## ____________________________________  END  ____________________________________


## ____________________________  Continue the TOAST Tutorial ____________________________

## First function creates a concatenated alignment
SuperAlign(alignment_directory, missing, filename = "Concat_Alignment.phy")

## second function generates your corresponding locus locations in nexus format
PartitionTable(alignment_directory, missing_df = missing,
filename = "Partition_block.nex")

#convert phylip file to fasta
PhyIntoFasta(phylipfile="Concat_Alignment.phy",
 filename="Concat_Alignment.fasta",externalfile=TRUE)

## _______________ Concatenated Alignment with missing data excluded _______________
#create a second set of fasta files with taxa that have less than 50% missing loci. In
#our case we have 351 loci, so 50% ~ 175 loci

#You need to make a new file on your HD called e.g."subsampled_50p_perTaxa"

#Set the path here
tff<-"~/Documents/Borag_353/Analyses/retrieved_dna_final/aligned/CIAlign/
Conca_Alignment_TOAST_50_missing_data"
#Change this number for different thresholds, this creates an object that maps #what to
#do for the next function
threshold_df<-ThresholdDataTable(missing, threshold=175)

#now run this and watch those fasta files appear!
ThresholdExtract(aligned_dir=alignment_directory, missing_df=threshold_df,
 threshold_fasta_folder=tff)

#This step is done intentionally for those wishing to only use these files for
#gene-tree species-tree analyses downstream. From here assemble as you just did before.
setwd(tff)
missing<-MissingDataTable(aligned_dir = tff)
write.csv(missing, file="50Percentmissing_data.csv")
SuperAlign(aligned_dir=tff, missing, filename="50pConcat_Alignment.phy" )
PartitionTable(aligned_dir=tff, missing, filename="50pPartitions.nex")
#This uses the missmap function to view missing data as we did before
par(mar = c(10, 2, 4, 4) + 0.1, oma= c(7, 1, 1, .5) + 0.1)
PhyloMissmap(missing, use_tree="FALSE", las=2, cex.axis=0.8)

## ___________Missing data and Advanced Filtration_____________

library(ape)
missing_data_path<-"~/Documents/Borag_353/Analyses/retrieved_dna_final/aligned/
CIAlign/Conca_Alignment_TOAST/missing_dataToastTutorial.csv"

#read in files
missing<-read.csv(missing_data_path)
#Visualize coverage
VisualizeCoverage(missing, threshold=175)
#You can also explore the impact of your threshold choice visually.
VisualizeThreshold(missing, threshold=175)

#summary statistics
dataSummary<-missingDataSummary(missing)
dataSummary[dataSummary[,2]<175,]

CompleteCoverage(missing)
```

Supplementary material S3. B. Bar plot produced with the TOAST R Script showing the missing data patterns in samples with more than 50% missing data.

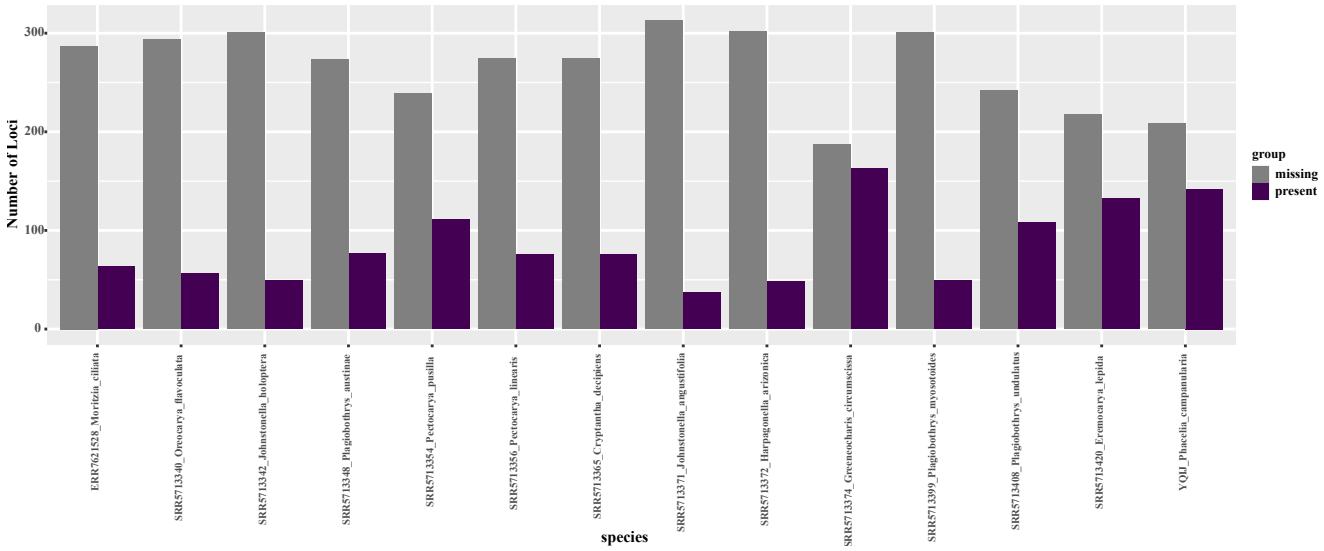


Figure S1. Phylogenetic relationships among the Boraginales. Concatenation analysis with IQTREE using the retrieved exonic regions of 353 nuclear loci from the Angiosperms353 probe set. Numerical values are ultrafast bootstrap approximation (UFBoot).

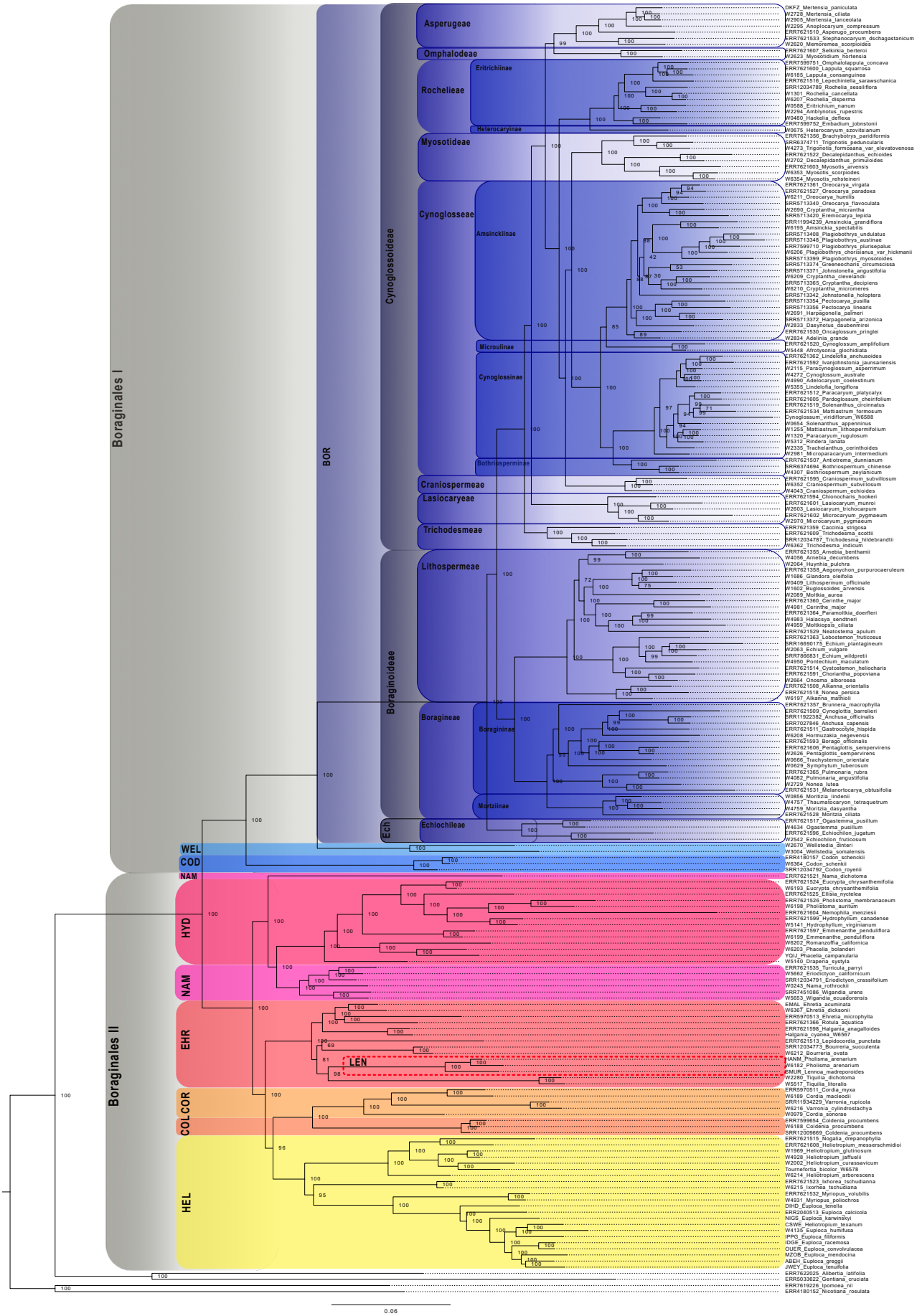


Figure S2. Phylogenetic relationships among the Boraginales. ASTRAL-III species tree generated using the retrieved supercontigs of 353 nuclear loci from the Angiosperms353 probe set. Numerical values are local posterior probabilities (LPP)

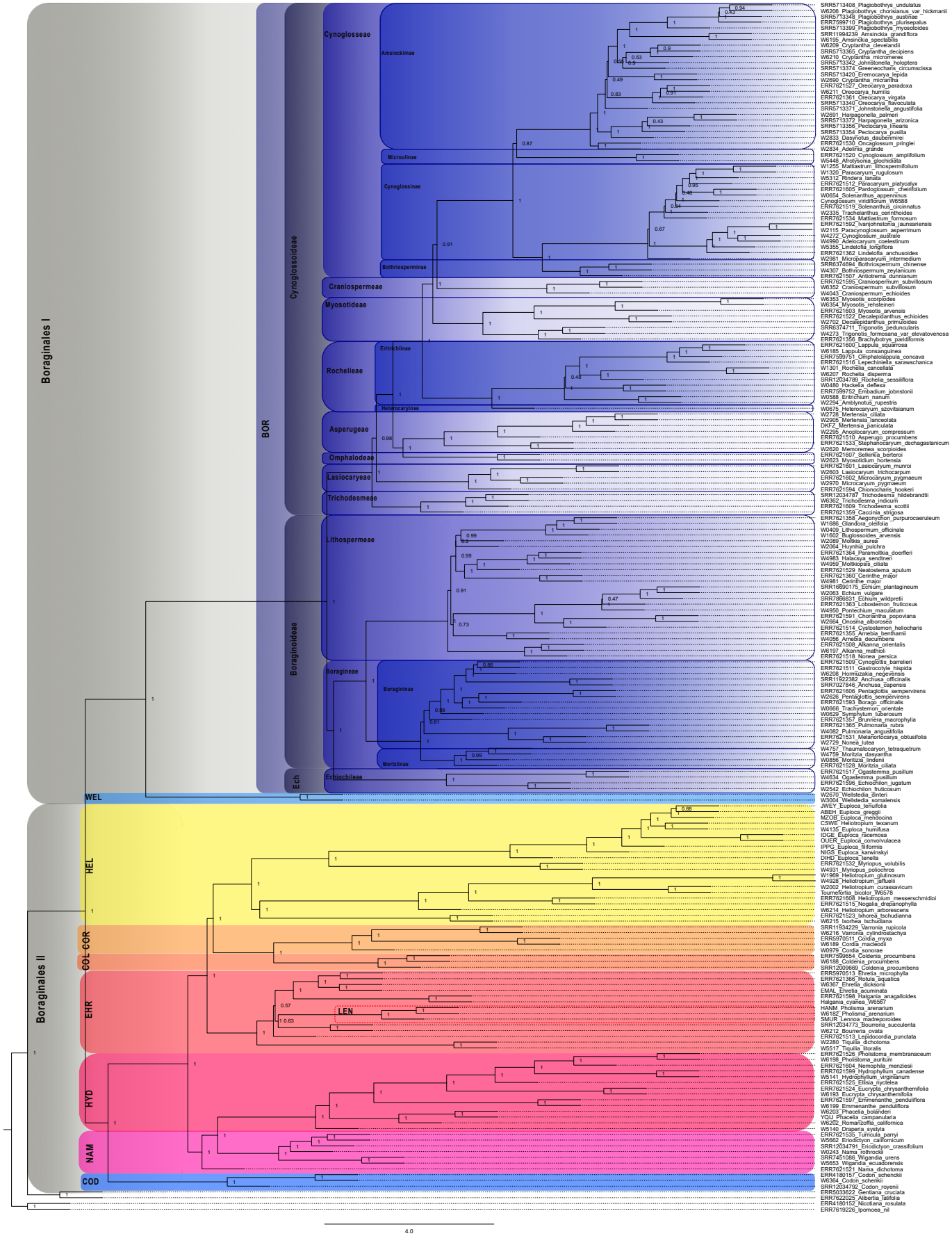


Figure S3. Phylogenetic relationships among the Boraginales. Concatenation analysis with IQTREE using the retrieved supercontigs of 353 nuclear loci from the Angiosperms353 probe set. Numerical values are ultrafast bootstrap approximation (UFBoot).

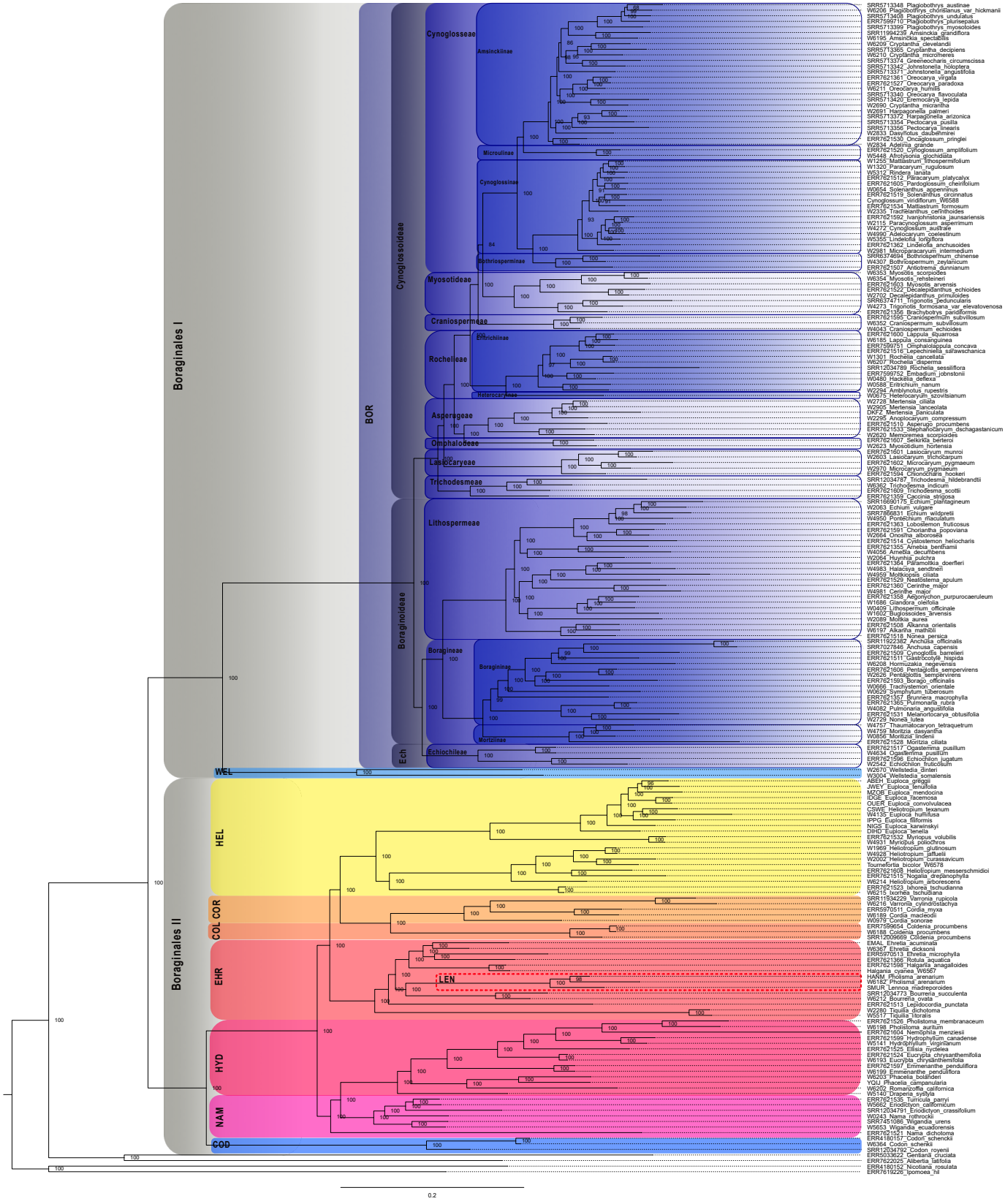


Figure S4. Phylogenetic relationships among the Boraginales. ASTRAL-III species tree generated using the retrieved exonic regions of 353 nuclear loci from the Angiosperms353 probe set, but with 26 potentially paralogous genes removed. Numerical values are local posterior probabilities (LPP).

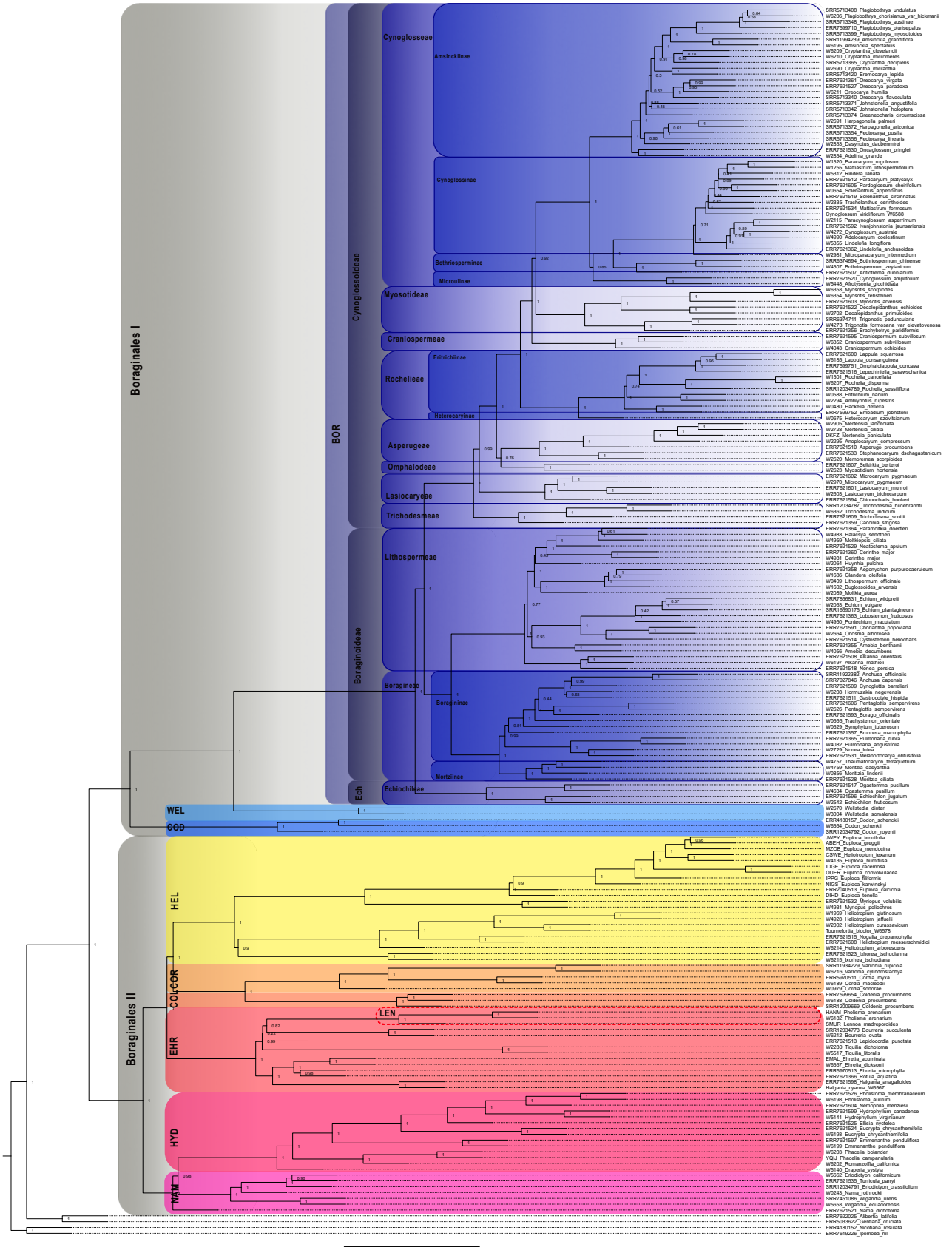


Figure S5. Phylogenetic relationships among the Boraginales. Concatenation analysis with IQTREE using the retrieved exonic regions of 353 nuclear loci from the Angiosperms353 probe set, but with 26 potentially paralogous genes removed. Numerical values are ultrafast bootstrap approximation (UFBoot).

