# Phylogenetics and Evolution of the Australian Camphorosmeae (Amaranthaceae)

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

der Johannes Gutenberg-Universität Mainz

(D77)

**Philipp Hühn**

**geb. am 02.10.1990 in Lauterbach (Hessen)**

**Mainz, März 2023**

Dekan: Herr Prof. Dr. Eckhard Thines

1. Berichterstatter: Frau Prof. Dr. Gudrun Kadereit

2. Berichterstatter: Frau Prof. Dr. Elvira Hörandl

Tag der mündlichen Prüfung: 31.08.2023

I dedicate this study to the generations of botanists
who put a lot of effort into unravelling the many mysteries
of this tremendously fascinating plant group.

“While working on some material from the Broken Hill District, I was impressed by the state of confusion which apparently existed in the Family Chenopodiaceae. The exigencies of the moment forced me more particularly into the genus *Bassia*, and I quickly came to the conclusion that a complete revision of the Australian members of the genus was necessary. During the past eighteen months I have labored to this end, and am now submitting a paper which, I trust, will do something towards straightening out a very chaotic genus.” - Robert Henry Anderson (1923)

# *CONTENTS*

# *1. INTRODUCTION*

The modern Australian continent is characterized by a unique composition of geologic and biotic features, which were shaped by an increasingly intensifying aridification process over the last 20 million years (Smith and Morton, 1990; Bowler et al., 2006; Morton et al., 2011; Byrne et al., 2018). The continent has relatively flat terrain, no major geologic barriers (except for the Great Dividing Range in the southeast), and most soils are highly weathered and nutrient-deficient (Mabbutt, 1988; Johnson, 2009; Crisp and Cook, 2013). Numerous ephemeral inland lakes fed by extensive drainage systems are scattered across the continents interior (Martin, 2006; Morton et al., 2011). The climate is strongly influenced by the surrounding oceans and weather systems (Suppiah et al., 2001; Timbal and Drosdowsky, 2013; Song et al., 2017). For most of the continent, precipitation is the dominant environmental factor (Keast, 1959; Mokany, 2022). Approximately 70 percent of the continent comprises a vast arid center surrounded by an extensive semi-arid periphery (Martin, 2006; Morton et al., 2011; Byrne et al., 2018). Little seasonality, major drought seasons, unpredictable rainfall events, average annual precipitation below 250-350 mm and high evaporation define the arid zone with its semi-arid periphery (Smith and Morton, 1990; Martin, 2006; Morton et al., 2011; Byrne et al., 2018). The arid-zone biome covers a range of environments such as sandy deserts, stony deserts and steppes, ranges and coastal plains (Mabbutt, 1988; Byrne et al., 2018; McDonald, 2020). These mosaic-like landscapes are home to a variety of sclerophyllous and xeromorphic vegetation types, such as (open) *Acacia* Mill. and Eucalypt shrub- and woodlands, spinifex, tussock and hummock grasslands, and chenopod shrubland (Martin et al., 2006; Morton et al., 2011; Crisp and Cook, 2013; Byrne et al., 2018). While this biome is a hotspot for reptiles and invertebrates, it hosts only 10% of Australia's plant diversity (Barker and Greenslade, 1982; Byrne et al., 2008; Anderson, 2016; Byrne et al., 2018). The plants of the arid zone, however, are highly specialized survival artists, having evolved and adapted over millions of years to

successfully cope with the harsh conditions of this habitat. Among other winner-lineages (e.g.: Asteraceae Bercht. & J.Presl, Fabaceae Lindl., Goodeniaceae R.Br., Malvaceae Juss., Poaceae (R.Br.) Barnhart, Proteaceae Juss., Rutaceae Juss.) (reviewed in Crisp et al., 2004; Byrne et al., 2008; 2018; Crisp and Cook, 2013) of the evolutionary race for persistence in this constantly transitioning, unreal habitat, are the Amaranthaceae Juss.

Amaranthaceae *sensu lato*, including Amaranthaceae *sensu stricto* and Chenopodiaceae Vent., constitute the most diverse lineage (ca. 180 genera and 2500 species) of the Caryophyllales (Kadereit et al., 2003; Müller and Borsch, 2005; The Angiosperm Phylogeny Group, 2017). Besides the phylogenetic evidence for the monophyly of the Amaranthaceae *sensu lato* (Kadereit et al., 2003, Müller and Borsch, 2005; Walker et al., 2018; Morales-Briones et al., 2021), the Chenopodiaceae are treated separately in this section because species of this former family are characteristic of the floristic circumscription of the "chenopod shrubland" vegetation, which is a ubiquitous component of Australia's arid-zone flora (Wilson, 1984; Martin, 2006; Byrne et al., 2008, 2018; Morton et al., 2011; Crisp and Cook, 2013). Amaranthaceae s.s. comprise ca. 70 genera and ca. 800-1,000 species and has a worldwide distribution in predominantly tropical, subtropical and warm temperate regions (Kadereit et al., 2003; Müller and Borsch, 2005; The Angiosperm Phylogeny Group 2017). Centers of diversity are the Neotropics, eastern and southern Africa and Australia. In Australia, the ca. 150-160 species of 15 recognized genera are most common in the arid zone and the wet tropics. The cosmopolitan Chenopodiaceae comprises ca. 110 genera with ca. 1,500-1,700 species that are predominantly distributed in arid to semi-arid, saline habitats of temperate and subtropical regions (Kadereit et al., 2003). In Australia, the ca. 300 species in ca. 30 recognized genera are particularly successful in semi-arid environments and in saline habitats (Wilson, 1984; Kadereit et al., 2005). The Chenopodiaceae have arrived in Australia at least nine times by long distance dispersal events (Kadereit et al., 2005). The divergence estimates of Australian lineages ranged

from late Pliocene as the youngest to late Eocene/early Oligocene as the oldest period, with a clear culmination of arrival events during the Late Miocene and the Pliocene (Kadereit et al., 2003, 2005). Since the first signs of increasing and progressively expanding aridity in Australia in the Middle Miocene, the dramatic aridification process favored the emergence of numerous arid-adapted taxa whose phylogenetic fingerprints reflect the desiccation of the continent (reviewed in Crisp et al., 2004; Byrne et al., 2008; Crisp and Cook, 2013; Byrne and Murphy, 2020). Large Australian Chenopodiaceae groups (Australian Camphorosmeae Moq., Salicornieae Moq. and *Atriplex* L.) radiated rapidly after their arrival during this period (Shepherd et al., 2004, 2005; Kadereit et al., 2005; Cabrera et al., 2011; McDonald, 2020; Zerdoner Calasan et al., 2022). Their pre-adaptation to dry and saline conditions likely promoted their rapid diversification in a period of major climatic change that transformed the once warm and wet continent into an arid one (Martin, 2006; Byrne et al., 2008).

Of the recognized Australian tribes, the Australian Camphorosmeae are the most species rich tribe with ca. 150 species in 12 genera, followed by Atripliceae with 60 species in the genus *Atriplex*, Chenopodieae with 50 species in 6 genera, and Salicornieae with 36 species in 6 genera (Wilson, 1984, Kadereit et al., 2003; 2005). Besides the ca. 150 Australian species, the Camphorosmeae comprise additional 30-40 non-Australian species that are centered in the Old World desert belt from the Canary Islands to Asia, with a few species distributed in North America and South Africa (Kadereit et al., 2005; Kadereit and Freitag, 2011). The Australian Camphorosmeae were found to form a monophyletic clade, sister to the Central Asian genus *Grubovia* (Kadereit et al., 2003, 2005, 2014; Kadereit and Freitag, 2011).

The Australian Camphorosmeae grow in all floristic regions of the continent, including the tropical north and the temperate and Mediterranean south (Wilson, 1984; Cabrera et al., 2009). They are not only a ubiquitous element of the arid zone flora but also dominate the landscape in many areas as component of the characteristic chenopod shrubland (Kershaw et

al., 1994; Martin, 2006; Morton et al., 2011). They are most abundant in the arid and semi-arid regions of the continents interior, with centers of diversity in the Yilgarn Plateau in Western Australia and in the Lake Eyre-Murray Basin in the Eastern Desert (Wilson, 1984; McDonald, 2020). They grow in a wide range of soils of various (mostly nutrient-poor) substrate composition (sandy, loamy, clayey, calcareous, gypseous, stony) and salinity (non-saline, slightly saline to saline) (Wilson, 1984; Harden, 1990; Walsh, 1996; Paczkowska and Chapman, 2000). They occur in diverse arid and non-arid habitats, such as Coast, Mesic Plains, ephemeral Riverine and Desert Lake habitats, as well as in the most extreme desert landscapes such as Sand and Stony Desert habitats (Mabbutt, 1988; McDonald, 2020). Because of their ability to survive long periods of drought and to accumulate salts and heavy metals, a number of Australian Camphorosmeae gained economic interest. Some species are valuable resources for livestock in zones of low-rainfall and high soil salinity (Leigh et al., 1979; Malcolm, 2000; Norman et al., 2010; Revell et al., 2013; El Shaer and Squires, 2015; Certain et al., 2021) and are the focus of sustainable agriculture strategies to protect soils and waters from degradation and desiccation (Barrett-Lennard et al., 2003; Blache et al., 2016; Certain et al., 2021). Several species are promising candidates or already established species for the restoration and revegetation of wasted mining sites (Brearley, 2003; Squires et al., 2012; Wu et al., 2021; Zhong et al., 2021; Valliere et al., 2022). Some species are the focus of species conservation as they or their habitats are threatened by deforestation, intensive mining of economically valuable soil resources, introduction of non-native taxa, and overgrazing by livestock (Fairfax and Fensham, 2000; Sharp et al., 2009; Mavromihalis, 2010 a, b; Linley et al., 2016; Fensham et al., 2018; Silcock and Fensham, 2018; Amor et al., 2020).

The species are predominantly woody perennials or small shrubs, rarely perennial herbs or annuals (Wilson, 1984; Kadereit and Freitag, 2011). The spreading to ascending branches, as well as the normally alternate leaves, are often covered with hairs or a more or less dense

indumentum. The leaves are normally small and succulent. The flowers are usually solitary and axillary, and the fruiting perianth is usually enlarged, hardened and bears appendages (Wilson, 1975, 1984) (Fig. 1). The perianth bears two series of veins, 5 tepaline and 5 intertepaline in position. As the nut-like fruit (utricle) develops, the basal portion of the perianth (the perianth tube, Fig. 2 A, G) usually enlarges and becomes variously thickened. The majority of species produce horizontal wing-like or spine-like appendages from one or other series of veins, or just below the lobes. Wing-like fruiting perianths develop usually a simple, horizontal wing (Fig. 1 C, D) or 5 separate wings which grow from the base of each lobe (Fig. 1 E). Additional vertical, wing-like structures along or from the base of the perianth tube can be present (Fig. 3 J, M, N). The portion of the perianth above the tube is formed by the perianth lobes, obscures the utricle and can bear accessory processes, such as erect lobes or floccose wool (Fig. 1 C, E, O and Fig. 3 Q, R). The seed is usually horizontal (Fig. 2 F). Spine-like fruiting perianths usually develop 2-6 intertepaline spines, of which a contiguous pair is borne opposite the radicle (Fig. 1 F-H, L, N, Fig. 2 A, B, E and Fig. 4). The spines are oriented horizontally to vertically, relative to the perianth tube. The 3-5 perianth lobes and the tube portion above the spines can be prominently developed. This division, called “limb”, corresponds the upper perianth in wing-like species (Fig. 2 A, G). The seed is horizontal to erect. In addition to the two "main" appendage types, there are fruiting perianths that do not develop prominent appendages (Fig. 1 M, V, W and Fig. 2 D, F, H), produce a particular star-like shape (Fig. 1 P and Fig. 2 C), are intermediate between typical wing- and spine-like appendages (Fig. 1 H, O, Fig. 2 B, Fig. 3 C, L), or form aggregated infructescences (Fig. 1 L, N, Fig. 2 E and Fig. 4 Q).

The taxonomic history of the Australian Camphorosmeae is characterized by numerous rearrangements using mainly characteristics of the fruiting perianth for generic delimitation (e.g. Mueller, 1882a, b; Ulbrich, 1934; Anderson, 1923; Ising, 1961, 1964; Wilson, 1975, 1984; Scott, 1978; Chinnock, 1980; Jacobs, 1988; reviewed in Anderson, 1923; Wilson, 1975; Scott,

1978). In particular, the shape, position (tepaline, intertepaline) and number of appendages were used to describe, delineate and revise genera, sections and species complexes in the Australian Camphorosmeae. Perhaps the most significant rearrangement was the separation of the Australian from the Eurasian, African and North American species. After many years of discussions with focus on distinguishing characters of the Australian species (e.g. Anderson, 1923; Ising, 1961, 1964; Wilson, 1975; Scott, 1978), the Australian species of *Kochia* and *Bassia* were revised and transferred to *Maireana* Moq. and *Sclerolaena* R.Br. by Wilson (1975) and Scott (1978), respectively. The Eurasian-Australian generic disjunction was justified by combinations of distinguishing characters (Wilson, 1975). Decisive were differences in life form and habit (Eurasian annual herbs vs. Australian shrubs and perennials with a woody stem), the number of flowers per bract (aggregated into condensed cymes vs. axillary and solitary or paired), the shape and structure of the fruiting perianth (papery and weak vs. variously enlarged, thickened and modified), the appendage position (back on the perianth lobes vs. tepaline and intertepaline) and the presence of a radicular modification (variously positioned radicle without slit in the perianth wall vs. radicle always in intertepaline position and presence of a perianth wall slit, rib, canal or tubercle).

In addition to the two species-rich genera *Maireana* (ca. 57 species with mostly wing-like fruiting perianths, Fig. 1 C-E) and *Sclerolaena* (ca. 65 species with mostly spine-like fruiting perianths, Fig. 1 F-H), 10 species-poor genera have been described (Table 1, Fig. 1). These include species with perianth appendage modifications and positions that do not strictly match the main perianth appendage types of the major genera (Wilson, 1975, 1984). In the monotypic genus *Didymanthus* Endl. (Fig. 1 K), the paired flowers are fused at their bases (unlike *Sclerolaena* or *Maireana*), the seed is erect (as in many *Sclerolaena* species) and the perianth develops 5 horizontal wings in fruit (as in *Maireana*). In the monotypic genus *Eriochiton* (R.H.Anderson) A.J.Scott (Fig. 1 O), the fruiting perianth develops two series of

appendages: 5 erect wing-like tepaline and 5 spinescent appendages arising from wing-like appendages. Species of the genus *Dissocarpus* F.Muell. (4 spp.) (Fig. 1 L) develop spinous appendages (as in *Sclerolaena*) but these are tepaline in position (as in *Maireana*) and the fruiting perianths are paired or aggregated into woody, ball-like infructescences. In *Eremophea* Paul G.Wilson (2 spp.) (Fig. 1 N), the fruiting perianths develop tepaline spines and become embedded in the woody branch axis. In *Neobassia* A.J.Scott (2 spp.) (Fig. 1 S), the fruiting perianth produces 5 tepaline spines that arise from the perianth base and unite into an apical cup. In *Threlkeldia* R.Br. (2 spp.) (Fig. 1 W), the perianth becomes succulent or dry in fruit and is without spines or wings but develops intertepaline knobs. In *Osteocarpum* F.Muell. (5 spp.) (Fig. 1 T), the fruiting perianth develops 1-5 intertepaline, vertical wings on top of the perianth. In *Malacocera* R.H.Anderson (4 spp.) (Fig. 1 P), the fruiting perianth bears 3-5 sub-cylindrical, tepaline appendages forming a Y or star-shaped configuration. In *Enchylaena* R.Br. (2 spp.) (Fig. 1 M), the perianth becomes succulent in fruit and, if present, develops a shallow, succulent wing-like cup at the perianth apex. In *Roycea* C.A.Gardner (3 spp.) (Fig. 1 U, V), the 5 tepals do not enlarge in fruit and the fruit is surrounded by the perianth at its base.

While the use of perianth characters has been suitable in differentiating the Australian from the Eurasian Camphorosmeae and in distinguishing the two species-rich Australian genera, it has been noted that the characters used to delimit Australian genera in many cases do not create a natural grouping of species (Anderson, 1923; Ising, 1964; Wilson, 1975, 1984; Scott, 1978). In addition, species sharing very similar morphological traits, hybridization between closely allied species and between genera as well as the evolution of homoplastic perianth traits further complicate the taxonomy (Wilson, 1975, 1984; Cabrera et al., 2009). Cabrera et al. (2009) conducted the first molecular phylogenetic study to clarify if the current taxonomy based on morphological characters fits the phylogenetic fingerprint of this group. This involved sampling 71 species of all genera recognized, surveying 15 morphological

characteristics for their systematic relevance and phylogenetic inference using seven molecular markers. Of the molecular markers tested, only the external transcribed spacer (ETS) and internal transcribed spacer (ITS) yielded sufficient sequence variation for phylogenetic inference. It turned out that the molecular-phylogenetic results did not support the current taxonomy of Australian Camphorosmeae and of the morphological traits investigated, only the fruiting perianth appendage type was of systematic relevance with respect to the molecular outcome. Although the utilized approach resulted in several statistically supported clades, it did not resolve generic boundaries and several decisive relationships remained unresolved. The authors suggested that Incomplete Lineage Sorting (as a consequence of rapid radiations) and ongoing hybridization in this relatively young group resulted in a lack of phylogenetic informative sequence data.

The biogeographic study by Cabrera et al. (2011) found that the Australian Camphorosmeae initially diversified at the end of the Miocene with subsequent radiation during the Pliocene. Diversification was likely driven by the intensifying aridification of the Australian continent during this time. The ancestors of the modern Australian Camphorosmeae probably arrived in southern west Australia by long-distance dispersal and were presumably already adapted to dry and saline conditions. A first inland migration may have taken place by species adapted to coastal, marshy conditions, providing a littoral connection along declining palaeodrainage systems to hypersaline inland habitats and was followed by a multidirectional dispersal across the continent (Cabrera et al., 2011; McDonald, 2020).

Although the molecular studies by Cabrera et al. (2009, 2011) confirmed that the current taxonomy based on morphological characters is in part artificial, resolution was insufficient across many parts of the phylogeny. This, and insufficient sampling, made it difficult to provide statistically supported clades and to characterize them with distinctive morphological data. In addition, it limited the ability to put the morphological evolution in context with landscape and

climatic changes along the aridification of the Australian continent. Further, it complicates the phylogenetic classification and interpretation of new data.

Hence, a major challenge in the phylogenetic study of the Australian Camphorosmeae was finding a suitable laboratory and data analysis method to tackle the complicated phylogeny and taxonomy caused by simultaneous and rapid diversification, ILS and ongoing hybridizations events of this relatively young plant group that previously hampered phylogenetic inference (Cabrera et al., 2009). In this study, a restriction-site associated DNA sequencing (RADseq; Miller et al., 2007; Baird et al., 2008) protocol developed by Hühn et al. (2022) was employed (Appendix 1).

The following RADseq introductory section was largely excerpted from Hühn et al. (2022). RADseq is a reduced-representation library (RRL) preparation protocol for high-throughput sequencing (HTS) that does not require prior genomic information and thus is suitable for studying non-model organisms. RADseq and other HTS-RRL technologies improved enormously during the last decade and provide the opportunity to generate extensive datasets for phylogenetic inference (reviewed in Hörandl and Appelhans, 2015; Andrews et al., 2016; McKain et al., 2018). The term RADseq comprises several methods that all rely on the enzymatic digestion of genomic DNA for complexity reduction, followed by adapter ligation, further reduction by size selection and HTS (Andrews et al., 2016). Thanks to the modular principle of RADseq, the individual wet lab steps, restriction endonucleases (REase/s) and adapters can be modified as required (McCormack et al., 2013; Andrews et al., 2016; McKain et al., 2018; Parchman et al., 2018).

Characteristically, RADseq datasets comprise relatively short loci (typically 100–250nt) and a high proportion of missing data (Ree and Hipp, 2015; Andrews et al., 2016; Eaton et al., 2017; Lee et al., 2018; McKain et al., 2018). This can negatively impact the phylogenetic estimation accuracy (Chou et al., 2015; Roch and Warnow, 2015; Xi et al., 2015, 2016; Xu and

Yang, 2016; Sayyari et al., 2017; Molloy and Warnow, 2018) and may theoretically lead to poorly resolved, incomplete, or positively misleading phylogenetic estimates (e.g. Degnan and Rosenberg, 2006, 2009; Roch and Steel, 2015; Roch and Warnow, 2015; Xu and Yang, 2016; Rannala et al., 2020), and can result in spuriously high bootstrap support values due to the sheer size of the dataset (e.g. Kubatko and Degnan, 2007; Rubin et al., 2012; Liu et al., 2015). Therefore, Hühn et al. (2022) aimed at an overall reduction of resulting RADseq fragments and sequences, while simultaneously increasing the fragment and sequence length. Two "modules" of classic RADseq protocols were modified: First, a combination of restriction endonucleases (REase/s) with a relatively long recognition site were used for DNA digestion (*BamHI* (G'GATCC) and *KpnI* (GGTAC'C)), which thus generate fewer and longer DNA fragments. Second, a direct size selection using an automated DNA fragment segregation technology with a target segregation range of 350-720 nucleotides (nt) fragment length was used. In addition to changes in laboratory workflow, high importance was also placed on assembling the resulting sequence data as accurately as possible, since the bioinformatics related to RADseq is often not straightforward and can heavily impact the assembly outcome, thus the resulting phylogeny. The modified approach by Hühn et al., (2022) has been successfully applied to a number of challenging plant groups, such as *Aeonium* Webb & Berthel. (Dos Santos et al., 2022; Messerschmid et al., 2023), *Aichryson* Webb & Berthel. (Hühn et al., 2022), *Viola* L. (Hühn et al., 2023) and *Vitis* L. (Rij et al., 2019).

In this study, seven research questions regarding phylogeny, taxonomy and evolution of morphological traits were re-examined using new molecular and data analysis methods and extended morphological and ecological data. 1) Can a modified NGS-based methodology improve the insufficient resolution of the phylogeny to reconstruct yet unrecognized species groups as statistically supported clades? 2) To what extent can the phylogenetic clades be circumscribed with morphological data? 3) Are newly surveyed ecological data, such as soil types, salinity and habitat types, of systematic relevance? 4) Do the newly characterized clades show potential to serve as a basis for a revised generic delimitation and taxonomy? 5) Do the postulated hypotheses regarding migration, rapid radiation and crown-group diversification correspond to the aridification of the Australian continent since the Miocene? 6) What influence did morphological traits of the fruiting perianth may have had on dispersal? 7) What causes may have led to the emergence of species groups of varying species richness?


**Figure 1**. Australian landscapes and fruiting perianths of the Australian Camphorosmeae. A: Desert Lake habitat with mats of Roycea pycnophylloides, Mortlock River salt lake system, Western Australia; B: Open Eucalyptus Woodland vegetation with Chenopod understory, Goldfields-Esperance region of Western Australia; C: *Maireana atkinsiana*; D: *Maireana georgei*; E: *Maireana lobiflora*; F: *Sclerolaena burbidgeae*; G: *Sclerolaena eurotioides*; H: *Sclerolaena alata*. Image Source: A, B, E by Philipp Hühn, Mainz; C, D, F-H by Christopher J. French, Perth.


**Figure 1 continued**. Australian landscapes and fruiting perianths of the Australian Camphorosmeae. I: Sand dune on the edge of a salt lake with low chenopodiaceous shrubs (*Tecticornia*, *Maireana*, *Atriplex*, *Malacocera*) in the Winninowie Conservation Park, South Australia; J: Stony Desert (gibber plain) habitat with *Maireana*, *Atriplex* and *Acacia* in the Flinders Ranges east of Lake Torrens, South Australia; K: *Didymanthus roei*; L: *Dissocarpus paradoxus*; M: *Enchylaena tomentosa*; N: *Eremophea aggregata*; O: *Eriochiton sclerolaenoides*; P: *Malacocera tricornis*. Image Source: I, J, M, O, P by Philipp Hühn, Mainz; K, N by Christopher J. French, Perth; L by Mark Marathon, CC BY-SA 3.0. Image sources and copyright information are given in section 9.


**Figure 1 continued**. Australian landscapes and fruiting perianths of the Australian Camphorosmeae. Q: Murray River with adjacent flood plains, open Eucalyptus and Acacia woodlands in the Chowilla Game Reserve, South Australia; R: Karst Plain habitat of the Nullarbor Plain with Chenopod shrubland (*Maireana* and *Atriplex*), S: *Neobassia proceriflora*, T: *Osteocarpum dipterocarpum*; U: *Roycea spinescens* (male) in flower; V: *Roycea pycnophylloides* in flower and fruit; W: *Threlkeldia diffusa*. Image Source: Q, U, V by Philipp Hühn, Mainz; R by State of Western Australia, Copyright Act 1968; S, T by Kym Nicolson, CC BY 4.0; W by Sue Jaggar, CC BY 4.0. Image sources and copyright information are given in section 9.

# *2. MATERIAL AND METHODS*

## *2.1 Sampling*

A total of 106 species representing all 12 recognized genera of the Australian Camphorosmeae, with two representatives of the extra-Australian Camphorosmeae (*Eokochia saxicola* (Guss.) Freitag & G.Kadereit from the Central Mediterranean area and *Grubovia dasyphylla* (Fisch. & C.A.Mey.) Freitag & G.Kadereit from Central Asia) were included in this study (Table 1, Supplement 1). Leaf material was preserved in silica gel for DNA extraction.

**Table 1**. Taxon sampling. Given are the genus, epithet and authority of all species sampled, as well as the voucher information (herbarium ID, collector, collection ID). In case of TERN (Terrestrial Ecosystem Research Network, University of Queensland, Brisbane) and BushBlitz (Australian Government, Department of Climate Change, Energy, the Environment and Water) samples, the voucher ID corresponds to the collection ID. The number of species sampled in this study relative to the number of species described for each genus are given in parentheses.

| *Taxon* | *Authority* | *Voucher ID* | *Collector/Collection ID* |
|---|---|---|---|
| *Didymanthus (1/1)* | Endl. | | |
| *Didymanthus roei* | Endl. | MJG 028809 | Hühn,P. et al. PS_PH_016 |
| *Dissocarpus (3/4)* | F.Muell. | | |
| *Dissocarpus biflorus* | (R.Br.) F.Muell. | MJG 019199 | McDonald,J.T. 1505/51 |
| *Dissocarpus fontinalis* | Paul G.Wilson | MJG 019223 | McDonald,J.T. 1506/9A-1 |
| *Dissocarpus paradoxus* | (R.Br.) F.Muell. | NSA07586 | TERN |
| *Enchylaena (2/2)* | R.Br. | | |
| *Enchylaena lanata* | Paul G.Wilson | MJG 028812 | Hühn,P. et al. PS_PH_020 |
| *Enchylaena tomentosa* | R.Br. | NSA013625 | TERN |
| *Eokochia (1/1)* | Freitag & G.Kadereit | | |
| *Eokochia saxicola* | (Guss.) Freitag & G.Kadereit | MJG 027671 | Kadereit,G. s.n. |
| *Eremophea (2/2)* | Paul G.Wilson | | |
| *Eremophea aggregata* | Paul G.Wilson | MJG 028773 | Hühn,P. et al. PS_PH_033 |
| *Eremophea spinosa* | (Ewart & O.B.Davies) Paul G.Wilson | NTA016323 | TERN |
| *Eriochiton (1/1)* | (R.H.Anderson) A.J.Scott | | |
| *Eriochiton sclerolaenoides* | (F.Muell.) A.J.Scott | MJG 027696 | McDonald,J.T. 1501/29A |
| *Grubovia (1/3)* | Freitag & G.Kadereit | | |
| *Grubovia dasyphylla* | (Fisch.&C.A.Mey.)Freitag&G.Kadereit | MJG 011708 | Miehe,G. & Miehe 96-203-02 |
| *Maireana (43/57)* | Moq. | | |
| *Maireana amoena* | (Diels) Paul G.Wilson | MJG 028836 | Hühn,P. et al. PK_PH_044 |
| *Maireana aphylla* | (R.Br.) Paul G.Wilson | MJG 026701 | McDonald,J.T. 1411/46 |
| *Maireana appressa* | (Benth.) Paul G.Wilson | MJG 026834 | McDonald,J.T. 1505/110 |
| *Maireana atkinsiana* | (W.Fitzg.) Paul G.Wilson | MJG 028837 | Hühn,P. et al. PK_PH_011 |
| *Maireana brevifolia* | (R.Br.) Paul G.Wilson | MJG 028772 | Hühn,P. et al. PK_PH_022 |
| *Maireana campanulata* | Paul G.Wilson | MJG 026092 | McDonald,J.T. 1506/50 |
| *Maireana cannonii* | (J.M.Black) Paul G.Wilson | MJG 025924 | McDonald,J.T. 1508/63 |

| | | | |
|---|---|---|---|
| *Maireana carnosa* | (Moq.) Paul G.Wilson | WAA012606 | TERN |
| *Maireana ciliata* | (F.Muell.) Paul G.Wilson | MJG 026736 | McDonald,J.T. 1409/31 |
| *Maireana convexa* | Paul G.Wilson | MJG 028831 | Hühn,P. et al. PK_PH_014 |
| *Maireana coronata* | (J.M.Black) Paul G.Wilson | NSA011389 | TERN |
| *Maireana dichoptera* | (F.Muell.) Paul G.Wilson | QDA000652 | TERN |
| *Maireana enchylaenoides* | (F.Muell.) Paul G.Wilson | MJG 028825 | Hühn,P. et al. PK_PH_055 |
| *Maireana eriantha* | (F.Muell.) Paul G.Wilson | SAT002476 | TERN |
| *Maireana erioclada* | (Benth.) Paul G.Wilson | VCA007109 | TERN |
| *Maireana eriosphaera* | Paul G.Wilson | MJG 028761 | Hühn,P. et al. PK_PH_020 |
| *Maireana excavata* | (J.M.Black) Paul G.Wilson | MJG 026124 | McDonald,J.T. 1511/35 |
| *Maireana georgei* | (Diels) Paul G.Wilson | MJG 027725 | McDonald,J.T. 0087096 |
| *Maireana glomerifolia* | (F.Muell. & Tate) Paul G.Wilson | MJG 028838 | Hühn,P. et al. PK_PH_009 |
| *Maireana humillima* | (F.Muell.) Paul G.Wilson | MJG 026106 | McDonald,J.T. 1511/37 |
| *Maireana integra* | (Paul G.Wilson) Paul G.Wilson | NSA07271 | TERN |
| *Maireana lanosa* | (Lindl.) Paul G.Wilson | QDA007093 | TERN |
| *Maireana lobiflora* | (F.Muell. ex Benth.) Paul G.Wilson | NSA01180 | TERN |
| *Maireana marginata* | (Benth.) Paul G.Wilson | MJG 028847 | Hühn,P. et al. PK_PH_004 |
| *Maireana oppositifolia* | (F.Muell.) Paul G.Wilson | MJG 019212 | McDonald,J.T. 1501/49 |
| *Maireana ovata* | (Ising) Paul G.Wilson | NTA004902 | TERN |
| *Maireana pentagona* | (R.H.Anderson) Paul G.Wilson | SAA005398 | TERN |
| *Maireana pentatropis* | (Tate) Paul G.Wilson | NSA010465 | TERN |
| *Maireana planifolia* | (F.Muell.) Paul G.Wilson | WAA004030 | TERN |
| *Maireana platycarpa* | Paul G.Wilson | MJG 028826 | Hühn,P. et al. PK_PH_035 |
| *Maireana polypterygia* | (Diels) Paul G.Wilson | MJG 028816 | Hühn,P. et al. PS_PH_027 |
| *Maireana pyramidata* | (Benth.) Paul G.Wilson | NSA013527 | TERN |
| *Maireana radiata* | (Paul G.Wilson) Paul G.Wilson | MJG 027694 | McDonald,J.T. 1501/78 |
| *Maireana rohrlachii* | (Paul G.Wilson) Paul G.Wilson | AUG000132 | TERN |
| *Maireana schistocarpa* | Paul G.Wilson | NSA013425 | TERN |
| *Maireana scleroptera* | (J.M.Black) Paul G.Wilson | MJG 027723 | McDonald,J.T. 0076456 |
| *Maireana sedifolia* | (F.Muell.) Paul G.Wilson | MJG 028820 | Hühn,P. et al. PK_PH_025 |
| *Maireana spongiocarpa* | (F.Muell.) Paul G.Wilson | SAA000753 | TERN |
| *Maireana suaedifolia* | (Paul G.Wilson) Paul G.Wilson | MJG 026188 | McDonald,J.T. 1601/93 |
| *Maireana tomentosa* | Moq. | NSA010831 | TERN |
| *Maireana trichoptera* | (J.M.Black) Paul G.Wilson | MJG 028808 | Hühn,P. et al. PS_PH_015 |
| *Maireana turbinata* | Paul G.Wilson | NSA013511 | TERN |
| *Maireana villosa* | (Lindl.) Paul G.Wilson | MJG 028819 | Hühn,P. et al. PK_PH_016 |
| *Malacocera (3/4)* | R.H.Anderson | | |
| *Malacocera albolanata* | (Ising) Chinnock | MJG 029141 | Hühn,P. JQG_516 |
| *Malacocera gracilis* | Chinnock | MJG 029143 | Hühn,P. APA_PH_043 |
| *Malacocera tricornis* | (Benth.) R.H.Anderson | NSA013825 | TERN |
| *Neobassia (1/2)* | A.J.Scott | | |
| *Neobassia proceriflora* | (F.Muell.) A.J.Scott | NSA01615 | TERN |
| *Osteocarpum (3/5)* | F.Muell. | | |
| *Osteocarpum acropterum* | (F.Muell. & Tate) Volkens | MJG 019190 | McDonald,J.T. 1505/86-1 |
| *Osteocarpum dipterocarpum* | (F.Muell.) Volkens | MJG 019226 | McDonald,J.T. 1506/15 |
| *Osteocarpum salsuginosum* | F.Muell. | SAS001910 | TERN |
| *Roycea (3/3)* | C.A.Gardner | | |

| | | | |
|---|---|---|---|
| *Roycea pycnophylloides* | C.A.Gardner | MJG 028821 | Hühn,P. et al. PK_PH_062 |
| *Roycea spinescens* | C.A.Gardner | MJG 028762 | Hühn,P. et al. PK_PH_052 |
| *Roycea divaricata* | Paul G.Wilson | MJG 028835 | Hühn,P. et al. PK_PH_017 |
| *Sclerolaena (40/65)* | R.Br. | | |
| *Sclerolaena alata* | Paul G.Wilson | MJG 028845 | Hühn,P. et al. PK_PH_012 |
| *Sclerolaena articulata* | (J.M.Black) A.J.Scott | NSABHC06184 | TERN |
| *Sclerolaena bicornis* | Lindl. | MJG 026086 | McDonald,J.T. 1506/34 |
| *Sclerolaena blackiana* | (Ising) A.J.Scott | SAA001813 | TERN |
| *Sclerolaena brachyptera* | (F.Muell.) S.W.L.Jacobs | SAS001876 | TERN |
| *Sclerolaena brevifolia* | (Ising) A.J.Scott | MJG 019214 | McDonald,J.T. 1501/67 |
| *Sclerolaena burbidgeae* | (Ising) A.J.Scott | MJG 028781 | Hühn,P. et al. PS_PH_054 |
| *Sclerolaena convexula* | (R.H.Anderson) A.J.Scott | WAA003994 | TERN |
| *Sclerolaena cornishiana* | (F.Muell.) A.J.Scott | QDA007001 | TERN |
| *Sclerolaena costata* | (R.H.Anderson) A.J.Scott | MJG 028802 | Hühn,P. et al. PS_PH_014 |
| *Sclerolaena cuneata* | Paul G.Wilson | MJG 019220 | McDonald,J.T. 1501/115 |
| *Sclerolaena decurrens* | (J.M.Black) A.J.Scott | NTA016548 | TERN |
| *Sclerolaena densiflora* | (W.Fitzg.) A.J.Scott | WAA009919 | TERN |
| *Sclerolaena deserticola* | Paul G.Wilson | WAA011226 | TERN |
| *Sclerolaena diacantha* | (Nees) Benth. | MJG 026848 | McDonald,J.T. 1501/84 |
| *Sclerolaena divaricata* | (R.Br.) Sm. | NSA013619 | TERN |
| *Sclerolaena drummondii* | (Benth.) Domin | WAA001796 | TERN |
| *Sclerolaena eriacantha* | (F.Muell.) Ulbr. | QDA007895 | TERN |
| *Sclerolaena eurotioides* | (F.Muell.) A.J.Scott | WAA001472 | TERN |
| *Sclerolaena fimbriolata* | (F.Muell.) A.J.Scott | MJG 026273 | McDonald,J.T. 1601/30 |
| *Sclerolaena fusiformis* | Paul G.Wilson | MJG 028842 | Hühn,P. et al. PK_PH_005 |
| *Sclerolaena glabra* | (F.Muell.) Domin | QDA007733 | TERN |
| *Sclerolaena holtiana* | (Ising) A.J.Scott | MJG 019222 | McDonald,J.T. 1506/10 |
| *Sclerolaena intricata* | (R.H.Anderson) A.J.Scott | MJG 026936 | McDonald,J.T. 1505/47 |
| *Sclerolaena johnsonii* | (Ising) A.J.Scott | BS1137-183 | BushBlitz |
| *Sclerolaena lanicuspis* | (F.Muell.) F.Muell | NSA013547 | TERN |
| *Sclerolaena limbata* | (J.M.Black) Ulbr. | MJG 025912 | McDonald,J.T. 1508/86 |
| *Sclerolaena minuta* | (Ising) A.J.Scott | NSA01109 | TERN |
| *Sclerolaena muricata* | (Moq.) Domin | MJG 026111 | McDonald,J.T. 1511/48 |
| *Sclerolaena napiformis* | Paul G.Wilson | MJG 026161 | McDonald,J.T. 1511/26 |
| *Sclerolaena obliquicuspis* | (R.H.Anderson) Ulbr | MJG 019203 | McDonald,J.T. 1411/8A |
| *Sclerolaena parallelicuspis* | (R.H.Anderson) A.J.Scott | NSA013599 | TERN |
| *Sclerolaena parviflora* | (R.H.Anderson) A.J.Scott | BS1137-221 | BushBlitz |
| *Sclerolaena patenticuspis* | (R.H.Anderson) Ulbr | MJG 026693 | McDonald,J.T. 1409/21 |
| *Sclerolaena sp. Yeltacowie* | ineditus | BS1097-408 | BushBlitz |
| *Sclerolaena stelligera* | (F.Muell.) S.W.L.Jacobs | MJG 026818 | McDonald,J.T. 1411/31 |
| *Sclerolaena stylosa* | (Ising) A.J.Scott | MJG 028817 | Hühn,P. et al. PS_PH_034 |
| *Sclerolaena tricuspis* | (F.Muell.) Ulbr. | MJG 025923 | McDonald,J.T. 1605/6 |
| *Sclerolaena uniflora* | R.Br. | SAA004308 | TERN |
| *Sclerolaena ventricosa* | (J.M.Black) A.J.Scott | MJG 026938 | McDonald,J.T. 1505/46 |
| *Threlkeldia (2/2)* | R.Br. | | |
| *Threlkeldia diffusa* | R.Br. | MJG 026729 | McDonald,J.T. 1407/28 A |
| *Threlkeldia inchoata* | (J.M.Black) J.M.Black | SAA000015 | TERN |

### *2.2 DNA extraction and RADseq lab protocol*

Lab work for DNA extraction and RADseq library preparation was carried out following Hühn et al. (2022) (Appendix 1). DNA-extraction was conducted using the DNeasy Plant Mini-Kit (QIAGEN, Venlo, Netherlands) according to the manufacturer's protocol for "Purification of Total DNA from Plant Tissue (Mini Protocol)" with a number of modifications outlined in Appendix 2. The DNA concentration and quality were evaluated using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and gel electrophoresis. A modified RADseq approach for NGS library preparation was used (Appendix 1 and 2). Genomic DNA (200ng per sample) was fragmented using the restriction endonucleases (REase/s) *BamHI* (restriction site: G'GATCC) and *KpnI* (restriction site: GGTAC'C) in a double digest reaction. Following digestion, the fragments were ligated to sample-specific barcode adapters containing equimolar amounts of the REases motif pairs. Reactions were incubated for 3 hours at 37°C. Subsequently, the libraries were multiplexed, purified (NucleoSpin Gel and PCR Clean-up Mini kit by Macherey-Nagel, Düren, Germany) and quantified using a Qubit 3.0 Fluorometer. A Pippin Prep (Sage Science, Beverly, MA, USA) was used for size selection with a segregation range of 350–720 nt. The size-selected product was amplified using a low-cycle 2-step PCR protocol, followed by column-based purification and fluorometric quantification. A second size selection was performed with SPRI magnetic beads (NucleoMag NGS kit, Macherey-Nagel, Düren, Germany) using a ratio of 0.8 bead suspension to one part library. Library quality was validated using fluorometric quantification and assessment of fragment length range with a Bioanalyzer (Agilent, Santa Clara, CA, USA) electropherogram. Sequencing was done using the Illumina MiSeq v3 kit for 300bp PE reads (San Diego, CA, USA) at the Macrogen NGS facility (Seoul, Republic of Korea).

### *2.3 Data assembly*

While RADseq is particularly appealing for studying non-model taxa, data assembly in the absence of genomic information as reference (*de novo* assembly) can be a major challenge (Hühn et al., 2022; Appendix 1). The bioinformatics effort related to RADseq data is often not straightforward and can heavily impact the assembly outcome regarding differentiation of orthologs and paralogs, as well as the quantity of recovered loci, sequence variation (VAR), single nucleotide polymorphisms (SNPs), and parsimony informative sites (PIS) (e.g. Rubin et al., 2012; Ilut et al., 2014; Lee et al., 2018; Hühn et al., 2022). Several assembly pipelines to facilitate data processing have been developed, among others "*ipyrad*" (Eaton and Overcast, 2020). These pipelines implement four main steps. 1) In-sample-clustering (ISC), in which reads within each sample are grouped by sequence similarity into putative loci. 2) Consensus calling of allele sequences from clustered reads. 3) Between-sample-clustering (BSC) of consensus sequences of all loci across all samples are clustered by sequence similarity to generate putatively homologous loci. 4) Data filtering based on given thresholds such as the number of samples per locus required (locus coverage) or the maximum proportion of shared heterozygous sites in a locus (detection of potential paralogs). To determine which reads represent the same genomic locus, a clustering threshold (CT) based on sequence similarity is used. Yet, genetic variation within the target genomes and across the studied taxa makes it difficult to find an appropriate CT that best ensures homology of the assembled loci (e.g. Rubin et al., 2012; Ilut et al., 2014; Harvey et al., 2015; Ilut et al., 2014; Lee et al., 2018; Springer and Gatesy, 2018; McCartney-Melstad et al., 2019). Approaches to facilitate this problem (CT selection approaches) aim at the determination of suitable CTs for homology assessment by analyzing trends of several assembly metrics over a wide range of tested CTs (e.g. Ilut et al., 2014; Mastretta-Yanes et al., 2015; Paris et al., 2017; McCartney-Melstad et al., 2019; Hühn et al., 2022). To determine CTs for ISC and BSC, Hühn et al. (2022) used several CT selection

approaches as guidance (Ilut et al., 2014; Mastretta-Yanes et al., 2015; Paris et al., 2017; McCartney-Melstad et al., 2019) and defined assumptions to select suitable CTs. 1) Over- and undermerging ranges have to be identified to avoid merging/splitting effects within these areas. 2) Overmerging is indicated by highly heterozygous alleles with a high proportion of filtered paralogs (Ilut et al., 2014; McCartney-Melstad et al., 2019). Hence, a suitable CT is expected in an increasing area of heterozygosity and a decreasing area of flagged paralogs, between the maxima of both metrics. 3) Undermerging of orthologs leads to an increased number of loci (and lower locus coverage in ISC, lower sample coverage per locus in BSC) while sequence divergence among taxa decreases (Mastretta-Yanes et al., 2015; McCartney-Melstad et al., 2019). Thus, sequence variation declines while missingness increases. This RADseq introductory section was largely excerpted from the study by Hühn et al. (2022), which describes the difficulty of data assembly in more detail (see Appendix 1).

RADseq data assembly was carried out following Hühn et al. (2022). Raw data was demultiplexed using *ipyrad* v0.9.52 (Eaton and Overcast, 2020). No mismatch in the barcode sequence was allowed. Adapter trimming was done using cutadapt version 2.3 (Martin, 2011). We used FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016) for read quality validation. For *de novo* data assembly, we used *ipyrad* v0.9.52 (Eaton and Overcast, 2020). We used default assembly settings except for the following settings: The assembly method was set to denovo-reference, using three Amaranthaceae plastomes to remove reads of plastid origin (*Amaranthus viridis* L., NCBI accession: PRJNA720789, Ding et al., 2021; *Beta vulgaris* L., NCBI accession: PRJEB45680, De Marchis et al., 2009; *Chenopodium acuminatum* Willd.NCBI accession: PRJNA671764, Wariss and Qu, 2021). The thresholds for statistical and majority rule base calling were raised from six to ten, the datatype was defined as “pairgbs”, we selected strict adapter filtering, required a minimum locus length of 200 nt and raised the maximum number of indels per locus from eight to 24 (due to the extended sequencing range).

For a detailed description of the assembly parameters and the selected thresholds see the *ipyrad* documentation (Eaton and Overcast, 2020; https://ipyrad.readthedocs.io, revision 1f88f521) and Hühn et al. (2022), respectively.

To determine suitable CTs for ISC and BSC, the customized CT selection approach by Hühn et al. (2022) was used. For ISC threshold selection, the demultiplexed samples were clustered using a CT range of 0.88–0.96. The heterozygosity and the proportion of detected paralogs were recorded for each sample individually, plotted across the CT range and evaluated with respect to a balance between both metric maxima, as suggested by Ilut et al. (2014) and McCartney-Melstad et al. (2019). Suitable thresholds for ISC were averaged to consensus values for each sample. Following ISC threshold selection, the samples were merged and clustered across all accessions for BSC threshold selection (CT range: 0.88–0.96). The assembly output statistics were evaluated with respect to the proportion of duplicates and putative paralogs detected, the proportion of missingness in the SNPs and sequence matrix, and the number of new polymorphic loci (NPL, Paris *et al*., 2017) between adjacent CTs. The BSC threshold was selected with respect to a minimized missingness in the data matrix within a CT-range of decreasing duplicates and paralogs detected and a maximized NPL proportion (Ilut et al., 2014; Paris et al., 2017; McCartney-Melstad et al., 2019; Hühn et al., 2022).

Three different assemblies were generated (Supplement 1): The first assembly contained all 104 ingroup samples plus two outgroup samples (Assembly 1). Based on the resulting phylogeny of the first assembly, the species were split into two smaller assemblies, the *Maireana* grade assembly (Assembly 2) and the *Sclerolaena* clade assembly (Assembly 3). This was intended to improve the resolution of the partial samplings. Assembly 2 contained the two outgroup samples, 66 samples of the *Maireana* grade and 11 representatives of the *Sclerolaena* clade. Assembly 3 contained 44 samples of the *Sclerolaena* clade and the 8 species from clade 11 as outgroup.

### *2.4 Phylogenetic inference*

Phylogenetic inference of concatenated super-matrices using maximum likelihood (CA-ML) is under criticism since it violates the multi-species coalescent (MSC) model can lead to poorly resolved, incomplete, or positively misleading phylogenetic estimates (e.g. Degnan and Rosenberg, 2006, 2009; Kubatko and Degnan, 2007; Roch and Steel, 2015; Xu and Yang, 2016; Rannala et al., 2020). However, a growing number of studies has shown that CA-ML can be comparably or more accurate than coalescent-based methods (CB-SM, which account for gene tree discordance caused by biological processes such as ILS, hybridization/introgression and gene duplication/loss events) under various conditions of locus linkage, locus length, phylogenetic information content, missingness and ILS, which usually hamper accurate phylogenetic inference (e.g. Chou et al., 2015; Roch and Warnow, 2015; Mirarab et al., 2016; Springer and Gatesy, 2016; Long and Kubatko, 2018; Molloy and Warnow, 2018). Moreover, inference of empirical data using CA-ML and coalescent-based methods generally yielded congruent results (e.g. Hosner et al., 2016; Sayyari et al., 2017; Curto et al., 2018; Rancilhac et al., 2019). Due to the size of the resulting data matrix and the high proportion of missing data (see 3.1), phylogenetic inference using CB-SM was not performed since high levels of missingness decrease the estimation accuracy of CB-SM (Chou et al., 2015; Roch and Warnow, 2015; Xi et al., 2015, 2016; Xu and Yang, 2016; Sayyari et al., 2017; Molloy and Warnow, 2018). This section was largely excerpted from the study by Hühn et al. (2022), which describes the applicability of different inference methods in more detail (see Appendix 1).

Maximum likelihood (ML) phylogenies of the concatenated sequence matrices were inferred using RAxML-NG v0.9.0 (Kozlov *et al*., 2019) under the GTR+GAMMA model. ML tree search was initiated with 10 random and 10 parsimony trees. Statistical support was assessed using 100 Felsenstein's bootstrapping replicates (Felsenstein, 1985) as implemented in RAxML-NG.

### *2.5 Molecular dating*

Divergence time estimation for the Camphorosmeae phylogeny was carried out using a previously published age estimate for the split of the Central Asian clade of *Grubovia dasyphylla* plus *Eokochia saxicola* from the Australian Camphorosmeae (i.e., 19.0 million years (myr); 95% HPD = 14.7–24.0 myr) as calibration point (Kadereit and Freitag, 2011; Kadereit et al., 2014). A second calibration point was pre-defined for the divergence of the clade comprising *M. ciliata* through *M. cannonii* from the clade of *N. proceriflora* and *Sclerolaena*, i.e., 5.0 myr (95% HPD = 3.9–6.2 myr; Cabrera et al., 2011).

The dating analysis was performed using BEAST v2.6.4 (Bouckaert et al., 2019). The topology of the Camphorosmeae phylogeny was fixed by setting the weight of the BEAST operators *subtree slide*, *narrow exchange*, *wide exchange* and *Wilson Balding* to zero. A lognormal relaxed clock and birth-death process were used for calculations. The ucld.mean parameter was defined as gamma-distributed with 0.001 and 1000 as lower and upper bounds, respectively, and the GTR+I+Γ substitution model was used. Four independent runs were performed with between 50 and 77 million generations per run, and the resulting log files were checked for convergence using Tracer v1.7.1 (Rambaut et al., 2018). Runs were combined using LogCombiner v2.6.3 (Bouckaert et al., 2019) pre-defining a burn-in of 10%, and a maximum clade credibility (MCC) tree was constructed using TreeAnnotator v.2.6.3 (Bouckaert et al., 2019) and inspected using FigTree v1.4.4 (Rambaut, 2018).

High levels of missing data can decrease the estimation accuracy of dating analyses using BEAST (Wiens and Morrill, 2011; Zheng and Wiens, 2015). To decrease the level of missing data and thus, to improve the accuracy of the BEAST analysis, the data matrix of Assembly 1 was filtered regarding a minimum locus coverage of 10 (minimum number of samples per locus).

### *2.6 Characters and character states*

Cabrera et al. (2009) investigated 15 morphological (vegetative and reproductive) characters for their systematic relevance. Of these, only the fruiting perianth appendage type showed a pattern partly supporting the phylogenetic clades found. Due to the insufficient resolution of the resulting phylogenies, however, it remained unclear if also other characters may be of systematic relevance. Hence, some of the already investigated characters were re-investigated in this study and complemented by more detailed fruiting perianth characters, as these have been used to describe, delineate and revise genera, sections and species complexes in the Australian Camphorosmeae (Mueller, 1882a, b; Ulbrich, 1934; Anderson, 1923; Ising, 1964; Wilson, 1975, 1984; Scott, 1978; Jacobs, 1988). In addition, ecological characteristics were also examined, as these had not yet been investigated for their systematic relevance or had only recently been collected (colonized habitat types, published by McDonald, 2020). A total of 25 vegetative, reproductive and ecological characters were examined for their systematic relevance (Table 2, Supplement 2). Of these, 13 "general" characteristics were surveyed for all species and 12 “appendage type specific” characters were surveyed with respect to the two large sub-groups “wing-like” and “spine-like” (Table 2).

The vegetative characters include the habit, growth height, pubescence of branches and leaves and phyllotaxy. For the evaluation of the habit, five habit descriptions were combined into three character states (Table 2, Supplement 2). The habit state “annual” was kept as “annual”. Descriptions of "perennials" (including just perennial, perennial herb and perennial with a woody base) were summarized under “perennial” as in these cases no clear separation is possible. The description “shrub” was kept as “shrub”. The information on the habit is inconsistent. However, it can be assumed that a basic distinction is made between presumably "short-lived perennials" (coded as perennials) and presumably "long-lived shrubs" (coded as shrubs; Scott, 1978; Wilson, 1984; McDonald, 2020). The reproductive characters include the

distribution of sexes, number of flowers per bract, fruiting perianth appendage type, shape and structure of the fruiting perianth tube, and the seed orientation. Ecological characteristics include the preferred soil (substrates) and habitat types. The information on preferred soil types and soil salinity is also inconsistent. In case of species that are not explicitly described to grow in saline soils and hence supposedly grow in non-saline soils, it can not be distinguished between a lack of information or an actual salinity preference. Therefore, such information was classified as “not applicable” (Supplement 2).

Vegetative and reproductive characteristics were recorded and classified based on the species descriptions by Wilson (1975, 1984), Scott (1978) and Chinnock (1980), using the generic classification by Wilson (1984) and by Jacobs (1988), who subsumed *Stelligera endecaspinis* A.J.Scott and *Sclerochlamys brachyptera* F.Muell. into *Sclerolaena*. Information on preferred soil types was taken from the "Flora of Australia, Volume 4, Phytolaccaceae to Chenopodiaceae” (Wilson, 1984), the "Flora of New South Wales, Volume 4" (Harden, 1990), the "Flora of Victoria, Volume 3" (Walsh, 1996) and “The Western Australian flora: a descriptive catalogue” by Paczkowska and Chapman (2000). Information on colonized habitats was taken from McDonald's "Biogeography of Australian chenopods" (McDonald, 2020). Classification and description of habitats is based on Mabbutt (1988) and McDonald (2020) (Table 3). Species distribution ranges were taken from the Flora of Australia (Wilson, 1984) and compared and extended using occurrence records maps (preserved specimens only) provided by the Atlas of Living Australia (http://www.ala.org.au; Belbin et al., 2021). Distribution areas were classified based on the terrestrial phytogeographic subregions by Ebach et al. (2015) (Supplement 2).

**Table 2.** Characterization. Characters and character states evaluated for the Australian Camphorosmeae.

| Character | Character state |
|---|---|
| general characteristics (all samples) | |
| Habit | annual, perennial, shrub |
| Height | decumbent, up to 50 cm, up to 100 cm, larger than 100 cm |
| Phyllotaxy | alternate, opposite |
| Branch pubescence | glabrous, pubescent |
| Leaf pubescence | glabrous, pubescent |
| Distribution of sexes | bisexual, dioecious, polygamodioecious |
| Flowers per bract | solitary, paired, aggregated |
| Perianth appendage type | wing, spine, no obvious appendage, wing-spine intermediate, Y-configuration |
| Perianth tube structure | smooth, ribbed |
| Seed orientation | horizontal, erect oblique |
| Soil type | rocky, sandy<br>heavy, loamy, clayey<br>calcareous, gypseous |
| Soil salinity | slightly saline, moderately saline, saline |
| Habitat type | Sand Desert, Desert Upland, Stony desert, Shield Plain, Riverine desert, Desert Clay Plain, Karst Plain, Desert Lake, mesic Plain and Range, Coast |
| wing-like appendages (*Maireana* grade) | |
| Perianth tube shape | flat, cylindrical, turbinate, convex, cup-shaped, hemispherical, sub-globular |
| Wing type | simple, lobed |
| Wing orientation | horizontal, vertical |
| Wing venation | absent, present |
| Radicular slit | absent, present |
| Upper perianth shape | flat, pentagonal or discoid, convex, prominently lobed or thickened |
| Upper perianth outgrowth | absent, present |
| spine-like appendages (*Sclerolaena* clade) | |
| Attachment shape | circular, elliptic, oblique |
| Perianth tube shape | cylindrical, oblong, dorsiventrally-compressed, urceolate, turbinate, convex, cup-shaped, hemispherical, sub-globular |
| Number of spines | two, three, four, five, six |
| Limb | absent, inconspicuous, prominent |
| Number of stamens | three, four, five |

### *2.6.1 Fruiting perianth morphology*

The traits of the fruiting perianth appendage type are of special interest, as these characters mainly have been used for generic delimitation (Mueller, 1882a, b; Ulbrich, 1934; Anderson, 1923; Ising, 1964; Wilson, 1975, 1984; Scott, 1978). These traits were surveyed and assigned with respect to the two general perianth appendage types "wing-like" or "spine-like" (Table 2, Supplement 2).

In species of the Australian Camphorosmeae (Fig. 1), the flowers are usually solitary and axillary, and the fruiting perianth is usually enlarged, hardened and bears appendages (Wilson, 1984). The perianth is normally 5-merous and consists of a saucer- to cup-shaped basal portion and a 5-lobed upper portion (Wilson, 1975). The perianth bears two series of veins, 5 tepaline and 5 intertepaline in position. As the fruit develops, the basal portion of the perianth usually enlarges and becomes variously thickened. Outgrowths often develop from one or the other series of veins, or just below the lobes, a wing-like structure may form which encircles the perianth. The embryo is always positioned in such a manner that the radicle is pointing at a particular point of the perianth at which the perianth normally develops a slit (radicular slit), bulge, ridge, canal, or one or a pair of tubercles and spines (radicular spines).

In *Maireana* species with a typical wing-like fruiting perianth appendage type, the fruiting perianth consists of two parts: the perianth tube, which develops the perianth appendage and the upper perianth (Fig. 2, G). The perianth greatly expands in fruit and becomes spongy, leathery, crustaceous or woody, thus forming the perianth tube (Wilson, 1984). The 5 perianth lobes are usually deeply divided facing the radicle, sometimes extending down the perianth tube (radicular slit). The variously modified perianth tube develops perianth appendages, usually a simple, horizontal wing or 5 separate wings (5-lobed wing) which grow from the base of each lobe. Some species bear additional vertical, wing-like structures along or from the base of the perianth tube. The part of the perianth above the tube (and between the wing-like

appendage) is here referred to as the “upper perianth”. The upper perianth is usually formed by the horizontal perianth lobes and obscures the utricle (seed usually horizontal). Accessory processes, such as erect lobes (Fig. 1 C, E and Fig. 3 R) or floccose wool (Fig. 3 Q), can be produced by the upper perianth. These processes can be tepaline or intertepaline in position. In the latter case, the processes presumably correspond to the spines found in *Sclerolaena*. For clarity, "process" (in a morphological context) refers to an additional appendage to the major (wing- or spine-like) fruiting perianth appendage (Wilson, 1975, 1984).

In *Sclerolaena* species with a typical spine-like perianth appendage type, the fruiting perianth consists of three parts: the perianth tube, the perianth appendage and the limb (Fig. 2, A). The crustaceous to woody fruiting perianth envelopes the utricle as a tube. The base of the tube is often hollow and filled with moist tissue when fresh. The connection between the tube base and the shoot axis is referred to as the “attachment”. The tube can be ribbed and usually has a slit or tubercle facing the radicle (seed usually erect to horizontal). The fruiting perianth produces usually 2-6 intertepaline spines, of which a contiguous pair is borne opposite the radicle (radicular spines). The 3-5 perianth lobes and the tube portion above the spines can be prominently developed (as a receptacle for the radicle when turned upwards). This division is called “limb” and corresponds to the upper perianth in *Maireana* species.

Besides the two major appendage modifications, wings and spines, there are comparatively rare other appendage modifications, as in *Dissocarpus*, *Enchylaena*, *Eremophea*, *Malacocera*, and *Roycea*. In *Roycea*, the perianth lobes do not enlarge in fruit (Fig. 1 V and Fig. 2, H). The (sub-) globular fruit is surrounded by the perianth at the base. In *Enchylaena*, the (sub-) globular fruiting perianth is succulent and the short, succulent perianth lobes obscure the ovary (Fig. 1 M and Fig. 2, F). A short, succulent corona can extend upwards the perianth. This succulent outgrowths corresponds to the wing-like appendage found in *Maireana*. In *Malacocera*, the flat fruiting perianth produces 3-5 tepaline processes attached to the length of

the perianth tube, forming an (inverted) Y or star-shaped configuration (Fig. 1 P and Fig 2, C). In contrast to the usually solitary and axillary flowers found in the majority of species, the perianths of two or many flowers of a leaf axil can be fused together at their bases to form woody fruiting perianth pairs or clusters (*Dissocarpus*, *Eremophea*) (Fig 1 L, N and Fig. 2, D and E). In these cases, spine-like perianth appendages arise opposite the perianth lobes (tepaline in position) or from the fused perianth bases.

In some cases, the perianth appendages are neither typical wing-like nor typical spine-like. Some species have a woody perianth with usually five (*S. brachyptera*) or ten (*S. stelligera*, Fig. 3, L) horizontal spines which may be united at their base into a short horizontal wing. The tepaline spines can be flattened and horizontal (*S. alata*, Fig. 1 H and Fig. 4 K,). Or species produce two series of appendages: five erect appendages which arise opposite the perianth lobes, alternating with horizontal appendages that are intertepaline in position (*S. fimbriolata*, Fig. 4, R) or five erect tepaline appendages that produce five spinescent appendages (*Eri. sclerolaenoides*, Fig. 1 O).

### *2.6.2 Characterization of the fruiting perianth*

For "wing-like" fruiting perianths, the shape of the wing (simple or lobed), the wing orientation relative to the perianth tube (horizontal or vertical), the presence of a prominent wing venation and a radicular slit, the shape of the upper perianth, and the presence of upper perianth processes (erect lobe-like processes or floccose wool arising from the upper perianth) were evaluated (Table 2; Fig. 3; Supplement 2). Rare tube shapes were assigned to more general shapes, i.e.: depressed barrel-shaped (*M. marginata*, *M. excavata*), broadly ovoid (*R. pycnophylloides*) and broadly obconical (*M. oppositifolia*) were summarized under the convex tube shape. For "spine-like" fruiting perianths, the shape of the attachment, the number of spines, the presence of a limb and the number of stamens were evaluated (Table 2; Fig. 4; Supplement 2). Rare perianth tube shapes were assigned to more general shapes, i.e. pear-

shaped (*S. diacantha*, *S. uniflora*), barrel-shaped (*S. ventricosa*, *S. blackiana*), ovoid (*T. diffusa*) and obovoid (*T. inchoata*) were summarized under the convex tube shape.

### *2.6.3 Habitat types*

Australian desert landscapes have been mapped and described by Mabbutt (e.g. 1969, 1977, 1984, 1988) and have been used to describe and explain associated vegetation, desert ecology and arid geomorphic processes (Cunningham, 1981; Fox 1999; Whitford, 2002; Yang and Goudie, 2007). The interaction of climate with base geology and geomorphology determines soil characteristics and impacts ecosystem processes and the nature and distribution of vegetation (Whitford, 2002). McDonald (2020) used the abundance of Australian chenopods on eight arid landscapes to describe the preferred habitat types of chenopod species. Two more habitats were added by McDonald (2020) to cover temperate to sub-tropical (Mesic Plain & Range) and coastal (Coast) species distributions in adjacent areas of the arid zone (Table 3). By mapping these land type occupations on the phylogenetic tree, clade-specific shifts and adaptations may be observed and may be put in relation to chronological diversification events. The species-specific habitat type occupations were taken from McDonald (2020). The definitions of habitat types (Table 3 and Appendix 3) were largely adopted from McDonald (2020), who based the descriptions on Stace et al. (1968), Mabbutt (1988) and McKenzie et al. (2004).

**Table 3**: Habitat types. Description of habitat types as defined by Mabbutt (1988) and McDonald (2020). Given are the habitat types with abbreviations in parentheses and characteristic geomorphological and floristic features, as well as a geochronological explanation of their origins. Descriptions were largely adopted from McDonald (2020). More detailed descriptions are given in Appendix 3.

| Habitat Type | Landforms | Soil/Salinity/Vegetation | Origin |
|---|---|---|---|
| Sand Desert (SD) | Aeolian dune fields and sandplains; discontinuous or in long, linear dunes; stable slopes under vegetation cover, with mobile sand on dune crests. | Hummock grassland, *Acacia* spp. tall shrubland or *Eucalyptus* spp. open woodland is major vegetation. Saline where thin sand sheets cover older landscapes. | Formation of dune fields is associated with substantial denundation of landscapes (Twidale and Wopfner, 1990; Laity, 2008). |
| Desert Upland (DU) | Rocky ranges, hillslopes and tablelands of sedimentary rock (sandstone and quartzite); includes major uplands such as Flinders Ranges and the Kimberley Plateau. | Sclerophyllous woodland, *Acacia* shrubland and hummock grassland. Thin, stony soils; non-saline and non-alkaline except on limestone substrates. Saline or alkaline sections downslope. | Uplands are younger than their constituent rock types. Major uplands in place through the Neogene. |
| Stony Desert (SP) | Dissected tablelands and plains of pebbly pavements (gibbers); complex drainage networks due to lack of infiltration that causes high surface run-off (McKenzie et al., 2004). | Low grassy/chenopod formations (Brandle, 1998) due to extreme exposure and very dry conditions. Clayey soils, moderately to strongly saline; much varied lime and gypsum. | Maximum Australian development occurred in areas of least rainfall during global dry phase (Fujioka and Chappell, 2009). |
| Shield Plain (SH) | Subdued, rocky landscapes of igneous and metamorphic rocks in Western, Central and South Australia, that has sandplain mantles. | Vegetation similarly varied with sclerophyllous woodland, shrubland and grassland. Varied soils being neutral, saline or calcareous, deep or shallow with surface sand or clay. | Cratonic platform that retains past elements of peneplanation and deep-weathering. |
| Riverine Desert (RL) | Desert drainage systems mainly in Lake Eyre and Murray Basins. Occasional channel flows with infrequent inundation of extensive floodplains. | Vegetation open and arborescent along channels and floodplains; grassy and chenopod formations in broad flats; extensive chenopod low shrubland on dry alluvial plains. Regularly inundated features less saline than more distant plains of mild to moderate salinity. | Emerged in the west in MiddleMiocene when palaeovalleys ceased to flow (van de Graaff et al., 1977). |
| Desert Clay Plain (CP) | Broad, alluvial plains of heavy soil in north-east Australia. | Open *Astrebla* spp. grasslands; chenopod species in groundcovers. Strongly alkaline or mildly saline dark clays. | Tertiary lacustrine and alluvial clay deposits, exposed in the later Neogene (Edgoose, 2003). |

| | | | |
|---|---|---|---|
| Karst Plain (KP) | Undulating plains of pedogenic sheet calcrete across southern Australia. Bunda Plateau on the Nullarbor Plain is a flat plain of low elevation on crystalline limestone. | Prominently chenopod (*Atriplex* and *Maireana*); limestone taxa dominate shrub layers and groundcovers. Shallow, alkaline loams over thin calcrete; saline near the coast. | The Nullarbor Plain has always been arid. The atypically flat surface of the Bunda Plateau is interpreted as evidence of little rainfall since the Late Miocene uplift (Benbow et al., 1995). |
| Desert Lake (DL) | Disconnected, saline complexes in pans, flats with marginal lunettes; vast lakes of regional lowlands; highly saline lowlands of temperate plains. | Vegetation zones of different chenopod taxa across sharp salinity gradients from bare salt crust to slight salinity on margins. | Desert Lake emerged from Riverine Desert landscapes, forming disconnected flats in the vast, flat Australian continent. |
| Mesic Plain & Range (MP) | Temperate to subtropical ranges and plains of eastern, southern and western continental margins; includes igneous peneplain, alluvial plain and fixed dunes. | Sclerophyllous vegetation dominant; chenopod taxa in very seasonal climates or on heavy soils in drier areas. Soils commonly acid-neutral trend, but alkaline in drier regions. | Present since the Early Miocene. |
| Coast (CS) | Beach-dune complexes; coastal lagoons and estuaries; rocky shores or cliffs; humid northern and eastern coast has more estuarine habitat than arid western and southern coasts. | Coastal shrubland, heath and saltmarsh. Marine chemistry determines saline characteristics; dominated by sodium chloride. | The arid and humid coastal contrast likely developed over the Neogene. |

# *3. RESULTS*

## *3.1 Sampling*

Circa 70% of all Australian Camphorosmeae species were sampled for this study (103 out of 147 described species; Wilson, 1984). Representatives of all genera were included (Table 1, Supplement 1). The sampling well reflects the morphological and ecological variation of Australian Camphorosmeae species (Supplement 2). However, some regions of Australia's arid zone are relatively under-represented. For instance, the north-western of Western Australia (including the Pilbara and Great Sandy Desert Interzone), the Central Desert spanning the north from Western Australia eastwards across the Northern Territory, and the eastern fringes of the Eastern Desert that expand into the subtropical and temperate areas of Central Queensland, Eastern Queensland and the Southeastern region. The phylogenetic classification of some species that have a distinctive fruiting perianth morphology may help to clarify the evolution of the fruiting perianth and associated life history traits in future studies. This involves: *Maireana luehmannii* (F. Muell.) Paul G. Wilson (5 woody, irregularly curved wings with spiny lobes, similar to *S. fimbriolata* and *S. alata*); *Maireana melanocoma* (F. Muell.) Paul G. Wilson (upper perianth covered with numerous erect, hair-like processes, similar to species in clades 2 and 3); *Maireana prosthecochaeta* (F. Muell.) Paul G. Wilson (upper perianth with erect, linear processes, similar to species in clades 2 and 3); *Maireana triptera* (Benth.) Paul G. Wilson (5 vertical wings attached to the length of the perianth tube fuse with horizontal wing at tube apex, similar to *M. polypterygia*, *M. erioclada* and *M. pentatropis*); *Neobassia astrocarpa* (F. Muell.) A. J. Scott (second *Neobassia* species, occasionally two flowers in the same leaf axil, spines arising from base of lobes, slender and spreading, united at base, similar to *Dissocarpus*); *Osteocarpum pentapterum* (F. Muell. & Tate) Volkens (unique morphology of the fruiting perianth: sub-globular, ribbed tube produces 5, erect, fan-shaped wings arising from side of tube, closely affiliated to remaining *Osteocarpum* species and *S. brachyptera*); *Sclerolaena*

*recurvicuspis* (W. Fitzg.) Domin (spines horizontal to recurved, united at base, similar to *Neobassia*); *Sclerolaena hostilis* (Diels) Domin (perianth base broadly inflated and attached to axis); *Sclerolaena forrestiana* (F. Muell.) Domin (perianth partly embedded in branch axis, similar to *Eremophea*); *Sclerolaena medicaginoides* Paul G. Wilson (unique morphology of the fruiting perianth: 5-9 slender, divaricate spines, perianths firmly attached to axis forming a condensed spike, similar to *Dissocarpus* and *Eremophea*); *Sclerolaena symoniana* (Ising) A. J. Scott (perianth produces two series of appendages, 5 erect, lanceolate, spiny appendages and 4-12 fan-shaped, fimbriate spines, similar to *S. fimbriolata*); *Sclerolaena anisacanthoides* (F. Muell.) Domin (cylindrical, hollow perianth tube bears 6 erect spines at the apex of which 2 are an extension of the radicular tubercle, similar to *N. proceriflora*, *Osteocarpum and S. recurvicuspis*); *Sclerolaena tetragona* Paul G. Wilson (4-angled, broadly turbinate tube extends into short, stout spines, similar to *Osteocarpum and S. brachyptera*); *Sclerolaena tridens* (F. Muell.) Domin (turbinate tube with hollow base produces 2+2 horizontal, strongly flattened spines, similar to species of clade 13); *Sclerolaena microcarpa* (R. Anderson) A. J. Scott (barrel-shaped, ripped tube produces 4+2 horizontal spines, similar to species of clade 13, *S. brachyptera* and *S. stelligera*); *Sclerolaena urceolata* (Ising) A. J. Scott (sub-globular, ribbed tube with hollow base produces 4 short, erect tubercles at apex, similar to *Osteocarpum*); and *Sclerolaena walkeri* (C. White) A. J. Scott (unique morphology of the fruiting perianth: sub-globular, ribbed tube with hollow base that produces 5 intertepaline, narrow wings and disc-like apex from which arise 4+2 erect, intertepaline processes and a narrow, horizontal ring, similar to *Osteocarpum*).

### *3.2 Sequencing and assembly statistics*

Sequencing resulted in 33,906,603 raw reads (319,873 reads per sample) of which 88% passed the quality filters. The selected thresholds for in-sample-clustering ranged from 0.93 to 0.95 (Supplement 3, ISCT selection). The metrics evaluated for between-sample-clustering threshold selection indicated a CT of 0.93 as most suitable for all assemblies (Supplement 3, BSCT Assembly 1-3). At this threshold, the proportion of detected duplicates decreased distinctly while the proportion of detected paralogs increased only slightly (plots: "detected duplicates" and "detected paralogs"). The plots of the new polymorphic loci showed a first strong increase at CT 93/94, which represents a CT value of 0.93 (plots: "NPL_PIS" and "NPL_VAR"). The SNPs and sequence matrix size was maximized (plots: "SNPs-" and "Sequence matrix size") and the proportion of missingness started to decrease (plots: "SNPs-" and "Sequences missing"). Assembly 1 (including all samples) contained 8,659 loci with 316,465 variable sites of which 72,009 sites were parsimony informative. The average sample coverage (retained loci per sample) was 560. The sequence matrix contained 93.59% missing data. The *Maireana* grade assembly (Assembly 2) contained 3,656 loci with 130,247 variable sites of which 29,780 sites were parsimony informative. The average sample coverage was 403. The sequence matrix contained 89.12% missing data. The *Sclerolaena* clade assembly (Assembly 3) contained 5,774 loci with 220,059 variable sites of which 50,093 sites were parsimony informative. The average sample coverage was 923. The sequence matrix contained 84.3% missing data. Assembly 4, used for the BEAST analysis, contained 1,084 loci with 56,794 variable sites of which 16,806 sites were parsimony informative. The sequence matrix contained 84.3% missing data. The concatenated supermatrices of assemblies 1-4 are provided as PHYLIP files in Supplement 4.

### *3.3 Morphological and ecological characterization of phylogenetic clades*

Of the analyzed traits, 10 could be surveyed for all species and were plotted against the phylogeny including all Camphorosmeae accessions (depicted in Fig. 2; Supplement 2). These more general traits included habit, maximum growth height, branch pubescence, phyllotaxy, leaf pubescence, distribution of sexes, number of flowers per bract, fruiting perianth appendage type, perianth tube structure, seed orientation (depicted in Fig. 2). More specific fruiting perianth traits were recorded for wing-like (depicted in Fig. 3) and spine-like (depicted in Fig. 4) fruiting perianth appendages. In addition to the morphological traits, soil types and soil salinity (depicted in Fig. 5), and habitat types (depicted in Fig. 6) could be surveyed for all species (provided the information was given unambiguously). The preferences for calcareous or gypseous soils, soil salinity (Fig. 5) and colonized habitat types (Fig. 6) were found informative.

The morphological traits of the fruiting perianth showed taxonomic relevance and were characterized in more detail. With respect to the general appearance of the perianth shapes, different traits were recorded for the *Maireana* grade (wing-like) and the *Sclerolaena* clade (spine-like) (Table 2, Supplement 2, “wing-like” and “spine-like” species).

Wing-like appendages: The winged appendages (e.g. *Maireana*, *Osteocarpum*) are produced either at the base of the petals as five separate wings or from the perianth tube as a simple horizontal or vertical wing (Fig. 1 C-E, K, O, T, Fig. 2 and 3). Additional wings may occur along the perianth tube (Fig. 3 J, M, N). The perianth lobes usually obscure the utricle (the actual fruit including the seed) and can develop erect appendages or emergences (floccose wool) on top of the fruiting perianth (upper perianth processes, Fig. 1 C, E, O, Fig. 2 Q, R). *Malacocera* species produce star-shaped or (inverted) Y-like tepaline processes arising from a depressed (flat) perianth (Fig. 1 P, Fig. 2 C). In *Enchylaena*, the succulent lobes cover the ovary and the perianth forms a succulent, globular fruiting body (Fig. 1 M, Fig. 2 F). In *Roycea*, the

tepals surround the fruit at the base and do not enlarge (Fig. 1 U, V, Fig. 2 H). For the species of the *Maireana* grade, seven wing-like appendage specific traits were recorded (Table 2, Fig. 3). Of these characters, the tube shape, the wing type, the wing orientation, the shape of the upper perianth as well as the presence of accessory upper perianth processes were found informative.

Spine-like appendages: The spinous appendages of *Sclerolaena* species arise between the perianth lobes along or on top of the fruiting perianth tube that envelopes the utricle (Fig. 1 F-H). The perianth lobes and upper portion of the tube are sometimes prominently developed as a limb (Fig. 2, A; Fig. 4, G, J). The spinous appendages of *Dissocarpus* arise either opposite of the perianth lobes or from a fused bases of the perianth (Fig. 1 L, Fig. 2, D; Fig. 4, Q). In *Eremophea* (Fig. 1 N, Fig. 2 E) and *Neobassia* (Fig. 1 S, Fig. 3 B), the spines arise from the base of the perianth lobes. For the species of the *Sclerolaena* clade, five spine-like appendage specific traits were recorded (Table 2; Fig. 4). Of these characters, the shape of the attachment and the tube, the number of spines, the appearance of the limb, and the seed orientation were found informative.

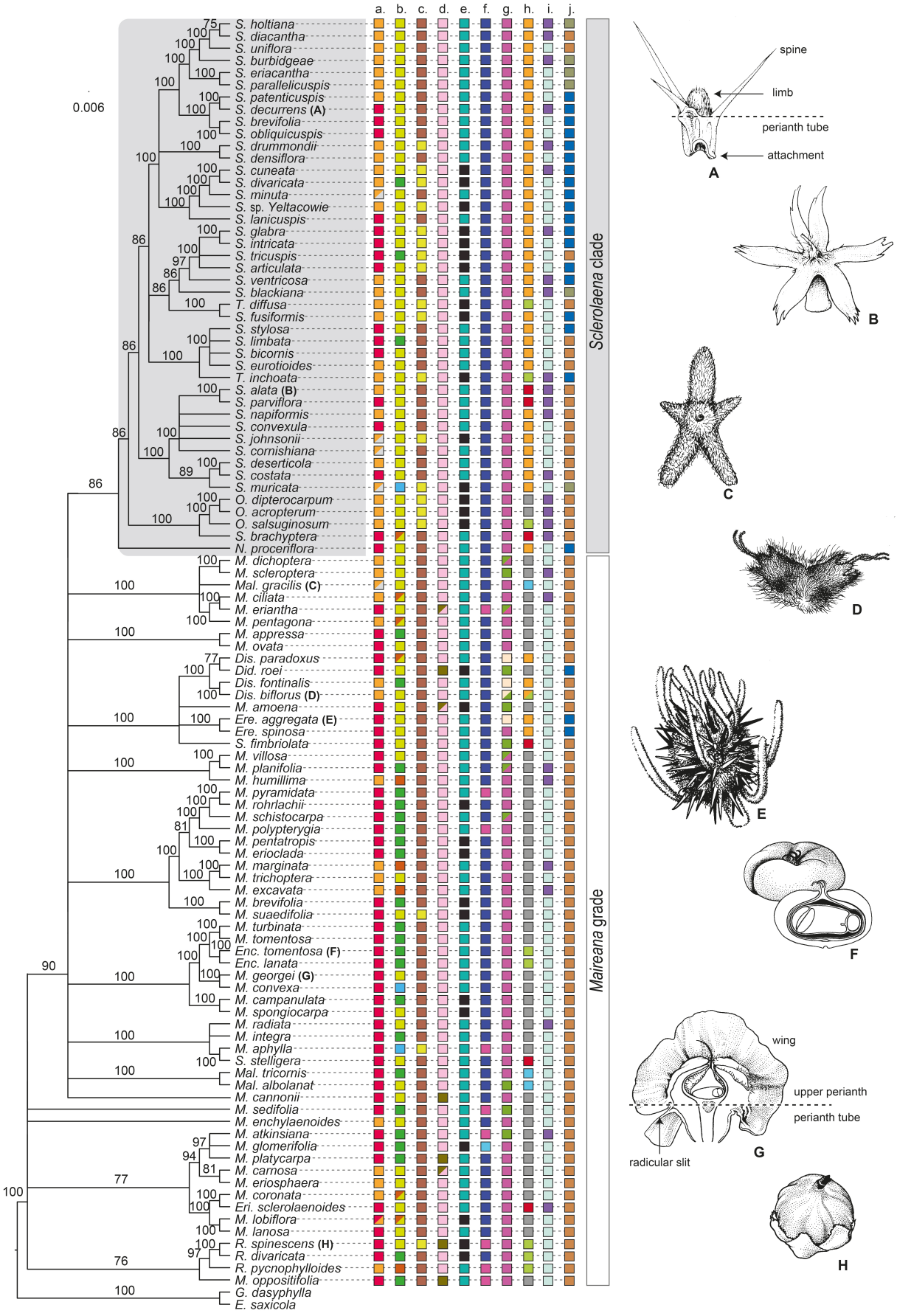



**Figure 2**. CA-ML Phylogeny Camphorosmeae (Assembly 1). Bootstrap support values are given above branches.

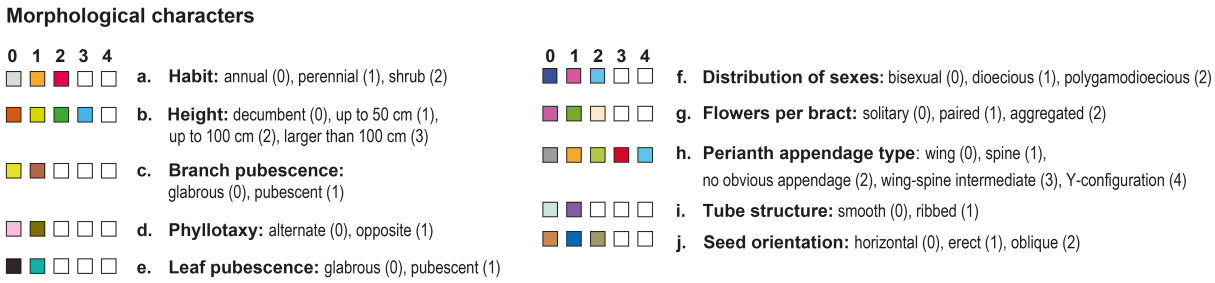



**Figure 2 continued**. CA-ML Phylogeny Camphorosmeae (Assembly 1). For each species are given the following character states: a. Habit, b. Height, c. Branch pubescence, d. Phyllotaxy, e. Leaf pubescence, f. Distribution of sexes, g. Flowers per bract, h. Perianth appendage type, i. Tube structure and j. Seed orientation. Magnification values of the fruiting perianth illustrations are given in parentheses. A: *Sclerolaena decurrens* (x4) is a shrub up to 30 cm high with pubescent branches. The alternate leaves are pubescent. The bisexual, solitary flowers produce a spiny fruiting perianth. The perianth tube is ribbed. The seeds are erect. B: *Sclerolaena alata* (x4) is a perennial up to 20 cm high with pubescent branches. The alternate leaves are pubescent. The bisexual, solitary flowers produces a wing-spine-intermediate fruiting perianth. The perianth tube is ribbed. The seeds are placed horizontally. C: *Malacocera gracilis* (x3.75) is an annual or perennial with a woody base up to 25 cm high with pubescent branches. The alternate leaves are pubescent. The bisexual, solitary flowers produce a fruiting perianth with Y- or starfish-like configuration. The tube is smooth. The seeds are placed horizontally. D: *Dissocarpus biflorus* (x3) is a perennial up to 25 cm high with pubescent branches. The alternate leaves are pubescent. The bisexual, paired or aggregated flowers produce fruiting perianths without obvious appendage. The perianth tube is smooth. The seeds are placed horizontally. E: *Eremophea aggregata* (x1.5) is a shrub up to 30 cm high with pubescent branches. The alternate leaves are pubescent. The bisexual flowers are aggregated and produce spiny fruiting perianths. The tube is smooth. The seeds are in erect position. F: *Enchylaena tomentosa* (x2.5) is a shrub up to 100 cm high with pubescent branches. The alternate leaves are pubescent. The bisexual, solitary flowers produce fruiting perianths without obvious appendages. The perianth tube is smooth. The seed is placed horizontally. G: *Maireana georgei* (x1.75) is a shrub up to 50 cm high with pubescent branches. The alternate leaves are pubescent. The bisexual, solitary flowers produce winged fruiting perianths. The perianth tube is smooth. The seed is placed horizontally. H: *Roycea spinescens* (x7.5) is a shrub up to 25 cm high with glabrous branches. The opposite leaves are glabrous. The unisexual, solitary flowers produce a fruiting perianth without prominent appendages. The seed is horizontal. A typical *Sclerolaena* perianth (A) is subdivided into the attachment (attachment site on the axis), the perianth tube, the spines, and the limb (remaining parts of the upper perianth). A typical *Maireana* perianth (G) is subdivided into the perianth tube, the wing and the upper perianth (above the tube and between the horizontal wing). A radicular slit can be present along the perianth tube and the wing. Image sources are given in section 9.

### *3.4 CA-ML phylogenies*

#### *3.4.1 Phylogeny of all sampled Camphorosmeae and distribution of morphological traits*

The CA-ML phylogeny of the entire Camphorosmeae sampling yielded 17 moderately to fully supported clades (Fig. 2). Four accessions were reconstructed on individual branches. The backbone was poorly resolved for the *Maireana* grade. The backbone topology of the *Sclerolaena* clade was well to fully supported. Collapsing branches below a BS support value of 75 resulted in two polytomies for the *Maireana* grade. The clades of the first polytomy contained the accessions of *Roycea*, *Eriochiton* and eleven *Maireana* species. The clades in the second polytomy comprised 32 *Maireana* species, two *Sclerolaena* species, and the accessions of *Didymanthus*, *Dissocarpus*, *Enchylaena, Eremophea* and *Malacocera*. The *Sclerolaena* clade contained 38 *Sclerolaena* species and the accessions of *Neobassia*, *Osteocarpum* and *Threlkeldia*. The CA-ML phylogeny showing all BS support values is provided in Supplement 4 (Assembly 1).

The species of the *Maireana* grade are shrubs (ca. 72 %) or perennials (ca. 27 %) and those of the *Sclerolaena* clade perennials (52 %) or shrubs (36 %) (Fig. 2, character a). The maximum growth height of the species tends to be larger in the *Maireana* grade (Fig. 2, character b). Glabrous branches and leaves occur more frequently in the *Sclerolaena* clade (36 and 34 %, respectively) than in the *Maireana* grade (5 and 22 %, respectively) (Fig. 2, char. c and e). The phyllotaxy is mostly alternating with a few species having opposite or alternate and opposite phyllotaxis (Fig. 2, char. d). Most species have bisexual flowers, with a few exceptions in the *Maireana* grade (Fig. 2, char. f). With a few exceptions in the *Maireana* grade, the flowers are solitary (Fig. 2, char. g). However, the clade including the accessions of *Didymanthus*, *Dissocarpus*, *Eremophea*, *M. amoena* and *S. fimbriolata* show mostly paired or aggregated flowers and infructescences. The appendage type of the fruiting perianth is predominantly spinous in the *Sclerolaena* clade and wing-like in the *Maireana* grade (Fig. 2, char. h).

Occasionally there are species with a wing-spine-intermediate appendage form or without a distinct perianth appendage (e.g. *S. alata* and *Enc. tomentosa*, Fig. 2 B and F, respectively). The tubes are more often ribbed in the *Sclerolaena* clade (ca. 41 %) than in the *Maireana* grade (15 %) (Fig. 2, char. i). The seed orientation in the *Maireana* grade is predominantly horizontal (95 %), and horizontal (43 %), erect (43 %) or oblique (ca. 14 %) in the *Sclerolaena* clade. Although the traits show tendencies between the species of the *Maireana* grade and the *Sclerolaena* clade, no trait is exclusively distributed.

By comparing the preferred habitats of the two polytomies (of the *Maireana* grade) and the *Sclerolaena* clade, a shift and change in habitats becomes apparent (Fig. 6, Table 4). A general adaptation to Riverine Desert (41 % of all included species) and Desert Lake habitats (33 %) is evident across the phylogeny. Stony Desert (27 %) and Karst Plain habitats (22 %) are also frequently colonized. Coastal (9 %) and Desert Clay Plain habitats (6 %) are the least frequently colonized habitats. Species of the first polytomy occur predominantly on sandy, saline soils in Desert Lake habitats (73 % of all species in the first polytomy), followed by Karst Plain and Riverine Desert habitats (both 27%) (Fig. 5 and 6, Table 4). This ratio shifts in favor of Riverine Desert (38 %), Desert Lake (33 %), and Karst Plain habitats (27 %) for the species in the second polytomy. Other arid habitats, such as Sand Desert, Desert Upland and Stony Desert landscapes become increasingly colonized. In addition, the occupied soils become less saline (Fig. 5). The species of the *Sclerolaena* clade occur predominantly in Riverine Desert (50 %) and Stony Desert habitats (41 %). As is the case with the *Maireana* clades of the second polytomy, other arid habitats are more frequently colonized.

**Table 4**: Habitat preferences. Comparison of species occurrence in defined habitats, given in percentages. Given are the percentages for all included species (Assembly 1), for the species of the *Sclerolaena* clade, for the species of the second polytomy (excluding the *Sclerolaena* clade) and for the species of the first polytomy. Habitat abbreviations: SD = Sand Desert, DU = Desert Upland, SP = Stony Desert, SH = Shield Plain, RL = Riverine Desert, CP = Desert Clay Plain, KP = Karst Plain, DL = Desert Lake, MP = Mesic Plain and Range, CS = Coast.

| | SD | DU | SP | SH | RL | CP | KP | DL | MP | CS |
|---|---|---|---|---|---|---|---|---|---|---|
| all species | | | | | | | | | | |
| count | 18 | 19 | 28 | 16 | 43 | 6 | 23 | 34 | 12 | 9 |
| %, n=104 | 17,31 | 18,27 | **26,92** | 15,38 | **41,35** | 5,77 | **22,12** | **32,69** | 11,54 | 8,65 |
| *Sclerolaena* clade | | | | | | | | | | |
| count | 7 | 8 | 18 | 8 | 22 | 3 | 7 | 8 | 3 | 4 |
| %, n=44 | 15,91 | 18,18 | **40,91** | 18,18 | **50,00** | 6,82 | 15,91 | 18,18 | 6,82 | 9,09 |
| *Maireana* grade 2nd Polytomy | | | | | | | | | | |
| count | 10 | 10 | 9 | 7 | 17 | 2 | 12 | 15 | 8 | 4 |
| %, n=45 | 22,22 | 22,22 | 20,00 | 15,56 | **37,78** | 4,44 | **26,67** | **33,33** | 17,78 | 8,89 |
| *Maireana* grade 1st polytomy | | | | | | | | | | |
| count | 1 | 1 | 1 | 1 | 4 | 1 | 4 | 11 | 1 | 1 |
| %, n=15 | 6,67 | 6,67 | 6,67 | 6,67 | **26,67** | 6,67 | **26,67** | **73,33** | 6,67 | 6,67 |

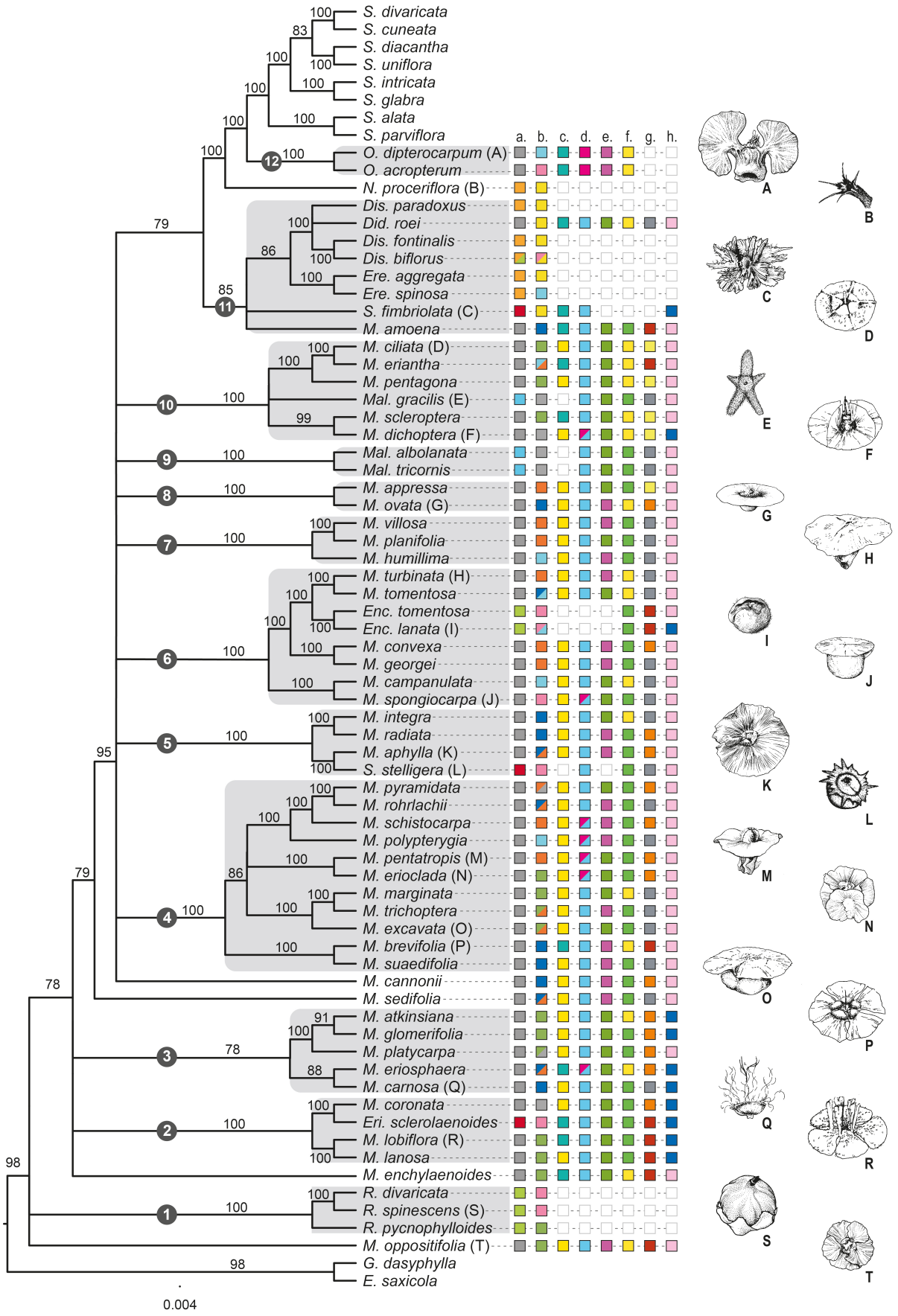



**Figure 3**. CA-ML Phylogeny of the *Maireana* grade (Assembly 2). Bootstrap support values are given above branches.

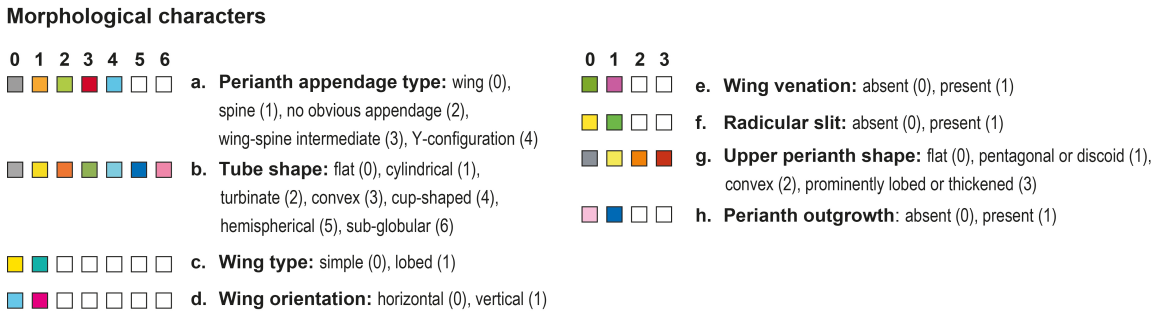



**Figure 3 continued**. CA-ML Phylogeny of the *Maireana* grade (Assembly 2). For each species are given the following character states: a. Perianth appendage type, b. Tube shape, c. Wing type, d. Wing orientation, e. Wing venation, f. Radicular slit, g. Upper perianth shape, h. Perianth outgrowth. Magnification values of the fruiting perianth illustrations are given in parentheses. A: The winged fruiting perianth of *Osteocarpum dipterocarpum* (x3.5) has a cup-shaped perianth tube. The entire, vertical wings are prominently veined. A radicular slit is absent. B: *Neobassia proceriflora* (x2) has a spiny fruiting perianth. The perianth tube is cylindrical. C: The wing-spine-intermediate fruiting perianth of *Sclerolaena fimbriolata* (x2) has a cylindrical perianth tube. The horizontal wing is lobed and fimbriate. The upper perianth produces appendages. D: *Maireana ciliata* (x3.5) has a winged fruiting perianth. The perianth tube is convex. The entire, horizontal wing has no prominent venation. A radicular slit is absent. The upper perianth has a prominent pentagonal layout an does not produce extra processes. E: The inverted star-like fruiting perianth of *Malacocera gracilis* (x2) has a flat perianth tube. The horizontal appendage has no prominent venation. A radicular slit is absent. The upper perianth is flat and produces no processes. F: *Maireana dichoptera* (x2.25) has a winged fruiting perianth. The tube is flat. The entire, horizontal and vertical wings are without obvious venation. A radicular slit is absent. The discoid upper perianth produces wing-like processes. G: The winged fruiting perianth of *Maireana ovata* (x2.25) has a hemispherical perianth tube. The entire, horizontal wing has a prominent venation (when dry). The upper perianth is convex and without extra processes. H: *Maireana turbinata* (x1.25) has a winged fruiting perianth with turbinate perianth tube. The entire, horizontal wing has a prominent venation. A radicular slit is absent. The flat upper perianth produces no processes., I: The fruiting perianth of *Enchylaena lanata* (x1.25) produces no prominent appendages. The succulent perianth tube is hemispherical to globular. A radicular slit is present. The upper perianth lobes are thickened and produce a corona-like outgrowth. J: *Maireana spongiocarpa* (x0.75) has a winged fruiting perianth with a sub-globular perianth tube. The entire horizontal wing has no obvious venation. An inconspicuous vertical wing is present along the tube. A radicular slit is present. The upper perianth is flat and without processes. K: The winged fruiting perianth of *Maireana aphylla* (x1.75) has a hemispherical to sub-globular perianth tube. The entire, horizontal wing has a prominent venation. A radicular slit is present. The upper perianth is (slightly) convex and without processes. L: *Sclerolaena stelligera* (x2) has a wing-spine-intermediate fruiting perianth. The perianth tube is sub-globular. The horizontal wing produces short, radiating spines. A radicular slit is present. The upper perianth is flat and without processes. M: The winged fruiting perianth of *Maireana pentatropis* (x1.25) has a turbinate perianth tube. The entire horizontal and vertical wings have no obvious venation. A radicular slit is present. The convex upper perianth produces no extra processes. N: *Maireana erioclada* (x1.75) has a winged fruiting perianth with a convex perianth tube. The entire, horizontal and vertical wings have no obvious venation. A radicular slit is present. The convex upper perianth produces no extra processes. O: The winged fruiting perianth of *Maireana excavata* (x2) has a turbinate to convex perianth tube. The entire, horizontal wing has no obvious venation. A radicular slit is present. The flat upper perianth is without extra processes. P: *Maireana brevifolia* (x2.25) has a winged fruiting perianth with a hemispherical perianth tube. The 5-lobed, horizontal wing has a prominent venation. A radicular slit is present. The upper perianth lobes are thickened and without extra processes. Q: The winged fruiting perianth of *Maireana carnosa* (x2.5) has a hemispherical perianth tube. The entire, horizontal wing has no obvious venation. A radicular slit is present. The flat upper perianth produces a long floccose wool enveloping the perianth., R: *Maireana lobiflora* (x1.5) has a winged fruiting perianth with a convex perianth tube. The 5-lobed, horizontal wing has no prominent venation. A radicular slit is present. The upper perianth is prominently lobed and produces erect processes. S: The fruiting perianth of *Roycea spinescens* (x5) has no obvious appendage. The lobes do not enlarge in fruit. The fruit is globular. T: *Maireana oppositifolia* (x2) has a winged fruiting perianth with a convex perianth tube. The 5-lobed, horizontal wing has a prominent venation. A radicular slit is absent. The prominently lobed upper perianth produces no extra processes. Image sources are given in section 9.

### *3.4.2 Phylogeny of the Maireana grade and distribution of wing-like perianth traits*

The CA-ML phylogeny of the *Maireana* grade sampling resulted in 11 well or fully supported clades (Fig. 3). The backbone of the phylogeny was poorly to moderately supported. Collapsing branches with BS support less than 75 resulted in three polytomies. Four taxa were reconstructed on individual branches: *Maireana oppositifolia*, *M. enchylaenoides*, *M. sedifolia* and *M. cannonii*.

The first polytomy contained *M. oppositifolia* and the three *Roycea* accessions of clade 1. The second polytomy contained *M. enchylaenaoides*, clade 2 and clade 3. *Maireana sedifolia* was reconstructed on an individual branch and sister to a polytomy of *M. canonnii* and clades 4 to clade 10. Clade 11 was sister to *Neobassia proceriflora* and the other representatives of the *Sclerolaena* clade. The CA-ML phylogeny showing all BS support values is provided in Supplement 4 (Assembly 2).

The characters of the following descriptions are depicted in Figure 2 (habit, height, distribution of sexes, flowers per bract, seed orientation), Figure 3 (perianth appendage type, perianth tube shape, wing type, wing orientation, wing venation, radicular slit, upper perianth shape, presence of upper perianth processes), Figure 5 (soil types and soil salinity), Figure 6 (habitat types) and Supplement 2 (species distribution).

***Maireana oppositifolia*** is a dioecious shrub up to 100 cm high. The unisexual flowers are solitary (Fig. 2). The perianth tube is convex and produces a 5-lobed, fan-shaped, horizontal wing with a prominent venation (Fig. 3, T). No radicular slit is present. The upper perianth is lobed and without erect processes. The seed is positioned horizontally. Its distribution ranges from the southern areas of the Western and Eastern Desert, to the southern, coastal ranges of the Southwestern region, the Nullarbor Plain, Eyre Peninsular and Adelaide (Supplement 2). It occurs on sandy, loamy or clayey saline soils (Fig. 5) in Desert Lake and Coastal habitats (Fig. 6).

**Clade 1** contained all three described *Roycea* species (Fig. 3). The species are perennial herbs (*R. pycnophylloides*) or shrubs (*R. spinescens*, *R. divaricata*) up to 60 cm high (Fig. 2, Supplement 2). The inconspicuous uni- or bisexual flowers are solitary with 5 imbricate tepals that do not enlarge in fruit (Fig. 2). The fruit is convex to sub-globular and surrounded by the perianth lobes at the base (Fig. 2, H and Fig. 3, S). No radicular slit is present. The seed is positioned horizontally and encloses the radicle. The species are endemic in the Western Desert and Southwestern region in Western Australia (Supplement 2). Their occurrence is restricted to sandy, loamy or clayey, saline soils (Fig. 5) of Desert Lake habitats (Fig. 6).

***Maireana enchylaenoides*** is a perennial herb up to 20 cm high. The bisexual flowers are solitary (Fig. 2). The convex perianth tube produces a 5-lobed, horizontal wing without a prominent venation (Fig. 3). No radicular slit is present. The upper perianth is lobed and without erect processes. The seed is horizontal. The distribution is centered in the south-eastern areas of the Eastern Desert, Eyre Peninsular and Adelaide, and has a second, isolated distribution in the Southwestern region of Western Australia (Supplement 2). It prefers heavy, loamy or clayey soil (Fig. 5) and occurs in Riverine Lake, Karst Plain and Mesic Plain habitats (Fig. 6).

**Clade 2** contained three *Maireana* species and the monotypic genus *Eriochiton* (Fig. 3). *Maireana coronata*, *M. lobiflora*, *M. lanosa* and *Eri. sclerolaenoides* are perennials with a woody base or shrubs up to 50 cm high (Fig. 2, Supplement 2). The bisexual flowers are solitary (Fig. 2). The perianth tube is flat (*M. coronata*), convex (*M. lobiflora* and *M. lanosa*) or sub-globular (*Eri. sclerolaenoides*) (Fig. 3). The horizontal wings are simple (*M. coronata* and *M. lanosa*) or 5-lobed (*Eri. sclerolaenoides* and *M. lobiflora*) and do not show a prominent venation. A radicular slit is present. The upper perianth is convex (*M. coronata*) or lobed and produces erect processes. In the case of *M. lobiflora* (Fig. 3, R) and *M. lanosa*, the six erect processes alternate with the perianth lobes. *Maireana coronata* produces a cup-shaped (corona) outgrowth and the five erect processes of *Eri. sclerolaenoides* arise from the base of the perianth

lobes. The seeds are oriented horizontally (Fig. 2). The distribution area of *Eri. sclerolaenoides* reaches along the southern parts of the Western Desert and Eastern Desert, south to the coastal lines of the Southwestern region, the Nullarbor, Eyre Peninsular and Adelaide (Supplement 2). The distribution range of *M. coronata* is centered in the Eastern Desert and Central Queensland. The distributions of *M. lobiflora* and *M. lanosa* are scattered sparsely from the Pilbara area and the Western Desert to the northern extensions of the Eastern Desert and southwards to the Nullarbor, Eyre Peninsular and Adelaide. Most species of this clade prefer sandy or heavy loamy, saline soils, with *M. lobiflora* also growing in calcareous soil (Fig. 5). The species occur primarily in Riverine Desert habitats (Fig. 6).

**Clade 3** contained 5 *Maireana* species (Fig. 3). The species are perennials with a woody base (*M. carnosa*, *M. eriosphaera*) or shrubs (*M. atkinsiana*, *M. glomerifolia*, *M. platycarpa*), 50 – 100 cm high (Fig. 2, Supplement 2). The flowers are usually solitary with *M. atkinsiana* having paired flowers, and are bisexual, unisexual (*M. atkinsiana;* as is *M. oppositifolia*, *and R. spinescens*, *R. pycnophylloides* (clade 1), *M. sedifolia*, and *M. aphylla* (clade 5)) or polygamodioecious (*M. glomerifolia*) (Fig. 2). The fruiting perianth develops a convex, hemispherical or sub-globular tube with a simple or lobed (*M. eriosphaera;* as is *M. oppositifolia, M. enchylaenoides,* and *Eri. sclerolaenoides* and *M. lobiflora*, clade 2, and *M. brevifolia*, clade 4), horizontal wing without a prominent venation (Fig. 3). A radicular slit is present in *M. glomerifolia*, *M. platycarpa* and *M. carnosa*. The upper perianth is convex, with *M. carnosa* having a flat upper perianth. With exception of *M. platycarpa*, the perianths produce either processes or a wooly indumentum. The perianth lobes of *M. atkinsiana* and *M. glomerifolia* produce erect, narrowly oblong processes (as the species in clade 2). In case of *M. eriosphaera* and *M. carnosa* (Fig. 3, Q) the perianths are covered in long silky wool forming a soft ball. The seeds are positioned horizontally. The distribution of this clade is centered in Western Australia, with *M. carnosa* also distributed in the Eastern Desert (Supplement 2). The

species grow in a variety of saline soils (Fig. 5), and are commonly found surrounding salt lakes in Desert Lake habitats (Fig. 6).

***Maireana sedifolia*** is a dioecious shrub up to 100 cm high (as is *M. oppositifolia*). The flowers are paired (as in *M. atkinsiana*, clade 3, *Mal. albolanata*, clade 9, *M. scleroptera*, clade 10, and species of clade 11). The fruiting perianth produces a convex to hemispherical tube with a simple, horizontal wing with a prominent venation (Fig. 3). A radicular slit is present. The upper perianth is flat and without processes. The seed is horizontal (Fig. 2). Its distribution ranges from the southern areas of the Western and Eastern Desert southwards to the Nullarbor Plain, Eyre Peninsular and Adelaide. It prefers sandy and calcareous soil (Fig. 5) and usually occurs in Desert Upland and Karst Plain habitats (Fig. 6).

***Maireana cannonii*** is a shrub up to 50 cm high. The bisexual flowers are solitary. The perianth tube is hemispherical and produces a simple, horizontal wing with a prominent venation (Fig. 3). A radicular slit is present. The convex upper perianth is without processes. The seed is horizontal (Fig. 2). The species occurs in the southern parts of the Eastern Desert, Eyre Peninsular and northern Adelaide, usually in Riverine Desert, Desert Lake or Coastal habitats (Fig. 6).

**Clade 4** contained eleven *Maireana* species (Fig. 3). The species are perennial (*M. excavata*), perennials with a woody base (*M. marginata*, *M. trichoptera*) or shrubs *(M. pyramidata*, *M. rohrlachii*, *M. schistocarpa*, *M. polypterygia*, *M. pentatropis*, *M. erioclada*, *M. brevifolia*, *M. suaedifolia*) (Fig. 2, Supplement 2). The majority of species have solitary, bisexual flowers, with exception of *M. schistocarpa* (bisexual, solitary or paired flowers), and *M. pyramidata* and *M. polypterygia* that have solitary, unisexual flowers. The perianth tube is mostly turbinate, convex or hemispherical (Fig. 3). The fruiting perianth produces a simple, horizontal wing with *M. polypterygia*, *M. erioclada* and *M. pentatropis* producing additional vertical wings. The wing of *M. brevifolia* is 5-lobed (as is *M. oppositifolia*, *M. enchylaenoides*

and *Eri. sclerolaenoides* and *M. lobiflora* in clade 2). A prominent wing venation can be present. With exception of *M. marginata* and *M. brevifolia*, a radicular slit is present. The upper perianth is mostly flat (*M. rohrlachii*, *M. polypterygia*, *M. marginata*, *M. trichoptera*, *M. excavata*, *M. suaedifolia*, thickened (*M. brevifolia*) or convex (*M. erioclada*, *M. pentatropis, M. pyramidata, M. schistocarpa*). The upper perianth produces no additional processes. The species *M. pyramidata, M. brevifolia*, and *M. suaedifolia* each have a disjunct distribution with occurences in the Western Desert and the Eastern Desert (Supplement 2). *Maireana brevifolia* is one of the few Australian Camphorosmeae that is distributed outside Australia (Chile - Marticorena 1997; Canary Islands – Brandes, 2002; South Africa - Mucina and Snijman, 2011). *Maireana erioclada*, *M. pentatropis* and *M. trichoptera* are distributed mainly along the south coast from the Southwestern Range across the Nullarbor, Eyre Peninsular and Adelaide, with *M. trichoptera* occurring also in the Western and Eastern Desert. The other species have much smaller ranges, which are restricted to either the Western (*M. polypterygia*) or Eastern Desert (*M. schistocarpa*, *M. excavata*), the Southwestern Range (*M. marginata*) or Eyre Peninsular and Adelaide (*M. rohrlachii*, *M. excavata*). The species grow in a variety of soil types of varying salinity (Fig. 5) and occur in various habitats, mostly in Mesic Plain and Desert Lake habitats (Fig. 6).

**Clade 5** contained three *Maireana* species and *Sclerolaena stelligera* (former *Stelligera endecaspinis*) (Fig. 3). The species are shrubs, 50 – 200 cm in height. The flowers are mostly solitary and bisexual with *M aphylla* having solitary, unisexual flowers (as is *M. rohrlachii* and *M. polytpterygia*, clade 4). The tube is either turbinate to hemispherical (*M. aphylla*), hemispherical (*M. integra*, *M. radiata*) or sub-globular (*S. stelligera*) (Fig. 3). The fruiting perianth produces a simple, horizontal wing with *S. stelligera* producing a rigid, multi-spined, horizontal wing. A prominent wing venation is present in *M. aphylla* and *M. radiata*. With exception of *M. integra*, a radicular slit is present. The upper perianth is flat (*M. integra*, *S.*

*stelligera*) or concave (*M. radiata*, *M. aphylla*) and without erect processes. The seeds are oriented horizontally. The distributions of *S. stelligera* and *M. aphylla* are centered in the eastern regions of the Eastern Desert, with *M. aphylla* having a second, isolated distribution in the northwestern Western Desert (Supplement 2). The distribution of *M. integra* is centered in the Eastern Desert with sparsely scattered occurrences in the Western Desert. The distribution of *M. radiata* ranges from the southern areas of the Western and Eastern Desert to the coastal lines of the Southwestern region, the Nullarbor Plain, Eyre Peninsular and Adelaide. The species grow in a variety of soils of varying salinity (Fig. 5), commonly in Riverine Desert and Karst Plain habitats (Fig. 6).

**Clade 6** contained six *Maireana* and the two *Enchylaena* accessions (Fig. 3). The species are shrubs, commonly up to 100 cm high, with *M. convexa* growing up to 200 cm high (Fig. 2, Supplement 2). The flowers are solitary and bisexual. While *Enc. lanata* and *Enc. tomentosa* produce a succulent, berry-like fruiting perianth with no obvious appendages, the fruiting perianths of *M. turbinata*, *M. tomentosa*, *M. georgei*, *M. convexa*, *M. campanulata*, and *M. spongiocarpa* produce winged appendages (Fig. 3). The perianth tube is either turbinate (*M. turbinata*, *M. georgei* and *M. convexa*), cup-shaped (*M. campanulata*), hemispherical (*M. tomentosa*) or sub-globular in *Enchylaena* and *M. spongiocarpa.* If present, the wing is simple, horizontal and can show a prominent venation (*M. turbinata*, *M. georgei* and *M. convexa*). A radicular slit is present in *M. turbinata*, *M. tomentosa*, *M. campanulata* and *M. spongiocarpa*. The upper perianth is mostly flat with exception of *Enchylaena* (fleshy) and *M. convexa* (convex) and without processes. The seeds are placed horizontally (Fig. 2). The distributions of *M. campanulata* and *M. spongiocarpa* are centered in the Eastern Desert (Supplement 2). *Maireana convexa* and *Enc. lanata* occur in the Western Desert and Southwestern region of Western Australia. *Maireana turbinata* is distributed in the southern areas of the Eastern Desert down to the coastal areas of the Nullarbor, Eyre Peninsular and Adelaide. The distribution of

*M. tomentosa* reaches from the Western Desert and Pilbara region eastwards to the Eastern Desert and Eyre Peninsular and Adelaide. *Maireana georgei* and *Enc. tomentosa* are widespread across all major deserts with *Enc. tomentosa* extending its range to the northern, eastern and southern coasts. The species grow in a variety of soils of varying salinity (Fig. 5), and occur in various habitats, commonly in Desert Upland habitats (Fig. 6). Of all Australian Camphorosmeae species analyzed, *Enc. tomentosa* occurs in the most diverse habitats and is one of the few species distributed outside Australia (New Caledonia; Wilson, 1984).

**Clade 7** contained three *Maireana* accessions (Fig. 3). The species are perennials with a woody base or shrubs, up to 100 cm high (Fig. 2, Supplement 2). The flowers are bisexual and solitary (*M. humillima*) or bisexual and solitary or paired (*M. villosa*, *M. planifolia* (as is *M. schistocarpa*, clade 4). The perianth tube is turbinate (*M. planifolia*, *M. villosa*) or cup-shaped (*M. humillima*) (Fig. 3). The wings are simple, horizontal and can be prominently veined (*M. villosa*, *M. humillima*). A radicular slit is present. The upper perianth is flat and without processes. The seeds are horizontal. The distribution of *M. humillima* is restricted to the southeastern extension of the Eastern Desert, Adelaide and Victoria (Supplement 2). *Maireana villosa* and *M. planifolia* are widely distributed across the Western and Eastern Desert, with *M. villosa* extending into the Central Desert and Central Queensland. The species grow in a variety of soils (Fig. 5) and are commonly found in Shield Plain and Desert Upland habitats (Fig. 6).

**Clade 8** contained *M. ovata* and *M. appressa* (Fig. 3). The species are shrubs up to 60 cm high (Fig. 2, Supplement 2). The flowers are solitary and bisexual. The perianth tube is hemispherical in *M. ovata* and turbinate in *M. appressa* (Fig. 3). The wing is simple and horizontal, and prominently veined in *M. ovata*. A radicular slit is present in *M. appressa*. The upper perianth is convex in *M. ovata* and pentagonal in *M. appressa* and without processes. The seeds are horizontal (Fig. 2). *Maireana appressa* has its center of distribution in the Eastern Desert, southwards to the Eyre Peninsular and Adelaide with a second range in the Western

Desert (Supplement 2) and prefers sandy or gypseous, saline soils (Fig. 5) in Desert lake habitats (Fig. 6), while *M. ovata* occurs only in the Eastern Desert and grows on sandy rises and eroded slopes of stony hills (Fig. 5) in Desert Upland and Shield Plain habitats (Fig. 6).

**Clade 9** contained two *Malacocera* species (Fig. 3). The species are shrubs up to 80 cm high (Fig. 2, Supplement 2). The flowers are bisexual and solitary in *Mal. tricornis* and bisexual and solitary or paired in *Mal. albolanata*. The fruiting perianths are depressed and produce flattened, horizontal appendages in a Y- or starfish-like configuration that are attached to the length of the perianth tube (similar to *Mal. gracilis*, clade 10) (Fig. 3). The upper perianth is flat. A radicular slit is present. The seed is placed horizontally (Fig. 2). Both species are distributed in the Eastern Desert (Supplement 2) in Riverine Desert and Desert Lake habitats (Fig. 6) and grow in loamy and clayey soils (Fig. 5).

**Clade 10** contained one *Malacocera* and five *Maireana* species (Fig. 3). The species are perennials (*M. dichoptera*, *M. scleroptera*), perennials with a woody base (*Mal. gracilis*, *M. ciliata*, *M. pentagona*) or shrubs (*M. eriantha*) up to 50 cm high (Fig. 2, Supplement 2). The flowers are mostly bisexual (with *M. eriantha* having unisexual flowers, as is *M. pyramidata* and *M. polypterygia*, clade 4, *M. aphylla*, clade 5, *M. atkinsiana*, clade 3, *R. spinescens* and *R. pycnophylloides*, clade 1, and *M. sedifolia* and M. oppositifolia) and solitary (*M. ciliata*, *M. pentagona*, *Mal. gracilis*), paired (*M. scleroptera*), or solitary or paired (*M. eriantha*, *M. dichoptera*, as is *M. villosa* and *M. planifolia*, clade 7, and *M. schistocarpa*, clade 4). The perianth tube can be flat (*Mal. gracilis*, *M. dichoptera*), convex (*M. ciliata*, *M. pentagona*, *M. scleroptera*) or turbinate to cup-shaped (*M. eriantha*) (Fig. 3). The fruiting perianth of the *Maireana* species produces wing-like appendages while the perianth of *Mal. gracilis* produces flattened, horizontal processes in an inverted Y-configuration (similar to the perianth of clade 9). The wings of *M. ciliata*, *M. pentagona* and *M. dichoptera* are simple and horizontal. The perianth wings of *M. eriantha* and *M. scleroptera* are horizontal and 5-lobed (as is *M. brevifolia*,

clade 4, *M. eriosphaera*, clade 3, *Eri. sclerolaenoides* and *M. lobiflora*, clade 2, and *M. enchylaenoides* and *M. oppositifolia*). The wings are without a prominent venation. A radicular slit is present in *Mal. gracilis*. The upper perianth is flat in *Mal. gracilis*, shows a pentagonal (*M. dichoptera*, *M. pentagona*, *M. ciliata*) to discoid layout (*M. scleroptera*) or is (deeply) lobed in *M. eriantha*. With exception of *M. dichoptera* (intertepaline, vertical wings, similar to species of clade 2 and 3), the upper perianth is without additional processes. The seed is oriented horizontally (Fig. 2). The distribution of *Mal. gracilis*, *M. eriantha* and *M. ciliata* is centered in the Eastern Desert reaching south to Eyre Peninsular (Supplement 2). *Maireana pentagona* occurs in the southeastern extensions of the Eastern Desert, *M. scleroptera* is distributed in the northwestern extension of the Eastern Desert, and *M. dichoptera* is known from the northeastern Eastern Desert and Central Queensland. The species occur on various soil types, with *M. ciliata* growing in calcareous and *Mal. gracilis* in gypseous soils (Fig. 5), and are commonly found in Riverine Desert habitats (Fig. 6).

**Clade 11** contained three *Dissocarpus* accessions, the monotypic genus *Didymanthus*, both described species of *Eremophea*, and one accessions of *Sclerolaena* and *Maireana*, each (Fig. 3). The species are perennials (*Dis. fontinalis*, *Dis. biflorus*) or shrubs, usually up to 60 cm high with *Dis. fontinalis* growing up to 100 cm (Fig. 2, Supplement 2). The bisexual flowers are either paired with a fused basis around the stem axis (*Did. roei*, *M. amoena*, *S. fimbriolata*) or clustered to more than two flowers per leaf axil and form a woody infructescence at maturity (*Dis. fontinalis*, *Dis. biflorus, Dis. paradoxus*, *Ere. aggregata*). The solitary flowers of *Ere. spinosa* are slightly sunken into the branch and become embedded in the woody branch axis in maturity, producing irregularly clustered fruits. The perianth tubes are mainly cylindrical (*Did. roei, Dis. fontinalis*, *Dis. paradoxus*, *Ere. aggregata*, *S. fimbriolata*), or can be cylindrical to sub-globular (*Dis. biflorus*), cup-shaped (*Ere. spinosa*) or hemispherical (*M. amoena*) (Fig. 3). The fruiting perianths produce either horizontal and 5-lobed wings (*Did. roei*, *M. amoena*), fan-

shaped, woody wing-spine-intermediates (*S. fimbriolata*), spines (*Ere. aggregata*, *Ere. spinosa*, *Dis. fontinalis*, *Dis. paradoxus*), or with or without spine-like appendages (*Dis. biflorus*). If present, the wings are without a prominent venation. A radicular slit is present in *M. amoena*. The upper perianth is flat in *Did. roei* and prominently lobed in *M. amoena*. The upper perianth of *S. fimbriolata* produces additional processes. *Didymanthus* and *Eremophea* have erect seeds while the other species have horizontal seeds (Fig. 2). *Didymanthus*, *M. amoena* and *S. fimbriolata* have their distribution centered in the Western Desert, with *Didymanthus* also occurring in the Southwestern region (Supplement 2). *Eremophea aggregata* is known only from the Shark Bay area in the Western Desert. The distribution of *Ere. spinosa* ranges from the Western Desert to the northwestern parts of the Eastern Desert. *Dissocarpus biflorus* is widespread in the Pilbara, the Western, Central and Eastern Desert, Central Queensland, Eyre Peninsular and Adelaide. *Dissocarpus paradoxus* occurs from Central Queensland, across the Eastern Desert, south to Eyre Peninsular and Adelaide. The distribution of *Dis. fontinalis* is centered in the Lake Eyre Basin in the Eastern Desert. The species grow in a variety of moderately saline to saline soils, with *Eremophea*, *M. amoena* and *S. fimbriolata* growing also in calcareous or gypseous soils (Fig. 5). The species usually occur in Riverine Desert and Desert Lake habitats (Fig. 6).

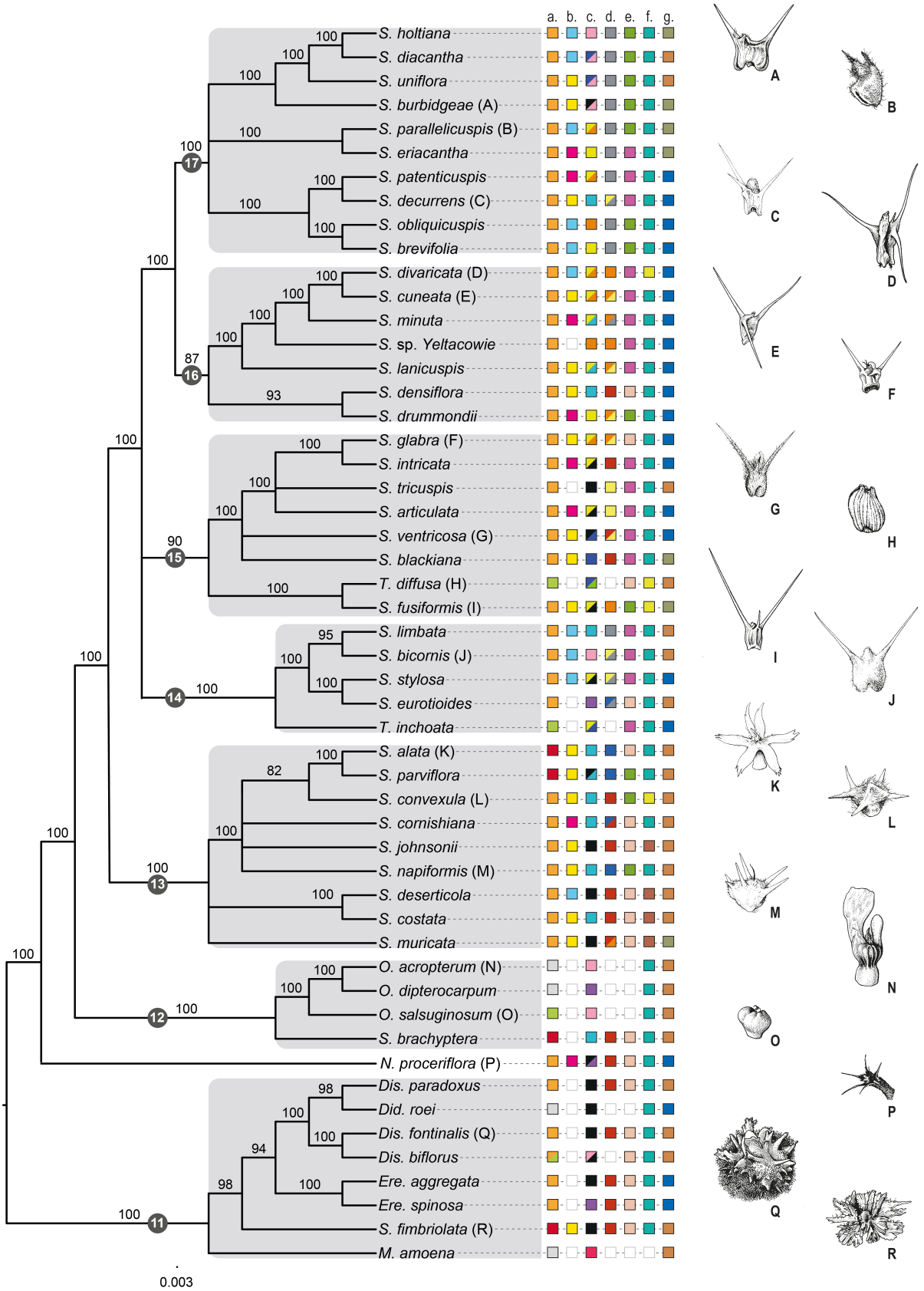



**Figure 4**. CA-ML Phylogeny of the *Sclerolaena* clade (Assembly 3). Bootstrap support values are given above branches. The encircled clade numbers are shaded in grey.

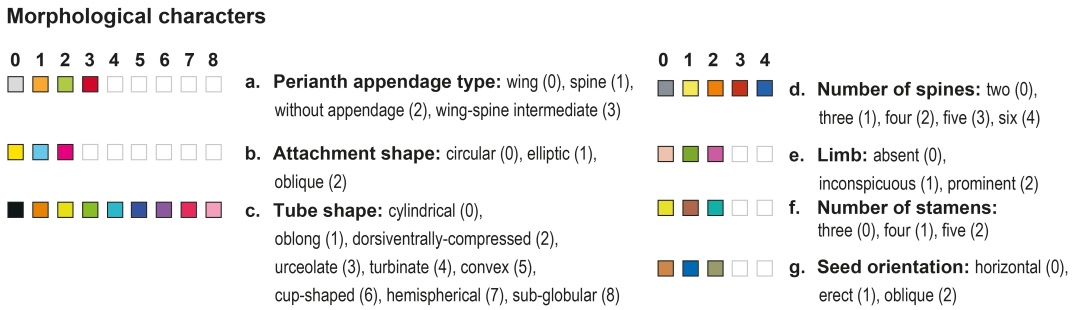



**Figure 4 continued**. CA-ML Phylogeny of the *Sclerolaena* clade (Assembly 3). For each species are given the following character states: a. Perianth appendage type, b. Attachment shape, c. Tube shape, d. Number of spines, e. Limb, f. Number of stamens, g. Seed orientation. Magnification values of the fruiting perianth illustrations are given in parentheses. A: *Sclerolaena burbidgeae* (x3) has a spiny fruiting perianth. The attachment is circular. The perianth tube is cylindrical to sub-globular. The perianth produces two spines. The limb is inconspicuous. The number of stamens is five. The seed is oriented obliquely. B: The fruiting perianth of *Sclerolaena parallelicuspis* (x2.5) is spiny. The attachment is elliptic. The tube is oblong and dorsiventrally-compressed. The perianth produces two spines. The limb is inconspicuous. The number of stamens is five. The seed is oriented obliquely. C: *Sclerolaena decurrens* (x2) has a spiny fruiting perianth. The attachment is circular. The tube is turbinate. The number of spines is two or three. A limb is prominent. The number of stamens is five. The seed is erect. D: The fruiting perianth of *Sclerolaena divaricata* (x2.5) is spiny. The attachment is elliptic. The tube is oblong and dorsiventrally-compressed. The number of spines is four. A limb is prominent. The number of stamens is three. The seed is erect. E: *Sclerolaena cuneata* (x1.75) has a spiny fruiting perianth with a circular attachment. The tube is oblong and dorsiventrally-compressed. The number of spines is three or four. A limb is prominent. The number of stamens is five. The seed is erect. F: The spiny fruiting perianth of *Sclerolaena glabra* (x1.5) has a circular attachment. The tube is oblong and dorsiventrally-compressed. The number of spines is three or four. There is no obvious limb. The number of stamens is five. The seed is erect. G: *Sclerolaena ventricosa* (x2) has a spiny fruiting perianth with a circular attachment. The perianth tube is cylindrical or sub-globular. The number of spines varies from three to five. A prominent limb is present. The number of stamens is five. The seed is erect. H: The fruiting perianth of *Threlkeldia diffusa* (x3) is without obvious appendages. The tube is convex to urceolate. A limb is absent. The number of stamens is three. The seed is placed horizontally. I: *Sclerolaena fusiformis* (x2.5) has a spiny fruiting perianth with circular attachment. The tube is cylindrical and dorsiventrally-compressed. The number of spines is four. The limb is inconspicuous. The number of stamens is three. The seed is oblique. J: The spiny fruiting perianth of *Sclerolaena bicornis* (x1) has an elliptic attachment. The perianth tube is sub-globular. The number of spines is two or three. The limb is prominently developed. The number of stamens is five. The seed is horizontal. K: *Sclerolaena alata* (x2.5) has a wing-spine-intermediate fruiting perianth with a circular attachment. The tube is turbinate. The number of (flattened) spines is six. There is no obvious limb. The number of stamens is five. The seed is horizontal. L: The fruiting perianth of *Sclerolaena convexula* (x3) is spiny and has a circular attachment. The tube is turbinate. The number of spines is five. The limb is inconspicuous. The number of stamens is three. The seed is oriented horizontally. M: *Sclerolaena napiformis* (x2) has a spiny fruiting perianth. The attachment is circular. The perianth tube is turbinate. The number of spines is six. The limb is inconspicuous. The number of stamens is five. The seed is horizontal. N: The winged fruiting perianth of *Osteocarpum acropterum* (x3.5) has a sub-globular perianth tube. The number of stamens is five. The seed is horizontal. O: The fruiting perianth of *Osteocarpum salsuginosum* (x3.5) has no obvious appendages. The perianth tube is globular. The number of stamens is five. The seed is horizontal. P: *Neobassia proceriflora* (x2) has a spiny fruiting perianth. The attachment is oblique. The perianth tube is cup-shaped. The number of spines is five. There is no obvious limb. The number of stamens is five. The seed is erect. Q: The fruiting perianth of *Dissocarpus fontinalis* (x2) is spiny. The tube is cylindrical. The number of spines is five. There is no obvious limb. The number of stamens is five. The seed is horizontal. The fruiting perianths fuse at their bases to form an aggregated infructescence. R: *Sclerolaena fimbriolata* (2x) produces a wing-spine-intermediate fruiting perianth. The attachment is circular. The tube is cylindrical. The number of (wing-like) spines is five. There is no obvious limb. The number of stamens is five. The seed is horizontal. Image sources are given in section 9.

### *3.4.3 Phylogeny of the Sclerolaena clade and distribution of spine-like perianth traits*

CA-ML inference of the *Sclerolaena* clade, using the accessions of clade 11 as root, resulted in six well to fully supported clades (Fig. 4). Clade 11 was sister to *N. proceriflora*, which in turn was sister to all other clades within the *Sclerolaena* clade. Clade 12, containing *Osteocarpum* and *S. brachyptera*, was sister to clade 13, which in turn was sister to a polytomy containing clades 14, 15 and clades 16 and 17, which were sister to another. The CA-ML phylogeny showing all BS support values is provided in Supplement 4 (Assembly 3).

The characters of the following descriptions are depicted in Figure 2 (habit, height, distribution of sexes, flowers per bract, seed orientation), Figure 4 (perianth appendage type, perianth tube shape, attachment shape, number of spines, appearance of a limb), Figure 5 (soil types and soil salinity), Figure 6 (habitat types) and Supplement 2 (species distribution).

***Neobassia proceriflora*** is a shrub up to 50 cm high. The bisexual flowers are solitary (Fig. 2). The cylindrical to cup-shaped perianth tube produces a shortly 5-spined structure (Fig. 1 S and Fig. 4 P). The attachment is oblique. The number of stamens is five and the seed is positioned obliquely. This species is distributed in the Eastern Desert and Central Queensland (Supplement 2). It usually grows in heavy, saline soils (Fig. 5) in Stony Desert and Riverine Desert habitats (Fig. 6).

**Clade 12** contained three *Osteocarpum* species and *S. brachyptera* (former *Sclerochlamys brachyptera*) (Fig. 4). The species are small shrubs (*S. brachyptera*) or perennials (*Osteocarpum*) up to 20 cm high (Fig. 2, Supplement 2). The flowers are bisexual and solitary. The perianth tube can be turbinate (*S. brachyptera*), cup-shaped (*O. dipterocarpum*) or sub-globular (*O. acropterum*, *O. salsuginosum*) (Fig. 3 and 4). The fruiting perianths of *O. acropterum* and *O. dipterocarpum* develop vertical wings with a prominent venation. The perianth of *O. salsuginosum* has no obvious appendage and *S. brachyptera* produces a hard, spine-like, 5-keeled wing. A radicular slit is absent in the *Osteocarpum* species

(Fig. 3, Supplement 2). The number of stamens is five (Fig. 4). The seeds are placed horizontally (Fig. 2 and 4). The distributions are centered and overlap widely in the Eastern Desert, Eyre Peninsular and Adelaide (Supplement 2). *Osteocarpum salsuginosum* has additional occurrences in the northeastern and southwestern parts of the Western Desert. The species grow mainly on heavy, loamy and clayey, moderately saline to saline soils (Fig. 5) near salt lakes in Riverine Desert and Desert Lake habitats (Fig. 6).

**Clade 13** contained nine *Sclerolaena* accessions (Fig. 4). The species are annuals (*S. johnsonii*, *S. cornishiana*, *S. muricata*), perennials *(S. alata*, *S. napiformis*, *S. deserticola*), perennial herbs (*S. cornishiana*) or shrubs (*S. parviflora*, *S. convexula*, *S. costata*) up to 50 cm high, with *S. muricata* growing up to 150 cm (Fig. 2, Supplement 2). The flowers are solitary and bisexual. The appendage type is mostly spinous, with *S. alata* and *S. parviflora* having wing-spine intermediate perianth appendages (Fig. 4). The attachment is usually circular (as is in clades 15 and 16), with exception of *S. cornishiana* and *S. deserticola* having an oblique and elliptic attachment, respectively. The perianth tube is cylindrical (as is in clade 15) to turbinate. The number of spines is mostly three or four with *S. muricata* producing a two- or three-spined fruiting perianth. An inconspicuous limb is present in *S. parviflora*, *S. convexula* and *S. napiformis*. The number of stamens is usually four or five, with *S. convexula* having three stamens. The seed is usually horizontal, with *S. muricata* having an oblique seed orientation (Fig. 2). The distributions of *S. convexula*, *S. cornishiana*, *S. costata*, *S. deserticola* and *S. johnsonii* are centered in the Western, Central and Eastern Desert (Supplement 2). *Sclerolaena alata* and *S. napiformis* are known only from the Western Desert and southeastern Eastern Desert, respectively. *Sclerolaena muricata* is widespread in the Eastern Desert, Central Queensland, Eastern Queensland and the Southeastern region. The distribution of *S. parviflora* reaches from the Southwestern region across the Western and Eastern Desert southwards to the Eyre Peninsular and Adelaide. With exception of *S. alata* and *S. napiformis*, the ranges overlap

in the interior of the arid zone (Western to Eastern Desert). The species predominantly grow in sandy, loamy to clayey soils of varying salinity (Fig. 5) in Sand Desert habitats (Fig. 6).

**Clade 14** contained four *Sclerolaena* accessions and *Threlkeldia inchoata* (Fig. 4). The species are perennials (*T. inchoata*), perennial herbs (*S. eurotioides*) or shrubs (*S. stylosa*, *S. limbata*, *S. bicornis*) up to 50 cm high, with *S. limbata* growing up to 100 cm (Fig. 2, Supplement 2). The flowers are solitary and bisexual. The appendage type spinous with *T. inchoata* having no obvious appendage (Fig. 2 and 4). The attachment is elliptic (Fig. 4). The perianth tube is variously shaped. The number of spines is two (*S. limbata*), two to three (*S. bicornis*, *S. stylosa*), or two to six (*S. eurotioides*). With exception of *S. eurotioides*, the fruiting perianth produces a prominent limb. The number of stamens is five. The seed is positioned horizontally (*S. bicornis*, *S. eurotioides*, *S. limbata*) or erect (*S. stylosa*, *T. inchoata*). *Sclerolaena eurotioides* is widespread in the Western Desert and Southwestern region (Supplement 2). *Sclerolaena stylosa* is known only from the North West Basin in the Western Desert. *Threlkeldia inchoata* occurs northwestern of the Lake Eyre Basin in the Eastern Desert. *Sclerolaena bicornis* is widespread across northern Western Australia, the Central and Eastern Desert, Arnhem and Central Queensland. The distribution of *S. limbata* is centered in the southeastern Eastern Desert. The species prefer mainly sandy or heavy clayey soils of varying salinity (Fig. 5) and usually occur in Riverine Desert habitats (Fig. 6).

**Clade 15** contained seven *Sclerolaena* species and *Threlkeldia diffusa* (Fig. 4). The species are perennials (*S. ventricosa*, *S. fusiformis*, *T. diffusa*), perennial herbs (*S. blackiana*) or shrubs (*S. glabra*, *S. intricata*, *S. tricuspis*, *S. articulata*) up to 40 cm high with *S. tricuspis* growing up to 100 cm. (Fig. 2, Supplement 2). The flowers are solitary and bisexual. The appendage type is spinous with *T. diffusa* having no obvious appendage (Fig. 4). The attachment is circular (*S. glabra*, *S. ventricosa*, *S. blackiana*) or oblique (*S. intricata*, *S. articulata*) (as is in clades 13 and 16). The perianth tube is variously shaped but often cylindrical (as is in clade

13) and dorsiventrally-compressed (as is in clade 16 and 17). The number of spines varies from two to five. A prominently developed limb is present in *S. intricata*, *S. tricuspis*, *S. articulata* and *S. ventricosa*. The number of stamens is three (*S. fusiformis*, *T. diffusa*) or five. The seed is oriented horizontally (*S. tricuspis*, *T. diffusa*), erect (*S. glabra*, *S. intricata*, *S. articulata*, *S. ventricosa*, *S. fusiformis*) or oblique (*S. blackiana*). The distributions of *S. articulata*, *S. blackiana*, *S. glabra*, *S. intricata*, *S. tricuspis* and *S. ventricosa* are centered in the Eastern Desert (Supplement 2). *Sclerolaena fusiformis* occurs in the Western Desert and *T. diffusa* is distributed along the coast of Western Australia, South Australia, Victoria and Tasmania. The species usually grow in heavy, clayey soils of varying salinity (Fig. 5) in Stony Desert habitats (Fig. 6).

**Clade 16** contained seven *Sclerolaena* accessions (Fig. 4). The species are annuals (*S. minuta*), perennials (*S. drummondii*, *S. densiflora*, *S. cuneata*, *S. divaricata*, *S.* sp. *Yeltacowie*) or shrubs (*S. lanicuspis*) up to 30 cm high, with *S. divaricata* growing up to 75 cm (Fig. 2, Supplement 2). The flowers are solitary and bisexual. The appendage type is spinous (Fig. 4). The attachment is circular (*S. cuneata*, *S. lanicuspis*, *S. densiflora*, elliptic (*S. divaricata*) or oblique (*S. minuta*, *S. drummondii*). The perianth tube is mainly oblong, turbinate and dorsiventrally-compressed. The number of spines varies from two to five. With exception of *S. densiflora* and *S. drummondii*, a prominent limb is present. The number of stamens is three (*S. divaricata*) or five. The seed is erect. *Sclerolaena drummondii* and *S. densiflora* are distributed in the Western Desert and the Southwestern region (Supplement 2). The distribution of *S. cuneata* and *S. lanicuspis* ranges from the Western Desert across the Central and Eastern Desert to Central Queensland. *Sclerolaena divaricata* is mainly distributed in the Eastern Desert and *S. minuta* occurs across the Central Desert, northern Eastern Desert and Central Queensland. *Sclerolaena* sp. *Yeltacowie* is known only from a collection site near Lake Torrens in the

Eastern Desert. The species grow in various soil types (Fig. 5), usually in Shield Plain habitats (Fig. 6).

**Clade 17** contained ten *Sclerolaena* accessions (Fig. 4). The species are perennial herbs (*S. uniflora*), perennials (*S. holtiana*, *S. diacantha*, *S. burbidgeae*, *S. eriacantha*, *S. parallelicuspis*, *S. patenticuspis*) or shrubs (*S. decurrens*, *S. brevifolia*, *S. obliquicuspis*) up to 30 cm high (Fig. 2, Supplement 2). The flowers are solitary and bisexual. The appendage type is spinous (Fig. 4). The attachment is circular (*S. uniflora*, *S. burbidgeae*, *S. decurrens*), elliptic (*S. holtiana*, *S. diacantha*, *S. parallelicuspis*, *S. obliquicuspis*, *S. brevifolia*) or oblique (*S. eriacantha*, *S. patenticuspis*). The perianth tube is variously shaped. The number of spines is two, with *S. decurrens* producing two or three spines. The limb is predominantly inconspicuous, except for *S. eriacantha*, *S. patenticuspis* and *S. decurrens*, which develop a prominent limb. The number of stamens is five. The seed can be horizontal (*S. diacantha*, *S. uniflora*), oblique (*S. holtiana*, *S. parallelicuspis*, *S. eriacantha*) or erect (*S. patenticuspis*, *S. decurrens*, *S. obliquicuspis*, *S. brevifolia*). The distributions of *S. holtiana*, *S. decurrens* and *S. parallelicuspis* are centered in the Eastern Desert (Supplement 2). *Sclerolaena obliquicuspis*, *S. diacantha*, *S. patenticuspis*, *S. eriacantha* and *S. uniflora* are widespread in the Western and Eastern Desert, and, with exception of *S. eriacantha*, the ranges extend to the southern coastline. *Sclerolaena brevifolia* is distributed from the Nullarbor to the Adelaide region and *S. burbidgeae* occurs in the Western Desert. The species grow in a variety of soils, often in calcareous (*S. diacantha*, *S. uniflora*, *S. brevifolia*, *S. obliquicupsi*, *S. patenticuspis*) or gypseous (*S. decurrens*) soil (Fig. 5) in Riverine Desert and Karst Plain habitats (Fig. 6).

### *3.5 Divergence time analysis with BEAST*

The filtered dataset used for the BEAST analysis was re-analyzed using CA-ML to check for any changes in topology and support values. The resulting BS support values were comparable to those of the unfiltered dataset and only minor topological differences were found (Supplement 4, Assembly 4). When collapsing branches below a BS support of 75, the topology was largely congruent with the topology of the unfiltered dataset (Fig. 2).

The time calibration using BEAST suggest that the Australian Camphorosmeae split from their Eurasian sister lineage during the Middle to Late Miocene ca. 12.99 Mya [9.96-15.97] (Fig. 5-7, Table 5, outgroup split). A first rapid diversification occurred in the Late Miocene ca. 8.34 [6.54-10.38] Mya (Fig. 6, Table 5, 1st polytomy). This first radiation includes clades 1-3, *M. oppositifolia*, *M. enchylaenoides* and *M. sedifolia*. A second rapid radiation occurred ca. 7.46 [5.85-9.23] Mya, and comprises clades 4-11 plus the species-rich *Sclerolaena* clade (*N. proceriflora* and clades 12-17, Fig. 6, Table 5, 2nd polytomy).

Crown group diversification of the first radiation started in a period of ca. 5.96 - 7.68 [4.22-9.67] Mya and was strongest during the Late Miocene (Messinian) (Fig. 6, Table 5). Crown group diversification of the second radiation (excluding the *Sclerolaena* clade) was strongest during the Pliocene in a period of 4.46 – 6.06 [3.09-7.63] Mya (Fig. 6, Table 5). Crown group diversification of the *Sclerolaena* clade started ca. 6.27 Mya [4.88-7.68] Mya and was strongest during the Pliocene in a period of 3.54 – 5.32 [2.53-6.56] Mya (Fig. 5 and 6, Table 5). Speciation slowed during the Pleistocene, but was relatively strong in some clades (Fig. 5 and 6, clades 2, 4, 6, 15, 16 and 17), while other clades showed no or only little divergence (clades 3, 5, 11, 13 and 14). The time estimates are largely congruent with the findings of Cabrera et al. (2011).

**Table 5.** Results of the BEAST divergence time analysis. Given are the node age estimates and 95% confidence intervals (HPD) for the outgroup split, the two polytomies plus the *Sclerolaena* clade, and the crown group estimates for the clades.

| | estimated age (Mya) | 95% HPD (Mya) |
|---|---|---|
| outgroup split | 12.99 | 9.96-15.97 |
| 1st polytomy (stem age) | 8.34 | 6.54-10.38 |
| clade 1 | 7.60 | 5.54-9.67 |
| clade 2 | 5.96 | 4.22-7.85 |
| clade 3 | 7.30 | 5.64-9.16 |
| clade 2+3 | 7.68 | 5.94-9.52 |
| *M. enchylaenoides* | 8.34 | 6.54-10.38 |
| *M. sedifolia* | 8.34 | 6.54-10.38 |
| 2nd polytomy (stem age) | 7.46 | 5.85-9.23 |
| *M. cannonii* | 7.46 | 5.85-9.23 |
| clade 4 | 6.08 | 4.63-7.63 |
| clade 5 | 5.47 | 3.87-7.11 |
| clade 6 | 4.46 | 3.09-5.92 |
| clade 7 | 5.53 | 3.87-7.23 |
| clade 8 | 0.38 | 0.16-0.68 |
| clade 9 | 1.93 | 0.97-3.03 |
| clade 10 | 5.04 | 3.76-6.47 |
| clade 11 | 5.75 | 4.33-7.26 |
| *Sclerolaena* clade | 6.27 | 4.88-7.68 |
| *N. proceriflora* | 6.27 | 4.88-7.68 |
| clade 12 | 4.80 | 3.55-6.12 |
| clade 13 | 5.32 | 4.09-6.56 |
| clade 14 | 4.29 | 3.25-5.42 |
| clade 15 | 4.72 | 3.64-5.87 |
| clade 16 | 3.54 | 2.53-4.6 |
| clade 17 | 3.90 | 2.95-4.9 |

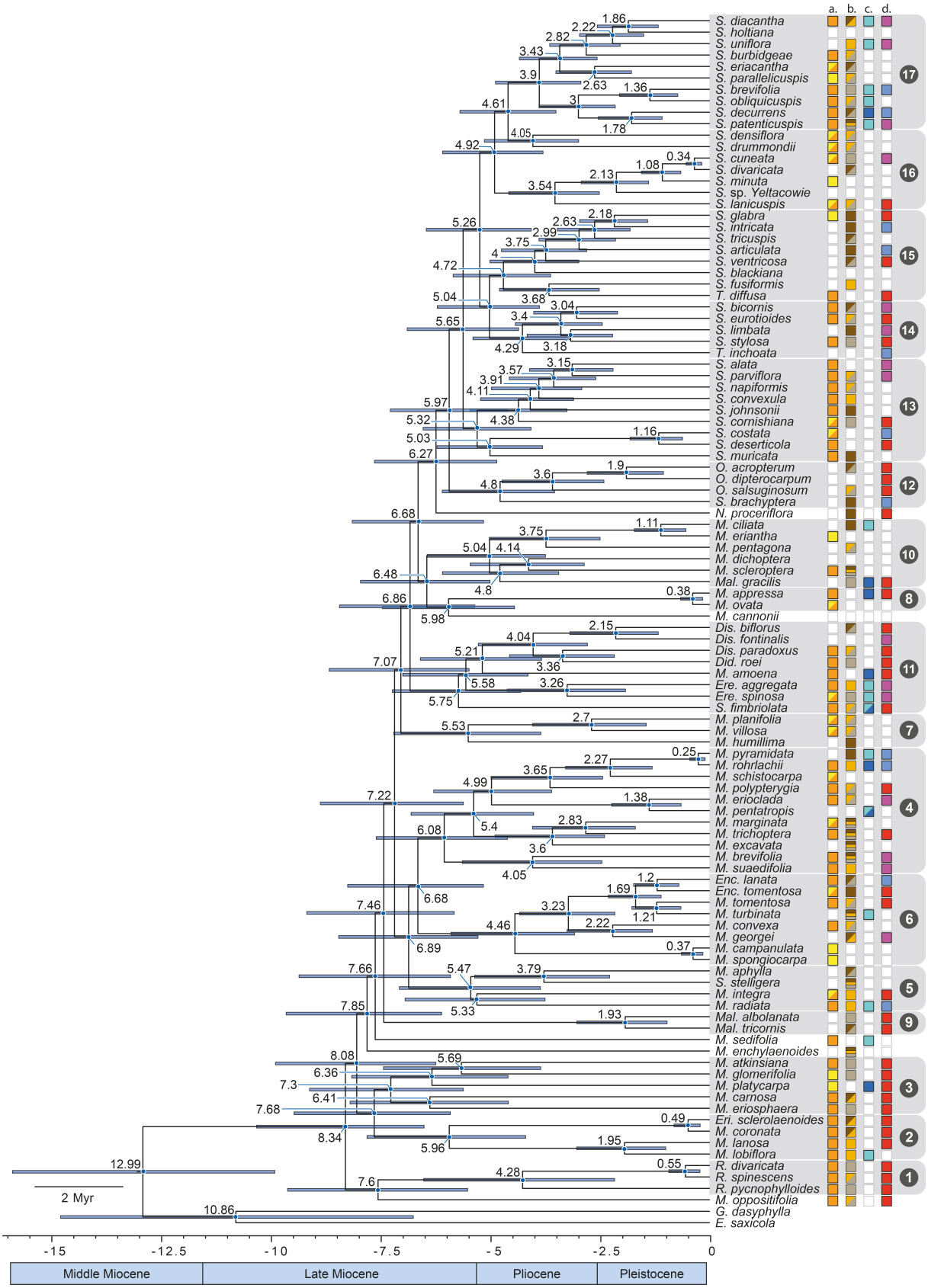



**Figure 5**. Chronogram with species preferences for soil substrates and salinity. Shown is the un-collapsed topology. Given are the node age estimates and 95% confidence intervals (blue bars).

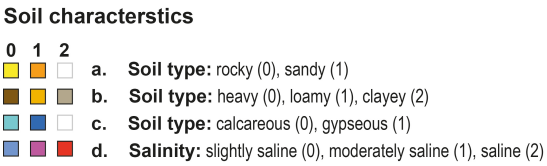



**Figure 5 continued**. Chronogram with species preferences for soil substrates and salinity. Timeline 16 Million years ago to present. The encircled clade numbers are shaded in grey. For each species are given: a.-c. soil substrate preferences, d. soil salinity preferences. Note: The information on preferred soil types and soil salinity is inconsistent. Missing information was classified as "not applicable" (Supplement 3). Confidence intervals are shown in Supplement 1, Assembly 4.

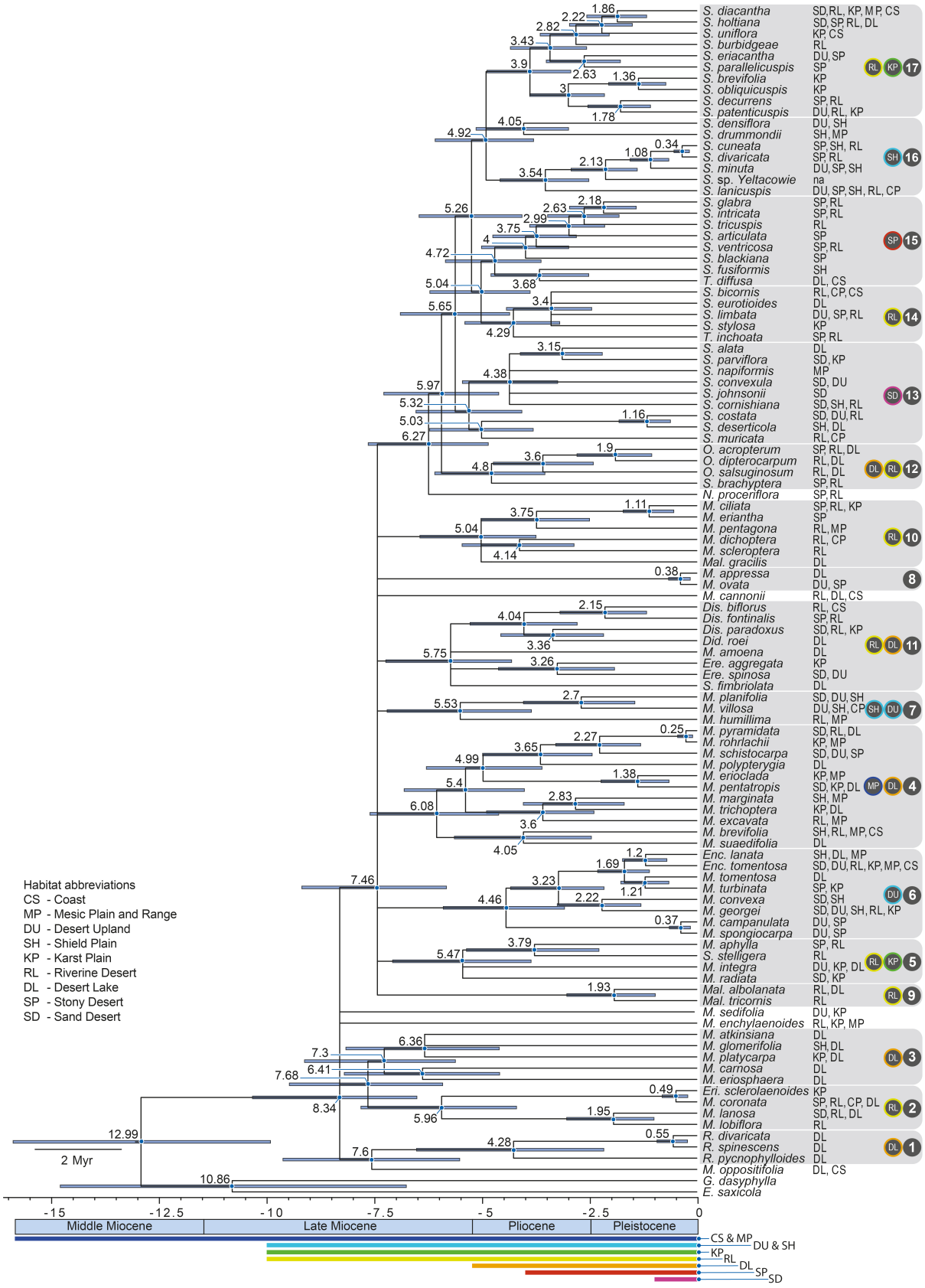



**Figure 6**. Chronogram with species habitat occurrence. Shown is the collapsed topology. Given are the node age estimates and 95% confidence intervals (blue bars). Timeline 16 Million years ago to present. Timeline of landscape (habitat) history based on McDonald (2020). The encircled clade numbers are shaded in grey. The occupied habitats are given for each species. The habitat occupations were taken from McDonald (2020). Missing information was classified as "not applicable" (Supplement 3). For each clade, the most frequently colonized

habitats are color-coded with respect to the habitat colors shown below the time scale. Confidence intervals are shown in Supplement 1, Assembly 4.

# *4. DISCUSSION*

## *4.1 Morphological and ecological characterization of phylogenetic clades*

The Australian Camphorosmeae have undergone numerous nomenclatural and taxonomic revisions using mainly characteristics of the fruiting perianth for generic delimitation (e.g. Mueller, 1882a, b; Ulbrich, 1934; Anderson, 1923; Ising, 1964; Wilson, 1975, 1984; Scott, 1978). The characters used to delimit the genera of the Australian Camphorosmeae do not always create a natural grouping of species and the current taxonomy is to some extent artificial (Wilson, 1975, 1984; Cabrera et al., 2009). Incomplete Lineage Sorting (as a consequence of rapid radiations), convergent evolution of homoplastic perianth traits, ongoing hybridization and species sharing very similar morphological traits complicate the taxonomy. Cabrera et al. (2009) found the perianth appendage type to be of systematic relevance. Based on this outcome, 25 traits were tested for taxonomic relevance in this study (Table 2). As already shown by Cabrera et al. (2009), some traits, such as the phyllotaxy and the leaf pubescence appear to be randomly distributed across the phylogeny (Fig. 2). Some other traits appear to have systematic relevance to the distinction between the *Maireana* grade and the *Sclerolaena* clade. It turned out that species of the *Maireana* grade tend to be larger and presumably longer-lived (annuals and short-lived perennials vs long-lived shrubs) than species of the *Sclerolaena* clade (Fig. 2, habit and growth height). This finding is in line with general assumptions to habit differences between *Maireana* (long-lived shrubs) and *Sclerolaena* (short-lived perennials) species (Scott, 1978; Wilson, 1984; McDonald, 2020). These differences in life history are probably related to the different perianth appendage types and dispersal syndromes (see 4.2.4). On a smaller scale, it appears that more specific features of the fruiting perianth may be of systematic relevance for circumscribing clades and small genera (Fig. 3 and 4). And to some extent, ecological features (such as preferred soil types and habitats) could also serve to circumscribe these clades and small genera (Fig. 5 and 6).

In this first section of the discussion, it is examined whether the resulting phylogenetic clades can be circumscribed using vegetative, reproductive, and ecological characters (in total 25 characters) and whether they appear plausible from a non-molecular view. Furthermore, the clades found are discussed in the context of previous taxonomic treatments. Finally, this section summarizes the recommendations for a combination of species and for a possible taxonomic revision.

For each clade, a brief description of the results is presented first. Then, each species is briefly described with the necessary characters of the fruiting perianth (and vegetative characteristics, when relevant). In addition, a description of potential variants and hybrids is given. This is followed by a description of the distribution, preferred soil types, and habitats colonized. If applicable, postulated assumptions regarding a closer relationship with other species and genera, as well as the taxonomic history, are then given. Each clade concludes with a brief summary.

The morphological descriptions were taken from the "Flora of Australia, Volume 4, Chenopodiaceae" (Wilson, 1984), unless indicated otherwise. Information on preferred soil types was also taken from this flora, as well as from the "Flora of New South Wales, Volume 4" (Harden, 1990), the "Flora of Victoria" (Walsh, 1996) and "The descriptive catalogue of the Western Australian flora" (Paczkowska and Chapman, 2000). Information on colonized habitats is from McDonald's "Biogeography of Australian chenopods" (McDonald, 2020).

### *4.1.1 Characterization of the Maireana grade (clades 1-11)*

**Clade 1** contains the three described *Roycea* species (Fig. 3). This clade may also include *M. oppositifolia* (Fig. 2). The perianth of the *Roycea* species surrounds the subglobular fruit at the base, but does not enlarge in fruit and thus does not envelope the utricle. The fruiting perianth of *M. oppositifolia* produces a tube that envelopes the utricle and a 5-lobed wing.

In *R. divaricata*, the subglobular, bisexual flowers are located in leafy fascicles. The tepals are free and do not enlarge in fruit. The fruit is depressed-spherical. In *R. spinescens* the cup-shaped (male, Fig. 1 U) and globular (female) flowers are located in the upper leaf-axils. The tepals are united (male) or free (female). The sub-globular fruit is surrounded by the persistent perianth at the base. In *R. pycnophylloides*, the cup-shaped (male) and sub-globular (female) flowers are located towards the apex of branches (Fig. 1 V). The fruit is broadly ovoid and surrounded by the persistent perianth at the base. In *M. oppositifolia*, the broadly obconical to convex tube envelopes the utricle, and a 5-lobed, fan-shaped, prominently veined wing is present.

The species of *Roycea* are endemic in temperate und subtropical south-western Western Australia and usually occur on saline sandy flats (*R. pycnophylloides*), in saline sand and sandy clay (*R. spinescens*) and heavy saline soil (*R. divaricata*) in Desert Lake habitats. *Maireana oppositifolia* is found along the south coast of Western Australia and South Australia, and also around inland salt lakes in western Victoria and southern Western Australia, in saline soil associated with salt lakes (Desert Lake habitats) or coastal areas (Coast habitats). In southern Western Australia, it is sympatric with *Roycea.*

Wilson (1975) assumed intergeneric relationship of *Roycea* spp. and *M. oppositifolia*. “The leaves, indumentum, and flowers of the three *Roycea* species correspond closely to that found in *M. oppositifolia*. The fruiting perianth is, however, quite different” (Wilson, 1975). *Maireana oppositifolia*, *R. pycnophylloides* and *R. spinescens* are dioecious, while *R. divaricata*

is hermaphrodite. The inland variant of *M. oppositifolia* (found in Western Australia) has divaricate branches with small alternate leaves clustered into compact glomerules, and strongly resembles the vegetative morphology of *M. glomerifolia* (clade 3) and *R. divaricata* (Wilson, 1975, 1984). The other variant (included in this study), which is found along the south coast of Western Australia and South Australia on mud flats, and also around inland salt lakes in western Victoria and southern Western Australia, strongly resembles the habit of *R. spinescens*.

Despite the quite different perianth morphology, the species share similar vegetative features and ecological preferences and thus this clade appears plausible from a non-molecular point of view.

***Maireana enchylaenoides*** produces a slightly convex perianth tube with five horizontal wings that only slightly exceed the tube (Wilson, 1984). The wings are broadly lunate, unevenly imbricate and the wing margins are incurved and somewhat inflated. This species has a disjunct distribution. It is found in southern Queensland, New South Wales, Victoria, south-eastern South Australia and south-western Western Australia on heavy, loamy soil (Wilson, 1984) in Riverine Desert, Karst Plain and Mesic Plain habitats (McDonald, 2020).

According to Wilson (1975), the form of the fruiting perianth is similar to that found in *M. brevifolia* (clade 4) and *M. diffusa* (not included). With this, he justified not to place this species in a separate genus. Given the general habit (small perennial with woody rootstock, flowers arranged in loose, leafy spikes) and the preferred soil types and habitats, this species could be assigned to clade 2.

**Clades 2** and **3** (Fig. 3) contain eight *Maireana* species and the monotypic genus *Eriochiton,* and may be sister to another (Fig. 2). All species have a winged or wing-spine intermediate (*Eri. sclerolaenoides*) fruiting perianth that can be entire (*M. atkinsiana*, *M. glomerifolia*, *M. platycarpa*, *M. carnosa*, *M. coronata*, *M. lanosa*) or 5-lobed (*M. eriosphaera*, *Eri. sclerolaenoides*, *M. lobiflora*). The most prominent features of these species are the

accessory processes present on the upper perianth above the horizontal wing. These additional processes appear as erect lobes or a woolly indumentum.

The morphology of *M. atkinsiana* (Fig. 1 C) and *M. glomerifolia* is quite similar. Both species produce a convex tube with a simple horizontal wing. A radicular slit is absent in *M. atkinsiana* and present in *M. glomerifolia*. The convex upper perianth is deeply 5-lobed. The processes of the upper perianth are unequally arranged. Two of the lobes produce a pair of erect, narrowly oblong processes, one lobe is has a single process, and the other two lobes have no processes. An important difference is found in the distribution of sexes, *M. atkinsiana* is dioecious while *M. glomerifolia* is polygamodioecious. The perianth morphology of *M. lobiflora* (Fig. 1 E) and *M. lanosa* is also quite similar. In *M. lobiflora,* the perianth tube is slightly convex and produces a horizontal wing of five fan-shaped lobes. In *M. lanosa*, the perianth tube is slightly convex and produces a horizontal, simple wing. A radicular slit is present. The upper perianth of both species bears 4+2 erect processes alternating with the perianth lobes, a contiguous pair is placed either side of the radicular slit. The flowers are arranged in rather loose spikes. In *M. carnosa* and *M. eriospharea*, a dense, woolly indumentum arises from the upper perianth, and covers the fruiting perianth. In *M. carnosa*, the shallowly hemispherical perianth tube produces a simple, horizontal wing and a radicular slit is present. In *M. eriosphaera*, the hemispherical to turbinate perianth tube produces five short rounded erect wing lobes that sometimes form a narrow horizontal wing. The flowers are arranged in congested leafy spikes. In *M. coronata*, the fruiting perianth is formed like an "inverted academic hat" (Wilson, 1984) and produces a narrow wing. A radicular slit is present. The upper perianth produces a cup-shaped outgrowth (corona). The flowers are arranged in congested leafy spikes. In *M. platycarpa*, the flat or slightly convex perianth tube produces a simple, horizontal wing. A radicular slit is present. The upper perianth is convex and somewhat woolly. In *Eri. sclerolaenoides*, the sub-globular fruiting perianth produces appendages in two

series (Fig. 1 O). The five inner appendages are erect, oblong, deeply divided and arise from the base of the perianth lobes. The five outer appendages are spreading, spinescent and arise from the base of the inner appendages. The fruiting perianth is covered in a woolly mass and a radicular slit is present. The flowers are arranged in congested leafy spikes.

*Maireana atkinsiana* is found in Western Australia from the Shark Bay area south to Watheroo and east to Laverton, in sandy clay, saline flats and claypans, usually near salt lakes in Desert Lake habitats. *Maireana glomerifolia* is found in the dry inland areas of Western Australia, in stony soils and clay, usually in saline or subsaline areas, rocky hillsides and claypans in Shield Plain and Desert Lake habitats. *Maireana lobiflora* is found from the southern half of Western Australia eastwards to western New South Wales and Victoria in sandy or loamy soils, river flats and limestone flats in Riverine Desert habitats. *Maireana lanosa* is found from north-western Western Australia to south-western Queensland and central New South Wales in sandy and loamy soils, saline flats and floodplains in Sand Desert, Riverine Desert and Desert Lake habitats. *Maireana carnosa* is found in south-eastern Queensland, northern South Australia and temperate und subtropical regions of Western Australia in sand, clay, loam and heavy soil in saline flats and often around salt lakes in Desert Lake habitats. *Maireana eriosphaera* is found in arid areas of Western Australia in clay, clayey sand, at the base of breakaways and near margins of salt lakes in Desert Lake habitats. *Maireana coronata* is found from the southern Northern Territory and eastern South Australia eastwards to central Queensland and New South Wales, usually in sandy loam, heavy soil and salt flats in Shild Plain, Desert Clay Plain, Riverine Desert and Desert Lake habitats. *Maireana platycarpa* is found in Western Australia from the Carnarvon area south-east to Norseman, usually in saline flats, gypsum dunes and rocky hillsides in Karst Plain and Desert Lake habitats. *Eriochiton sclerolaenoides* is found in temperate Western Australia eastwards to western New South Wales

and north-western Victoria on loam, clay or sand, heavy alluvial soils and saline flats, usually in Karst Plain habitats.

It is little surprise that *Eriochiton* was found in this group, given the morphological features of the fruiting perianth and the taxonomic history of this monotypic genus. *Eriochiton sclerolaenoides* had been placed in *Bassia* sect. *Eriochiton* by Anderson (1923) and Ising (1964) as the only species. The decisive factor here was the five erect membranous appendages (in addition to the five spinescent appendages), resembling the perianth morphology of some *Maireana* species (*Kochia*). Wilson (1975) treated *Eri. sclerolaenoides* as *Maireana* and saw morphological similarities to other *Maireana* species with wing-like appendages and erect processes arising from the upper perianth (*M. lobiflora*). However, he noted that *Eri. sclerolaenoides* does not fit easily into any of the genera described, and pointed to the spiny wings and the erect processes that are placed opposite the perianth lobes as important differences from other *Maireana* species. Scott (1978) interpreted the 5-lobed perianth appendages as spines, not wings, and placed the species in the monotypic genus *Eriochiton*, which was accepted by Wilson (1984). In a phylogenetic study by Cabrera et al. (2009), *Eriochiton* was placed in a statistically supported clade with *Didymanthus* and *Dissocarpus*. The authors noted that the species do not appear to be particularly closely related to each other.

The accessory processes present on the upper perianth above the horizontal wing are characteristic for this clade (with exception of *M. platycarpa*). In addition, the species prefer similar habitats and soil types and the distribution areas overlap widely, especially in Western Australia. Accordingly, this clade appears plausible from a non-molecular point of view.

***Maireana sedifolia*** is a dioecious shrub, covered with a close bluish grey woolly tomentum (hence, the common name “blue bush”). The flowers are paired but usually only one matures (Fig. 2 and 3). The fruiting perianth produces a hemispherical to turbinate tube with a simple, horizontal wing. A prominent venation and a radicular slit are present. The upper

perianth is flat and without erect processes. This species is found in south-western New South Wales and north-western Victoria, South Australia, the southern Northern Territory and south-eastern Western Australia in calcareous soils in Desert Upland and Karst Plain habitats. It is one of the principal shrub species over much of the southern arid zone of Australia (Wilson, 1975).

Wilson (1975, 1984) noted that *M. sedifolia* and *M. astrotricha* (not included) have been confused but can readily distinguished by the dendritic hairs of the latter. *Maireana astrotricha* is a polygamodioecious shrub (as is *M. glomerifolia*, clade 3) with a dense indumentum and solitary flowers. The fruiting perianth produces a turbinate tube with a simple, horizontal wing. A radicular slit is present, the upper perianth is convex and without extra processes. It is found in western New South Wales, central South Australia and the southern portion of the Northern Territory in well drained gravelly soil or on calcareous rises.

Although this species has some characteristics in common with other species, such as the paired flowers of which only one matures (*M. planifolia*, clade 7), the strict dioecy (*M. oppositifolia*, *M. atkinsiana*) or the general habit (*M. astrotricha*) it is difficult to assign to another clade.

***Maireana cannonii*** is a small shrub with opposite leaves that are shortly spurred at the base and recurved at the apex. The shortly hemispherical perianth tube produces a simple, horizontal wing (Fig. 2 and 3). A radicular slit and a prominent venation are present. The upper perianth is convex and without extra processes. This species occurs only in South Australia between Port Pirie and Leigh Creek in Riverine Desert, Desert Lake and Coast habitats.

It is difficult to assign this species to any clade because the features of the fruiting perianth are homoplastic and found in many other clades. The closest similarity is to *M. oppositifolia*. The habits are reminiscent of another (compact shrub, opposite, spurred leaves) (Wilson, 1984). Both species occur in similar habitats and the distribution of *M. cannonii* is within the range of *M. oppositifolia*. The morphology of the fruiting perianth, however, is quite different.

**Clade 4** contains eleven *Maireana* species (Fig. 2 and 3). This group shows quite heterogeneous morphological characters and is difficult to summarize. The fruiting perianth tube is usually turbinate to convex. Most species develop a simple, horizontal winged perianth appendage (with exception of the 5-lobed wing in *M. brevifolia*). Additional vertical wings along the perianth tube can be present (*M. polypterygia*, *M. pentatropis*, *M. erioclada*). A prominent wing venation can be present (*M. rohrlachii*, *M. schistocarpa*, *M. polypterygia*, *M. trichoptera*, *M. brevifolia*, *M. suaedifolia*). A radicular slit is present in most species, with exception of *M. marginata* and *M. brevifolia*). The upper perianth is usually flat, or can be convex *(M. pyramidata*, *M. schistocarpa*, *M. pentatropis*, *M. erioclada*) or thickened (*M. brevifolia*), and is without extra processes.

In *M. pyramidata*, the flat to shortly turbinate perianth tube produces a simple, horizontal wing. A radicular slit is present and the upper perianth is pyramidal (convex). In *M. rohrlachii*, the broadly turbinate to hemispherical perianth tube produces a simple, horizontal wing with prominent venation. A radicular slit is present and the upper perianth is flat. In *M. brevifolia*, the shallowly hemispherical perianth produces five horizontal, fan-shaped wings. A prominent wing venation is present. In *M. suaedifolia*, the shortly hemispherical perianth produces a simple, horizontal wing. A prominent wing venation and a radicular slit are present. In *M. schistocarpa*, the broadly turbinate perianth tube produces a simple, horizontal wing with a prominent venation. A radicular slit is present from whose margins the wing passes down the tube (vertical wing). The upper perianth is convex. In *M. polypterygia*, the cup-shaped perianth tube produces a wing at the base and at the apex and bears several imperfectly developed vertical wings. A radicular slit is present and the upper perianth is flat. In *M. pentatropis*, the turbinate perianth tube produces a simple, horizontal wing with three to five vertical wings along the lower half of the tube. A radicular slit is present and the upper perianth is convex. In *M. erioclada*, the narrowly funnel-shaped perianth tube produces a simple, horizontal wing at

the apex with five vertical wings attached to the horizontal wing and to the tube throughout its length. A radicular slit is present and the upper perianth is convex. In *M. marginata*, the depressed barrel-shaped (convex) perianth tube produces a narrow horizontal wing. When dry, the lower half of the tube forms an expanded spongy base, the upper half becomes broadly turbinate. Prominent venation and a radicular slit are absent. The upper perianth is flat. In *M. trichoptera*, the lower half of the perianth tube is expanded into a fleshy hollow stipe, the upper half is shortly turbinate. When dry, the lower half becomes spongy and the upper half becomes convex. The simple, horizontal wing at the perianth apex is prominently veined. A radicular slit is present and the upper perianth is flat. In *M. excavata*, the depressed barrel-shaped perianth tube is fleshy when fresh. When dry, the lower half becomes expanded into a spongy hollow base, the upper half becomes broadly turbinate to convex. The simple, horizontal wing at the apex is prominently veined. A radicular slit is present and the upper perianth is flat.

*Maireana pyramidata* has a disjunct distribution, separated by the central Western Plateau and the Nullarbor Plain. It is found in inland Western Australia and eastwards through South Australia to western New South Wales and north-western Victoria in calcareous and slightly saline heavy soil in Sand Desert, Riverine Desert and Desert Lake habitats. In Victoria, it occurs often in areas prone to flooding (Walsh, 1996). *Maireana rohrlachii* occurs from northern Eyre Peninsular in South Australia eastwards to western Victoria on slightly saline or gypseous soil, sandy loam, and often around lakes or in seasonally wet areas (Walsh, 1996), usually in Karst Plain and Mesic Plain habitats. *Maireana brevifolia* has a disjunct distribution, separated by the central Western Plateau and the Nullarbor Plain. It is found in western and south-western Western Australia, and from southern Queensland over inland New South Wales to southern South Australia and north-western Victoria. It occurs on loam, sandy clay, sub-saline soils, and in heavy, seasonally wet soils, in Shield Plain, Mesic Plain, Riverine Desert and Coast habitats. This species is an early colonizer of disturbed lands. *Maireana suaedifolia*

has a disjunct distribution, separated by the central Western Plateau and the Nullarbor Plain. It is found in the western inland areas of Western Australia and in southern South Australia on loam and sand, in sand dunes around salt lakes and on alluvial plains, usually in Desert Lake habitats. *Maireana schistocarpa* has a disjunct distribution, separated by the Lake Eyre – Lake Torrens Basin and the Flinders Ranges. It is found in central Australia in south-western Queensland, north-western New South Wales, and south-eastern Northern Territory and north-eastern South Australia on non-saline soils such as stony hills and rises, sand dunes and gibber plains, in Sand Desert, Desert Upland and Stony Desert habitats. *Maireana polypterygia* is found only in north-western Western Australia between Exmouth and Shark Bay on sandy clay, loam and salt marshes, usually in Desert Lake habitats. *Maireana pentatropis* occurs in central and south-eastern Western Australia, the southern Northern Territory, South Australia, and western New South Wales and Victoria on calcareous soil, hilltops or flats, lake margins and gypsum dunes, in Sand Desert, Karst Plain and Desert Lake habitats. *Maireana erioclada* is found from southern Western Australia, eastwards to western New South Wales and Victoria on sand, loam, clay and sometimes sub-saline soil, in Karst Plain and Mesic Plain habitats. This species is a frequent invader of disturbed areas (alongside roads). *Maireana marginata* is found only in south-western Australia from Morawa to Ravensthorpe in clay, sandy to heavy loam, gravel and scree slopes, in Shield Plain and Mesic Plain habitats. Its occurrence is often associated with eucalyptus woodlands. *Maireana trichoptera* is widespread in western New South Wales, western Victoria, South Australia, temperate Western Australia and the southern portion of the Northern territory. It occurs in a variety of soils such as sandy, clayey and (heavier) loamy soils, on plains, rises, salt lakes and in eucalyptus woodlands, usually in Karst Plain and Desert Lake habitats. *Maireana excavata* has a disjunct distribution, separated by the southern Central Lowlands. It is found in south-eastern South Australia, and southern New South Wales and western Victoria in heavy, loamy and clay soils, commonly in grassland or open woodlands, usually in Riverine Desert and Mesic Plain habitats.

The species of this group have their main distributions in the south-western, southern and south-eastern portions of Australia. Many of them are sympatric over wide areas (e.g. *M. brevifolia*, *M. erioclada*, *M. pyramidata*, *M. pentatropis*, *M. trichoptera*), with some species restricted to rather small ranges in Western Australia (*M. marginata*, *M. polypterygia*). The species occur in a variety of soil types and habitats. The morphological characters are quite heterogeneous and not exclusive to this group, but shared with other *Maireana* clades (e.g. clades 5 - 8). Thus, it is difficult to separate this group from the other *Maireana* clades. However, two morphological subgroups stand out. Four species develop additional vertical wings along the perianth tube (*M. schistocarpa*, *M. polytperygia*, *M. pentaropis* and *M. erioclada*), and three species have a perianth tube dissected in the middle (*M. marginata*, *M. trichoptera* and *M. excavata*).

Some of Wilson's (1975) assumptions regarding relatedness, based on morphological similarities and putative hybrids, can be confirmed. Wilson pointed out that *M. brevifolia* (and *M. diffusa,* not included) appear to be most closely related to *M. suaedifolia* (and *M. decalvans,* not included). The former two species may be distinguished by differences in habit, leaf shape and by the absence of bracteoles (in *M. diffusa*). The latter two species may be distinguished by the completely glabrous branches, leaf shape and the presence of bracteoles (in *M. suaedifolia*). Further, he assumed a closer relationship of *M. trichoptera*, *M. marginata* and *M. excavata*, which all have an inflated, convex to cup- or barrel-shaped perianth tube and mainly differ in the shape of the horizontal wing. He also suggested a closer relationship of *M. polypterygia*, *M. pentatropis*, and *M. erioclada* (and *M. triptera*, not included), all of which develop additional vertical wings along the perianth tube, and described a putative hybrid between *M. pentatropis* and *M. erioclada*. Lastly, he described *M. schistocarpa* and *M. pyramidata* as most closely allied, which can be distinguished by the leave shape and the wing that passes down the perianth tube in *M. schistocarpa*.

Although some assumptions by Wilson (1984) regarding closer relationship can be confirmed, this clade does not appear plausible from a non-molecular perspective since the morphological features are homoplastic and occur in several other clades of the *Maireana* grade.

**Clade 5** contains three *Maireana* species and *S. stelligera* (Fig. 2 and 3). The species produce a horizontal wing (*M. integra*, *M. radiata*, *M. aphylla*) or a horizontal, wing-spine intermediate perianth appendage (*S. stelligera*). The tube is turbinate to sub-globular. A radicular slit and prominent wing venation can be present. The upper perianth is flat (*M. integra*, *S. stelligera*) or convex (*M. radiata*, *M. aphylla*) and without extra processes.

In *M. radiata*, the hemispherical perianth tube produces a simple, horizontal wing. A prominent venation and a radicular slit are present. The upper perianth is slightly convex. In *M. integra*, the hemispherical perianth tube produces a simple, horizontal wing. A prominent venation and a radicular slit are absent. The upper perianth is flat. In *M. aphylla*, the hemispherical to turbinate perianth tube produces a simple, horizontal wing. A prominent venation and a radicular slit are present. The upper perianth is thickened and convex. In *S. stelligera*, the sub-globular perianth tube produces a narrow horizontal wing with a crenate-dentate margin from which protrude about eleven short radiating spines (five tepaline, five intertepaline, one spine opposite the radicle). A radicular slit is present. The upper perianth is flat.

*Maireana radiata* is found in southern Australia from Western Australia east to north-western Victoria and south-western New South Wales, in sandy loam, calcareous soils, slightly saline or alkaline situations, usually in Sand Desert or Karst Plain habitats. *Maireana integra* is found from western New South Wales westwards through South Australia and the southern Northern Territory into south-central Western Australia. It occurs in clay pans, saline flats and well-drained loamy soil, in Desert Upland, Karst Plain and Desert Lake habitats. *Maireana aphylla*

has a disjunct distribution. It occurs in the Carnarvon-Willuna area in north-western Western Australia, and from the southern portion of the Northern Territory south-eastwards through South Australia and southern Queensland to New South Wales and northern Victoria. It usually occurs in seasonally waterlogged heavy clayey soil, in Shield Plain and Desert Lake habitats. *Sclerolaena stelligera* is found in eastern temperate Australia in southern Queensland, central and western New South Wales, eastern South Australia and north-western Victoria. It usually occurs in heavy, loamy or clayey seasonally waterlogged soil, usually in Riverine Desert habitats.

The species are sympatric over wide ranges and prefer similar soil conditions. However, this clade also lacks exclusive morphological characters. Wilson (1975) noted morphological affiliations and intermediate collections of *M. integra* and *M. villosa* (clade 7), with which it has been confused, but is readily distinguishable by the radicular slit present in *M. villosa*. The intermediate forms may be of hybrid origin and may have arisen following land disturbance by agricultural activities. The putative hybrid is found principally on rocky hillsides. He further mentioned similarities to *M. tomentosa* (clade 6) with which it may integrates. The position of *S. stelligera* within a *Maireana* grade clade is worth highlighting, given the taxonomic history of this species. In the treatment of the Australian *Bassia* species, Scott (1978) described the new genus *Stelligera* to include one species, *Stelligera endecaspinis*. He pointed to the close relationship to *Sclerolaena* (with intertepaline spines), from which it differs by possessing also tepaline spines connected by a short, thickened wing. Wilson (1984) accepted the new monotypic genus and suggested a close relationship to *Sclerolaena walkeri*, *Sclerolaena brachyptera* (*Sclerochlamys brachyptera*) and *Maireana microcarpa*. Jacobs (1988) saw little justification for maintaining the genus *Stelligera* as distinct due to the strong morphological affinities to other *Sclerolaena* species and made a new combination, *Sclerolaena stelligera* (F. Muell) S.W.L. Jacobs. In a phylogenetic study by Cabrera et al. (2009), *S. stelligera* was

resolved in a polytomy dominated by *Maireana* species, and the authors questioned the inclusion in *Sclerolaena* and the resurrection of a separate genus. Inclusion in *Maireana* appears reasonable, based on the sub-globular to hemispherical perianth tube and the horizontal, wing-like perianth appendage with a dentate to spiny wing margin. Apart from that, this clade does not appear plausible from a non-molecular perspective due to the homoplastic perianth features that are also present in several other *Maireana* clades (e.g. 3, 4, 6-8).

**Clade 6** contains six *Maireana* species and the two species of *Enchylaena* (Fig. 3). The perianth tube is turbinate (*M. turbinata*, *M. convexa*, *M. georgei*), cup-shaped (*M. tomentosa*, *M. campanulata*) or sub-globular (*Enc. tomentosa*, *Enc. lanata*, *M. spongiocarpa*). The *Maireana* species develop a simple, horizontal wing, while *Enchylaena* produces a succulent fruiting perianth that can extend upwards as a short corona, corresponding to the wing of *Maireana* (Wilson, 1975, 1984). A radicular slit and prominent wing venation can be present. The upper perianth can be flat (*M. turbinata*, *M. tomentosa*, *M. georgei*, *M. campanulata*, *M. spongiocarpa*), convex (*M. convexa*) or thickened (*Enchylaena*).

In *M. turbinata*, the turbinate, crustaceous perianth tube produces a simple, faintly veined horizontal wing. A radicular slit is absent. The upper perianth is flat and without extra processes. *Maireana tomentosa* has two subspecies that differ in the morphology of the fruiting perianth and the distribution range. *M. tomentosa* subsp. *tomentosa* (included) has a hemispherical, thickly crustaceous and dull perianth tube. *M. tomentosa* subsp. *urceolata* (not included) has a cup-shaped, woody and glossy perianth tube. The fruiting perianth of both subspecies produces a simple, horizontal wing with indistinct venation. A radicular slit is absent. The upper perianth is flat and without extra processes. *Enchylaena tomentosa* has two varieties that differ in the branch pubescence. In *Enc. tomentosa* var. *tomentosa* (included, with pubescent branches and leaves), the depressed-globular, succulent fruiting perianth has a flat or sunken center (upper perianth) (Fig. 1 M) and a wing can be present as an (inconspicuous)

incurved corona extending upwards the succulent perianth. In *Enc. tomentosa* var. *glabra* (not included, with glabrous branches and leaves), the succulent fruiting perianth is flat or slightly sunken in the center. The upper perianth is without a wing-like, incurved corona. In *Enc. lanata*, the globular to cup-shaped and succulent fruiting perianth extends upwards as a prominent short corona (wing). In *M. georgei*, the very thick-walled and fleshy, turbinate perianth tube produces a simple, faintly veined, horizontal wing (Fig. 1 D). A radicular slit is present and the upper perianth is flat. In *M. convexa*, the thick and hard, broadly turbinate perianth tube produces a simple, faintly veined, horizontal wing. A radicular slit is present and the upper perianth is convex. In *M. campanulata*, the cup-shaped, spongy perianth tube produces a simple, horizontal wing. A radicular slit is absent and the upper perianth is flat. In *M. spongiocarpa*, the subglobular, spongy and swollen perianth tube produces a simple, horizontal wing. The wing is shortly decurrent along the radicular slit as a fused vertical membrane on the tube.

*Maireana turbinata* is found in central Queensland, western New South Wales, north-western Victoria to south-eastern Western Australia, usually in reddish soils of flats, loam or clay, in Stony Desert and Karst Plain habitats. *Maireana tomentosa* subsp. *tomentosa* is distributed in Western Australia and the southern portion of the Northern Territory. It occurs on clayey sand, loam, saline plains and salt lakes, usually in Desert Lake habitats. *Maireana tomentosa* subsp. *urceolata* is geographically isolated from the other subspecies. It is distributed in north-western New South Wales, central South Australia and west of the Flinders ranges. E*nchylaena tomentosa* var. *tomentosa* is widely distributed in all mainland states. It occurs in scrub- and woodlands on heavy alluvial soils, saline soils, on rocky and sandy soil along the coast, and disturbed areas in a variety of habitats. *Enchylaena lanata* occurs only in south-western Western Australia, on sandy and clayey soils, and heavy slightly saline soils, usually in Shield Plain, Desert Lake and Mesic Plain habitats. *Maireana georgei* is widespread across the arid zone form Western Australia eastwards to Queensland, New South Wales and north-

western Victoria. It occurs in a variety of soils, such as heavy loamy soil, woodlands, swales and plains, and in various habitats. *Maireana convexa* is found only in Western Australia. It usually grows in non-saline soils, sandy clay or loam, in mulga (*Acacia*) communities, usually in Sand Desert and Shield Plain habitats. *Maireana campanulata* is found in central Australia from the southern halt of the Northern Territory to northern South Australia and south-western Queensland. It usually occurs on rocky slopes and ranges, and in other well-drained situations, usually in Desert Upland and Stony Desert habitats. *Maireana spongiocarpa* is found in central Australia from the southern half of the Northern Territory south and east to western New South Wales, usually on open plains and stony ranges, usually in Desert Upland and Stony Desert habitats.

While there is no exclusive character to delineate this group, the nature of the perianth is worth noting. Six out of eight species develop an either thick-walled and somewhat fleshy (*M. convexa*, *M. georgei*), spongy (*M. campanulata*, *M. spongiocarpa*) or succulent (*Enchylaena* spp.) fruiting perianth. However, this perianth character also occurs in clade 4 (*M. marginata*, *M. trichoptera*, *M. excavata*). It is notable that the species in this group occupy a great variety of soil types and habitats. Three species (*Enc. tomentosa*, *M. georgei*, *M. tomentosa*) have vast ranges that overlap with the ranges of *M. spongiocarpa* and *M. campanulata* in the Eastern Desert, with *Enc. lanata* and *M. convexa* in Western Australia, and with *M. tomentosa* and *M. turbinata* in the southern portion of the continent. The position of *Enchylaena* in this *Maireana* clade is unsurprising, given the morphological similarities (the succulent perianth of *Enchylaena* corresponds to the winged fruiting perianth of *Maireana*) and putative hybrids described (Wilson, 1975, 1984). *Enchylaena tomentosa* and *M. georgei* hybridize in Western Australia and *Enc. tomentosa* and *M. turbinata* hybridize in South Australia and Victoria (Wilson 1984). The hybrid forms produce fleshy fruits with short erect wings. Further, *M. georgei* exhibits considerable morphological variation across its range and

shows morphological similarities to *M. turbinata* and *M. convexa*, from which it differs by the presence of a radicular slit and the flat upper perianth, respectively. And a closer relationship of the morphologically similar species *M. campanulata* and *M. spongiocarpa* to other species with a fleshy, spongy or succulent fruiting perianth is also reasonable.

Although this clade appears plausible based on the predominantly fleshy to succulent perianth tube and the hybrid forms described, the morphological features are homoplastic and not suitable for delineation.

**Clade 7** contains three *Maireana* species (Fig. 3). The perianth tube is turbinate (*M. villosa*, *M. planifolia*) or cup-shaped (*M. humillima*) and produces a simple, horizontal wing. A prominent wing venation can be present. A radicular slit is present in all species. The upper perianth is flat and without extra processes.

In *M. villosa*, the flowers are solitary or paired. The broadly turbinate perianth tube produces a simple, horizontal wing. A prominent wing venation and a radicular slit are present. The upper perianth is flat. In *M. planifolia*, the flowers are solitary or paired, usually only one maturing. The broadly turbinate perianth tube produces a simple, horizontal wing. A prominent wing venation and a radicular slit are present. The upper perianth is flat. In *M. humillima*, the flowers are solitary. The cup-shaped perianth tube has a rounded base and produces a simple, horizontal wing. A prominent wing venation and a radicular slit are present. The upper perianth is flat.

*Maireana villosa* is found in all mainland states, except Victoria, on sandy, clayey or loamy and often stony soils, usually in Desert Upland, Shield Plain and desert Clay Plain habitats. *Maireana planifolia* occurs from Western Australia eastwards to the southern portion of the Northern Territory and South Australia on sandy, clayey or loamy soils, rocky hillslopes, stony plains and in mulga (*Acacia*) scrub, usually in Sand Desert, Desert Upland and Shield

Plain habitats. *Maireana humillima* is found from central New South Wales to southern Victoria, usually in heavy soil, in Riverine Desert and Mesic Plain habitats.

Wilson (1975, 1984) assumed a close relationship of *M. villosa* and *M. planifolia* that are sympatric for a large part of their range, and intermediate plant forms (possible hybrids) have commonly been collected. Hybridization is also suspected between *M. villosa* and *M. integra* (clade 5). Further, he described *M. humillima* and *M. excavata* (clade 4) to be similar in habit but easily distinguishable by the fruiting perianth. There are no characters exclusive to this group. The species have much in common with the *Maireana* clades 4, 5 and 8. Accordingly, this clade does not appear plausible from a non-molecular point of view.

**Clade 8** contains two *Maireana* species (Fig. 3). The perianth tube is turbinate (*M. appressa*) or hemispherical (*M. ovata*) and produces a simple, horizontal wing. A prominent wing venation is present. A radicular slit can be present. The upper perianth is discoid (*M. appressa*) or convex (*M. ovata*) and without extra processes.

In *M. appressa*, the turbinate perianth tube produces a simple, faintly veined, horizontal wing. A radicular slit is present. The upper perianth is discoid with five ridges extending radially between the lobes. In *M. ovata*, the hemispherical perianth tube produces a simple, prominently veined, horizontal wing. A radicular slit is absent. The upper perianth is convex.

*Maireana appressa* is found in semi-arid areas of all mainland states of Australia, often in saline or gypseous soils, subsaline flats and margins of salt lakes, usually in Desert Lake habitats. *Maireana ovata* occurs in north-western New South Wales westwards to South Australia, on sandy rises and eroded slopes of stony hills, usually in Desert Upland and Stony Desert habitats.

Wilson (1975) described morphological similarities between the fruiting perianth of *M. ovata* and *M. tomentosa* (clade 6) subsp. *urceolata* (not included), and similarities between *M.*

*appressa* and *M. radiata* (not included). A close phylogenetic relation of these species cannot be confirmed. The discoid layout of the upper perianth in *M. ovata* also occurs in clade 10. There characters are homoplastic and not useful to delineate this group. Accordingly, this clade does not appear plausible from a non-molecular perspective.

**Clade 9** contains two *Malacocera* species (Fig. 3). The fruiting perianth is depressed (flat) and produces three to five horizontal, tepaline appendages forming a Y- (*Mal. tricornis*) or Y- to star-shaped configuration (*Mal. albolanata*). A prominent venation is absent. A radicular slit is present. The upper perianth is flat and without extra processes.

In *Mal. tricornis*, the flat fruiting perianth produces three (rarely four) tepaline, subcylindrical, horizontal processes, forming a Y-configuration (Fig. 1 P). In *Mal. albolanata*, the flat perianth produces three major and two minor, slender processes with a Y- or star-shaped configuration.

*Malacocera tricornis* occurs in inland southern Western Australia, the southern Northern Territory, central and eastern South Australia, southern Queensland, western New South Wales and north-western Victoria. This species is common on alluvial flood plains, along major drainage systems, and around saline depressions on heavy clay soils (Chinnock, 1980), usually in Riverine Desert habitats. *Malacocera albolanata* occurs in north-eastern South Australia, south-western Queensland and western New South Wales, in depressions, clay pans and saline areas (Wilson, 1984), usually in Riverine Desert and Desert Lake habitats.

*Malacocera* is most common in the arid areas of eastern South Australia and in western New South Wales (Chinnock, 1980). All *Malacocera* species favor heavy, clayey soils and are most frequently encountered along river systems on alluvial flats or around salt lakes or saline depressions in chenopodiaceous shrublands. The morphological characters of the fruiting perianth used to circumscribe the genus *Malacocera* are unique among the Australian Camphorosmeae (Chinnock, 1980; Wilson, 1984). The four *Malacocera* species are

vegetatively very similar in having striate, densely lanate branches with spiral, villous leaves. However, *Mal. gracilis* differs from the other three species in its more slender branches and appressed leaves. Chinnock (1980) assumed *Mal. albolanata* and *Mal. tricornis* to be more closely related than to other *Malacocera* species. The three species analyzed in this study turned out polyphyletic, with *Mal. gracilis* nested within a clade composed of *Maireana* species (clade 10). This supports Chinnock's suggestion that *Mal. gracilis* is not particularly closely related to the other species. Further, Cabrera et al. (2009) found *Mal. tricornis* with *Mal. biflora* (not included) in a well-supported clade. Hence, this clade appears plausible from both a molecular and a non-molecular point of view.

**Clade 10** contains five *Maireana* species and *Malacocera gracilis* (Fig. 3). The perianth tube can be flat (*Mal. gracilis*, *M. dichoptera*), convex (*M. ciliata*, *M. pentagona*, *M. scleroptera*) or turbinate to cup-shaped (*M. eriantha*), and produces a simple (*M. ciliata*, *M. pentagona*, *M. dichoptera*) or five-lobed (*M. eriantha*, *M. scleroptera*), horizontal wing. The fruiting perianth of *Mal. gracilis* produces horizontal appendages in an inverted-Y configuration. A prominent wing venation is absent in all species. A radicular slit is present in *Mal. gracilis*. The upper perianth shows a pentagonal to discoid layout (*M. ciliata*, *M. pentagona*, *M. scleroptera*, *M. dichoptera*) or is prominently lobed and thickened (*M. eriantha*) or flat and forms a plate-like structure (*Mal. gracilis*). The upper perianth of *M. dichoptera* produces vertical wings.

In *M. dichoptera*, the flat perianth tube produces a simple, horizontal wing with 4+2 vertical, intertepaline wings on the upper side of the perianth (the contiguous pair placed either side of the radicular slit) attached to the horizontal wing. The upper perianth forms a convex ring with short inflexed membranous lobes. In *M. pentagona*, the slightly convex fruiting perianth produces a narrow horizontal rim (wing) with a pentagonal layout. The upper perianth is raised into a thick, pentagonal ridge. In *M. ciliata*, the slightly convex wing-like fruiting

perianth is pentagonal in outline and with five radiating intertepaline ribs extending to the angles of the wing. In *M. scleroptera*, the convex perianth tube produces five horizontal wings. The thick and hard upper perianth forms a discoid rim. In *M. eriantha*, the broadly turbinate to cup-shaped perianth tube produces a horizontal, irregularly 5-lobed wing. The upper perianth is deeply lobed. In *Mal. gracilis*, the flat fruiting perianth produces three to five processes, with the three major processes forming an inverted Y-configuration. The two minor ones are irregularly shaped and fused with the major ones to form a plate-like expansion. A radicular slit is present.

*Maireana dichoptera* is found in central and southern Queensland, usually in Riverine Desert and Desert Clay Plain habitats. *Maireana pentagona* is distributed in central and south-western New South Wales, western Victoria and south-eastern South Australia. It usually occurs in heavy soil, and seasonally wet, alluvial clays, in Riverine Desert and Mesic Plain habitats. *Maireana ciliata* is found in central and north-eastern South Australia and far western New South Wales in heavy and calcareous soils, in Stony Desert, Riverine Desert and Karst Plain habitats. *Maireana scleroptera* is found in central Australia, in the southern portion of the Northern Territory, northern South Australia and eastern Western Australia. It occurs on heavy loam, plains and claypans, usually in Riverine Desert habitats. *Maireana eriantha* is distributed in south-western Queensland, north-western New South Wales, eastern South Australia and the southern portion of the Northern Territiry, and usually occurs on stony plains or on rocky hills, usually in Stony Desert habitats. *Malacocera gracilis* is known only from South Australia where it occurs on the coastal strip between Port Augusta and Mt Grainger, Yalata Harbour and on the southern shores of Lake Callabonna. This species is restricted to gypseous mounds and saline clay soils (Chinnock, 1980), and found in Desert Lake habitats.

Chinnock (1980) noted that *Mal. gracilis* does not appear to be closely related to the other *Malacocera* species as the (major) processes are oriented in a reverse direction to form

an “inverted-Y” configuration, and the radicular slit is situated between the two lower major processes, while in the others it is adjacent to the base of the lowermost lobe. Further, occasionally in *Mal. gracilis* two or more processes may fuse to form a plate-like structure. This plate-like structure is somewhat similar in appearance to the disc-like upper perianth of *M. scleroptera*. Wilson (1975) highlighted similar vegetative and fruiting characters in *M. pentagona*, *M. dichoptera* and *M. ciliata*. The vertical wings on top of the perianth in *M. dichoptera* correspond to the horizontal rim and vertical ridges found in *M. pentagona* and *M. ciliata*. Apart from this, Wilson (1975) assumed *M. scleroptera* to be closely related to *M. amoena* (clade 11), and *M. dichoptera* to be closely related to *M. coronata* (clade 2).

The species of this clade have their main distribution in the central and eastern Eastern Desert, grow in similar soil types (heavy and clayey soil) and habitats (Riverine Desert). The most prominent morphological feature of these species is the discoid or plate-like upper perianth with a pentagonal layout (only *M. eriantha* does not show this character). Accordingly, this clade appears plausible from a non-molecular perspective. However, the descriptive perianth features are also present in clades 8 and 9.

**Clade 11** contains the monotypic genus *Didymanthus*, the two species of *Eremophea*, the three included species of *Dissocarpus*, and one species each of *Sclerolaena* and *Maireana* (Fig. 2-4). The morphological characteristics of this clade are very heterogeneous. The perianth tube is mostly cylindrical (*Did. roei*, *Dis. paradoxus*, *Dis. fontinalis*, *Ere. aggregata*, *S. fimbriolata*), or cup-shaped in *Ere. spinosa*, and hemispherical in *M. amoena*. The fruiting perianth produces either spine-like appendages (*Dis. paradoxus*, *Dis. fontinalis*, *Ere. aggregata*, *Ere. spinosa*), 5-lobed, wing-like appendages (*Did. roei*, *M. amoena*), a wing-spine intermediate appendage type (*S. fimbriolata*) or can be with or without spiny appendages (*Dis. biflorus*). All species have paired or aggregated flowers and fruits that usually are fused at the perianth bases (with exception of *M. amoena* and *S. fimbriolata*).

In *Did. roei*, the axillary paired flowers are fused at their bases (Fig. 1 K). The cylindrical fruiting perianth produces five tepaline, horizontal wings arising from the base of the lobes. In *Dis. biflorus* the 2-7 flowers are united into a small woolly ball and the mature fruiting perianths are united in the lower half into a globular, woody infructescence. The perianth produces spines arising from the perianth lobes or is without appendages. The fruiting perianth shows considerable variation in the shape of the fused portion of the infructescence. In *Dis. paradoxus*, the flowers are connate in groups (8-16) in dense woolly glomerules and the mature perianths fuse in the lower two-third into a woody spherical mass producing a woolly ball (Fig. 1 L). The perianth produces five erect spines that arise at the base of each lobe. In *Dis. fontinalis*, the flowers are connate in groups (8-12) in dense woolly clusters and the mature perianths produce a woody ball. Between the aggregated perianths, irregularly-shaped flattened or spine-like appendages emerge from the fused perianth bases. In *M. amoena*, the paired flowers are not fused at the bases. The fruiting perianth produces five, horizontal, tepaline wings (fan-shaped). In *Ere. aggregata*, the flowers are in 3-flowered axillary cymes congested into globular inflorescences (owing to the closely aggregated floral leaves). The woody fruiting perianths are closely aggregated in groups up to 30 and partly sunken into the branch, forming a globular infructescence (Fig. 1 N). The perianth produces five tepaline, erect spines arising from the base of each lobe. In *Ere. spinosa*, the flowers are solitary and slightly sunken into the branch. During maturity, the perianths become embedded in the woody branch axis and cluster irregularly. The tepaline spines arise from below the perianth lobes and usually fuse to produce a stout, adaxial three-pronged spine and a two-pronged, abaxial spine with its fused bases flattened. In *S. fimbriolata*, the paired flowers are not fused at the base. The cylindrical perianth tube produces two series of appendages. The first series consists of five antetepaline, spiny, rough and dentate appendages. The second series consist of five intertepaline, horizontal, fan-shaped and rough appendages with a dentate margin.

*Didymanthus roei* is found in the drier inland south-western portion of Western Australia. It usually occurs on clay and sand, on salt flats and around salt lakes, in Desert Lake habitats. *Dissocarpus biflorus* is widespread in arid and semi-arid regions of the Northern Territory, South Australia, Queensland, New South Wales and Victoria. It occurs on heavy, slightly saline soil in Riverine Desert and Coast habitats. *Dissocarpus paradoxus* is widespread from north-western Victoria, western New South Wales and south-western Queensland, westwards to the Shark Bay region in north-western Western Australia. It often occurs in clayey loam, clayey sand, on floodplains and saline flats in Sand Desert, Riverine Desert and Karst Plain habitats. *Dissocarpus fontinalis* is found in north-eastern South Australia and the neighboring western areas of Queensland and New South Wales. It often occurs near springs or creeks, in sub-saline soils, in Stony Desert and Riverine Desert habitats. *Maireana amoena* is widespread in Western Australia and usually occurs on sandy and gypseous soil around salt lakes in Desert Lake habitats. *Eremophea aggregata* occurs only between Geraldton and Carnarvon in north-western Western Australia on calcareous sandy loam in Karst Plain habitats. *Eremophea spinosa* is distributed in north-western South Australia, the southern Northern Territory westwards to north-western Western Australia. It occurs in mulga scrub communities, and on calcareous hills and in sandy soil on eroded flats, usually in Sand Desert and Desert Upland habitats. *Sclerolaena fimbriolata* is found in the desert of central Western Australia on margins of gypsum salt lakes in Desert Lake habitats.

*Dissocarpus paradoxus* is the most widespread species of this clade. It overlaps widely with the distributions of *Ere. spinosa* (in the western and inland deserts), *Dis. biflorus* (in the eastern desert areas), and *Dis. fontinalis* (in the south-eastern desert areas of South Australia and neighboring areas in Queensland and New South Wales), but the species prefer different soil types and habitats. *Dis. paradoxus* and *Ere. aggregata* prefer different habitats and soil types, and their ranges are spatially separated. *Eremophea aggregata* and *Ere. spinosa* prefer

similar soil types but occur in different habitats. *Didymanthus* and *Dis. paradoxus* prefer similar soil types but different habitats and their distributions overlap only marginally in Western Australia. *Didymanthus*, *M. amoena* and *S. fimbriolata* prefer similar soil types and habitats around salt lakes and their ranges overlap partly in Western Australia. The ranges of *Dis. paradoxus*, *M. amoena* and *S. fimbriolata* overlap widely in Western Australia, and the species prefer in part similar soil types.

Wilson (1984) assumed the genus *Eremophea* to be most closely related to *Sclerolaena* and *Dissocarpus*. *Eremophea* may be distinguished from *Sclerolaena* by the tepaline spines that arise from below the base of the perianth lobes, and from *Dissocarpus* by the infructescences that are borne directly on the branches and embedded in the axis (Wilson 1984). *Eremophea aggregata* resembles *Dis. paradoxus* superficially but differs from that species in having flowers in triads that are subtended by a single leaf. The triads themselves are clustered in leafy infructescences along the branch. In *Dis. paradoxus*, the flowers are arranged in axillary glomerules without leaves. *Dissocarpus fontinalis* resembles *Dis. paradoxus* but differs in not having tepaline spines arising from the base of each lobe, and producing many flattened, spine-like processes that arise from the fused bases of the perianths (Wilson, 1984).

Wilson (1975) assumed intergeneric relationship between *M. amoena*, *M. luehmanii* (not included), *S. symoniana* (not included) and *S. fimbriolata*, and questioned the placement in two different genera. All species have flowers in pairs and exceptional perianth appendages. In *M. amoena* and *M. luehmannii*, the perianth base can be flattened and appressed to the axis (this is reminiscent of the fused perianth bases found in *Didymanthus*, *Dissocarpus* and *Eremophea*). *Maireana amoena* exhibits great variability in the shape of the perianth tube, which can be short with a broad spongy base appressed to the axis, the perianth wings, which can be very short to oblong, coriaceous, irregularly shaped and fimbriate, and the upper perianth, which can bear five prominent ridges (and in this case is reminiscent of the pentagonal

layout present in clade 8 and 10). *Maireana luehmannii* produces five woody, irregularly curved wings that each are divided into two or more spiny lobes. *Sclerolaena symoniana* and *S. fimbriolata* produce two series of rough to woody wing-spine intermediate appendages. The arrangement of the perianth appendages in *S. fimbriolata* is very similar to the perianth morphology of *S. symoniana* and in part to *M. luehmannii*. From *S. symoniana* it differs mainly in having the perianth appendage margins only slightly dissected and not fimbriate. The forms of *M. amoena* producing coriaceous, irregularly shaped and fimbriate wings appear similar to the wing morphology of *S. fimbriolata* and *S. symoniana*.

The genus *Dissocarpus* was distinguished by Mueller (1882a, b) and recognized by Ulbrich (1934) on the basis of the flowers (and fruiting perianths) being united in clusters (Wilson, 1975). This feature was not considered to warrant generic delimitation by Anderson (1923) and Ising (1964), and thus *Dissocarpus* was placed as a section (in synonym) under "*Bassia*" (*Sclerolaena*). According to Wilson (1975), *Dissocarpus* differs markedly from other *Sclerolaena* species in having spines which develop from the base of the perianth lobes (not from an intertepaline position) and grow upwards to overtop and obscure them, and in not having a radicular protuberance or slit. Ising (1964) highlighted the paired flowers of *S. symoniana* (*Bassia symoniana* Ising. sp.nov) and *M. luehmannii* (*Bassia luehmanni* F.Muell.) and made a new combination (*Bassia* sect. *Spinosissinae*) to included species with paired flowers (and fruits). In the phylogenetic study by Cabrera et al. (2009), *Eremophea* formed a monophyletic clade (plus *Neobassia astrocarpa*, not included), and *Dissocarpus* was found in clade with *Didymanthus*.

The species of this clade prefer various yet similar soil types and habitats (sandy, calcareous, or gypseous soils on margins of salt lakes in Riverine Desert and Desert Lake habitats) and the distributions overlap widely in the Western Desert of central Western Australia and the neighboring areas of the Northern Territory and South Australia. The morphological

characters of the fruiting perianth appear quite diverse, but the fused perianth base may be a character that unites this group morphologically, as it has already been under taxonomic discussion by Mueller (1882a, b), Anderson (1924), Ulbrich (1934), Ising (1964) and Wilson (1975, 1984). Accordingly, this clade appears plausible from a non-molecular perspective.

### *4.1.2 Characterization of the Sclerolaena clade (clades 12-17)*

***Neobassia proceriflora*** produces a tubular (or cylindrical) to bluntly ellipsoidal fruiting perianth with an initially succulent outside (and woody within) perianth wall (Fig. 1 S and Fig. 3 and 4). The oblique base is firmly attached to the branch. The apex is produced into a thin, shortly 5-spined cup-shaped structure. The short spines (or teeth) arise opposite the perianth lobes. This species occurs in the southern Northern Territory and northern South Australia to eastern New South Wales and Queensland in heavy, saline soils (Wilson, 1984; Harden 1990;, Cunningham et al., 1981), usually in Riverine Desert and Stony Desert habitats (McDonald, 2020).

The second species of *Neobassia*, *N. astrocarpa* (not included), produces a cylindrical fruiting perianth with five slender, spreading spines at the apex (arising from the base of the lobes). The spines are shortly united at the base by a saucer-shaped rim. The base is truncate to sunken. The perianth is crustaceous to thinly woody and expanded at the apex. This species occurs from the western coast of Western Australia eastwards to the western Northern Territory. It grows in sandy loam, limestone and saline soils, in tidal mud flats, salt marshes, and around salt lakes, (Wilson, 1984; Harden, 1990; Cunningham et al., 1981), and usually occurs in Desert Lake and Coast habitats (McDonald, 2020).

Wilson (1975) highlighted the close morphological similarities of *N. proceriflora* (former *Threlkeldia proceriflora* F. Muell.) and *N. astrocarpa* (former *Bassia astrocarpa* F. Muell) and suggested both species to be segregated as a distinct genus. Scott (1978) combined both species in *Neobassia* using the tepaline spines and vertical seeds for generic delimitation. These characters, however, are not unique and shared with *Eremophea*, *Dissocarpus* and many *Sclerolaena* species.

Wilson (1975, 1984) and Scott (1978) noted that *Neobassia* has strong affinities with *Dissocarpus*. The species have a similar perianth shape (cylindrical to cup-shaped) and spine

arrangement (tepaline spines arising from the perianth lobes) to that found in *Dissocarpus*, and differ principally in having solitary flowers in each leaf axil. Wilson (1984) suggested *N. astrocarpa* to be closely related to *Dis. biflorus*. Interestingly, Wilson (1975, 1984) and Scott (1978) used different descriptions for the perianth structures of this cryptic genus. In 1975, Wilson described the appendages at the cup-shaped perianth apex as five "teeth", arising opposite the perianth lobes. In 1978, Scott described these structures as five short lobes or short, needle-like spines, arising from the middle of each perianth segment. In 1984, Wilson described five spines united into an apical cup, arising from the perianth lobe bases. Further, Wilson (1975) described the perianth wall of *N. proceriflora* as "dry (not succulent)", while in Wilson (1984), the perianth wall is "initially succulent and later spongy". A clear characterization of the nature of the perianth (and its appendages) appears to be necessary.

Anderson (1923) noted that *N. astrocarpa* shows affinities to *S. recurvicuspis* (not included), from which it differs by the softer character of the perianth, and the number and nature of the spines. *Sclerolaena recurvicuspis* has a 3-merous perianth (Wilson, 1984). The cylindrical perianth tube has a truncate base, and produces four (two major and two minor radicular) horizontal to recurved spines, that are united at the base into a cup-shaped structure continuous with the tube. This species is sympatric with *N. astrocarpa*, and is found in north-western Western Australia in slightly saline soils, in claypans and floodplains (Wilson, 1984), usually in Riverine Desert habitats (McDonald, 2020).

In the molecular study by Cabrera et al. (2009), the two *Neobassia* species were found in two different clades. *Neobassia proceriflora* resolved as sister to the *Sclerolaena* clades. *Neobassia astrocarpa* was placed as sister to *Eremophea*. The non-sister relationship of the two *Neobassia* species was not surprising to the authors, giving their taxonomic history.

Given the indistinct morphological characters of the fruiting perianth, similarities to other genera and species, and the taxonomic history, the reconstruction as an individual species

appears reasonable, even from a non-molecular point of view. However, the perianth traits are not exclusive. Further, it is questionable whether the second *Neobassia* species (*N. astrocarpa*), as well as *S. recurvicuspis*, are affiliated with this genus.

**Clade 12** contains the three species of *Osteocarpum* and *Sclerolaena brachyptera* (Fig. 2-4). The perianth tube can be sub-globular (*O. acropterum*, *O. salsuginosum*), cup-shaped *(O. dipterocarpum*) or turbinate (*S. brachyptera*) and produces vertical wings (*O. acropterum*, *O. dipterocarpum*), a 5-angled (or keeled) wing (*S. brachyptera*), or is without prominent appendages (*O. salsuginosum*). All species have a ribbed perianth tube, five stamens and a horizontal seed.

In *O. acropterum*, the apex of the smooth or ribbed perianth tube is usually produced into two opposite erect tubercles which bear one or two vertical wings (var. *acropterum* (F. Muell. & Tate) Volkens, included, Fig. 4 N), or the wings are absent but the tubercle is prominent and radially flattened (var. *dimenutum* (J. Black) Paul G. Wilson, not included). In *O. salsuginosum*, the sub-globular, smooth or ribbed tube is keeled and gibbous on the side with the radicular protuberance that sometimes is produced in a rounded tubercle (Fig. 4 O). In *O. diperocarpum*, the apex of the ribbed perianth tube bears two or three vertical wings (Fig. 1 T). In *S. brachyptera*, the fruiting perianth tube has five prominent tepaline and five less prominent intertepaline ribs. The apex is extended into a hard narrow entire horizontal 5-angeled (keeled) wing.

*Osteocarpum acropterum* var. *acropterum* occurs from north-western Western Australia through the southern Northern Territory and north-western South Australia to south-western Queensland and western New South Wales, and grows in heavy, loamy, saline soils, usually in Riverine Desert, Desert Lake and Stony Desert habitats. The other variant, *O. acropterum* var. *deminutum*, occurs in eastern South Australia, western New South Wales and Victoria, usually in heavy periodically waterlogged soil. *Osteocarpum salsuginosum* occurs

from inland southern Western Australia, eastwards through South Australia and the southern Northern territory to central and western New South Wales and western Victoria, usually in clay soil subject to periodic flooding or on margins of salt lakes, in Riverine Desert and Desert Lake habitats. *Osteocarpum dipterocarpum* occurs from the southern Northern territory to western New South Wales on margins of salt lakes or on clay pans, in Riverine desert and Desert Lake habitats. *Sclerolaena brachyptera* occurs in south-western Queensland, central and western New South Wales, western Victoria, South Australia, the southern Northern Territory and north-central Western Australia usually in heavy somewhat saline soils, in Riverine Desert and Stony Desert habitats.

According to Wilson (1975, 1984), the members of *Osteocarpum* (former *Babbagia* F.Muell.) are principally characterized by the intertepaline tubercles or vertical wings at the perianth apex and are apparently linked by variants intermediate in form. The varieties of *O. acropterum* grade into each other, and *O. acropterum* var. *deminutum* appears to grade into *O. salsuginosum*. In the Northern Territory and northern South Australia, a variant of *O. salsuginosum* with a stipitate perianth and a strongly ribbed tube is intermediate in morphology between *O. salsuginosum* and *S. urceolata* (not included). *Sclerolaena urceolata* produces a prominent erect radicular tubercle and four shorter erect tubercles at the perianth apex. This species is found in the southern Northern Territory on heavy soil subject to flooding. Further, putative hybrids between *Osteocarpum* and *Sclerolaena* have been described as a distinct species (*Osteocarpum scleropterum* (F. Muell.) Volkens) (Wilson, 1984. *Osteocarpum acropterum* is assumed to hybridize with *S. calcarata* (Ising) A. J. Scott (not included) and *S. anisacanthoides* (F. Muell.) Domin (not included) to produce a form with two to five erect wings at the perianth apex, or an erect, narrowly oblong radicular tubercle and four smaller intertepaline tubercles with or without short erect wings (*Osteocarpum x scleropterum*) is distributed in north-western New South Wales and south-western Queensland. *Sclerolaena*

*calcarata* and *S. anisacanthoides* are in part sympatric with *O. acropterum* and *Osteocarpum x scleropterum* and prefer similar soil types (usually heavy soil). The fruiting perianth of both species produces intertepaline spines at the perianth apex of which a pair is borne on an ascending radicular tubercle.

Due to the flattened, hard, winged and pentagonal perianth apex, Wilson (1984) assumed *S. brachyptera* to be most closely related to *S. stelligera* (clade 5) and *S. tetragona* (not included). The first assumption can not be confirmed by this study. The morphological similarity to *S. tetragona*, which produces a broadly turbinate, ribbed perianth tube with a 4-angled apex (or short stout spines), is undoubted. Further, the 5-keeled perianth apex of *S. brachyptera* appears similar to the pentagonal, somewhat keeled perianth tube apex of *Osteocarpum pentapterum* (F. Muell. & Tate) Volkens (not included). This species occurs in the southern Northern Territory, north-eastern South Australia to south-western Queensland on heavy soil subject to flooding.

In his revision of the Camphorosmoideae, Scott (1978) described the monotypic genus *Sclerochlamys* to include the species *Sclerochlamys brachyptera*. It was characterized by the fruiting perianth being ribbed along the tepaline and intertepaline veins, with an entire, thickened wing and five intertepaline ribs extending into the wing with short spines at the apex edges. He admitted the close relationship to *Sclerolaena*, but delimited it from this genus by having the spines connected by a wing. Wilson (1984) accepted Scott's combination and highlighted the assumed close relation to *S. tetragona* (and *S. stelligera*). In 1988, Jacobs subsumed *Sclerochlamys brachyptera* into *Sclerolaena.* Cabrera et al. (2009) found the sampled *Osteocarpum* species to form a monophyletic clade, which in turn was polyphyletic with the *Sclerolaena* clades.

The close relationship of *Osteocarpum* to the *Sclerolaena* clade is obvious. However, given the morphological features of the fruiting perianth (ribbed perianth tube with the apex

bearing prominent tubercles, keels or vertical wings), the intermediate species variants that grade into each other, as well as common ecological preferences (heavy, clayey, saline soils, subject to periodic flooding in Riverine Desert habitats), this clade appears reasonable from a non-molecular point of view, and could thus be delineated from *Sclerolaena*.

**Clade 13** contains nine *Sclerolaena* species (Fig. 2 and 4). The attachment is mostly circular (*S. alata*, *S. parviflora*, *S. convexula*, *S. johnsonii*, *S. napiformis*, *S. costata*, *S. muricata)* and the perianth tube is cylindrical (*S. johnsonii*, *S. deserticola*, *S. muricata)* to turbinate (*S. alata*, *S. parviflora*, *S. convexula*, *S. cornishiana*, *S. napiformis*, *S. costata)*. The perianth produces five to six spines. In general, there are three or four major intertepaline spines and two short spines opposite the radicle (radicular spines). The spines are mostly horizontal, radiating or spreading. The limb is either absent (*S. parviflora*, *S. convexula*, *S. napiformis*) or inconspicuous (*S. alata*, *S. cornishiana*, *S. johnsonii*, *S. deserticola*, *S. costata*, *S. muricata*). The number of stamens varies from three to five. The seed is oblique (*S. muricata*) or horizontal (all other species). *Sclerolaena alata* and *S. parviflora* produce flattened, radiating spines (wing-spine intermediate appendage type).

In *S. alata*, the perianth apex bears 4+2 (four major and two radicular spines) wing-like, hard, horizontally radiating appendages (Fig. 1 H and Fig 4 K). In *S. parviflora*, the perianth apex produces 4+2, horizontally radiating spines. In *S. convexula*, the apex bears 3+2 radiating spines. In *S. cornishiana*, the perianth apex produces 4+2 radiating spines. In *S. johnsonii*, the apex produces 3+2 horizontally spreading spines. In *S. napiformis*, the perianth apex produces 4+2 radiating, slightly ascending spines. In *S. deserticola*, the perianth apex bears 3+2 spreading spines. In *S. costata*, the perianth apex produces 3+2 horizontal spines. In *S. muricata*, the perianth apex bears 3+2 or 3+1 spines (two lateral, one abaxial and a short radicular pair of which sometimes only one spine develops).

*Sclerolaena alata* occurs in central Western Australia on sandy soil and outer margins of salt lakes, in Desert Lake habitats. *Sclerolaena parviflora* is widespread in the drier central and southern portions of Australia and occurs on sandy loam or clay, on plains over limestone, and dunes around salt lakes, in Sand Desert and Karst Plain habitats. *Sclerolaena convexula* is found in central Australia and eastern to southern Queensland and northern New South Wales on sandy or loamy soils, in Sand Desert and Desert Upland habitats. *Sclerolaena cornishiana* is found in central Australia and western Queensland in stony or rocky saline, sandy or clayey soils, in Riverine Desert, Desert Upland and Shield Plain habitats. *Sclerolaena johnsonii* occurs in the southern Northern Territory and adjacent regions of Western Australia, South Australia, Queensland and New South Wales, on sandplains or in sandy swales between dunes, in Sand Desert habitats. *Sclerolaena napiformis* occurs in south-central New South Wales and north-central Victoria on sandy loam or clay, in Mesic Plain habitats. *Sclerolaena deserticola* is found in central Western Australia, southern Northern Territory and north-western South Australia on sandy soil around salt lakes and sand dunes, in Shield Plain and Desert Lake habitats. *Sclerolaena costata* occurs from north-western Western Australia to southern Northern Territory and north-western South Australia in slightly saline soil or rocky hills, in Riverine Desert, Desert Upland and Sand Desert habitats. *Sclerolaena muricata* has three, mostly sympatric varieties that mainly differ in the pubescence of branches and leaves: *S. muricata* var. *muricata* (included in this study) occurs in eastern Australia from central Queensland south to northern Victoria and eastern South Australia. It grows in heavy sandy soil in Riverine desert and Desert Clay Plain habitats and its occurrence is often associated with creeks. *S. muricata* var. *semiglabra* (not included) is found from south-eastern Queensland to northern Victoria and south-eastern South Australia. *S. muricata* var. *villosa* (not included) is found from south-eastern Queensland, south to northern Victoria and south-eastern South Australia.

Some assumptions regarding closer relationships based on the fruiting perianth morphology can be confirmed. With respect to the (generally) horizontally radiating and spreading spines, Anderson (1923) and Wilson (1984) assumed *S. convexula*, *S. parviflora*, and *S. costata* to be closely related. The close affiliation of *S. johnsonii* to *S. muricata* var. *muricata* with respect to the 3+2 horizontally spreading spines was highlighted. Further, Wilson (1984) pointed to the similar appearance of *S. alata* and *S. tridens* (F. Muell.) Domin (not included), but the flowers of *S. tridens* are 3-merous and the spines have a 2+2 arrangement. Further, he suggested *S. napiformis* to be of hybrid origin involving *S. parviflora*.

Given the similar soil and habitat preferences (sandy, somewhat saline soil in predominantly Sand Desert habitats), the nature of the attachment (mostly circular) and of the perianth tube (cylindrical to turbinate), the generally inconspicuously produced limb, the (relatively high) number of spines (five to six) that radiate horizontally, and the mainly horizontally oriented seed, this clade appears reasonable from a non-molecular point of view. However, the morphological features are not exclusive to this clade.

**Clade 14** contains four *Sclerolaena* species and *Threlkeldia inchoata* (Fig. 2 and 4). The attachment is elliptic and the shape of the perianth tube varies greatly. The perianth produces two to three spines, with *S. eurotioides* producing up to six spines, and with *T. inchoata* not producing spines. The limb is prominently developed (*S. limbata*, *S. bicornis*, *S. stylosa*, *T. inchoata*) or absent (*S. eurotioides)* and the number of stamens is five. The seed is horizontal (*S. limbata*, *S. bicornis*, *S. eurotioides*) or erect (*S. stylosa*, *T. inchoata*).

In *S. limbata*, the perianth produces two divergent spines. The limb is erect and cylindrical. In *S. bicornis*, the perianth bears two, rarely three erect to spreading spines. The limb is erect. In *S. stylosa*, the perianth produces 2+1 or 2+2 spines with the radicular spines being shorter than the lateral pair, the limb is erect. In *S. eurotioides*, the perianth produces two to 4+2 slender and weak, spreading spines, of which the radicular pair is shorter than the

principal spines (Fig. 1 G). A limb is absent. In *T. inchoata*, the obovoid, laterally compressed, 5-ribbed fruiting perianth is without spines, a radicular rib is prominently developed. The membranous perianth lobes (limb) are erect.

*Sclerolaena limbata* is found in eastern South Australia and far western New South Wales with isolated records from north-western Western Australia, southern Northern Territory and central Queensland. It occurs on heavy slightly saline soil, in Desert Upland, Stony Desert and Riverine Desert habitats. *Sclerolaena bicornis* has two varieties that mainly differ in the spine length: *S. bicornis* var. *bicornis* Lindley (included) has spines up to 20 mm long, and occurs on heavy soil, clay, floodplains and tidal flats from north-western Western Australia eastwards through southern Northern Territory and South Australia to Queensland and western New South Wales, in Riverine Desert, Desert Clay Plain and Coast habitats. *S. bicornis* var. *horrida* Domin (not included) has spines up to 5 mm length, and is found in northern New South Wales and in central and southern Queensland in heavy soil. *Sclerolaena stylosa* occurs only in the in the Shark Bay area in north-western Western Australia, in sandy, clayey, saline soils, usually in Karst Plain habitats. *Sclerolaena eurotioides* occurs in drier areas of subtropical Western Australia in sandy or clayey soil, along salt lakes, floodplains and hillslopes, usually in Desert Lake habitats. *Threlkeldia inchoata* occurs in the southern Northern Territory and northern South Australia in rocky soil and slightly saline areas near creeks, in Riverine Desert and Stony Desert habitats.

According to Wilson (1984), *Threlkeldia* has close affinities with *Sclerolaena*, from which they were formally excluded due to the absence of spines (Wilson, 1984). He pointed out that the *Threlkeldia* species do not appear to be closely related and indeed, they are related to different *Sclerolaena* clades (*T. diffusa* in clade 15). Further, he highlighted the close affiliation of *S. limbata* to *S. bicornis* var. *horrida* (not included), from which it differs by having a dendritic indumentum and a turbinate perianth tube. He assumed *S. stylosa* to be of

hybrid origin involving *Sclerolaena forrestiana* (F. Muell.) Domin (not included). This species occurs in the inland areas of north-western Western Australia on clayey sand and sandplains. The fruiting perianth is partly embedded in the branch axis, produces 2+2 ascending spines, of which one radicular spine can be reduced to a tubercle, the limb is erect and the seed is horizontal. The embedded fruiting perianth, however, is reminiscent of *Ere. spinosa* (clade 11).

The species grow in a variety of soil and habitat types and the distributions show an odd pattern. The range of *S. bicornis* overlaps with the distributions of *T. inchoata* and *S. limbata* in central and south-eastern Australia, respectively. The distribution of *S. eurotioides* overlaps with *S. stylosa* in north-western Western Australia. In this area, the range of *S. eurotioides* is in contact with *S. bicornis*. This clade lacks exclusive morphological features. The elliptic attachment and prominent limbs are also present in clades 15-17. The various perianth tube characteristics occur in all other *Sclerolaena* clades. The number of spines and the seed orientation are variable. Only the number of stamens is constant (five) but also present in clade 17. Besides the affiliations described by Wilson (1984), which can be confirmed, this clade does not appear plausible from a non-molecular perspective.

**Clade 15** contains seven *Sclerolaena* species and *Threlkeldia diffusa* (Fig. 2 and 4). The attachment is mainly circular, with *S. intricata* and *S. articulata* having an oblique attachment, and the variously shaped perianth tubes are often dorsiventrally compressed. The perianth produces three to five spines, with *T. diffusa* not having spines. The limb is absent (*T. diffusa*), or inconspicuous (*S. glabra*, *T. diffusa*) but predominantly prominently developed (*S. intricata*, *S. tricuspis*, *S. articulata*, *S. ventricosa*, *S. blackiana*). The number of stamens is three (*T. diffusa*, *S. fusiformis*) or five (all other species). The seed is horizontal (*S. tricuspis*, *T. diffusa*), oblique (*S. blackiana*, *S. fusiformis*) or erect (*S. glabra*, *S. intricata*, *S. articulata*, *S. ventricosa*).

In *S. glabra*, the dorsiventrally compressed fruiting perianth produces 2+1 or 2+2 spines (an opposite laterally placed pair is ascending) and the limb is erect with an inflexed tip. In *S.*

*intricata*, the semicylindrical, dorsiventrally compressed perianth produces 3+2 spreading to recurved spines (the inner of the radicular pair is shorter) and the limb is erect. In *S. tricuspis*, the cylindrical tube produces three symmetrically spreading spines and the limb is conical. In *S. articulata*, the semicylindrical, dorsiventrally compressed fruiting perianth produces three spreading to recurved spines and the limb is erect. In *S. ventricosa*, the barrel-shaped (convex) to cylindrical fruiting perianth produces 2+1, 2+2 or rarely 3+2 spines, of which the lateral major spines are ascending to divergent and the shorter radicular spines can be reduced to tubercles, the conical limb is erect. In *S. blackiana*, the barrel-shaped (convex) fruiting perianth produces five recurved spines (the radicular one the shortest), the limb is conical. In *T. diffusa*, the ovoid to urceolate, succulent fruiting perianth does not produce prominent appendages (Fig. 1 W), however, the fleshy tube apex forms three to four intertepaline knobs on drying. In *S. fusiformis*, the cylindrical, dorsiventrally compressed fruiting perianth produces 2+2 equidistantly arranged spines, the limb is very small and slightly sunken.

*Sclerolaena glabra* is found from north-western Western Australia eastwards trough the Northern Territory and northern Southern Australia to western Queensland and north-western New South Wales, usually in rocky or heavy, saline soil, in Riverine Desert and Stony Desert habitats. *Sclerolaena intricata* is found in the southern Northern Territory, south-western Queensland, western New South Wales and north-eastern South Australia, usually in heavy, slightly saline soil, in Riverine Desert and Stony Desert habitats. *Sclerolaena tricuspis* is found in New South Wales, south-eastern Queensland, eastern South Australia and north-western Victoria on heavy or clayey soil, in Riverine Desert habitats. *Sclerolaena articulata* is found in central Western Australia, north-eastern South Australia, western New South Wales and north-western Victoria in heavy, slightly saline soil, in Stony Desert habitats. *Sclerolaena ventricosa* is found in southwestern Queensland and western New South Wales, westwards to central South Australia in heavy or clayey soil, often in saline areas, in Riverine Desert and Stony Desert

habitats. *Sclerolaena blackiana* is found in northern South Australia, south-western Queensland and western New South Wales, in Stony Desert habitats. *Threlkeldia diffusa* is found along the coastal areas from northern to southern Western Australia, eastwards south along the coast to Victoria and Tasmania, and also at scattered inland localities, usually on sandy soils in saline coastal conditions and also near inland salt lakes and saline flats, in Desert Lake and Coast habitats. *Sclerolaena fusiformis* is found in the central-southern area of Western Australia in loamy soil of eucalyptus woodland, usually in Shield Plain habitats.

Wilson (1984) highlighted the close affiliation of *S. articulata* to *S. intricata*, from which it differs only in having three, not five spines. He noted that *S. drummondii* (clade 16) has been confused with *S. fusiformis*, from which it may be distinguished by the radicular spines being placed equidistantly between the larger spines. Further, he assumed hybridization to take place between *S. articulata* and *S. cuneata* (clade 16), and between *T. diffusa* and *S. uniflora* (clade 17). Wilson (1975) noted that the intertepaline knobs (of which a pair is placed opposite the radicle) formed by the drying perianth of *T. diffusa* are comparable to the intertepaline spines and radicular spines found in many *Sclerolaena* species.

Most species in this clade grow in heavy, somewhat saline soil in Stony Desert habitats. Their main distribution is in the central-eastern deserts in inland Australia (*S. articulata*, *S. intricata*, *S. blackiana*, *S. tricuspis*, *S. ventricosa*, *S. glabra*). *Sclerolaena glabra* also occurs in coastal and inland Western Australia and the northern Northern Territory. *Sclerolaena fusiformis* is relatively isolated in Western Australia, in contrast to the other species which occur sympatrically in inland Australia. *Threlkeldia diffusa* dominates the coastal areas and is in part sympatric with *S. fusiformisa* and *S. glabra* in Western Australia, and with *S. ventricosa*, *S. intricata* and *S. tricuspis* in southern South Australia. With respect to the morphological features, the species in this clade have much in common with the species of clade 16 (circular attachment, dorsiventrally-compressed perianth tube, prominently developed limb, five

stamens). However, a few features are relatively dominant such as the mainly cylindrical and dorsiventrally compressed perianth tube that produces three to five spines, the prominently developed limb and the five stamens per flower.

With respect to the perianth characteristics of *T. diffusa*, which are comparable to those of *Sclerolaena* (Wilson, 1984), its phylogenetic placement in a *Sclerolaena* clade is of little surprise. Given the similar ecological preferences and the dominating perianth features of the species in this clade, this clade appears plausible from a non-molecular perspective. The dominant characteristics of the fruiting perianth, however, are also present in clade 16.

**Clade 16** contains seven *Sclerolaena* species. The attachment is circular (*S. cuneata*, *S. lanicuspis*, *S. densiflora*), elliptic (*S. divaricata*) or oblique (*S. minuta*, *S. drummondii*). The mostly oblong to turbinate and dorsiventrally compressed perianth tube produces predominantly three to four spines, with *S. densiflora* having five spines. A prominently developed limb is present in most species, with exception of *S. densiflora* and *S. drummondii* that have an inconspicuous and absent limb, respectively. The number of stamens is three (*S. divaricata*) or five. The seed is erect.

In *S. divaricata*, the oblong dorsiventrally compressed fruiting perianth produces four spreading spines (the radicular spine is much smaller), the limb is incurved. In *S. cuneata*, the oblong dorsiventrally compressed fruiting perianth produces three or four horizontally spreading spines, the limb is incurved. In *S. minuta*, the turbinate dorsiventrally compressed fruiting perianth produces two or four ascending spines (the radicular pair is present as minute tubercles or developed to short spines), the limb is erect. In *S. lanicuspis*, the turbinate, dorsiventrally compressed fruiting perianth produces three to rarely four erect or divaricate spines, the limb is erect. In *S. densiflora*, the turbinate fruiting perianth produces 3+2 ascending spines (the closely collateral radicular pair often appears to be of one spine), the apex is truncate. In *S. drummondii*, the dorsiventrally compressed fruiting perianth produces 2+2 or 2+1

ascending spines (one of the radicular spines can be reduced to a tubercle), the small limb is erect.

*Sclerolaena divaricata* is found in western and central New South Wales and eastern South Australia on heavy clayey soils, in Riverine Desert and Stony Desert habitats. *Sclerolaena cuneata* is widespread in inland Australia on sandy, clayey or stony soils on sub-saline flats, in Riverine Desert, Shield Plain and Stony Desert habitats. *Sclerolaena minuta* is found in north-eastern Western Australia, the southern Northern Territory and western Queensland on eroded granite and rocky hillsides, in Desert Upland, Shield Plain and Stony Desert habitats. *Sclerolaena lanicuspis* is found in central and northern Western Australia, the Northern territory, South Australia, central and western Queensland, western New South Wales and north-western Victoria on stony loam and sandy clay, on alluvial and saline flats, in Desert Upland, Stony Desert, Shield Plain, Riverine Desert and Desert Clay Plain habitats. *Sclerolaena densiflora* is found in central and northern desert areas of Western Australia on sandy clay or loam or stony soils on rocky hills and floodplains, in Desert Upland and Shield Plain habitats. *Sclerolaena drummondii* is found in southern Western Australia in red sandy soil and clayey gravely loam, in Shield Plain and Mesic Plain habitats.

The morphological features of this clade have much in common with clades 14 and 15. However, the results for this clade support assumptions of closer relationship. Wilson (1984) described *S. drummondii* and *S. fusiformis* as closely allied (of which the latter differs by having glabrous leaves and by the equidistantly arranged spines). Further, he noted that *S. cuneata* appears to be closely related to *Sclerolaena blakei* (Ising) A. J. Scott (not included) from which it differs in the shape of the perianth tube and the orientation of spines, and he assumed hybridization to occur between *S. divaricata* and *O. acropterum* (clade 12). Wilson (1984) and Ising (1964) noted that *S. lanicuspis* may encompass a number of distinct taxa and that some variants of *S. lanicuspis* resemble *S. densiflora*, which can be distinguished by the square apex

which produces five spines. The ranges of these species overlap in Western Australia. Ising (1964) pointed to the morphological similarities of *S. minuta* and *S. lanicuspis* that can be distinguished by the number of spines and the shape of the limb.

The assumptions of Wilson (1984) and Ising (1964) regarding closer relationship of some species in this clade can be confirmed. Given the similar soil preferences and dominant perianth characteristics, this clade appears plausible from a non-molecular perspective. However, the dominant traits of this clade are also present in clades 14 and 15.

**Clade 17** contains ten *Sclerolaena* species (Fig. 2 and 4). The attachment is circular (*S. uniflora*, *S. burbidgeae*, *S. decurrens*), elliptic (*S. holtiana*, *S. diacantha*, *S. parallelicuspis*, *S. obliquicuspis*, *S. brevifolia*) or oblique (*S. eriacantha*, *S. patenticuspis*). The fruiting perianth produces two spines, with *S. decurrens* having two or three spines. The limb is prominently (*S. eriacantha*, *S. patenticuspis*, *S. decurrens*) or inconspicuously developed (all other species). The number of stamens is five. The seed is horizontal (*S. diacantha*, *S. uniflora*), erect (*S. patenticuspis*, *S. decurrens*, *S. obliquicuspis*, *S. brevifolia*) or oblique (*S. holtiana*, *S. burbidgeae*, *S. parallelicuspis*, *S. eriacantha*).

In *S. holtiana*, the sub-globular fruiting perianth produces two lateral, divergent spines, the limb is incurved and inconspicuous. In *S. diacantha*, the depressed-spherical to pear-shaped fruiting perianth produces two lateral, divergent spines, and the incurved limb is inconspicuous. In *S. uniflora*, the depressed-globose to pear-shaped fruiting perianth produces two lateral, erect spines (one or both can be absent), and the incurved limb is inconspicuous. In *S. burbidgeae*, the shortly cylindrical to depressed-globose fruiting perianth produces two divergent spines (one or both can be reduced and knob-like), and the incurved limb is inconspicuous (Fig. 1 F). In *S. parallelicuspis*, the oblong, dorsiventrally compressed perianth tube produces two lateral, parallel and erect spines, and the short, erect limb is obscured by an indumentum. In *S. eriacantha*, the dorsiventrally compressed perianth tube produces two, obliquely erect spines,

and the inflexed limb is erect. In *S. patenticuspis*, the shortly oblong perianth tube produces two lateral ascending spines (that can be very reduced or absent), and the limb is erect. In *S. decurrens*, the shortly turbinate perianth tube produces two or three spines (of which two are opposite and divergent), and the oblong limb is erect. . In *S. obliquicuspis*, the broadly oblong perianth tube produces two obliquely spreading spines, and the inconspicuous limb is erect. In *S. brevifolia*, the dorsiventrally compressed fruiting perianth produces two obliquely spreading spines, and the incurved limb is inconspicuous.

*Sclerolaena holtiana* is found in the southern Northern Territory and northern South Australia, in Riverine Desert, Desert Lake, Sand Desert and Stony Desert habitats. *Sclerolaena diacantha* is widespread in subtropical and temperate Australia from Western Australia, eastwards to Queensland, New South Wales and Victoria and occurs on heavier, sometimes subsaline, sandy loams, sand plains and often on calcareous, in Riverine Desert, Sand Desert, Karst Plain, Mesic Plain and Coast habitats. The typical variant of *Sclerolaena uniflora* is coastal and occurs on limestone in South Australia and Western Australia (Wilson, 1984; Paczkowska and Chapman, 2000), in Karst Plain and Coast habitats. Other variants occur on loamy, subsaline soils (rich in limestone or gypsum rubble) in inland South Australia and New South Wales (Walsh, 1996). *Sclerolaena burbidgeae* is found in southern central Western Australia on clay, sandy loam and floodplains, in Riverine Desert habitats. *Sclerolaena parallelicuspis* is found from the southern Northern territory, through north-eastern South Australia to north-western New South Wales and south-western Queensland on clay in stony plains or hillsides, in Stony Desert habitats. *Sclerolaena eriacantha* is found in all mainland states except Victoria and occurs in heavy soil, in sand overlying clay (Wilson, 1984), on rocky hills and plains (Paczkowska and Chapman 1995), in Desert Upland and Stony Desert habitats. *Sclerolaena patenticuspis* is found in central New South Wales, north-western Victoria, extending westwards to central and south-eastern Western Australia on heavy soils, sandy clay

or loam, sub-saline flats or clay depresions and limestone plains, in Desert Upland, Riverine Desert and Karst Plain habitats. *Sclerolaena decurrens* is found from southern Northern Territory to south-western Queensland, western New South Wales and north-western Victoria usually in heavy slightly saline soil, gypseous clayey sand, in Riverine Desert and Stony Desert habitats. *Sclerolaena obliquicuspis* is found in south-central and eastern Western Australia, south-western Northern territory, South Australia, western New South Wales and north-western Victoria on sandy clay or loam, limestone and calcareous soils, in Karst Plain habitats. *Sclerolaena brevifolia* is found from south-eastern Western Australia, eastwards to the northern Eyre peninsular in South Australia, in sandy clay over limestone and slightly saline or calcareous soil, in Karst Plain habitats.

Ising (1964) and Wilson (1984) recognized close morphological affiliations between the species of this clade and named this species group the "*S. uniflora* complex" (or "*S. diacantha* – *S. uniflora* complex"). The fruiting perianths usually produce two spines and an inconspicuous to prominently developed limb. This complex includes several species with variants and hybrids grading into each other (*S. uniflora*, *S. diacantha*, *S. holtiana*, *S. burbidgeae*, *S. parallelicuspis*, *S. constricta*, *S. obliquicuspis*, *S. patenticuspis*, *S. brevifolia*). Wilson noted morphological similarities between *S. uniflora* and *S. diacantha* (which may hybridize), and highlighted close affiliations of *S. holtiana* to the *S. diacantha* – *S. uniflora* complex, from which it is not clearly distinguishable. *Sclerolaena burbidgeae* may be readily distinguished from other species in the *S. uniflora* complex by the glabrous, dark red fruits (Wilson, 1984). *Sclerolaena diacantha* includes several variants (e.g. *S. diacantha* var. *longispina* Benth.) that are sometimes recognized as a distinct species (*S. obliquicuspis*, *S. patenticuspis*), however, these intergrade morphologically into each other and their circumscription is arbitrary. In addition, some variants of *S. obliquicuspis* intergrade into *S. brevifolia* and *S. patenticuspis*. *Sclerolaena patenticuspis* is a polymorphic species with several

variants that are variable in the development of the spines (Ising, 1964), and some variants intergrade into *S. obliquicuspis* (Wilson, 1984). *Sclerolaena decurrens* has not yet been suggested to be a member of this complex, likely due to the occasionally three spines produced by the fruiting perianth. Wilson suggested *S. decurrens* to be closely related to *Sclerolaena bicuspis* (F. Muell.) Domin (not included) from which it mainly differs by the woolly branches and the orientation of the spines.

Given the common preference for sandy to clayey, calcareous or gypseous soil, the occurrence in Karst Plain (and Riverine Desert) habitats, the fruiting perianth that generally produces two spines, and the close morphological affiliations described by Ising (1964) and Wilson (1984), which are confirmed by the phylogenetic results, this clade appears plausible also from a non-molecular point of view.

### *4.1.3 Summary: Morphological and ecological characterization of phylogenetic clades*

As assumed by Wilson (1975, 1984) and for the first time proven with molecular data by Cabrera et al. (2009), the characters used to delimit the genera of the Australian Camphorosmeae do not always create a natural grouping of species and the current taxonomy is to some extent artificial. The different genera show strong vegetative and reproductive affinities with one another (e.g. *Threlkeldia* + *Sclerolaena* + *Neobassia* + *Dissocarpus*, *M. amoena* + *M. luehmannii* + *S. symoniana* + *S. fimbriolata*, *M. oppositifolia* + *Roycea*, *Enchylaena* + *Maireana*), so that a distinct circumscription of genera remains problematic. As found by Cabrera et al. (2009), the traits of the fruiting perianth appendages emerged independently in different clades and are extreme homoplastic. This homoplasy is well illustrated by hybridization events. Species of different perianth appendage characteristics hybridize to produce modified, intermediate perianth forms (e.g. *Enc. tomentosa* x *M. georgei*, *Enc. tomentosa* x *M. turbinata*, *O. salsuginosum* x *S. urceolata*, *O. acropterum* x *S. calcarata*, *O. acropterum* x *S. anisacanthoides*, *M. lanosa* x *M. lobiflora*, *M. integra* x *M. villosa*) (Wilson, 1975, 1984). In addition, there is ongoing hybridization within morphologically well-characterized clades, resulting in intergrading forms that are considered valid species (e.g. clade 17, "*S. uniflora* complex"). Further, *S. napiformis* (clade 13) and *S. stylosa* (clade 14) are putative hybrids resulting from crosses involving *S. parviflora* and *S. forrestiana*, respectively. However, these hypothesized hybridization events between species and clades has yet to be tested on a molecular scale in future studies. The convergent evolution of homoplastic perianth features complicates the search for useful morphological characters for generic delimitation, and resulted in numerous nomenclatural and taxonomic revisions (e.g. Mueller, 1882a, b; Ulbrich, 1934; Anderson, 1923; Ising, 1964; Wilson, 1975, 1984; Scott, 1978).

While the analysis of in total 25 vegetative, reproductive, and ecological characters in this study did not result in a "universal recipe" for taxonomic revision of the Australian Camphorosmeae, some clades can be described using distinctive characters or a combination of characters. For still other clades, a combination with clades of similar characteristics appears to be reasonable.

The strong vegetative affiliations of *M. oppositifolia* to the genus *Roycea* have already been described by Wilson (1975, 1984), who assumed intergeneric relationship. The two variants are similar to *R. divaricata* and *R. spinescens*. All four species are sympatric in south-western Western Australia and have similar ecological preferences (sandy to clayey, saline soil in Desert Lake habitats). When including all Camphorosmeae samples in the phylogeny, *M. oppositifolia* is sister to the *Roycea* species (Fig. 2). Hence, despite the homoplastic perianth traits of *M. oppositifolia*, a combination appears reasonable.

Clades 2 and 3 are sister to another in the phylogeny including all samples (Fig. 2). The species are well described by a combination of characters. The most prominent feature are the accessory processes produced by the upper perianth. A radicular slit is present in most species and the horizontal wing is without prominent venation in all species. Most species are relatively small shrubs or perennials with a woody rootstock (decumbent to 50 cm high) that predominantly occur in sandy, saline soil in Riverine Desert and Desert Lake habitats. Disregarding the intricate description of the perianth features of *Eriochiton* and considering them instead as a 5-lobed horizontal wing with erect processes, the grouping within this clade seems to make sense. Taking the 5-lobed horizontal wing and the prominently lobed upper perianth into account, which are comparatively frequently present in clade 2, *M. enchylaenoides* could also be added. Accordingly, a combination appears reasonable.

The *Maireana* affiliated clades of the second polytomy present a major taxonomic challenge, since most traits are shared by almost all clades. Only a few characteristics appear to be comparatively frequently present in some clades.

Within clade 4, there are two morphological sub-groups. Four species (*M. schistocarpa*, *M. polypterygia*, *M. pentatropis*, *M. erioclada*) produce vertical wings along the perianth tube in addition to the horizontal wing at the perianth apex. Three species (*M. marginata*, *M. trichoptera*, *M. excavata*) have a perianth tube dissected in the middle. Despite that, this clade has much in common with clades 5-8.

Although *S. stelligera* has already been found within a *Maireana* clade in the molecular phylogeny of Cabrera et al. (2009), the position of this species in a *Maireana* clade (clade 5) is to some extent of surprise given the taxonomic history of this species (Scott, 1978; Wilson, 1984; Jacobs, 1988). However, considering the "short radiating spines (five tepaline, five intertepaline, one spine opposite the radicle)" simply as a narrow horizontal wing with a crenate-dentate margin (with acuminate teeth), the phylogenetic outcome is only of little surprise. Accordingly, a new combination appears plausible.

The species of clade 6 may be well described by the fleshy, spongy or succulent fruiting perianth developed by the majority of species. This characteristic is not exclusive but shared with species of clade 4 (*M. marginata*, *M. trichoptera*, *M. excavata*). However, the perianth tube of the species in clade 4 is dissected in the middle, but not in species in clade 6. Several species of this clade have vast distributions and occupy a great variety of soil types and habitats. The placement of *Enchylaena* in this *Maireana* clade is not surprising, given the close morphological affiliation (succulent perianth of *Enchylaena* corresponds to the wing of *Maireana*) and the hybrids described between *Enc. tomentosa* with *M. georgei* and with *M. turbinata* (Wilson, 1975, 1984). A combination appears to be necessary, but is difficult due to the homoplastic features, with the exception of the nature of the perianth tube.

Clade 9 includes two (out of four) *Malacocera* species, *Mal. albolanata* and *Mal. tricornis*. The third species included in this study, *Mal. gracilis*, was placed in *Maireana* clade 10. The fourth species, *Mal. biflora* was found in a clade with *Mal. tricornis* by Cabrera et al. (2009). This genus is characterized by the peculiar shape of the perianth, which produces three to five horizontal appendages forming a Y-configuration. It is not surprising that the three species turned out in different clades, as already Chinnok (1980) suggested that *Mal. gracilis* is not particularly closely related to the other species. In contrast to the distinctive Y-configuration of the other three *Malacocera* species, the three (major) processes are oriented in a reverse direction to form an "inverted-Y" configuration, while the two minor processes are fused with the major ones to form a plate-like expansion, in *Mal. gracilis*. The placement of *Mal. gracilis* in clade 10 appears reasonable, given the discoid to plate-like upper perianth with a pentagonal layout that circumscribes the species of this clade. Hence, a combination using the discoid to plate-like nature of the upper perianth as descriptive character, and as delimiting character to the genus *Malacocera*, appears reasonable.

Clade 11 includes species of five genera: *Didymanthus*, *Dissocarpus*, *Eremophea*, *Maireana* and *Sclerolaena*. The species have exceptional perianth appendages that appear quite diverse. The perianths produce spine-like, wing-like or wing-spine-intermediate appendage types and are greatly variable. However, the fused perianth base that results in paired and aggregated infructescences may be a character that circumscribes this group morphologically. Mueller (1882a, b), Anderson (1924), Ulbrich (1934), Ising (1964) and Wilson (1975) have discussed this feature repeatedly as a descriptive character. Despite the morphological similarities to *Maireana* and *Sclerolaena*, a new combination appears to be necessary.

Within the species-rich *Sclerolaena* clade, the phylogenetic clades overall appear reasonable from a non-molecular perspective. The clades composed of (predominantly) *Sclerolaena* species (clades 13-17) can be well circumscribed using one or several distinctive

characters (with exception of clade 14) and do not necessarily require taxonomic treatment. Only the smaller genera *Neobassia*, *Osteocarpum* and *Threlkeldia* require taxonomic treatment.

The genus *Neobassia* will remain cryptic until further research. Given the indistinct perianth morphology, similarities to other genera and species, and the taxonomic history, the reconstruction of *N. proceriflora* as an individual species appears reasonable. It is questionable whether the second *Neobassia* species (*N. astrocarpa*) is affiliated with this species on a molecular scale.

*Osteocarpum* was placed with *S. brachyptera* in a well-supported clade. Using the nature of the fruiting perianth (ribbed perianth tube with the apex bearing prominent tubercles, keels or vertical wings) as descriptive characters, the four species can be well circumscribed and delineated from *Sclerolaena*, to which a close relationship is obvious.

The two *Threlkeldia* species were assumed to be not closely related by Wilson (1984), and indeed, they are related to different *Sclerolaena* clades. *Threlkeldia inchoata* and *T. diffusa* were placed in clade 14 and 15, respectively. Given the close affinities with *Sclerolaena* (intertepaline knobs at the apex of the dry perianth in *Threlkeldia* correspond to the spines in *Sclerolaena*) (Wilson, 1984), a subsumption in *Sclerolaena* appears reasonable.

In congruence with the findings by Cabrera et al. (2009), the results of this study show that several genera do not merit generic status and may be included as species of other genera. This finding is in line with assumptions regarding an artificial taxonomy of this group by Wilson (1975). In particular, the species of the *Maireana* grade need further research and re-evaluation of generic and species concepts. The morphological characters analyzed in this study are in part suitable for circumscription of the phylogenetic clades.

### *4.2 Diversification and perianth trait evolution of the Australian Camphorosmeae*

The Australian Camphorosmeae split from their Eurasian sister group in the Middle Miocene ca. 12.99 [9.96-15.97] Mya and arrived in Australia sometime between the split and the onset of diversification in the Late Miocene (Fig. 5-7; Table 5). Diversification happened rapidly, as evident in the chronogram of the collapsed phylogeny (Fig. 6). Two rapid radiation events overlap in time. The first occurred ca. 8.34 [6.54-10.38] Mya and includes clades 1-3, *M. oppositifolia*, *M. enchylaenoides* and *M. sedifolia*. The second rapid radiation occurred ca. 7.46 [5.85-9.23] Mya ago and includes all other clades of the *Maireana* grade and the *Sclerolaena* clade. Species-richness differs between the *Maireana* clades (with winged fruiting perianths) and the *Sclerolaena* clade (with spiny fruiting perianths). The *Maireana* species of the first and second polytomy and the species of the *Sclerolaena* clade show differing habitat preferences. Numerous phylogenetic and biogeographic studies have shown that the evolution of arid-adapted taxa in Australia has paralleled the aridification of the continent since the Miocene (reviewed in Crisp et al., 2004; Byrne et al., 2008; Crisp and Cook, 2013; Byrne and Murphy, 2020). Further, diversification is probably promoted by biome shifts (e.g. Toon et al., 2015; Cardillo et al., 2017; Renner et al., 2020), which in turn depends on the mobility of plants by means of seed dispersal (Howe and Smallwood, 1982; Garant et al., 2007, Räsänen and Hendry, 2008; Vander Wall and Beck, 2012). Two overarching questions are discussed in this section: 1) How does the phylogenetic pattern fit the development of arid habitats in Australia? 2) What effect on diversification may the different perianth types and associated dispersal syndromes have had?

First, climatic and floristic features of Australia (4.2.1) and the aridification of Australia since the Miocene (4.2.2) are outlined. Next, arrival, migration and diversification (4.2.3) and the perianth types and dispersal syndromes of Australian Camphorosmeae (4.2.4) are discussed with respect to the two overarching questions.

### *4.2.1 Climatic and floristic features of modern Australia*

Modern Australia is the driest inhabited continent and has a wide range of climatic and floristic zones due to its large geographical size and north-south extension from the tropics to the mid-latitudes (Mabbutt, 1988; Crist ep al., 1999, 2004; Suppiah et al., 2001; Martin, 2006; Crisp and Cook, 2013; Bird et al., 2016). The continent is dominated by relatively flat landscapes lacking major geological barriers, with exception of the Great Dividing Range in the south-east (Mabbutt, 1988; Johnson, 2009; Pillans, 2018). In addition to the geographical features, the climate is strongly influenced by the surrounding oceans and weather systems, such as the subtropical high-pressure belt or sub-tropical ridge (Suppiah et al., 2001; Timbal and Drosdowsky, 2013; Song et al., 2017). The northern part of the continent is dominated by summer rain events while the southern part generally has rainfall in the winter. Temperature and precipitation vary considerably across the continent and from season to season. Precipitation is the overriding environmental factor for most of the continent (Keast 1959; Martin, 2006).

The tropical north experiences a well-marked dry and wet season (Martin, 2006; Bowman et al., 2010). Intense seasonal rainfalls and weather extremes during the summer wet season, such as tropical cyclones, are frequent. Annual precipitation commonly exceeds 2000 mm (Pillans, 2018). The northern tropical areas are dominated by eucalyptus woodlands and tall shrublands (Crisp et al., 1999; Martin, 2006). Rainforests are found in the north-eastern and eastern tropics and in more montane regions along the east coast (Specht and Specht, 1999; Martin, 2006).

The south-west and the south-east of the continent have a temperate climate (Crisp et al., 1999; Martin, 2006). The Mediterranean experiences a dry summer and marked wet winter with an annual rainfall of 400-1500 mm (Pillans, 2018). The south-east has wet winters and low summer rainfall or uniform rainfall across the year with an annual precipitation of 400-1500 mm

(Pillans, 2018). Temperate eucalypt woodland and forest are dominant (Crisp et al., 1999; Martin, 2006).

The largest part of Australia is arid or semi-arid with an annual rainfall less than 250 mm and 250-350 mm, respectively (Crisp et al., 1999; Martin, 2006; Crisp and Cook, 2013; Pillans, 2018). This so-called "arid-zone" covers 70% of the continent. Irregular major and seasonal droughts, heat waves, unpredictable rainfall events and only moderate rainfall seasonality are characteristic of this area. In addition, fire is an important component of arid areas (Crisp et al., 2011), although fire events tend to be more common in semi-arid peripheral areas due to the fragmentary and slow-growing vegetation of the arid zone (van Etten and Burrows, 2018). The arid zone is a relatively flat landscape without major geological barriers (Mabbutt, 1988; Pillans, 2018). Arid landscapes (Table 3, Appendix 3) are distributed across the arid zone in a mosaic pattern (Mabbutt, 1988; Fujioka and Chappell, 2010). The most common land type of the arid zone is Sand Desert that covers ca. 38 % of the area, followed by Upland Desert (16 %), Stony Desert (15 %), Shield Plain (14 %) and Riverine Desert (9 %). Less common are Clay Plain (4 %), Karst Plain (3 %) and Desert Lakes (1 %). The soils are highly weathered and unusually nutrient-poor (Smith and Morton, 1990).

An important feature of the arid zone are the numerous inland lakes fed by vast inland drainage systems that are dry for very long periods and fill only occasionally after exceptional rainfall (ephemeral saline lakes; Mabbutt, 1988; Martin, 2006; Bowen and Benison, 2009). The largest ephemeral lake systems are located in the Artesian Basin (Lake Eyre, Lake Torrens, Lake Frome) and in the Gibson Desert (Lake Mackay). Many more salt lake and drainage systems are scattered across central and western Australia. The geomorphological characteristics of the soils, such as surface form, substrates, water storage capacity and water permeability, have a direct influence on the vegetation found (Mabbutt, 1988).

The arid-zone provides habitat for 10% of Australian plant species and is dominated by arid plant formations, such as (open) *Acacia* and Casuarinaceae woodland, chenopod shrubland and hummock and tussock grasslands (Crisp et al., 1999; Groves, 1999; Martin, 2006). Open woodlands are defined by well-spaced trees with a rich shrub groundcover (including Camphorosmeae species). *Acacia* woodlands dominate drier parts of the western, central and eastern part of the arid zone. *Eucalyptus* woodlands are less dominant in the central arid zone but become more prominent towards the periphery in semi-arid and temperate areas. Arid grasslands are dominant in the northern arid zone. Chenopod shrublands are composed of drought- and salt-tolerant species and dominate vast areas in the central and southern arid zone (Leigh, 1994; Martin, 2006).

The majority of Camphorosmeae species are adapted to arid and semi-arid environments (Wilson, 1984). However, many species are also distributed in the temperate periphery and some also colonize tropical and coastal habitats. Species richness of the Australian Camphorosmeae is highest in the Yilgarn Plateau in Western Australia and in the Central Lowlands of the Lake Eyre-Murray Basin in south-eastern Australia (McDonald, 2020). The Yilgarn Plateau is characterized by sandy soils in the uplands and sand plains, and is dominated by extensive river desert and desert lake habitats with open *Eucalyptus* and *Acacia* woodlands (Martin, 2006; McDonald, 2022). The Lake Eyre-Murray Basin is dominated by sandy and stony deserts, Riverine Deserts and Desert Lake habitats with open *Acacia* woodlands, arid grassland and chenopod shrubland vegetation. *Maireana* species dominate the western center of diversity and *Sclerolaena* is most species rich in the eastern diversity center (McDonald, 2022). The two centers of diversity are connected by the Central Australian Ranges and the Sandland South province.

### *4.2.2 Aridification of Australia since the Miocene*

The arrival of Australian Camphorosmeae dates to the Middle Miocene with subsequent initial diversification in the Late Miocene and major crown group diversification during the Pliocene (Fig. 6 and 7). Diversification of multiple arid-adapted taxa, including the Camphorosmeae, was coincident with and likely promoted by progressive aridification of the continent (Byrne et al., 2008, 2018; Byrne and Murphy, 2020). The following section outlines the geologic, climatic, and biotic changes in Australia since the Miocene that led to the aridification of the continent.

Aridification of the continent started with the breakup of the South American plate from the Antarctic plate in the late Cenozoic, which initiated the Antarctic Circular Current (ACC) (McGowran et al., 2004; Martin, 2006; Byrne et al., 2008; Fujioka and Chappell, 2010). As a result, ocean temperatures dropped and the warm currents that brought moist air to Australia ceased. The warm and humid climate of Australia, which had persisted until the early Cenozoic and favored a vegetation of warm to cool temperate rainforests, ended. The Australian continent drifted steadily northward into warmer latitudes, thus widening the rift to the Antarctic plate, and steadily intensifying the ACC. The lower ocean temperatures resulted in cooler and drier west winds. The stage was set for Australian aridification.

The Early to Middle Miocene was a warm and humid period with high sea levels (Fig. 7) and flooding of shallow basins, such as the Murray basin in the south-east and the Eucla basin in the south (Martin, 2006). Seasonality with a well-marked dry season became apparent in south-eastern and western Australia. Central Australia was home to extensive, shallow alkaline lakes. The continent was occupied by diverse forest types. Rainforest covered the south-east Murray Basin (Martin, 1993), while the eastern Highlands were occupied by *Eucalyptus* forests (Martin, 2006). Swamp (Cyperaceae), ephemeral swamp (*Wilsonia*) and sclerophyll (*Acacia*, *Eucalyptus*, *Casuarina*) communities were present around extensive lake systems in central

Australia (Martin, 2000, 2006). Western Australia was covered by *Eucalyptus* and *Casuarinaceae* forests. Rainforest was restricted to small patches of wetter habitats (Benbow et al., 1995). The cooler and drier west winds resulted in progressively declining rainfall and the evaporation rate increased (Martin, 2006). The western part of the continent was affected by this development earlier than the interior. The drainage systems in central Australia persisted into the Middle Miocene (Quilty, 1994). By the Middle Miocene the regular flows in the palaeodrainage systems of Western Australia ceased (van der Graaff et al., 1977; Fig. 7). Once perennial river systems reaching the coast progressively became disconnected ephemeral river systems feeding floodplains. These systems became early precursor landscapes of Riverine Desert and Desert Lake habitats (Mabbutt, 1988; McDonald, 2020; Fig. 7).

The Middle to Late Miocene was marked by global cooling, falling sea levels, and a strong expansion of the Antarctic Ice Sheet (McGowran et al., 2004; Bowler et al., 2006; Fig. 7). The Late Miocene was a time of major climatic, landscape, and biotic transition in Australia, known as the "Hill Gap" (Hill, 1994; Byrne et al., 2008). This period (ca. 10-6 Mya, Fig. 7) is characterized by the end of warm and humid Early Miocene environments, unprecedented levels of erosion in a landscape that had been stabilized by an effective vegetation cover, the expansion of dry, open woodlands and shrublands, and the demise of once extensive lakes in central Australia (Martin, 2006; Bowler, 2006; Byrne et al., 2008). While Australia drifted northwards, the widening rift between Australia and Antarctica intensified the Antarctic Circumpolar Current (ACC) (Bowler, 1982; McGowran et al., 2004; Fujioka and Chappell, 2010; Fig. 7). The larger land mass and cooler, drier west winds decreased continental precipitation and resulted in a major transformation of the dominant vegetation types (Fig. 7). Rainforest communities contracted and sclerophyllous vegetation expanded (Martin, 2006). The north-east and south-east of the continent were dominated by rainforest and sclerophyll forest. Open woodland with isolated rainforest patches dominated southern Australia. In

Western Australia, sclerophyll forest expanded. Central Australia became increasingly desiccated and was largely dominated by open woodlands and chenopod shrubland. The Sub-Tropical High Pressure system (STHP) moved northwards and resulted in seasonal aridity (Bowler, 1982; Fujioka and Chappell, 2010; Fig. 7). Seasonality with a strong dry period allowed regular fire g to become an ecological feature of sclerophyll forest and woodlands. At the end of the Late Miocene, the Nullarbor region was uplifted after the final retreat of the sea (Webb and James, 2006; Fig. 7). The Nullarbor Limestone was exposed to weathering and the calcareous soils were deposited across the inland plains. During this period, Karst Plain landscapes spread across the southern continent, and Riverine Lake and Desert Lake landscapes spread into central Australia (Mabbutt, 1988; Fujioka and Chappell, 2010) and numerous arid-adapted taxa started to diversify (Byrne et al., 2008; Byrne and Murphy, 2020).

The Pliocene was a period of intensifying aridity throughout the continent (Martin, 2006; Byrne et al., 2008). The early Pliocene was marked by a short period of wet and warm conditions in which rainforest communities temporarily expanded in suitable river slopes (Martin, 2006; Fig. 7). However, during the remainder of the Pliocene, the generally cooler and drier conditions became more intense, leading to a contraction of rainforest elements and an expansion of sclerophyllous forests, woodlands, arid shrublands, and grasslands (Fig. 7). The continent was characterized by a precipitation gradient from a drier west to a wetter east (Kershaw et al., 1994; Fig. 7). Australian landscapes and vegetation types increasingly diversified towards more arid landscapes and vegetation communities. Arid-adapted taxa increasingly diversified (Byrne et al., 2008; Byrne and Murphy, 2020). Desert chenopod shrubland gradually replaced sclerophyllous forests (Martin and McMinn, 1994; Hekel, 1972; Fig. 7). The expansion of arid shrubland and grassland concurs with an increase of grazing animals, such as kangaroos and wombats (Archer et al., 1994; Couzens and Prideaux, 2018; Fig. 7) and diversification of arid-adapted ants and termites (Ward et al., 2010; Heimburger et

al., 2022). Central Australia (Lake Frome to Lake Eyre) was home to freshwater and saline swamps surrounded by sclerophyll forest and arid shrubland in the dryland areas. In western Australia, heath and sclerophyllous woodland expanded across the drying palaeodrainage system (Bint, 1981; Dodson and Ramrath, 2001). Most Australian arid landscapes (Desert Upland, Shield Plain Desert, Karst Plain, Riverine Desert) were present since the Late Miocene (Fig. 6 and 7, Table 3) and progressively expanded during the Pliocene (Mabbutt, 1988; Fujioka and Chappell, 2010; McDonald, 2020). The steadily west-to-east expanding ephemeral desert river systems (Riverine Desert habitats) of Western Australia favored the emergence of Desert Lake habitats (Bowen and Benison, 2009; Fujioka and Chappell, 2010; Fig. 7). In central Australia, the first stony deserts appeared by 4–2 Myr (Fujioka et al., 2005; Fujioka and Chappell, 2010). By the end of the Pliocene, Australia approached its current latitude and the STHP became positioned at the center of the continent and caused an arid climate across a broad area (Fujioka and Chappell, 2010; Fig. 7). The modern climate regime had established, but it was still more humid than today (Martin, 1986, 2006).

The general trend of a cooler and drier climate of the Pliocene continued in the Pleistocene. The beginning of the Pleistocene represents the onset of major oscillations between glacial and interglacial climates, which developed in response to the growth of the Northern Hemisphere ice caps (Williams et al., 1998; Martin, 2006; Byrne et al., 2008). Along the expanding and contracting ice caps, sea level and temperature fluctuated (Fig. 7). During interglacial periods, sea level and temperature were higher, leading to a land mass decrease, increase in humidity and precipitation, and higher water tables of inland basins. As a result, mega lakes fed by ground water discharge or monsoonal precipitation formed and Eucalyptus forests expanded. During glacial periods, sea level and temperature were lower, leading to an increased land mass, decreasing precipitation and lower water tables of inland basins. Thus, an expansion of drier climates and arid landscapes. The arid zone expanded during glacial periods. Extreme weather

events caused by the STHP, such as long-lasting droughts and strong, extensive cyclones, became common (Martin, 2006; Fujioka and Chappell, 2010). Throughout the continent, the tree cover decreased and grasses and desert shrubs increased (Martin and McMinn, 1994). *Eucalyptus* woodlands rich in grasses transformed to dry open woodlands rich in chenopod shrubland (Fig. 7). The onset of extreme aridity in the Pleistocene promoted the expansion of extreme arid habitats. Stony Desert landscapes expanded in the eastern arid-zone (Fujioka et al., 2005; Fujioka and Chappell, 2010). Sandy Deserts (Great Sandy Desert, Gibson Desert, Great Victoria Desert, Simpson Desert) with extensive sand plains and dunefields, formed during the late Pleistocene (Fujioka and Chappell, 2010; Fig. 7). Due to the very low annual precipitation and the highly weathered soils of low water storage capacity, Stony and Sandy Desert habitats are the most challenging landscapes within the arid-zone (Mabbutt, 1988; Fujioka and Chappell, 2010).

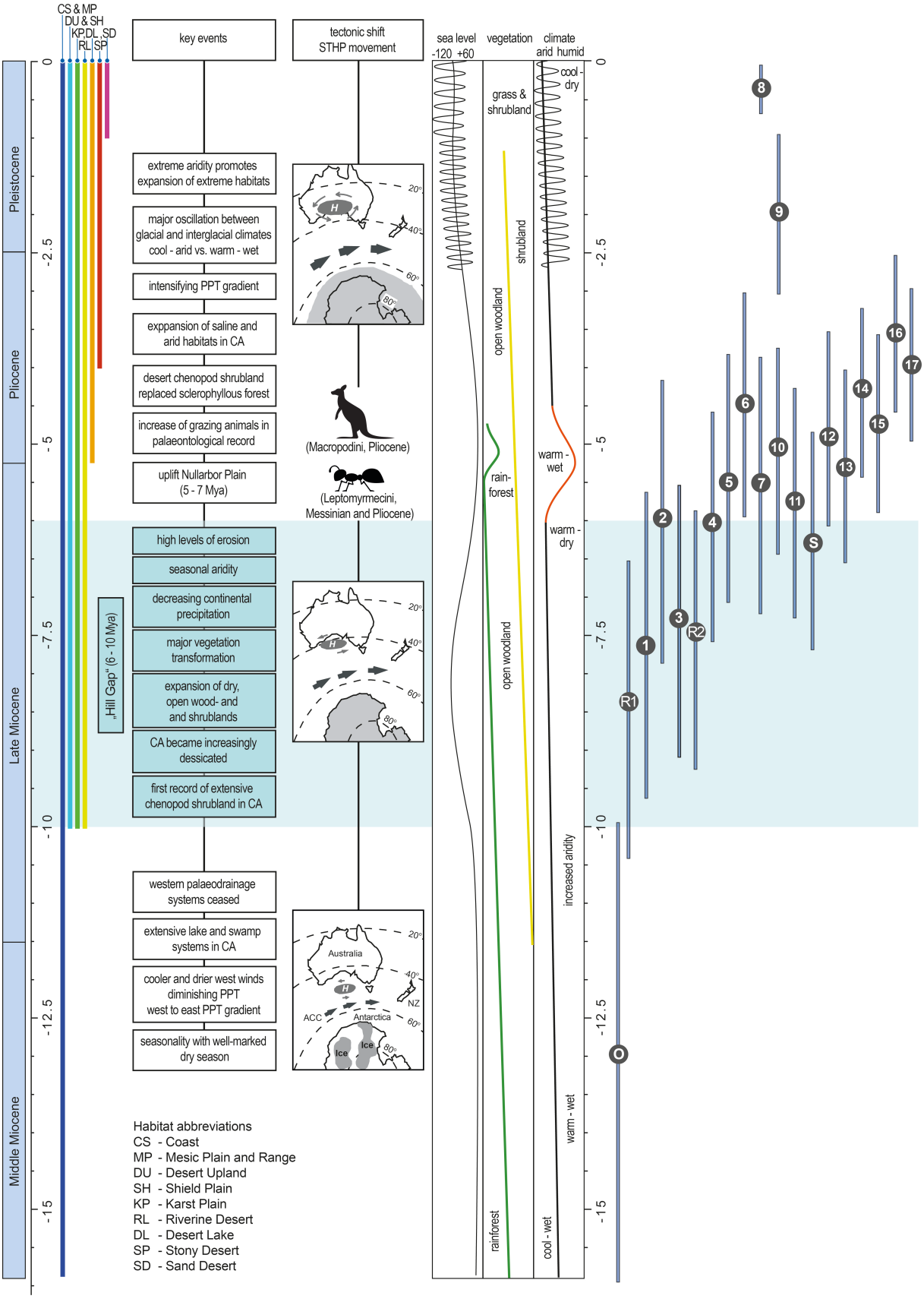



**Figure 7**. Overview of climatic, tectonic, landscape, and biotic changes in Australia since the Middle Miocene with respect to the diversification of the Australian Camphorosmeae. Given from left to the right: Timeline 16 Million years ago to present. Timeline of landscape (habitat) history based on McDonald (2020). Key climatic, geological and biotic events of Australia's aridification. Tectonic shift of the Australian plate and movement of the

Sub-Tropical High Pressure system, according to Bowler (1982) and Fujioka and Chappell (2010). Changes in sea level based on Martin (2006) and Byrne et al. (2008). Changes in vegetation composition based on Martin (2006) and Fujioka and Chappell (2010). Climatic changes based on Byrne et al. (2008) and Fujioka and Chappell (2010). Diversification events of the Australian Camphorosmeae taken from Table 5 and Figure 6. The encircled clade numbers are shaded in grey: O: outgroup split, R1: rapid radiation of the first polytomy (clades 1-3), R2: rapid radiation of the second polytomy (clades 4-11), S: diversification of the *Sclerolaena* clade (clades 12-17).

Aridification of Australia became evident at the end of the Middle Miocene: Cooler and drier west winds resulted in a diminishing west to east precipitation (PPT) gradient and the western palaeodrainage systems ceased. The Late Miocene was characterized by major environmental changes leading to the development of more arid conditions and an early expansion of the arid zone. This period is known as the "Hill Gap". The Pliocene was a period of intensifying aridity throughout the continent in which characteristic arid-zone habitats greatly expanded. The beginning of the Pleistocene represents the onset of major oscillations between glacial and interglacial climates. Along with these major climatic oscillations, extreme aridity promoted the expansion of extreme arid habitats such as Stony and Sand Desert habitats. Initial diversification coincides with an early expansion of arid climates and landscapes during the "Hill Gap". Grown group diversification increases with increasing aridification of the continents interior at the end of the Late Miocene and in particular during the Pliocene. Diversification of the *Sclerolaena* clade coincides with the occurrence and diversification of large herbivores, such as kangaroos (Macropodini) and wombats (Archer et al., 1994; Couzens and Prideaux, 2018), and with rapid radiations of arid-adapted ants (Leptomyrmecini) and termites (Ward et al., 2010; Heimburger et al., 2022) during the early Pliocene. Image sources are given in section 9.

### *4.2.3 Arrival, migration and diversification of Australian Camphorosmeae*

#### *4.2.3.1 Arrival and first inland migration*

The arrival of Australian Camphorosmeae dates to the Middle and early Late Miocene (12.99 [9.96-15.97] Mya] with subsequent initial diversification (8.34 [6.54-10.38] Mya) in the Late Miocene (Fig. 5-7, Table 5). As suggested by Cabrera et al. (2011) and Kadereit and Freitag (2011), the ancestors of the Australian Camphorosmeae must have arrived via long-distance dispersal, as the Australian continent was not connected to the original Gondwana landmasses since the beginning of the Cenozoic. The group probably originated from temperate semi-arid to arid parts of continental Eurasia, the area where the closest relatives are distributed today. The Eurasian Camphorosmeae taxa are predominantly annuals, less often subshrubs (Zhu et al., 2003; Kadereit and Freitag, 2011). The Central Asian genus *Grubovia* Freitag & G.Kadereit contains annual xerophytes that grow in open steppe communities with summer rain, mainly in sandy, slightly saline soils. The 5-merous perianths produce horizontal, wing-like or spine-like appendages in fruit. The species of the Eurasian genus *Bassia* All. grow in dry steppe and semi-desert communities, mainly in sandy, non-saline soils. The 5-merous perianth produces wing-like or spine-like appendages or is without appendages. The 5-merous perianth of the Central Mediterranean genus *Eokochia* Freitag & G.Kadereit produces a 5-lobed wing in fruit.

The biogeographic analysis by Cabrera et al. (2011) showed that the most likely area of arrival and ancestral diversification is located in the south and west of the continent. This estimation coincides with the presence of chenopod pollen fossils in west and south Australia since the Late Miocene (McDonald, 2020). The group then expanded its range easterly across the interior of the continent and colonized the Eastern Desert including the Lake Eyre-Murray basin during the middle Late Miocene, which again is supported by the pollen fossil record in central Australia and the Murray Basin (McDonald, 2020). Subsequently, the peripheries of

today's arid zone and finally today's tropical areas were colonized. McDonald (2020) investigated the migration routes along arid land types of Australian chenopods and suggested the Sandland province, with its extensive network of disconnected drainage systems and river-desert and desert-lake landscapes, as a possible migration corridor through the interior of the continent. Further, Camphorosmeae migration was likely multidirectional, with steady movement of species between newly colonized areas (Cabrera et al., 2011; McDonald, 2020). Cabrera et al. (2011) suggested that the progenitors of Australian Camphorosmeae were already adapted to arid and saline environments and thus were able to quickly colonize newly emerging arid landscapes. Successful colonization of newly emerging arid habitats may have been favored by niche pre-emption.

At the time of arrival, suitable landscapes for the colonization of the continent gradually formed as the climate became increasingly cooler and drier (Mabbutt, 1988; Martin, 2006; Byrne et al., 2008; Fujioka and Chappell, 2010; Fig. 6 and 7). Ecological adaptation to coastal, marshy conditions may have enabled the progenitors to colonize coastal and marshy habitats of flooded basins, such as the Eucla Basin at the south coast (Burbidge, 1960; Barlow, 1994; McDonald, 2020). A first inland migration may have taken place by species adapted to coastal, marshy conditions, providing a littoral connection along declining palaeodrainage systems to hypersaline inland habitats (Shmida 1985; Saintilan, 2009; Cabrera et al., 2011; Lomolino et al. 2010; McDonald, 2020). The ecologically diverse coastal habitats have inland analogues, such as inland playas of saline lake systems and inland marsh habitats surrounding lake systems and swamps. Adaption to coastal habitats is rare in Australian Camphorosmeae but present across the phylogeny (Fig. 6, clade 1, 4, 6, 10, 11, 15, 17). Various Camphorosmeae species (e.g. *Dis. biflorus*, *Enc. tomentosa*, *M. brevifolia*, *M. oppositifoila*, *N. astrocarpa*) are found in both saltmarsh habitats and in hypersaline Desert Lake habitats (Saintilan, 2009). The low number of species adapted to coastal conditions may be explained by extinction or progressive

replacement of coastal species by species adapting to arid conditions (Shmida, 1984). Moreover, other Amaranthaceae lineages, such as *Atriplex* L. and Salicornioideae Ulbr., are major competitors in coastal habitats (Kadereit et al., 2006; Zerdoner Calasan et al., 2022) and may have displaced Camphorosmeae species from these habitats. Hence, the habitation hypothesis by means of a littoral connection appears plausible.

#### *4.2.3.2 Inland migration and initial diversification (1st radiation)*

A first rapid radiation occurred in the Late Miocene, ca 8.34 [6.54-10.38] Mya (Fig. 6 and 7, Table 5). Crown group diversification of clades 1-3 started in the Late Miocene, ca. 7.6-7.68 [5.94-9.67] Mya (Fig. 5-7, Table 5, clades 1-3). The species of the first rapid radiation are mainly distributed in western and south Australia and have clear preference for sandy, saline soils and Desert Lake habitats (73 % of all species in the first polytomy), followed by Karst Plain and Riverine Desert habitats (both 27%) (Fig. 5 and 6, Table 4). The rapid radiation and beginning crown group diversification of the first polytomy likely have been driven by the progressively expansion of drier climates and drier and increasingly saline landscapes during this period (Fig. 7). This supports the assumption proposed by Byrne et al. (2008) that the "Hill Gap" was an early period of diversification of arid zone taxa. The multidirectional expansion into the arid zone (from south-west easterly across the interior and then south-east and towards the arid-zone periphery) (Cabrera et al., 2011; McDonald, 2020) may have taken place by species adapted to various habitats (shifting habitat preferences, Fig. 6; Table 4), and relatively quick, as there are no major geological barriers. The species of the first polytomy (including clades 1-3, *M. oppositifolia*, *M. sedifolia* and *M. enchylaenoides*) have soil and habitat preferences that support three west to east migration scenarios.

The first migration scenario may have occurred from southwest to northwest and then to the easterly interior and was probably favored by adaptation to ephemeral river and lake habitats. Clades 1 and 3 have their distribution center in Western Australia (with *M. carnosa*

also occurring in the eastern arid-zone) and are predominantly adapted to Desert Lake habitats (Fig. 6). The species of clade 2 are predominantly adapted to Riverine Lake habitats (Fig. 6) and their distributions range from Western Australia to the Eastern Desert and the Lake Eyre-Murray Center. Riverine Lake habitats were present in Western Australia since the Miocene and progressively expanded with ongoing cessation of the palaeodrainage system. Progenitors of the first migration scenario may have migrated northwards along increasingly saline desert river systems and diversified into vacant habitats that progressively turned into Desert Lake habitats. Easterly inland migration was likely promoted by the west-to-east gradient of declining precipitation and increasing aridity that resulted in an easterly expansion of Riverine Desert habitats throughout the continent's interior (Kershaw et al., 1994; Martin, 2006). This easterly expansion started in the upper Late Miocene and increased during the early Pliocene. As suggested by McDonald (2020), the Sandland provinces may have been the main migration corridor from the western to the eastern deserts. Today, the Riverine Desert and Desert Lake habitats of the Sandland provinces are disconnected by extensive sand plains and dune field formations of the Great Sandy, Gibson and Victoria Desert that formed in central Australia during the Pleistocene (Fujioka and Chappell, 2010). Assuming vicariance, this dissected landscape is also reflected in the disjunct distributions of *M. carnosa*, *M. lobiflora* and *M. lanosa*.

For the second and third scenario, the origin of the Nullarbor Plain may be of importance. Prior to the uplift of the Nullarbor Plain in the upper Late Miocene (Fig. 7), the coastline was along the southern portion of the Sandland South province (Webb and James, 2006; McDonald, 2020). With the retreating sea, the coastline slowly moved to its today's southern margin and exposed the Nullarbor Limestone to weathering. The surface slopes gently towards the sea and is traversed by old river courses. The plain terminates abruptly in a cliff line that extends ca. 900 km, with coastal plains in the center and in the west.

The second migration scenario along southern coastlines may have taken place by species adapted to coastal and marshy conditions. *Maireana oppositifolia* is found in Coastal and Desert Lake habitats (Fig. 6) and is distributed along the south coast from Western Australia to South Australia, spreads into the southern Yilgarn Plateau in Western Australia and across Eyre Peninsular northwards up to Lake Torrens in South Australia. Noticeably, this species occurs only along the coast at the Nullarbor, avoids the limestone plain, but can be found at the north-eastern and north-western margins, which reflect the coastline prior to the uplift of the Nullarbor Plain at the end of the Late Miocene (Webb and James, 2006). The habitat preferences of this species to both saline coastal habitats and saline desert lake habitats, represents an ecological niche in which progenitors could migrate along the coastline and progressively colonize inland habitats.

The third migration scenario may have taken place by species adapting to calcareous Karst Plain habitats. *Maireana sedifolia* and *Eri. sclerolaenoides* prefer calcareous soils of Karst Plain habitats. Both species have a southern distribution from the Western Desert areas across the Nullarbor Plain. The emergence of *M. sedifolia* (ca. 7-8 Mya) coincides roughly with the uplift of the Nullarbor Plain at the end of the Late Miocene (Fig. 6 and 7). *Eriochiton sclerolaenoides* emerged much later during the Pleistocene. It appears plausible that the uplift of the Nullarbor Plain and spread of limestone across the interior favored ecological adaptation to these calcareous soils and provided a southern migration path across this newly elevated plain.

#### *4.2.3.3 Radiation into diverse habitats (2nd radiation)*

A second rapid radiation event occurred ca. 7.46 [5.85-9.23] Mya (Fig. 6 and 7, Table 5). Crown group diversification of the *Maireana* affiliated clades (clades 4-11) started in the Late Miocene and early Pliocene (Fig. 5-7, Table 5). Clade diversification continued throughout the Pliocene. Clades 8 and 9 diversified during the Pleistocene. The species are widespread across the continent. Compared to the species of the first polytomy, the preferred habitat types shift in favor of Riverine Desert (38 %), Desert Lake (33 %), and Karst Plain habitats (27 %), and other arid habitats, such as Sand Desert, Desert Upland and Stony Desert landscapes, are becoming increasingly colonized (Table 4). In addition, the occupied soils become less saline (Fig. 5). The high degree of diversification during this period (Fig. 6 and 7, Table 5) was probably driven by the intensifying aridification and consequent major expansion of interior arid habitats, in particular in central and eastern areas of the arid-zone that became increasingly saline and arid (Fig. 7).

The relatively species-poor clades of the *Maireana* grade (compared to the species-rich *Sclerolaena* clade) and longer branches of these clades indicate either lower speciation or higher extinction rates (Crisp et al., 2004). There are several plausible scenarios for this pattern. First, the short return to a warm and humid climate at the beginning of the Pliocene (Fig. 7) with accompanying expansion of rainforest communities in suitable river slopes may have contributed to increased extinction rates in *Maireana* lineages that were adapted to Desert Lake and Riverine Desert habitats. Second, colonized habitats may have been quickly saturated due to rapid radiations of other arid-adapted taxa (Fig. 7) and ecological adaptation and expansion to new, vacant habitats was slow. Third, a rapid multidirectional colonization and constant migration between habitats may have resulted in increased gene flow, which in turn decreased divergence among populations and thus reduced speciation of these lineages (Hamrick and Godt, 1996, 1997; Räsänen and Hendry, 2008; Kisel and Barraclough, 2010). Fourth, during

the extremely arid periods of glacial maxima in the Pleistocene, the central Australian desert was much larger in extent (desert dunes covered up to 85% of the continent during peak aridity) and much more hostile to plant growth (Jones and Bowler, 1980; Crisp et al., 2001; Bowler et al., 2006; Byrne et al., 2008; Andersen, 2016). This may have led to the contraction of species distributions and extinction events of narrowly endemic species. Fifth, the emergence of large herbivores in the early Pliocene may have favored extinction events (Fig. 7). However, these herbivores prefer grasses as a primary food source (Couzens and Prideaux, 2018) and feed on the sodium rich *Maireana* species (Revell et al., 2013) mainly during drought periods or in the presence of fresh water (Wilson, 1984). Whether this herbivory pressure led to increased extinction rates is doubtful, but grazing animals may have had an effect on perianth trait evolution and thus on dispersal syndromes and dispersal speed (see 4.2.4).

#### *4.2.3.4 Diversification into extreme habitats (Sclerolaena clade)*

The diversification of the *Sclerolaena* clade (including *N. proceriflora* and clades 12-17) started in the Late Miocene, ca. 6.27 Mya [4.88-7.68] with a continuous speciation trend throughout the Pliocene and some speciation during the Pleistocene (Fig. 5-7, Table 5). Diversification coincides with an increase in chenopod pollen fossils in central Australia and the Murray Basin at the end of the Late Miocene and beginning Pliocene (McDonald, 2020). The species of the *Sclerolaena* clade are also widespread across the continent but have their center of diversity in the Lake Eyre-Murray basin of the Eastern Desert. The species of the *Sclerolaena* clade have a clear preference for Riverine Desert (50 %) and Stony Desert habitats (41 %) (Table 4). As is the case with the *Maireana* clades of the second polytomy, other arid habitats are more frequently colonized. Diversification was likely favored by the increasing aridity across Australia's interior and expansion of Stony Desert habitats (Fig. 7). The species-richness and continuous cladogenesis of this clade may be explained by the appearance of spiny perianths (see 4.2.4).

The colonization of two newly emerging arid habitats stands out. The species of clades 13 and 15 are predominantly found in Sandy Desert habitats and Stony Desert habitats, respectively (Fig. 6 and 7). The crown group diversification of clade 15 fits the formation of stony pavements in the eastern arid zone, across the Great Artesian Basin and the Lake Eyre Basin in the Pliocene (Fujioka and Chappell, 2010). At species level (*S. glabra*, *S. intricata*, *S. tricuspis, S. articulata*), it appears that the Riverine Desert landscapes may again have taken an important role as migration corridor. Four species (*S. glabra*, *S. intricata*, *S. tricuspis, S. articulata)* are sympatric across a large area in the Eastern Desert. *Sclerolaena glabra* has a disjunct distribution with occurrences in Riverine Desert habitats in north-western Australia and a main distribution in Riverine Desert and Stony Desert habitats in the eastern desert. The two range parts are separated by Sand Desert formations of the Great Sandy Desert, Gibson Desert and Great Victoria Desert. *Sclerolaena intricata* is distributed across the Riverine Desert and Stony Desert habitats of the Great Artesian Basin and the Lake Eyre Basin. *Sclerolaena tricuspis* is sympatric with *S. intricata* but found only in Riverine Desert habitats, avoiding the Stony Desert gibber plains of the Eastern Desert. *Sclerolaena articulata* is distributed only across the Stony Desert landscapes of the Eastern Desert, avoiding Riverine Desert habitats. The prevailing Stony Desert habitat preference suggests a directed adaptation to these habitats with emergence during the middle Pliocene. The persistence of Riverine Desert habitat preferences indicates a possible migration corridor between western and eastern ranges prior to or along the formation of Stony Desert and Sand Desert landscapes. However, the migration may have also been bidirectional in time and space. For instance, from easterly Stony Deserts to westerly Riverine Desert habitats, after the emergence of pebbly pavements and prior to the formation of Sand Deserts.

The species of clade 13 are predominantly found in Sand Desert habitats that emerged during the late Pleistocene, ca. 1 Mya (Fig. 6) (Fujioka and Chappell, 2010). The crown age of

this clade (ca. 5.32 Mya, Table 5) considerably predates the formation of sand plain and dunefield landscapes. This drastic difference between the diversification of this clade and the emergence of the preferred habitat might be explained by two scenarios. 1) The discrepancy may be owed to dating methods that yield high uncertainties for desert dune formation prior to the late Pleistocene (Fujioka and Chappell, 2010). 2) Taken together, the species of this clade have a large west-east inland range and colonize not only sandy desert landscapes but also arid habitats less characterized by sand dune formations and sandy plains, such as Riverine Desert habitats. These landscapes emerged earlier than Sand Deserts, during the Late Miocene and Pliocene. Thus, these species could originally have been adapted to Riverine Desert habitats and are only found in sand-dominated landscapes as a consequence of the formation of sandy habitats in north to central Australia during the last 1 Million years (e.g., Great Sandy Desert, Gibson Desert, Simpson Desert). Assuming vicariance, this second scenario is further supported by the disjunct distribution ranges of *S. costata* and *S. deserticola*. The lineage of these two species emerged ca. 5 Mya and diversified ca. 1.2 Mya (Fig. 6, clade 13), slightly prior to the formation of sand deserts. As indicated by the preferred habitats (Sand Desert, Desert Upland, Riverine Desert, Desert Lake and Shield Plain) and their chronological emergence, the distribution ranges may have ranged continuously from the western to the eastern extant along ephemeral river and lake systems in the middle Pleistocene and were only separated by sand deserts formations during the late Pleistocene.

#### *4.2.3.5 Pleistocene speciation slowdown*

The apparent slowdown in speciation during the Pleistocene (Fig. 5 and 6) has also been observed in other arid-adapted taxa (Crisp et al., 2004; Byrne et al., 2008, 2018; Crisp and Cook, 2013; Byrne and Murphy, 2020). In this regard, several scenarios are being discussed: 1) Repeated cycles of localized habitat contraction and expansion, along with climate oscillation and extreme drought, may have resulted in "species diversity maintenance" (Chesson, 2000; Byrne et al., 2008). The extreme climatic fluctuations did not lead to a higher speciation rate, but to a finer scale phylogeographic structuring and speciation by hybridization, parthenogenesis, and polyploidy (Byrne et al., 2008). Alternatively, the extreme climatic fluctuations may have led to stabilizing species diversity (Chesson, 2000): If density of species gets low due to harming environmental conditions during extreme drought periods, the species tend to recover during periods that are more favorable. However, species of similar ecology in the same spatial region mutually limit resource availability, hence, cause interspecific competition. Extreme climatic fluctuations and strongly limited resources impair speciation and result in maintained or stable species richness over long timescales. 2) The expansion of the central Australian desert during the extremely arid periods of glacial maxima in the Pleistocene may have led to the contraction of species distributions and extinction of narrowly endemic species (Jones and Bowler, 1980; Crisp et al., 2001; Bowler et al., 2006; Byrne et al., 2008; Andersen, 2016). 3) Newly emerging habitats were increasingly and progressively colonized by numerous arid-adapted taxa during the upper Late Miocene and Pliocene (Crisp et al., 2004; Shepherd et al., 2004; Kadereit et al., 2005; Byrne et al., 2008; Cabrera et al., 2011; Crisp and Cook, 2013; McDonald, 2020; Zerdoner Calasan et al., 2022). Not seldom, this was followed by rapid radiations. Thus, there was probably high interspecific competition between lineages and consequently exclusion of a habitat and less speciation in the inferior lineages.

In contrast to the extreme climatic oscillations that are usually described for this period, a study by Herrando-Moraira et al. (2022) found that the climate in Australia during the Pleistocene was relatively stable with low climatic fluctuations. Thus, the stable climate may have contributed to higher species richness of arid-adapted taxa (García-Palacios et al., 2018; Herrando-Moraira et al., 2022).

Following the hypothesis by Byrne et al. (2008), Cabrera et al. (2011) assumed that evolving lineages may have been reduced in their distribution to relictual areas as the arid zones contracted. In addition, frequent shifts in distribution may have resulted in admixture between species (or clades) that were previously separate, complicating the recent phylogenetic history of the Australian Camphorosmeae.

In this study, the diversification trend among clades varies (Fig. 5 and 6). Some clades show no or only little speciation during the Pleistocene (Fig. 5 and 6, clades 3, 5, 11, 13 and 14), while other clades continuously diversified (Fig. 5 and 6, clades 2, 4, 6, 15, 16 and 17). These varying trends may have been caused by a mixture of the hypotheses given above. After successful establishment in arid habitats during the Pliocene, some clades (with no or little Pleistocene diversification) may have experienced a kind of species diversity maintenance and diversified on a finer phylogenetic scale, or high interspecific competition resulted in seemingly stable species richness (Chesson, 2000; Byrne et al., 2008). Speciation mechanisms were not investigated in this study and are to be tested in future studies. In particular, the effects of hybridization on species richness and morphological variation are worthy of future investigation. Some clades may have been affected by the radiation of other lineages in specific habitats (e.g., Desert Lake habitats colonized by Camphorosmeae and Salicornieae). The strong expansion of extreme arid habitats during glacial maxima may have resulted in increased extinction rates of lineages that may have been endemic to once larger ranges. Other clades may have benefited from the formation and expansion of extremely arid habitats (Sandy and Stony

Deserts) and diversified. Without further molecular analysis, however, explanations for the apparent speciation slowdown during the Pleistocene remain highly speculative.

#### *4.2.3.6 Radiation of other Australian arid-adapted taxa*

The rapid radiations and subsequent diversification of crown groups in the Australian Camphorosmeae reflects the continuous aridification of the Australian continent since the Miocene. The intensifying aridification during the Pliocene likely favored crown group diversification by migration into and adaptation to newly emerging habitats. This diversification trend is in line with other xeric and halophytic taxa that occupied the arid zone with the beginning of Australian aridification (Crisp et al., 2004; Byrne et al., 2008, 2018; Crisp and Cook, 2013; Byrne and Murphy, 2020). In addition, a growing number of studies has shown that lineage diversification is probably promoted by frequent biome shifts (e.g. Toon et al., 2015; Cardillo et al., 2017; Renner et al., 2020).

Molecular phylogenetic studies by Shepherd et al. (2004, 2005) showed the endemic Australian Salicornioideae Ulbr. (Amaranthaceae) to have rapidly radiated during the Late Miocene and Pliocene as Australian aridity intensified. Diversification was likely driven by adaptation to saline inland lake systems. Salt tolerance, wind pollination and small, easily dispersed seeds promoted the rapid diversification across newly emerging arid habitats spreading across the continent.

Australian *Atriplex* L. (Amaranthaceae) has undergone a rapid radiation during the Late Miocene, followed by intensifying diversification during the Pliocene (Kadereit et al., 2010; McDonald, 2020). A wide and rapid range expansion was likely promoted by dispersal along drainage axes of Riverine Desert habitats that connected coast and inland habitats (McDonald, 2020). Diversification was likely driven by several migration events and subsequent adaptation to varied inland habitats (e.g. Karst Plain, Stony Desert). The formation of Karst Plain habitats and Stony Desert habitats coincides with the emergence of species and clades adapted to these

habitats. Further, tolerance of coastal habitats likely pre-adapted *Atriplex* progenitors to colonize arid inland regions by migration along littoral connections.

Diverse *Eucalyptus* L'Her (Myrtaceae) clades have adapted to a drying environment starting with the onset of Australian aridity during the Miocene (Thornhill, 2019). The lineages radiated and diversified over the emerging semi-arid landscapes dominated by open woodlands in the recent past, as evident by a steady increase in diversification rates during the Pliocene and a significant increase during the Pleistocene.

The Australian spinifex grasses, Triodiinae Benth (Poaceae), diversified rapidly coincident with the global cooling during the Middle Miocene and aridification when drier habitats expanded in Australia (Toon et al., 2015). Niche shifts into savannah habitats (open woodland with grassy understory) were coincident with the expansion of woodland, shrubland and grassland during the Late Miocene and Pliocene. Biome shifts correlated with changes in leaf anatomy, from amphistomatous to epistomatous leaves. The authors suggested this trait shift of being adaptive for taxa shifting in the savannah biome. Further, drought tolerance and a quick growth respond to rainfall might have provided a pre-adaptation to migrate into highly seasonal savannah habitats. Diversification was relatively constant during the Pliocene and slowed towards the present, likely due to a progressive saturation of niches.

Phylogenetic and biogeographic studies of *Hakea* (Cardillo et al., 2017) and *Acacia* (Renner et al., 2020) revealed the role of biome shifts during diversification on a continental scale. In *Hakea*, species distribution expanded with a transition to more open, drought and fire-prone habitats promoted by aridification during the Miocene and diversification of semi-arid arid adapted species increased during the Pliocene. Diversification appeared to be driven by frequent biome shifts between semi-arid, arid and savannah biomes. In *Acacia*, changing climate profiles (during the Miocene and Pliocene) may have affected diversification processes

and promote radiation across the lineage, and glacial–interglacial cycles during the Pleistocene may have promoted more recent speciation.

#### *4.2.3.7 Summary: Arrival, migration and diversification of Australian Camphorosmeae*

As previously concluded by Cabrera et al. (2009, 2011), the Camphorosmeae radiated concurrently with the development of arid habitats starting in the Late Miocene. Initial diversification, rapid radiation and subsequent diversification of Australian Camphorosmeae is in line with the evolution of various arid-zone taxa that evolved in response to the continent's intensifying aridity since the Miocene (Crisp et al., 2004; Byrne et al., 2008, 2018; Crisp and Cook, 2013; Byrne and Murphy, 2020). The arrival of this group dates to a period of major climatic, landscape and biotic transition. Establishment and inland migration may have been favored by adaptation. Progenitors adapted to coastal, marshy conditions may have migrated along Riverine Desert landscapes that developed as a consequence of ceasing palaeodrainage systems in south and western Australia. Initial diversification was likely promoted by major climatic and geologic changes during the "Hill Gap". Despite the lack of phylogenetic resolution, a shift in habitat preferences is evident between the two rapid radiations and the *Sclerolaena* clade. Originating from Riverine Desert habitats, progressively forming arid landscapes may have favored rapid range expansion across the continent and diversification into newly emerging arid habitats. Migration was probably multidirectional and likely followed a west-east aridification trend of the continent. Crown group diversification was strongest during the Pliocene and likely promoted by the west-to-east expansion of Riverine Desert habitats and subsequent expansion and colonization of newly developing arid habitats. However, the diversification pattern is not uniform across the phylogeny. While some clades appear to be quite conservative with respect to their habitat, several clades show a diversification of colonized habitats. Rapid range expansion, fast habitat saturation, as well as periodic expansion, contraction and replacement of arid habitats may have caused the rather species-poor clades of the *Maireana* grade compared to the continuously diversifying *Sclerolaena* clade.

### *4.2.4 Perianth types and dispersal syndromes*

The dispersal of seeds is a critical step in the life cycle of a plant and has important implications for spatial distribution, colonization, gene flow and evolution (Howe and Smallwood, 1982; Garant et al., 2007; Räsänen and Hendry, 2008; Vander Wall and Beck, 2012). The evolution of fruit characters is thought to be highly linked to specific dispersal syndromes and thus to be highly responsive to interactions with animals (Howe and Smallwood, 1982; Primack, 1987; Strauss and Irwin, 2004). Dispersal syndromes and animal-plant interactions (zoochory, herbivory) affect the fitness of individual plants and thus the evolutionary success of lineages. In the Australian Camphorosmeae, the fruiting perianth enclosing the seed functions as diaspore (with exception of *Roycea*).

The Australian Camphorosmeae have evolved a wide range of perianth characters (Fig. 1) that can be assigned to a number of syndromes and provide suggestions for the observable phylogenetic trends. *Maireana* species of the first and second polytomy predominantly have horizontal, winged perianth appendages (Fig. 2 and 3). The perianth of *Roycea* does not enlarge in fruit. *Enchylaena* and *Threlkeldia diffusa* produce fleshy, berry-like fruiting perianths. The berry-like fruits of *Enchylaena* are usually red, pink, or yellow, while the fruits of *T. diffusa* are reddish to black. The infructescences of *Dissocarpus* and *Eremophea* are fused at the perianth base and form spiny balls (Fig. 2 and 4). The species of the *Sclerolaena* clade produce predominantly spiny perianth appendages, with *Osteocarpum* producing vertical wings, and some species produce wing-spine intermediate perianth appendages (*Eri. sclerolaenoides*, *S. stelligera*, *S. fimbriolata*, *S. brachyptera*, *S. alata*, *S. parviflora*).

Wind-dispersed plants are common in dry habitats while animal-dispersed plants are more frequent in wet habitats (Howe and Smallwood, 1982). Wind dispersal does not rely on animals for diaspore dispersal, and strong winds and weather events result in wide dispersal of suitable diaspores. However, dispersal is non-directional and can reduce the fitness by reducing

the chances of seedling establishment in suitable areas for germination. Zoochory increases the fitness of plants by increasing the chances of seedling establishment on sites with high probability of survival (Wenny, 2001). Herbivory, however, decreases the fitness of plants by reducing plant growth, survival or reproduction (Strauss & Agrawal, 1999).

#### *4.2.4.1 Perianth appendage types*

The predominant perianth appendages of Australian Camphorosmeae, wings and spines, emerged independently in different clades and thus represent a homoplastic feature (Fig. 2-4). This convergent evolution may relate to different dispersal syndromes. Wing-like appendages dominate the *Maireana* grade and are present in all *Maireana* clades (Fig. 2 and 3, clades 1-11) and in *Osteocarpum* (clade 12). The vertical wings of *Osteocarpum* are not homologous to the horizontal *Maireana* wings but are more similar to the spines of *Sclerolaena* species (Wilson, 1984; Cabrera et al., 2009). Wing-like perianth appendages are also present in Camphorosmeae members in Eurasia, North America and in the central Asian sister group *Grubovia* (Scott, 1978; Zhu et al., 2003; Cabrera et al., 2009; Kadereit and Freitag, 2011). Wing-like perianth appendages are thought to be the plesiomorphic character within Camphorosmeae (Kadereit et al., 2003, 2005; Akhani et al. 2007; Cabrera et al., 2009), which is confirmed by this study. Spine-like appendages dominate the *Sclerolaena* clade but are also present in clade 11 (*Eremophea*, *Dissocarpus*). The spines of *Eremophea* and *Dissocarpus* are probably not homologous, as they differ in their position on the fruiting perianth (Wilson, 1984; Cabrera et al., 2009). Noticeably, the *Maireana* clades have fewer species than the species-rich *Sclerolaena* clade, even though all clades of the second polytomy have the same stem age (Fig. 6, Table 5). This raises the question of whether the appearance of spines (or loss of wings) was a key innovation for the diversification of the *Sclerolaena* clade.

#### *4.2.4.2 Perianths without prominent appendages*

The perianth of the three *Roycea* species does not enlarge in fruit, thus does not develop any appendages. This loss of appendages may represent an ecological adaptation (Cabrera et al., 2009). *Roycea* species are edaphic specialists in heavy, saline soils of Desert Lake habitats in south-western Australia. Interspecific competition is probably low, since only few species are adapted to such extreme sites. Once the species became adapted to this particular habitat, dispersal structures no longer provided an evolutionary advantage. The small fruits may be dispersed either by gravity or by water flow after rainfall events.

Fleshy, berry-like fruiting perianths arose two times independently in *Enchylaena* (clade 6) and in *Threlkeldia diffusa* (clade 15). The emergence of this fruit type may be considered an ecological adaptation to animal dispersion (Howe and Smallwood, 1982). The berry-like fruits of *Enchylaena* are dispersed by birds (Rogers, 1993) and the fruits of *T. diffusa* are assumed to be dispersed by birds (Milewski, 1982; Kühn et al., 1993). Dispersal by birds increases the distribution distance and chance of colonizing new habitats (Milewski, 1982; Vander Wall and Beck, 2012). Seedling establishment from feces, however, is sometimes poor as the soil surface can be a harsh environment (Vander Wall and Beck, 2012). *Threlkeldia diffusa* is distributed along the coast from north-western Australia to south-eastern Australia and is also found on Tasmania. This species is mainly found in coastal habitats but also occurs in Desert Lake habitats in south-western and South Australia. *Enchylaena tomentosa* is one of the most widespread Australian Camphorosmeae species and occupies various soil types and habitats across all mainland states (Wilson, 1984; Fig. 5 and 6, clade 6). The wide distribution and adaptation to diverse soils and habitats may have been promoted by endozoochory by birds.

#### *4.2.4.3 Wing-like perianth appendages*

Wing-like perianth appendages are present in ca. 40 % of all described Australian Camphorosmeae species (Wilson, 1984) and dominate the *Maireana* clades of the first and second polytomy. Wing-like fruit appendages are associated with aerial dispersal (Howe and Smallwood, 1982; van der Pijl, 1982; Sorensen, 1986). The emergence of wind-dispersed fruits is thought to be promoted by consistent strong winds (Howe and Smallwood, 1982). Wind dispersal commonly occurs in open habitats and promotes long distance dispersals (Nathan et al., 2002; Traveset et al., 2014). As the winged-perianth type is thought to be the plesiomorphic character within Camphorosmeae, the progenitors with wing-like appendages likely benefited from the climatic changes during the Late Miocene (see 4.2.4 and Fig. 7). In particular, the stronger westerly winds (driven by the intensifying ACC), the seasonal aridity (caused by the northwards moving STHP) and extreme weather events (anti-clockwise cyclones) may have enabled a fast and wide dispersal of species with a light, winged perianth across Australia's flat and open interior. These species were able to quickly colonize the progressively expanding arid landscape and subsequently diversified, probably driven by ecological adaptation.

Many diaspores adapted to wind dispersal also float and can be widely distributed by floods along drainage systems (Ridley, 1930; McDonald, 2020). Winged diaspore appendages are also present in the arid-adapted Australian *Atriplex* species (Parr-Smith, 1982; Wilson, 1984). *Atriplex* radiated in the Late Miocene-Pliocene along with the Camphorosmeae (McDonald, 2020). The fruits possess bracteoles or winged bracteoles that are effective in water and wind dispersal. Wind dispersal probably promotes range expansion better than dispersal by water. The vegetation in dry, saline, and alkaline valleys of Riverine Desert systems is open and low, enhancing the chances of dispersal over long distances. Dispersal by water confines range extension to water features. However, heavy rainfall events may have promoted a spread over broad flood plains. A combination of wind and water dispersal is thought to have promoted

a fast range expansion and colonization of new habitats by migration through drainage systems (McDonald, 2020).

Successful dispersal by water depends on the buoyant capacity, which in turn depends on the perianth surface area and weight. Some *Maireana* species produce fruiting perianths with a potential for dispersal by water. Papery, inflated, hollow or spongy perianths with sometimes large wings are present in *Maireana sedifolia*, *M. georgei*, *M. campanulata*, *M. spongiocarpa* and *M. trichoptera.* It is conceivable that water-dispersal by regular flows of ephemeral desert rivers and flooding events may have resulted in colonization of additional habitats, such as flood plains of Riverine Deserts habitats and Desert Clay Plains.

The lack of geological barriers and the fast, undirectional dispersal by wind (and water) probably allowed for higher gene flow over long distances. This ease of long-distance migration is well illustrated by the extensive ranges of many *Maireana* species (Wilson, 1984). The suspected increased gene flow is evident in the numerous hybrids and morphologic intermediate species and variants described. For instance, the species of clades 4, 6, 11, and 17 frequently hybridize, and hybrids between clades have commonly been described (e.g. *M. integra*, clade 5, and *M. planifolia*, clade 7). Hamrick and Godt (1996, 1997) investigated the effects of life history traits on genetic diversity in plants and found that long-lived perennial species with wind-dispersed seeds tended to have higher genetic diversity and less divergence between populations, compared to short-lived species with seeds dispersed by epizoochory or gravity that showed low genetic diversity and high divergence between populations. Hence, the increased mobility of wind-dispersed species and consequent increased gene flow between recently diverged species may have resulted in low divergence, reduced speciation and relatively species-poor *Maireana* clades (Hamrick and Godt, 1996, 1997; Garant et al., 2007; Räsänen and Hendry, 2008; Kisel and Barraclough, 2010). Interestingly, the differences in life form and genetic diversity observed by Hamrick and Godt (1996, 1997) are also evident

between the species of the *Maireana* Grade and the *Sclerolaena* clade (Fig. 2). The *Maireana* grade species are predominantly high-growing shrubs (up to 100 cm and long-lived), while the *Sclerolaena* clade species are predominantly low-growing perennials with a woody base (up to 50 cm high and short-lived).

#### *4.2.4.4 Spine-like perianth appendages*

Spine-like perianth appendages are present in ca. 45 % of all described Australian Camphorosmeae species (Wilson, 1984) and dominate the *Sclerolaena* clade. The spine-like perianth appendage is a derived character (Cabrera et al., 2009) and arose at least two times in clade 11 and in the *Sclerolaena* clade (Fig. 2-4). Spines provide effective protection against herbivory (Belowsky et al., 1991; Hanley et al., 2007). With respect to dispersal syndromes, spines are associated with animal-dispersal (Howe and Smallwood, 1982; Sorensen, 1986). In general, the spiny diaspores of *Sclerolaena* are associated with adhesive dispersal (epizoochory) (van der Pijl 1982; Cabrera et al., 2009). Endozoochory by emus has been observed for a few species (*S. bicornis*, *S. divaricata*) (Rogers et al., 1993) and some species are dispersed by ants (Davidson and Morton, 1981). In the absence of successful animal dispersal, the diaspores merely fall to the ground (Jurado et al., 1991; Peakall and Beattie, 1995) or are wind-dispersed as a tumbleweed (*S. muricata*) (Cunningham et al., 1981; McDonald, 2020). Dispersal syndromes in *Sclerolaena* have not yet been studied in their entirety, thus the ratio of species dispersed by animals to species with unassisted diaspore dispersal is unknown.

Assuming epizoochory, the spiny diaspores may adhere to animal fur or foot and be dispersed from the mother plant and detach randomly from the dispersing animal (van der Pijl 1982; Sorensen, 1986). On the Australian continent, these most likely are large herbivores such as kangaroos. The onset of the *Sclerolaena* crown diversification coincides with the emergence and rapid radiation of modern kangaroos at the beginning of the Pliocene ca. 5 Mya (Couzens and Prideaux, 2018, Fig. 7). The rapid radiation of kangaroos is associated with an abrupt

increase in grassland habitats that expanded during the Pliocene (Toon et al., 2015; Couzens and Prideaux, 2018; Anderson et al., 2019). Since kangaroos feed primarily on grasses (Couzens and Prideaux, 2018) and avoid ingestion of plants with spines (Belovsky et al., 1991), dispersal of spiny *Sclerolaena* diaspores is likely to occur randomly and undirected. Moreover, the adhesion capacity and dispersal distance may be limited due to the lack of typical adhesive fruit features such as burrs and barbs in *Sclerolaena* species. In addition, species producing relatively large and heavy diaspores (e.g. *Dis. paradoxus*, *S. bicornis*, *S. cornishiana*) were found to be unable to adhere to animals (Jurado et al., 1991).

Some *Sclerolaena* species, such as *S. diacantha* (clade 17), *S. ventricosa* (clade 15) and *S. convexula* (clade 13), develop food bodies in the hollow perianth base to which ants are attracted (Davidson and Morton, 1981a). Ant dispersal may also be present in other *Sclerolaena* species with a hollow perianth base, such as *S. brachyptera* (clade 12), *S. costata*, *S. alata*, *S. parviflora* (clade 13), *S. glabra* (clade 15), *S. parallelicupis*, *S. burbidgeae*, *S. uniflora*, *S. holtiana*, *S. eriacantha* (clade 17), and in *Osteocarpum* (clade 12). Myrmecochory appears to be an adaptation to highly directional dispersal to favorable microsites (Davidson and Morton, 1981). The ants harvest the diaspore and transport it to the nest and then remove the food masses from the perianth. As nutrients, such as mineral nitrogen and phosphorous, are concentrated on the ant mound, the seed is in a favorable microsite for germination and growth. Plants growing randomly in the area around the mound were smaller and notably less vigorous than their counterparts on ant mounds (Davidson and Morton, 1981b). The maximum transport distance is usually less than 10 m, with exception of one ant taxon observed (Iridomyrmex sp. D.) that transported the diaspore up to 77 m. Interestingly, differences between soil properties were also found. In habitats characterized by crusty, alluvial loam soils, ant-dispersed species grew almost exclusively on ant mounds. The same species usually grow in leachy, nutrient-poor sandy soils. This restricted seed dispersal may result in a depletion of the gene pool as gene flow is

minimized (Garant et al., 2007). However, ant dispersal has been shown to enhance small-scale genetic structure of ant-dispersed *S. diacantha* populations (Peakall and Beattie, 1995). This is because ants harvest numerous plant individuals and accumulate thousands of seeds on the long-lived ant mounds, thus increasing the available genetic resources. Myrmecochory has also been observed in *Dissocarpus* (Ising 1964; Davidson and Morton, 1981b). Australian ant taxa (Leptomyrmecini Emery) underwent a rapid radiation in the Early and Middle Miocene (Ward et al., 2010). The dominant arid-adapted genera *Anonychomyrma* Donisthorpe and *Iridomyrmex* Mayr emerged and diversified during the Late Miocene (Messinian) and Pliocene (Fig. 7). A rapid diversification of Australian arid-adapted termites (Amitermes group) also coincides with the emergence of spine-like appendages in *Sclerolaena* (Heimburger et al., 2022). The emergence of spines may have also be beneficial in terms of water management. Compared to wing-like perianths, the slender, spiny perianths of *Sclerolaena* species have a reduced surface. This may help to reduce transpiration in hot environments (Peguero-Pina et al., 2020). In addition, the presence of leaf and branch pubescence is more abundant in species of the *Sclerolaena* clade than in the *Maireana* grade (Fig. 2). This pubescence may reduce light absorptance and results in a reduced heat load, and thus in lower leaf temperatures and lower transpiration rates (Ehleringer and Mooney, 1978).

In contrast to the wind-dispersed *Maireana* species, colonization of new habitats by *Sclerolaena* species may have been slower and on a smaller scale. Unassisted seed dispersal by gravity restricts the dispersal distance to an area close to the mother plant. Exozoochory by grazing animals probably occurs randomly over relatively small distances. Myrmecochory restricts seed dispersal to small areas but counteracts inbreeding by accumulating seeds of numerous plants on the ant mounds. In addition, ant dispersal may favor adaptation to different soil types and colonization of new habitats. The decreased mobility of species with a spiny perianth and consequent decreased gene flow between recently diverged species may have

increased divergence and thus speciation of the predominantly low-growing, short-lived perennial species in the *Sclerolaena* clade (Hamrick and Godt, 1996, 1997; Garant et al., 2007; Räsänen and Hendry, 2008; Kisel and Barraclough, 2010), leading to a continuously diversifying *Sclerolaena* clade. Moreover, the spiny perianths are an effective protection against herbivores and diversification of the species-rich *Sclerolaena* clade coincides with the emergence of large grazing animals and in the early Pliocene.

#### *4.2.4.5 Summary: Perianth types and dispersal syndromes*

The predominant fruiting perianth types, wings and spines, of the Australian Camphorosmeae are a homoplastic feature that emerged independently in different clades. The wing-like appendages represent the pleisomorphic feature from which apomorphic appendage types derived. The perianth appendages are a product of convergent evolution and arose variously modified in different clades. The fruiting perianths act as diaspore and relate to different dispersal syndromes. The winged perianths of *Maireana* are dispersed by wind (and partly probably also by water) and favor fast, non-directional dispersal over long distances, across Australia's relatively flat interior. This likely promoted rapid colonization of vacant habitats throughout the expansion of arid landscapes, in particular with respect to the strengthened wind regime during the Late Miocene. Increased mobility and consequent increased gene flow probably resulted in lower speciation rates, which is now apparent in the species-poor clades of the *Maireana* grade, compared to the species-rich *Sclerolaena* clade. The spiny fruiting perianths of *Sclerolaena* appear to be animal-dispersed. However, the proportion of animal-dispersed species to species with unassisted diaspore dispersal is unknown. The onset of *Sclerolaena* diversification coincides with radiations of large herbivores and ants. Dispersal distance by large herbivores is likely limited due to a lack of adhesive diaspore features. Myrmecochory is a highly directional dispersal to favorable microsites and may promote colonization of extreme arid habitats. In contrast to the wind-dispersed *Maireana* species,

colonization of new habitats and migration between habitats of *Sclerolaena* species may have been slower and on a smaller scale. This reduced mobility results in lower gene flow and consequently may have increased speciation, which is now apparent in the continuous cladogenesis of the species-rich *Sclerolaena* clade. Further, the success of spiny perianths may have been promoted by their protective properties against herbivores and by their beneficial properties in terms of water management.

# *5. CONCLUSION AND OUTLOOK*

In this study, seven research questions regarding phylogeny, taxonomy and evolution of morphological traits were re-examined using new molecular and data analysis methods and extended morphological and ecological data. In the following, the results for each hypothesis are briefly summarized and complemented by an outlook for future studies.

*1) Can a modified NGS-based methodology improve the insufficient resolution of the phylogeny to reconstruct yet unrecognized species groups as statistically supported clades?*

The NGS method used in this study included a modified RADseq laboratory protocol and an assembly workflow tailored to the yielded sequence data. This approach has already been successfully tested for a number of taxonomically challenging plant groups (Rij et al., 2019; Dos Santos et al., 2022; Hühn et al., 2022, 2023; Messerschmid et al., 2023). This study showed a RADseq-inherent limitation: The relatively large divergence of the Australian Camphorosmeae (ca. 10-16 million years) considerably increases the proportion of missing data (almost 94 % in Assembly 1). A high proportion of missing data is generally unproblematic in RADseq analyses (e.g. Huang and Knowles, 2016; Eaton et al., 2017; Tripp et al., 2017; Graham et al., 2020), but it does limit the applicable methods for phylogenetic inference (e.g. Chou et al., 2015; Xu and Yang, 2016; Molloy and Warnow, 2018; Hühn et al., 2022). Still, the methodology used here has confirmed the previously assumed biological processes that led to the observable phylogenetic signal. Further, it yielded several statistically well-supported clades of yet unrecognized species groups.

A putative positive effect of data filtering on phylogenetic inference should be investigated in future studies (e.g. Lanier et al., 2014; Molloy and Warnow, 2018; Hühn et al., 2022). In addition, complementation of sampling could have a positive effect on data assembly and phylogenetic inference (e.g. Huang and Knowles, 2016; Eaton et al., 2017).

*2) To what extent can the phylogenetic clades be circumscribed with morphological data?*

Phylogenetic inference resulted in 17 statistically supported clades. Many of them may be characterized by combinations of multiple morphological traits (clades 1, 2+3, 6, 9, 10, 11, 12, 13, 15+16, 17). In addition, there were also morphological subgroups within well-supported clades (clade 4). In particular, the fruit characters, such as the general appendage type, the shape and orientation of wing-like appendages, the shape of the perianth tube, the shape of the upper perianth as well as the occurrence of additional processes on top of it, the number of flowers and fruits per bract, the number and orientation of spines, the appearance of a limb (in species with spine-like appendages), and the orientation of the seed, showed systematic relevance. The morphological characters evaluated appear highly homoplastic, likely because of convergent evolution (Wilson, 1975; Cabrera et al., 2009). However, the devil is in the detail: For instance, *Sclerolaena* spines are not homolog to the *Dissocarpus* and *Eremophea* spines, and *Maireana* wings of the first radiation are not the same as wings in the second radiation and in *Osteocarpum*. *Sclerolaena* spines are produced by the intertepaline veins of the fruiting perianth tube, while the spines of *Dissocarpus* and *Eremophea* are tepaline and arise from the base of the fruiting perianth. The horizontal, 5-lobed wing produced by *Maireana* species of the first radiation usually consists of 5 separate wings which grow from the base of each lobe, while the horizontal, entire wing produced by *Maireana* species of the second radiation is produced by the perianth tube, and the vertical wings in *Osteocarpum* are produced as extension of the intertepaline tubercles at the perianth tube apex.

The vegetative traits evaluated were found useful to describe general trends across the phylogeny. Species of the *Maireana* grade tend to be larger in growth height and longer-lived than species of the *Sclerolaena* clade. Species of the *Sclerolaena* clade more often have glabrous branches and leaves compared to species of the *Maireana* grade that are more often pubescent. The fruiting perianth tubes are generally slender, somewhat dorsiventrally

compressed and more often ribbed in species of the *Sclerolaena* clade, compared to rather widely thickened and smoot fruiting perianth tubes in species of the *Maireana* grade.

In future studies, the nature of the perianth tube at different stages of maturity (for instance; spongy, succulent, papery, leathery, crustaceous, woody), as well as ontogeny and position of the perianth appendages need to be elucidated in more detail. Further, it appears reasonable to review the life forms and habits, as they are inconsistently described.

*3) Are the newly surveyed ecological data, such as soil types, salinity and habitat types, of systematic relevance?*

The soil types, salinity and preferred habitats surveyed were found to be of partial systematic relevance. Species of the first radiation (clades 1-3) are predominantly found in sandy, saline soils of Riverine Desert and Desert Lake habitats. Species of the second radiation (including the *Sclerolaena* clade) are more often found in soils of slight to moderate salinity, in contrast to species of the first radiation. A preference for calcareous and gypseous accumulated in clades 11 and 17. Species of clades 13 and 15 are predominantly found in Sand and Stony Desert habitats, respectively. However, these data are more useful for the interpretation of presumed migration routes with respect to changing landscapes along Australia's aridification.

For future studies, a complementation of the preferred soil types would be beneficial, as the species descriptions are inconsistent with respect to soil compositions and the degree of salinity.

*4) Do the newly characterized clades show potential to serve as a basis for a revised generic delimitation and taxonomy?*

Yes, to some extent. This study confirms the concerns regarding an artificial generic delimitation of natural species groups. *Sclerolaena* and *Maireana* turned out polyphyletic and the species-poor genera were found nested within larger genera (e.g. *Threlkeldia* nested within

*Sclerolaena* clades 14 and 15; *Eriochiton* nested within *Maireana* clade 2; *Enchylaena* nested within *Maireana* clade 6) or to form statistically supported clades with other small genera (e.g. clade 10, *Didymanthus*, *Dissocarpus*, *Eremophea* plus *S. fimbriolata* and *M. amoena*). Further, this study clarified the phylogenetic relationship of species that were previously treated as monotypic genera: *Eri. sclerolaenoides* in clade 2, *S. stelligera* (former *Stelligera endecaspinis*) in clade 5, and *S. brachyptera* (former *Sclerochlamys brachyptera*) in clade 12. Moreover, assumptions regarding affiliated species groups based on either morphological similarities or described hybridization events, can now be confirmed on a molecular level. For instance, Wilson (1975, 1984) suggested: 1) *M. oppositifolia* to be closely related to *Roycea*, 2) the species of clade 4 to be closely affiliated due to close interspecific morphological affiliations, 3) the species of clade 6 to be closely related due to morphological affinities and interspecific and intergeneric crossings, and 4) the species of clade 17 to form a morphologically intergrading species complex. Further, the presumed separation of *Mal. gracilis* from the other *Malacocera* species (Chinnock, 1980) is now supported by molecular data. Moreover, previously debated characteristics were found useful in circumscribing an apparently highly diverse species group (clade 11, paired or aggregated perianths fused at their bases) (Ising, 1964; Wilson, 1975). Hence, new taxonomic combinations based on the phylogenetic outcome using combinations of morphological traits appear reasonable. However, not for the entirety of the species studied. Within the *Maireana* grade there are several clades that are difficult to circumscribe and can only be combined with other clades using fairly general characteristics (clades 4, 5, 7 and 8). In addition, several species were reconstructed on individual branches and are hard to assign to other species groups (e.g. *M. cannonii*, *M. sedifolia*).

This study could be a good basis for a taxonomic revision, but distinctive characters need to be found for several clades and missing species need to be assigned to phylogenetic clades based on morphological characters. This has already proved difficult in numerous

taxonomic and nomenclatural studies. Therefore, an expanded sampling would be desirable to add species to existing clades or to recover new clades that are morphologically distinct. In addition, reanalyzing the NGS data generated in this study using approaches developed for species delimitation may shed new light on natural species groups (e.g. Duminil and Di Michele, 2009; Tobias et al., 2010; Puillandre et al., 2012; Luo et al., 2018). Using these approaches for generic delimitation may appear unconventional at a first glance, however, most clades diversified within the last 5 Million years and yield large amounts of sequence data on species level. Due to the extended locus length, these data are sufficiently informative for tree-based applications and at the same time comparable to SNP datasets that are commonly used for species delimitation. Hence, analyzing subsets of the here generated datasets using a combination of approaches may help to extend our knowledge of natural species groups within the Australian Camphorosmeae and consequently may help to conceptualize a taxonomic framework.

*5) Do the postulated hypotheses regarding migration, rapid radiation and crown-group diversification correspond to the aridification of the Australian continent since the Miocene?*

Overall, the phylogenetic results and the outcome of the divergence time estimation are consistent with the previously postulated hypotheses. The Australian Camphorosmeae likely arrived in Australia via long distance dispersal during the Middle to Late Miocene. Colonization was likely promoted by pre-adaptation to arid and saline conditions. A littoral connection may have taken place by species adapted to coastal conditions that migrated along declining palaeodrainage systems to hypersaline inland habitats. Along with the increasing and progressively expanding and intensifying aridity, the Australian Camphorosmeae diversified rapidly. Diversification was likely promoted by the major climatic and geologic changes during the “Hill Gap”. Two rapid radiation events overlap in time during the Late Miocene. The first rapid radiation may have taken place by species migrating form south-western to north-western

Australia that relatively quick colonized vacant habitats and subsequently diversified. The second rapid radiation may have been caused by multidirectional migration of species diversifying into various habitats across the continents interior. Riverine Desert and Desert Lake habitats that expanded easterly during this period may have been the main migration corridors. Most crown group diversification took place during the Pliocene, corresponding the intensifying aridification and expanding arid habitats during this period. Concurrently with the emergence and expansion of extreme habitats during the Pliocene and Pleistocene, lineages started to diversify in Sand and Stony Desert habitats. Diversification trends among clades vary. Some clades show no or only little speciation since the Pliocene crown group diversification, while other clades continuously diversified even throughout the Pleistocene. This pattern may have been caused by repeated cycles of localized habitat contraction and expansion, along with climate oscillation and extreme drought periods during the Pleistocene, which were in favor for species of some clades but hostile to species of other clades.

Although the events evident in the dated phylogenetic topology appear to be in concordance with the major climate and landscape transitions of Australia since the Miocene, the divergence time estimation yielded relatively large confidence intervals. This makes a chronological interpretation of events relatively vague. However, the aridification processes of the Australian continent were also relatively vaguely known and gained substantial accuracy due to the numerous phylogenetic studies that confirmed and extended our knowledge of these processes (reviewed in Crisp et al., 2004; Byrne et al., 2008, 2018; Crisp and Cook, 2013; Byrne and Murphy, 2020). Still, a reduced proportion of missing data in the dataset may improve the estimation accuracy, narrow the confidence intervals and thus advance our knowledge of processes that shaped the diversity of Australian Camphorosmeae. A reduction of missing data may be achieved by filtering for high coverage loci or by application of a method that is less prone to locus dropout.

*6) What influence did morphological traits of the fruiting perianth may have had on dispersal?*

The fruiting perianths of the Australian Camphorosmeae act as diaspore and relate to different dispersal syndromes. The winged perianths of *Maireana* are dispersed by wind and favor fast, non-directional dispersal over long distances, across Australia's relatively flat interior. This likely promoted rapid colonization of vacant habitats throughout the expansion of arid landscapes during the Late Miocene and Pliocene. Increased mobility and consequent increased gene flow probably resulted in lower speciation rates, which is now apparent in the species-poor clades of the *Maireana* grade. The spiny fruiting perianths of *Sclerolaena* appear to be animal-dispersed. The onset of *Sclerolaena* diversification coincides with radiations of large herbivores, termites and ants. Myrmecochory is a highly directional dispersal to favorable microsites and may promote colonization of extreme arid habitats. In contrast to the wind-dispersed *Maireana* species, colonization of new habitats and migration between habitats of *Sclerolaena* species may have been slower and on a smaller scale. This reduced mobility results in lower gene flow and consequently may have increased speciation, which is now apparent in the continuous cladogenesis of the species-rich *Sclerolaena* clade.

The dispersal of wing-like fruiting perianths by wind appears reasonable. However, little is known about the dispersal of spine-like perianths. In order to improve our understanding of the evolution of perianth traits and their impact on dispersal and speciation, the species need to be studied more closely in terms of their dispersal strategies, in particular with respect to myrmecochory.

*7) What causes may have led to the emergence of species groups of varying species richness?*

Clearly, the current generic delimitation does not reflect the species richness of natural species groups revealed by this study. As mentioned above, repeated cycles of localized habitat contraction and expansion, extreme climate oscillation during the Pleistocene, as well as different dispersal strategies resulting in different levels of gene flow, may have resulted in varying degrees of species richness among clades. In addition, hybridization may have resulted in a finer scale phylogenetic structuring. However, introgression is yet to be tested in future studies.

Extinction events other than habitat contraction may also have led to a loss of diversity. Although there are some studies on the extinction of important faunal components during the Cenozoic (e.g. Miller et al., 1999; Van Der Kaars et al., 2017), little is known about putative extinction events of Australia's plant world in this period (e.g. Jordan, 1997; Lopes dos Santos et al., 2013). It is conceivable that fire as a ubiquitous component of arid landscapes (Martin, 2006; Crisp et al., 2011) and perhaps the emergence of large herbivores (Archer et al., 1994; Couzens and Prideaux, 2018) could have led to local extinction events. In addition, the short return to wet and warm conditions during the early Pliocene may have caused extinction. However, due to a lack of evidence possible extinction events will remain highly speculative until further research.

# *6. REFERENCES*

**Akhani, H., Edwards, G. and Roalson, E.H.,** 2007. Diversification of the old world Salsoleae sl (Chenopodiaceae): molecular phylogenetic analysis of nuclear and chloroplast data sets and a revised classification. International Journal of Plant Sciences, 168(6), pp.931-956. https://doi.org/10.1086/518263.

**Amor, M.D., Walsh, N.G. and James, E.A.,** 2020. Genetic patterns and climate modelling reveal challenges for conserving Sclerolaena napiformis (Amaranthaceae sl) an endemic chenopod of Southeast Australia. Diversity, 12(11), p.417. https://doi.org/10.3390/d12110417.

**Andersen, A.N.,** 2016. Ant megadiversity and its origins in arid Australia. Austral Entomology, 55(2), pp.132-137. https://doi.org/10.1111/aen.12203.

**Anderson, R.H.,** 1923. A revision of the Australian species of the genus Bassia. Proceedings of the Linnean Society of New South Wales, 48:317–355.

**Anderson, B.M., Thiele, K.R., Grierson, P.F., Krauss, S.L., Nevill, P.G., Small, I.D., Zhong, X. and Barrett, M.D.,** 2019. Recent range expansion in Australian hummock grasses (Triodia) inferred using genotyping-by-sequencing. AoB Plants, 11(2), p.plz017. https://doi.org/10.1093/aobpla/plz017.

**Andrews, S.**, 2010. FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

**Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. and Hohenlohe, P.A.,** 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17(2), pp.81-92. https://doi.org/10.1038/nrg.2015.28.

**Angiosperm Phylogeny Group, The,** 2017. The Catalogue of Life Partnership. APG IV: Angiosperm Phylogeny Group classification for the orders and families of flowering plants. Checklist dataset https://doi.org/10.15468/fzuaam accessed via GBIF.org on 2023-02-18.

**Archer, M., Hand, S.J. and Godthelp, H.,** 1994. Patterns in the history of Australia's mammals and inferences about palaeohabitats. History of the Australian vegetation: Cretaceous to Recent, pp.80-103. University of Adelaide Press. ISBN: 9781925261479. https://www.jstor.org/stable/10.20851/j.ctt1sq5wrv.

**Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A.,** 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one, 3(10), p.e3376. https://doi.org/10.1371/journal.pone.0003376.

**Barrett-Lennard, E.G., Bathgate, A.D. and Malcolm, C.V.,** 2003. Saltland pastures in Australia, a practical guide. WA Government. Dept. of Agriculture and Food. ISBN: 192086007

**Barker, W.R. and Greenslade, P.J.M.,** 1982. Evolution of the flora and fauna of arid Australia. Peacock Publications in Association with the Australian Systematic Botany Society and ANZAAS, South Australia Div., inc. ISBN: 0909209626.

**Barlow, B.A.,** 1994. Phytogeography of the Australian region. In: "Australian vegetation." Editor: Groves, R.H. pp. 3-35. Cambridge University Press, Cambridge, UK. ISBN: 0521414202.

**Belbin, L., Wallis, E., Hobern, D. and Zerger, A.,** 2021. The Atlas of Living Australia: History, current state and future directions. Biodiversity Data Journal, 9. https://doi.org/10.3897%2FBDJ.9.e65023. <http://www.ala.org.au>.

**Belovsky, G.E., Schmitz, O.J., Slade, J.B. and Dawson, T.J.,** 1991. Effects of spines and thorns on Australian arid zone herbivores of different body masses. Oecologia, 88, pp.521-528. https://doi.org/10.1007/BF00317715.

**Benbow, M.C., Lindsay, J.M. and Alley, N.F.,** 1995. Eucla Basin and palaeodrainage. The geology of South Australia, 2, pp.178-186. In: Drexel, J.H., Preiss, W.V., 'The geology of South Australia, Volume 2. The Phanerozoic. South Australia. Geological Survey. Bulletin, 54. ISBN: 0730806219.

**Bint, A.N.,** 1981. An early Pliocene pollen assemblage from Lake Tay, south-western Australia, and its phytogeographic implications. Australian Journal of Botany, 29(3), pp.277-291. https://doi.org/10.1071/BT9810277.

**Bird, M.I., O'Grady, D. and Ulm, S.,** 2016. Humans, water, and the colonization of Australia. Proceedings of the National Academy of Sciences, 113(41), pp.11477-11482. https://doi.org/10.1073/pnas.1608470113.

**Blache, D., Vercoe, P.E., Martin, G.B. and Revell, D.K.,** 2016. Integrated and innovative livestock production in drylands. Innovations in dryland agriculture, pp.211-235. https://doi.org/10.1007/978-3-319-47928-6_8.

**Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J.,** 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS computational biology, 15(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

**Bowen, B.B. and Benison, K.C.,** 2009. Geochemical characteristics of naturally acid and alkaline saline lakes in southern Western Australia. Applied Geochemistry, 24(2), pp.268-284. https://doi.org/10.1016/j.apgeochem.2008.11.013.

**Bowler, J.M.,** 1982. Aridity in the late Tertiary and Quaternary of Australia. In: Barker, W.R., and Greenslade, P.J.M., Evolution of the flora and fauna of arid Australia. Peacock Publications, South Australia. ISBN: 0909209626.

**Bowler, J.M., Kotsonis, A. and Lawrence, C.R.,** 2006. Environmental evolution of the Mallee region, western Murray Basin. Proceedings of the Royal Society of Victoria, 118(2), pp.161-210.

**Bowman, D.M., Brown, G.K., Braby, M.F., Brown, J.R., Cook, L.G., Crisp, M.D., Ford, F., Haberle, S., Hughes, J., Isagi, Y. and Joseph, L.,** 2010. Biogeography of the Australian monsoon tropics. Journal of Biogeography, 37(2), pp.201-216. https://doi.org/10.1111/j.1365-2699.2009.02210.x.

**Brandle, R.,** 1998. A biological survey of the stony deserts, South Australia, 1994-1997. Adelaide: Department for Environment, Heritage and Aboriginal Affairs.

**Brearley, D.,** 2003. Developing completion criteria for rehabilitation areas on arid and semi-arid mine sites in Western Australia (Doctoral dissertation, Curtin University). http://hdl.handle.net/20.500.11937/745.

**Burbidge, N.T.,** 1960. The phytogeography of the Australian region. Australian Journal of Botany, 8(2), pp.75-211. https://doi.org/10.1071/BT9600075.

**Byrne, M., Yeates, D.K., Joseph, L., Kearney, M., Bowler, J., Williams, M.A.J., Cooper, S., Donnellan, S.C., Keogh, J.S., Leys, R. and Melville, J.,** 2008. Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. Molecular ecology, 17(20), p.4398. https://doi.org/10.1111/j.1365-294X.2008.03899.x.

**Byrne, M., Joseph, L., Yeates, D.K., Roberts, J.D., Edwards, D.,** 2018. Evolutionary History. In: Lambers, H. (editor), On the Ecology of Australia's Arid Zone. Springer, Cham. https://doi.org/10.1007/978-3-319-93943-8_3.

**Byrne, M. and Murphy, D.J.,** 2020. The origins and evolutionary history of xerophytic vegetation in Australia. *Australian Journal of Botany*, *68*(3), pp.195-207. https://doi.org/10.1071/BT20022.

**Cabrera, J.F., Jacobs, S.W. and Kadereit, G.,** 2009. Phylogeny of the Australian Camphorosmeae (Chenopodiaceae) and the taxonomic significance of the fruiting perianth. International Journal of Plant Sciences, 170(4), pp.505-521. https://doi.org/10.1086/597267.

**Cabrera, J., Jacobs, S.W. and Kadereit, G.,** 2011. Biogeography of Camphorosmeae (Chenopodiaceae): tracking the Tertiary history of Australian aridification. Telopea, 13(1-2), pp.313-326. https://doi.org/10.7751/telopea20116023.

**Cardillo, M., Weston, P.H., Reynolds, Z.K., Olde, P.M., Mast, A.R., Lemmon, E.M., Lemmon, A.R. and Bromham, L.,** 2017. The phylogeny and biogeography of Hakea (Proteaceae) reveals the role of biome shifts in a continental plant radiation. Evolution, 71(8), pp.1928-1943. https://doi.org/10.1111/evo.13276.

**Certain, C., Della Patrona, L., Gunkel-Grillon, P., Léopold, A., Soudant, P. and Le Grand, F.,** 2021. Effect of salinity and nitrogen form in irrigation water on growth, antioxidants and fatty acids profiles in halophytes Salsola australis, Suaeda maritima, and Enchylaena tomentosa for a perspective of biosaline agriculture. Agronomy, 11(3), p.449. https://doi.org/10.3390/agronomy11030449.

**Chesson, P.,** 2000. Mechanisms of maintenance of species diversity. Annual review of Ecology and Systematics, 31(1), pp.343-366. https://doi.org/10.1146/annurev.ecolsys.31.1.343.

**Chinnock, R.J.,** 1980. The genus Malacocera RH Anderson (Chenopodiaceae). Journal of the Adelaide Botanic Garden, pp.139-149. https://www.jstor.org/stable/23873780.

**Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S. and Warnow, T.,** 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. BMC genomics, 16(10), pp.1-11. https://doi.org/10.1186/1471-2164-16-S10-S2.

**Couzens, A.M. and Prideaux, G.J.,** 2018. Rapid Pliocene adaptive radiation of modern kangaroos. Science, 362(6410), pp.72-75. https://doi.org/10.1126/science.aas8788.

**Crisp, M.D., West, J.G. and Linder, H.P.,** 1999. Biogeography of the terrestrial flora. Flora of Australia, 1, pp.321-367.

**Crisp, M.D., Laffan, S., Linder, H.P. and Monro, A.,** 2001. Endemism in the Australian flora. Journal of Biogeography, 28(2), pp.183-198. https://doi.org/10.1046/j.1365-2699.2001.00524.x.

**Crisp, M., Cook, L. and Steane, D.,** 2004. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present–day communities? Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359(1450), pp.1551-1571. https://doi.org/10.1098/rstb.2004.1528.

**Crisp, M.D., Burrows, G.E., Cook, L.G., Thornhill, A.H. and Bowman, D.M.,** 2011. Flammable biomes dominated by eucalypts originated at the Cretaceous–Palaeogene boundary. Nature Communications, 2(1), p.193. https://doi.org/10.1038/ncomms1191.

**Crisp, M.D. and Cook, L.G.,** 2013. How was the Australian flora assembled over the last 65 million years? A molecular phylogenetic perspective. Annual Review of Ecology, Evolution, and Systematics, 44, pp.303-324. https://doi.org/10.1146/annurev-ecolsys-110512-135910.

**Cunningham, G.M., Mulham, W.E., Milthorpe, P.L. and Leigh, J.H.,** 1981. Plants of western New South Wales.− Sydney. NSW: NSW Government Printing Office, Soil Conservation Service of NSW, 662. http://hdl.handle.net/102.100.100/292240?index=1.

**Curto, M., Schachtler, C., Puppo, P., Meimberg, H.,** 2018. Using a new RAD-sequencing approach to study the evolution of Micromeria in the Canary islands. Molecular Phylogenetics and Evolution, 119, 160–169. https://doi.org/10.1016/j.ympev.2017.11.005.

**Davidson, D.W. and Morton, S.R.,** 1981a. Competition for dispersal in ant-dispersed plants. Science, 213(4513), pp.1259-1261. https://doi.org/10.1126/science.213.4513.1259.

**Davidson, D.W. and Morton, S.R.,** 1981b. Myrmecochory in some plants (F. Chenopodiaceae) of the Australian arid zone. Oecologia, 50, pp.357-366. https://doi.org/10.1007/BF00344976.

**De Marchis, F., Wang, Y., Stevanato, P., Arcioni, S. and Bellucci, M.,** 2009. Genetic transformation of the sugar beet plastome. Transgenic research, 18, pp.17-30. https://doi.org/10.1007/s11248-008-9193-4.

**Degnan, J.H., Rosenberg, N.A., Wakeley, J.,** 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2 (5), e68. https://doi.org/10.1371/journal.pgen.0020068.

**Degnan, J.H., Rosenberg, N.A.,** 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24 (6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009.

**Ding, D.B., Pang, Q.H., Han, X.J. and Fan, S.J.,** 2021. Characterization and phylogenetic analysis of the complete chloroplast genome of Amaranthus viridis (Amaranthaceae). Mitochondrial DNA, 6(9), pp.2610-2612. https://doi.org/10.1080/23802359.2021.1961631.

**Dodson, J.R. and Ramrath, A.,** 2001. An Upper Pliocene lacustrine environmental record from south-Western Australia—preliminary results. Palaeogeography, Palaeoclimatology, Palaeoecology, 167(3-4), pp.309-320. https://doi.org/10.1016/S0031-0182(00)00244-3.

**Dos Santos, P., Brilhante, M.Â., Messerschmid, T.F., Serrano, H.C., Kadereit, G., Branquinho, C. and de Vos, J.M.,** 2022. Plant growth forms dictate adaptations to the local climate. Frontiers in Plant Science, p.4658. https://doi.org/10.3389/fpls.2022.1023595.

**Duminil, J. and Di Michele, M.,** 2009. Plant species delimitation: a comparison of morphological and molecular markers. Plant Biosystems, 143(3), pp.528-542. https://doi.org/10.1080/11263500902722964.

**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. Syst. Biol. 66 (3), 399–412. https://doi.org/10.1093/sysbio/syw092.

**Eaton, D.A., Overcast, I.,** 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics, 36(8): 2592-2594. doi: 10.1093/bioinformatics/btz966.

**Ebach, M.C., Gonzalez-Orozco, C.E., Miller, J.T. and Murphy, D.J.,** 2015. A revised area taxonomy of phytogeographical regions within the Australian Bioregionalisation Atlas. Phytotaxa, 208(4), pp.261-277. https://doi.org/10.11646/phytotaxa.208.4.2.

**Edgoose, C.J.,** 2003. Barkly Tableland Region, Northern Territory. In: Anand, R.R. and De Broekert, P., 2005. Regolith landscape evolution across Australia: A compilation of regolith landscape case studies with regolith landscape evolution models. Perth, WA, CRC LEME. ISBN: 1921039302.

**Ehleringer, J.R. and Mooney, H.A.,** 1978. Leaf hairs: effects on physiological activity and adaptive value to a desert shrub. Oecologia, 37, pp.183-200. https://doi.org/10.1007/BF00344990.

**El Shaer, H.M. and Squires, V.R.,** 2015. Halophytic and salt-tolerant feedstuffs: impacts on nutrition, physiology and reproduction of livestock. CRC Press. https://doi.org/10.1201/b19862.

**Ewels, P., Magnusson, M., Lundin, S., Kaller, M.,** 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32 (19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

**Fairfax, R.J. and Fensham, R.J.,** 2000. The effect of exotic pasture development on floristic diversity in central Queensland, Australia. Biological Conservation, 94(1), pp.11-21. https://doi.org/10.1016/S0006-3207(99)00169-X.

**Felsenstein, J.,** 1985. Confidence limits on phylogenies: an approach using the bootstrap. evolution, 39(4), pp.783-791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.

**Fensham, R.J., Laffineur, B. and Silcock, J.L.,** 2018. In the wake of bulldozers: Identifying threatened species in a habitat decimated by rapid clearance. Biological Conservation, 219, pp.28-34. https://doi.org/10.1016/j.biocon.2017.12.008.

**Flora of Australia,** 1984. Volume 4, Phytolaccaceae to Chenopodiaceae. Bureau of flora and fauna, Canberra. Australian Government Publishing Service. Copyright: Commonwealth of Australia, 1984. ISBN: 0644034424.

**Fox, M.D.,** 1999. Present environmental influences on the Australian flora. Flora of Australia, 1, pp.205-249. Australian Biological Resources Study/CSIRO Publishing. ISBN: 0643059652.

**Fujioka, T., Chappell, J., Honda, M., Yatsevich, I., Fifield, K. and Fabel, D.,** 2005. Global cooling initiated stony deserts in central Australia 2–4 Ma, dated by cosmogenic 21Ne-10Be. Geology, 33(12), pp.993-996. https://doi.org/10.1130/G21746.1.

**Fujioka, T., Chappell, J., Fifield, L.K. and Rhodes, E.J.,** 2009. Australian desert dune fields initiated with Pliocene–Pleistocene global climatic shift. Geology, 37(1), pp.51-54. https://doi.org/10.1130/G25042A.1.

**Fujioka, T. and Chappell, J.,** 2010. History of Australian aridity: chronology in the evolution of arid landscapes. Geological Society, London, Special Publications, 346(1), pp.121-139. https://doi.org/10.1144/SP346.8.

**Garant, D., Forde, S.E. and Hendry, A.P.,** 2007. The multifarious effects of dispersal and gene flow on contemporary adaptation. Functional Ecology, pp.434-443. https://www.jstor.org/stable/4540042.

**García-Palacios, P., Gross, N., Gaitán, J. and Maestre, F.T.,** 2018. Climate mediates the biodiversity–ecosystem stability relationship globally. Proceedings of the National Academy of Sciences, 115(33), pp.8400-8405. https://doi.org/10.1073/pnas.1800425115.

**Graham, M.R., Santibáñez-López, C.E., Derkarabetian, S. and Hendrixson, B.E.,** 2020. Pleistocene persistence and expansion in tarantulas on the Colorado Plateau and the effects of missing data on phylogeographical inferences from RADseq. Molecular Ecology, 29(19), pp.3684-3701. https://doi.org/10.1111/mec.15588.

**Groves, R.H.,** 1999. Present vegetation types. Flora of Australia, volume 1. In: Flora of Australia Volume 1, Introduction. 2nd Edition. Australian Biological Resources Study/CSIRO Publishing. ISBN: 0643059652.

**Hamrick, J.L. and Godt, M.W.,** 1996. Effects of life history traits on genetic diversity in plant species. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 351(1345), pp.1291-1298. https://doi.org/10.1098/rstb.1996.0112.

**Hamrick, J.L. and Godt, M.J.W.,** 1997. Allozyme diversity in cultivated crops. Crop science, 37(1), pp.26-30. https://doi.org/10.2135/cropsci1997.0011183X003700010004x.

**Hanley, M.E., Lamont, B.B., Fairbanks, M.M. and Rafferty, C.M.,** 2007. Plant structural traits and their role in anti-herbivore defence. Perspectives in Plant Ecology, Evolution and Systematics, 8(4), pp.157-178. https://doi.org/10.1016/j.ppees.2007.01.001.

**Harden, G.J.,** 1990. Flora of New South Wales (Vol. 4). NSW University Press. ISBN: 0868401889.

**Harvey, M.G., Judy, C.D., Seeholzer, G.F., Maley, J.M., Graves, G.R., Brumfield, R.T.,** 2015. Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. PeerJ 3:e895. doi: 10.7717/peerj.895.

**Heimburger, B., Schardt, L., Brandt, A., Scheu, S. and Hartke, T.R.,** 2022. Rapid diversification of the Australian Amitermes group during late Cenozoic climate change. Ecography, 2022(9), p.e05944. https://doi.org/10.1111/ecog.05944.

**Hekel, H.,** 1972. Pollen and spore assemblages from Queensland Tertiary sediments. Queensland Geological Survey, Brisbane. ISBN: 0724200568.

**Herrando-Moraira, S., Nualart, N., Galbany-Casals, M., Garcia-Jacas, N., Ohashi, H., Matsui, T., Susanna, A., Tang, C.Q. and López-Pujol, J.,** 2022. Climate Stability Index maps, a global high resolution cartography of climate stability from Pliocene to 2100. Scientific Data, 9(1), p.48. https://doi.org/10.1038/s41597-022-01144-5.

**Hill, R.S.,** 1994. History of the Australian vegetation: Cretaceous to Recent. Cambridge University Press. http://library.oapen.org/handle/20.500.12657/31444. ISBN: 9781925261479.

**Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. and Kimball, R.T.,** 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Molecular biology and evolution, 33(4), pp.1110-1125. https://doi.org/10.1093/molbev/msv347.

**Hörandl, E. and Appelhans, M.S.,** 2015. Next - Generation Sequencing in Plant Systematics. 2015. (Regnum Vegetabile, 158). 7 col. figs. 294 p. ISBN: 9783874294928.

**Howe, H.F. and Smallwood, J.,** 1982. Ecology of seed dispersal. Annual review of ecology and systematics, 13(1), pp.201-228. https://www.jstor.org/stable/2097067.

**Huang, H. and Knowles, L.L.,** 2016. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Systematic biology, 65(3), pp.357-365. https://doi.org/10.1093/sysbio/syu046.

**Hühn, P., Dillenberger, M.S., Gerschwitz-Eidt, M., Hörandl, E., Los, J.A., Messerschmid, T.F., Paetzold, C., Rieger, B. and Kadereit, G.,** 2022. How challenging RADseq data turned out to favor coalescent-based species tree inference. A case study in Aichryson (Crassulaceae). Molecular Phylogenetics and Evolution, 167, p.107342. https://doi.org/10.1016/j.ympev.2021.107342.

**Hühn, P., Dillenberger, M.S., Krause, S. and Kadereit, J.W.,** 2023. Polyploid hybrid speciation in the Calcarata species complex of Viola section Melanium (Violaceae): relating hybrid species to parent species distribution and ecology. Botanical Journal of the Linnean Society, 201(3), pp.309-328. https://doi.org/10.1093/botlinnean/boac056.

**Ilut, D.C., Nydam, M.L., Hare, M.P.,** 2014. Defining loci in restriction-based reduced representation genomic data from nonmodel species: sources of bias and diagnostics for optimal clustering. BioMed research international, 2014, 1–9. https://doi.org/10.1155/2014/675158.

**Ising, E.H.,** 1961. Bassia Uniflora R. Br. FvM. Chenopodiaceae and Allies in Australia. Transactions of the Royal Society of South Australia, 84, pp. 87-98.

**Ising, E.H.,** 1964. The species of Bassia All. (Chenopodiaceae) in Australia. Transactions of the Royal Society of South Australia, 88, pp. 63-110.

**Jacobs, S.W.L.,** 1988. Notes on Aizoaceae and Chenopodiacae. Telopea, 3(2), pp.139-143. https://biodiversitylibrary.org/page/60677704.

**Johnson, D.,** 2009. The geology of Australia. Cambridge University Press. https://doi.org/10.1017/CBO9781139194853.

**Jones, R. and Bowler, J.,** 1980. Struggle for the savanna: northern Australia in ecological and prehistoric perspective. In: Jones, R., (editor), Northern Australia: options and implications , Research School of Pacific Studies, Australian National University, Canberra, 3-32.

**Jordan, G.J.,** 1997. Evidence of Pleistocene plant extinction and diversity from Regatta Point, western Tasmania, Australia. Botanical Journal of the Linnean Society, 123(1), pp.45-71. https://doi.org/10.1111/j.1095-8339.1997.tb01404.x.

**Jurado, E., Westoby, M. and Nelson, D.,** 1991. Diaspore weight, dispersal, growth form and perenniality of central Australian plants. The Journal of Ecology, pp.811-828. https://doi.org/10.2307/2260669.

**Kadereit, G., Borsch, T., Weising, K. and Freitag, H.,** 2003. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis. International journal of plant sciences, 164(6), pp.959-986. https://doi.org/10.1086/378649.

**Kadereit, G., Gotzek, D. and Freitag, H.,** 2005. Origin and age of Australian Chenopodiaceae. Organisms Diversity & Evolution, 5(1), pp.59-80. https://doi.org/10.1016/j.ode.2004.07.002.

**Kadereit, G., Mavrodiev, E.V., Zacharias, E.H. and Sukhorukov, A.P.,** 2010. Molecular phylogeny of Atripliceae (Chenopodioideae, Chenopodiaceae): implications for systematics, biogeography, flower and fruit evolution, and the origin of C4 photosynthesis. American Journal of Botany, 97(10), pp.1664-1687. https://doi.org/10.3732/ajb.1000169.

**Kadereit, G. and Freitag, H.,** 2011. Molecular phylogeny of Camphorosmeae (Camphorosmoideae, Chenopodiaceae): Implications for biogeography, evolution of C4-photosynthesis and taxonomy. Taxon, 60(1), pp.51-78. https://doi.org/10.1002/tax.601006.

**Kadereit, G., Lauterbach, M., Pirie, M.D., Arafeh, R. and Freitag, H.,** 2014. When do different C4 leaf anatomies indicate independent C4 origins? Parallel evolution of C4 leaf types in Camphorosmeae (Chenopodiaceae). Journal of Experimental Botany, 65(13), pp.3499-3511. https://doi.org/10.1093/jxb/eru169.

**Keast, A.,** 1959. The Australian environment. In: Keast, A., Crocker, R.L., Christian, C.S. (Eds.), Biogeography and ecology in Australia. Springer.

**Kershaw, A.P., Martin, H.A. and Mason, J.M.,** 1994. 13 The Neogene: a period of transition. History of the Australian vegetation: Cretaceous to Recent, p.299. University of Adelaide Press. http://www.jstor.org/stable/10.20851/j.ctt1sq5wrv.17.

**Kisel, Y. and Barraclough, T.G.,** 2010. Speciation has a spatial scale that depends on levels of gene flow. The American Naturalist, 175(3), pp.316-334. https://doi.org/10.1086/650369.

**Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A.,** 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics, 35(21), pp.4453-4455. https://doi.org/10.1093/bioinformatics/btz305.

**Kubatko, L.S. and Degnan, J.H.,** 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic biology, 56(1), pp.17-24. https://doi.org/10.1080/10635150601146041.

**Kühn, U., Bittrich, V., Carolin, R., Freitag, H., Hedge, I.C., Uotila, P. and Wilson, P.G.,** 1993. Chenopodiaceae. Flowering Plants· Dicotyledons: Magnoliid, Hamamelid and Caryophyllid Families, pp.253-281. https://doi.org/10.1007/978-3-662-02899-5_26.

**Laity, J.J.,** 2009. Deserts and desert environments (Vol. 3). John Wiley & Sons. ISBN: 9781577180333.

**Lanier, H.C., Huang, H. and Knowles, L.L.,** 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Molecular Phylogenetics and Evolution, 70, pp.112-119. https://doi.org/10.1016/j.ympev.2013.09.006.

**Lee, K.M., Kivelä, S.M., Ivanov, V., Hausmann, A., Kaila, L., Wahlberg, N., Mutanen, M.,** 2018. Information dropout patterns in restriction site associated DNA phylogenomics and a comparison with multilocus Sanger data in a species-rich moth genus. Systematic Biology, 67(6):925–939. doi: 10.1093/sysbio/syy029.

**Leigh, J.H., Wilson, A.D. and Mulham, W.E.,** 1979. A study of sheep grazing a Belah (Casuarina cristata)–Rosewood (Heterodendrum oleifolium) shrub woodland in western New South Wales. Australian Journal of Agricultural Research, 30(6), pp.1225-1236. https://doi.org/10.1071/AR9791225.

**Leigh, J.H.,** 1994. Chenopod shrublands. In: Australian vegetation. Groves, R.H., (editor), pp. 345–367. Cambridge University Press. ISBN: 0521414202.

**Linley, G.D., Moseby, K.E. and Paton, D.C.,** 2016. Vegetation damage caused by high densities of burrowing bettongs (Bettongia lesueur) at Arid Recovery. Australian Mammalogy, 39(1), pp.33-41. https://doi.org/10.1071/AM15040.

**Liu, L., Xi, Z., Wu, S., Davis, C.C. and Edwards, S.V.,** 2015. Estimating phylogenetic trees from genome-scale data. Annals of the New York Academy of Sciences, 1360(1), pp.36-53. https://doi.org/10.1111/nyas.12747.

**Lomolino, M.V., Brown, J.H. and Sax, D.F.,** 2010. Island biogeography theory. The theory of island biogeography revisited, 13. Sinauer Associates. Sunderland, Massachusetts.

**Long, C., Kubatko, L.,** 2018. The effect of gene flow on coalescent-based species-tree inference. Systematic Biology, 67(5):770–785. doi: 10.1093/sysbio/syy020.

**Lopes dos Santos, R.A., De Deckker, P., Hopmans, E.C., Magee, J.W., Mets, A., Sinninghe Damsté, J.S. and Schouten, S.,** 2013. Abrupt vegetation change after the Late Quaternary megafaunal extinction in southeastern Australia. Nature Geoscience, 6(8), pp.627-631. https://doi.org/10.1038/ngeo1856.

**Luo, A., Ling, C., Ho, S.Y. and Zhu, C.D.,** 2018. Comparison of methods for molecular species delimitation across a range of speciation scenarios. Systematic Biology, 67(5), pp.830-846. https://doi.org/10.1093/sysbio/syy011.

**Mabbutt, J.A.,** 1969. Landforms of arid Australia. In: Arid Lands of Australia, editors: Slatyer, R.O., Perry, R.A. pp.11-32. ISBN: 9780708106938.

**Mabbutt, J.A.,** 1977. Desert Landform. Australian National University Press, Canberra. ISBN: 0708104371.

**Mabbutt, J.A.,** 1984. Landforms of the Australian deserts. Deserts and arid lands, pp.79-94. Remote Sensing of Earth Resources and Environment, vol 1. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-6080-0_4.

**Mabbutt, J.A.,** 1988. Australian desert landscapes. GeoJournal, 16, pp.355-369. https://doi.org/10.1007/BF00214394.

**Malcolm, B.,** 2000. Farm management economic analysis: a few disciplines, a few perspectives, a few figurings, a few futures. Agribusiness Perspectives 42. Agribusiness Association of Australia Inc.

**Martin, H.A.,** 1986. Cainozoic history of the vegetation and climate of the Lachlan River region, New South Wales. In Proceedings of the Linnean Society of New South Wales.

**Martin, H.A.,** 1993. The palaeovegetation of the Murray Basin, Late Eocene to Mid-Miocene. Australian Systematic Botany, 6(6), pp.491-531. https://doi.org/10.1071/SB9930491.

**Martin, H.A.,** 2000. Re-assignment of the affinities of the fossil pollen type Tricolpites trioblatus Mildenhall and Pocknall to Wilsonia (Convolvulaceae) and a reassessment of the ecological interpretations. Review of Palaeobotany and Palynology, 111(3-4), pp.237-251. https://doi.org/10.1016/S0034-6667(00)00027-0.

**Martin, H.A.,** 2006. Cenozoic climatic change and the development of the arid vegetation in Australia. Journal of Arid Environments, 66(3), pp.533-563. https://doi.org/10.1016/j.jaridenv.2006.01.009.

**Martin, H.A. and McMinn, A.,** 1994. Late Cainozoic vegetation history of north-western Australia, from the palynology of a deep sea core (ODP Site 765). Australian Journal of Botany, 42(1), pp.95-102. https://doi.org/10.1071/BT9940095.

**Martin, M.,** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17 (1), 10. https://doi.org/10.14806/ ej.17.110.14806/ej.17.1.200

**Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D. and Emerson, B.C.,** 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular ecology resources, 15(1), pp.28-41. https://doi.org/10.1111/1755-0998.12291.

**Mavromihalis, J.,** 2010a. National Recovery Plan for the Chariot Wheels, Maireana Cheelii. Department of Sustainability and Environment. ISBN: 9781742089591.

**Mavromihalis, J.,** 2010b. National Recovery Plan for the Turnip Copperburr Sclerolaena napiformis. Victorian Government Department of Sustainability and Environment. ISBN: 9781742089713.

**McCartney-Melstad, E., Gidiş, M. and Shaffer, H.B.,** 2019. An empirical pipeline for choosing the optimal clustering threshold in RADseq studies. Molecular ecology resources, 19(5), pp.1195-1204. https://doi.org/10.1111/1755-0998.13029.

**McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T.,** 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution 66 (2), 526–538. https://doi.org/10.1016/j.ympev.2011.12.007.

**McDonald, J.T.,** 2020. Biogeography of Australian chenopods: landscape in the evolution of an arid flora (Doctoral dissertation). https://hdl.handle.net/2440/126029.

**McGowran, B., Holdgate, G.R., Li, Q. and Gallagher, S.J.,** 2004. Cenozoic stratigraphic succession in southeastern Australia. Australian Journal of Earth Sciences, 51(4), pp.459-496. https://doi.org/10.1111/j.1400-0952.2004.01078.x.

**McKain, M.R., Johnson, M.G., Uribe-Convers, S., Eaton, D. and Yang, Y.,** 2018. Practical considerations for plant phylogenomics. Applications in plant sciences, 6(3), p.e1038. https://doi.org/10.1002/aps3.1038.

**McKenzie, N., Jacquier, D., Isbell, R. and Brown, K., 2004.** Australian soils and landscapes: an illustrated compendium. CSIRO publishing. ISBN: 0643069585.

**Messerschmid, T.F., Abrahamczyk, S., Bañares Baudet, Á., Brilhante, M.A., Eggli, U., Hühn, P., Kadereit, J.W., Dos Santos, P., de Vos, J.M. and Kadereit, G.,** 2023. Inter-and intra-island speciation and their morphological and ecological correlates in Aeonium (Crassulaceae), a species-rich Macaronesian radiation. Annals of Botany. https://doi.org/10.1093/aob/mcad033.

**Milewski, A.V.,** 1982. The occurrence of seeds and fruits taken by ants versus birds in mediterranean Australia and southern Africa, in relation to the availability of soil potassium. Journal of Biogeography, pp.505-516. https://doi.org/10.2307/2844617.

**Miller, G.H., Magee, J.W., Johnson, B.J., Fogel, M.L., Spooner, N.A., McCulloch, M.T. and Ayliffe, L.K.,** 1999. Pleistocene extinction of Genyornis newtoni: human impact on Australian megafauna. Science, 283(5399), pp.205-208. https://doi.org/10.1126/science.283.5399.205.

**Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. and Johnson, E.A.,** 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome research, 17(2), pp.240-248. http://www.genome.org/cgi/doi/10.1101/gr.5681207.

**Mirarab, S., Bayzid, M.S., Warnow, T.,** 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Systematic Biology, 65 (3), 366–380. https://doi.org/10.1093/sysbio/syu063.

**Mokany, K., McCarthy, J.K., Falster, D.S., Gallagher, R.V., Harwood, T.D., Kooyman, R. and Westoby, M.,** 2022. Patterns and drivers of plant diversity across Australia. Ecography, 2022(11), p.e06426. https://doi.org/10.1111/ecog.06426.

**Molloy, E.K. and Warnow, T.,** 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. Systematic Biology, 67(2), pp.285-303. https://doi.org/10.1093/sysbio/syx077.

**Morales-Briones, D.F., Kadereit, G., Tefarikis, D.T., Moore, M.J., Smith, S.A., Brockington, S.F., Timoneda, A., Yim, W.C., Cushman, J.C. and Yang, Y.,** 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae sl. Systematic Biology, 70(2), pp.219-235. https://doi.org/10.1093/sysbio/syaa066.

**Morton, S.R., Smith, D.S., Dickman, C.R., Dunkerley, D.L., Friedel, M.H., McAllister, R.R.J., Reid, J.R.W., Roshier, D.A., Smith, M.A., Walsh, F.J. and Wardle, G.M.,** 2011. A fresh framework for the ecology of arid Australia. Journal of Arid Environments, 75(4), pp.313-329. https://doi.org/10.1016/j.jaridenv.2010.11.001.

**Mueller, F.J.H.,** 1882a. Fragmenta Phytographiae Australiae 12:12-14, Melbourne, Auctoritate Gubern. Coloniae Victoriae, Ex Officina Joannis Ferres. https://doi.org/10.5962/bhl.title.287.

**Mueller, F.J.H.,** 1882b. Systematic Census of Australian Plants. p. 30, 140. Melbourne, Printed for the Victorian Government by M'Carron, Bird & Co. https://doi.org/10.5962/bhl.title.54034.

**Müller, K. and Borsch, T.,** 2005. Phylogenetics of Amaranthaceae based on matK/trnK sequence data: evidence from parsimony, likelihood, and Bayesian analyses. Annals of the Missouri Botanical Garden, pp.66-102. http://www.jstor.org/stable/3298649.

**Nathan, R., Katul, G.G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W. and Levin, S.A.,** 2002. Mechanisms of long-distance dispersal of seeds by wind. Nature, 418(6896), pp.409-413. https://doi.org/10.1038/nature00844.

**Norman, H.C., Wilmot, M.G., Rintoul, A.J., Hulm, E., Revell, D.K.,** 2010. Nutritive value of 5 genera of West Australian chenopods for livestock. genus, 8(9), p.8.

**Paczkowska, G. and Chapman, A.R.,** 2000. The Western Australian flora: a descriptive catalogue. Wildflower Society of Western Australia. Perth, WA. ISBN: 0646402439.

**Parchman, T.L., Jahner, J.P., Uckele, K.A., Galland, L.M., Eckert, A.J.,** 2018. RADseq approaches and applications for forest tree genetics. Tree Genetics & Genomes. 14 (3), 39. https://doi.org/10.1007/s11295-018-1251-3.

**Paris, J.R., Stevens, J.R. and Catchen, J.M.,** 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution, 8(10), pp.1360-1373. https://doi.org/10.1111/2041-210X.12775.

**Parr-Smith, G.A.,** 1982. Biogeography and evolution in the shrubby Australian species of Atriplex (Chenopodiaceae). In: Evolution of the flora and fauna of arid Australia, Barker, W.R. and Greensdale, P.J.M. Peacock Publications, South Australia. ISBN: 0909209626.

**Peakall, R. and Beattie, A.J.,** 1995. Does ant dispersal of seeds in Sclerolaena diacantha (Chenopodiaceae) generate local spatial genetic structure?. Heredity, 75(4), pp.351-361. https://doi.org/10.1038/hdy.1995.146.

**Peguero-Pina, J.J., Vilagrosa, A., Alonso-Forn, D., Ferrio, J.P., Sancho-Knapik, D. and Gil-Pelegrín, E.,** 2020. Living in drylands: Functional adaptations of trees and shrubs to cope with high temperatures and water scarcity. Forests, 11(10), p.1028. https://doi.org/10.3390/f11101028.

**Pillans, B.J.,** 2018. Seeing red: some aspects of the geological and climatic history of the Australian arid zone. On the ecology of Australia's arid zone, pp.5-43.

**Primack, R.B.,** 1987. Relationships among flowers, fruits, and seeds. Annual review of ecology and systematics, 18(1), pp.409-430. https://doi.org/10.1146/annurev.es.18.110187.002205.

**Puillandre, N., Modica, M.V., Zhang, Y., Sirovich, L., Boisselier, M.C., Cruaud, C., Holford, M. and Samadi, S.,** 2012. Large-scale species delimitation method for hyperdiverse groups. Molecular ecology, 21(11), pp.2671-2691. https://doi.org/10.1111/j.1365-294X.2012.05559.x.

**Quilty, P.G.,** 1994. The background: 144 million years of Australian palaeoclimate and palaeogeography. History of the Australian vegetation: Cretaceous to Recent, 14, p.43. https://www.jstor.org/stable/10.20851/j.ctt1sq5wrv.7.

**Rambaut, A.,** 2018. FigTree, version 1.4.4. Program distributed by the author. http://tree.bio.ed.ac.uk/software/figtree.

**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.

**Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M.R., Elmer, K.R., Vences, M. and Steinfarz, S.,** 2019. Phylogeny and species delimitation of near Eastern Neurergus newts

(Salamandridae) based on genome-wide RADseq data analysis. Molecular Phylogenetics and Evolution, 133, pp.189-197. https://doi.org/10.1016/j.ympev.2019.01.003.

**Rannala, B., Edwards, S.V., Leach´e, A., Yang, Z.,** 2020. The Multispecies Coalescent Model and Species Tree Inference. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 3.3, pp. 3.3:1–3.3:21. No commercial publisher. Authors open access book. The book is freely available at https://hal.inria.fr/PGE. HAL Id: hal-02535622.

**Räsänen, K. and Hendry, A.P.,** 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. Ecology letters, 11(6), pp.624-636. https://doi.org/10.1111/j.1461-0248.2008.01176.x.

**Ree, R.H. and Hipp, A.L.,** 2015. Inferring phylogenetic history from restriction site associated DNA (RADseq). Next-generation sequencing in plant systematics, 181, p.204. In: Horandl, E., Appelhans, M.S. (Eds.), Regnum Vegetabile Vol. 158: Next-Generation Sequencing in Plant Systematics. Koeltz Scientific Books, Oberreifenberg, pp. 181–204.

**Renner, M.A., Foster, C.S., Miller, J.T. and Murphy, D.J.,** 2020. Increased diversification rates are coupled with higher rates of climate space exploration in Australian Acacia (Caesalpinioideae). New Phytologist, 226(2), pp.609-622. https://doi.org/10.1111/nph.16349.

**Revell, D.K., Norman, H.C., Vercoe, P.E., Phillips, N., Toovey, A., Bickell, S., Hulm, E., Hughes, S. and Emms, J.,** 2013. Australian perennial shrub species add value to the feed base of grazing livestock in low-to medium-rainfall zones. Animal Production Science, 53(11), pp.1221-1230. https://doi.org/10.1071/AN13238.

**Ridley, H.,** 1930. The dispersal of plants throughout the world. Reeve and Ashford, Kent, United Kingdom.

**Rij, L.M., Grundler, S., Kumar, K., Hühn, P., Cousins, P., Rühl, E., Schmid, J., Kadereit, G. and Frotscher, J.,** 2019, June. Classification of a Texan wild grapevine population within the genus Vitis. In XV EUCARPIA Symposium on Fruit Breeding and Genetics 1307 (pp. 131-140). https://doi.org/10.17660/ActaHortic.2021.1307.21.

**Roch, S., Steel, M.,** 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Populat. Biol. 100, 56–62. https://doi.org/10.1016/j.tpb.2014.12.005.

**Roch, S., Warnow, T.,** 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Systematic Biology, 64 (4), 663–676. https://doi.org/10.1093/sysbio/syv016.

**Rogers, R.W., Butler, D. and Carnell, J.,** 1993. Dispersal of germinable seeds by emus in semi-arid Queensland. Emu-Austral Ornithology, 94(2), pp.132-134. https://doi.org/10.1071/MU9940132.

**Rubin, B.E.R., Ree, R.H., Moreau, C.S., Kolokotronis, S.O.,** 2012. Inferring phylogenies from RAD sequence data. PLoS ONE, 7 (4), e33394. https://doi.org/10.1371/journal. pone.0033394.

**Saintilan, N.,** 2009. Biogeography of Australian saltmarsh plants. Austral Ecology, 34(8), 929-937. https://doi.org/10.1111/j.1442-9993.2009.02001.x.

**Sayyari, E., Whitfield, J.B., Mirarab, S.,** 2017. Fragmentary gene sequences negatively impact gene tree and species tree reconstruction. Molecular Biology and Evolution, 34 (12), 3279–3291. https://doi.org/10.1093/molbev/msx261

**Scott, A.J.,** 1978. A revision of the Camphorosmioideae (Chenopodiaceae). Feddes Repertorium, 89(2-3), pp.101-119. https://doi.org/10.1002/fedr.19780890202.

**Sharp, G., Jannusch, N. and Stone, H.,** 2009. Management of the vulnerable plant species, Sclerolaena walkeri (C. White) AJ Scott in a remote electricity corridor, Currawinya National Park. Ecological Management & Restoration, 10(2), pp.158-166. https://doi.org/10.1111/j.1442-8903.2009.00480.x.

**Shepherd, K.A., Waycott, M. and Calladine, A.,** 2004. Radiation of the Australian Salicornioideae (Chenopodiaceae)—based on evidence from nuclear and chloroplast DNA sequences. American Journal of Botany, 91(9), pp.1387-1397. https://doi.org/10.3732/ajb.91.9.1387.

**Shepherd, K.A., Macfarlane, T.D. and Waycott, M.,** 2005. Phylogenetic analysis of the Australian Salicornioideae (Chenopodiaceae) based on morphology and nuclear DNA. Australian Systematic Botany, 18(1), pp.89-115. https://doi.org/10.1071/SB04031.

**Shmida, A.,** 1985. Biogeography of the desert floras of the world. In: "Hot deserts and arid shrublands." (Editors: Evenari, M., Noy-Meir, I., Goodall, D.W.), Vol. 12A pp. 23-77. Elsevier, Amsterdam.

**Silcock, J.L. and Fensham, R.J.,** 2018. Using evidence of decline and extinction risk to identify priority regions, habitats and threats for plant conservation in Australia. Australian Journal of Botany, 66(7), pp.541-555. https://doi.org/10.1071/BT18056.

**Song, L., Chen, S., Chen, W. and Chen, X.,** 2017. Distinct impacts of two types of La Niña events on Australian summer rainfall. International Journal of Climatology, 37(5), pp.2532-2544. https://doi.org/10.1002/joc.4863.

**Sorensen, A.E.,** 1986. Seed dispersal by adhesion. Annual Review of Ecology and Systematics, 17(1), pp.443-463. https://doi.org/10.1146/annurev.es.17.110186.002303.

**Specht, R.L. and Specht, A.,** 1999. Australian plant communities: Dynamics of structure, growth and biodiversity. Oxford University Press. ISBN: 0195516540.

**Springer, M.S. and Gatesy, J.,** 2016. The gene tree delusion. Molecular phylogenetics and evolution, 94, pp.1-33. https://doi.org/10.1016/j.ympev.2015.07.018.

**Springer, M.S. and Gatesy, J.,** 2018. On the importance of homology in the age of phylogenomics. Systematics and biodiversity, 16(3), pp.210-228. https://doi.org/10.1080/14772000.2017.1401016.

**Squires, H., Priest, M., Sluiter, I. and Loch, R.,** 2012, September. Leading practice waste dump rehabilitation at the Ginkgo mineral sands mine. In Mine Closure 2012: Proceedings of

the Seventh International Conference on Mine Closure (pp. 59-71). Australian Centre for Geomechanics. https://doi.org/10.36487/ACG_rep/1208_07_Loch_Squires.

**Stace, H.C.T., Hubble, G.D., Brewer, R., Northcote, K.H., Sleeman, J.R., Mulcahy, M.J. and Hallsworth, E.G.,** 1968. A handbook of Australian soils, Rellim Technical Publications, Glenside, South Australia.

**Stafford, S.D.,** 1990. A framework for the ecology of arid Australia. Journal of Arid Environments, 18, pp.255-278.

**Strauss, S.Y. and Agrawal, A.A.,** 1999. The ecology and evolution of plant tolerance to herbivory. Trends in ecology & evolution, 14(5), pp.179-185. https://doi.org/10.1016/S0169-5347(98)01576-6.

**Strauss, S.Y. and Irwin, R.E.,** 2004. Ecological and evolutionary consequences of multispecies plant-animal interactions. Annu. Rev. Ecol. Evol. Syst., 35, pp.435-466. https://doi.org/10.1146/annurev.ecolsys.35.112202.130215.

**Suppiah, R., Collins, D. and Della-Marta, P.,** 2001. Observed changes in Australian climate. Bureau of Meteorology and CSIRO Division of Atmospheric Research Tech. Rep, 6.

**Thornhill, A.H., Crisp, M.D., Külheim, C., Lam, K.E., Nelson, L.A., Yeates, D.K. and Miller, J.T.,** 2019. A dated molecular perspective of eucalypt taxonomy, evolution and diversification. Australian Systematic Botany, 32(1), pp.29-48. https://doi.org/10.1071/SB18015.

**Timbal, B. and Drosdowsky, W.,** 2013. The relationship between the decline of Southeastern Australian rainfall and the strengthening of the subtropical ridge. International Journal of Climatology, 33(4), pp.1021-1034. https://doi.org/10.1002/joc.3492.

**Tobias, J.A., Seddon, N., Spottiswoode, C.N., Pilgrim, J.D., Fishpool, L.D. and Collar, N.J.,** 2010. Quantitative criteria for species delimitation. Ibis, 152(4), pp.724-746. https://doi.org/10.1111/j.1474-919X.2010.01051.x.

**Toon, A., Crisp, M.D., Gamage, H., Mant, J., Morris, D.C., Schmidt, S. and Cook, L.G.,** 2015. Key innovation or adaptive change? A test of leaf traits using Triodiinae in Australia. Scientific Reports, 5(1), pp.1-12. doi: 10.1038/srep12398.

**Traveset, A., Heleno, R. and Nogales, M.,** 2014. The ecology of seed dispersal. In: Seeds: the ecology of regeneration in plant communities (pp. 62-93). Wallingford UK: CABI. https://doi.org/10.1079/9781780641836.0062.

**Tripp, E.A., Tsai, Y.H.E., Zhuang, Y. and Dexter, K.G.,** 2017. RAD seq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia. Ecology and Evolution, 7(19), pp.7920-7936. https://doi.org/10.1002/ece3.3274.

**Twidale, C.R., Wopfner, H.,** 1990. Dunefields. In: Tyler, M.J., Twidale, C.R., Davies, M. and Wells, C.B., 1990. Natural History of the North East Deserts. Royal Society of South Australia Inc., Adelaide. ISBN: 0959662758.

**Ulbrich, E.,** 1934. Chenopodiaceae. Pages 379–584 in: Engler, A., Prantl, K., Pax, F.A., Harms, H.A.T., Die Natürlichen Pflanzenfamilien. Vol 16c. Engelmann, Leipzig. https://doi.org/10.5962/bhl.title.4635.

**Valliere, J.M., Ruscalleda Alvarez, J., Cross, A.T., Lewandrowski, W., Riviera, F., Stevens, J.C., Tomlinson, S., Tudor, E.P., Wong, W.S., Yong, J.W. and Veneklaas, E.J.,** 2022. Restoration ecophysiology: an ecophysiological approach to improve restoration strategies and outcomes in severely disturbed landscapes. Restoration Ecology, 30, p.e13571. https://doi.org/10.1111/rec.13571.

**Van de Graaff, W.E., Crowe, R.W.A., Bunting, J.A., Jackson, M.J.,** 1977. Relict early Cainozoic drainages in arid Western Australia. Zeitschrift für Geomorphologie, 21(4), pp.379-400.

**Van Der Kaars, S., Miller, G.H., Turney, C.S., Cook, E.J., Nürnberg, D., Schönfeld, J., Kershaw, A.P. and Lehman, S.J.,** 2017. Humans rather than climate the primary cause of Pleistocene megafaunal extinction in Australia. Nature Communications, 8(1), pp.1-7. https://doi.org/10.1038/ncomms14142.

**Van der Pijl, L.,** 1982. Principles of dispersal in higher plants (Vol. 214). Berlin: Springer-Verlag. https://doi.org/10.1007/978-3-642-96108-3.

**Vander Wall, S.B. and Beck, M.J.,** 2012. A comparison of frugivory and scatter-hoarding seed-dispersal syndromes. The Botanical Review, 78, pp.10-31. https://doi.org/10.1007/s12229-011-9093-9.

**van Etten, E.J. and Burrows, N.D.,** 2018. On the Ecology of Australia's Arid Zone: 'Fire Regimes and Ecology of Arid Australia'. On the ecology of Australia's arid zone, pp.243-282. https://doi.org/10.1007/978-3-319-93943-8_10.

**Walker, J.F., Yang, Y., Feng, T., Timoneda, A., Mikenas, J., Hutchison, V., Edwards, C., Wang, N., Ahluwalia, S., Olivieri, J. and Walker-Hale, N.,** 2018. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. American Journal of Botany, 105(3), pp.446-462. https://doi.org/10.1002/ajb2.1069.

**Walsh, N.G.,** 1996. Flora of Victoria: Dicotyledons, Winteraceae to Myrtaceae. Edited by Walsh, N.G., and Entwisle, T.J. Inkata Press, Melbourne. ISBN: 0409308528.

**Ward, P.S., Brady, S.G., Fisher, B.L. and Schultz, T.R.,** 2010. Phylogeny and biogeography of dolichoderine ants: effects of data partitioning and relict taxa on historical inference. Systematic Biology, 59(3), pp.342-362. https://doi.org/10.1093/sysbio/syq012.

**Wariss, H.M. and Qu, X.J.,** 2021. The complete chloroplast genome of Chenopodium acuminatum Willd. (Amaranthaceae). Mitochondrial DNA Part B, 6(1), pp.174-175. https://doi.org/10.1080/23802359.2020.1860716.

**Webb, J.A. and James, J.M.,** 2006. Karst evolution of the Nullarbor plain, Australia. https://doi.org/10.1130/2006.2404(07).

**Wenny, D.G.,** 2001. Advantages of seed dispersal: a re-evaluation of directed dispersal. Evolutionary Ecology Research, 3(1), pp.37-50.

**Wiens, J.J. and Morrill, M.C.,** 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Systematic biology, 60(5), pp.719-731. https://doi.org/10.1093/sysbio/syr025.

**Williams, M., Dunkerley, D., De Deckker, P., Kershaw, P. and Chappell, J.,** 1998. Quaternary Environments, 2nd edition. Arnold Publishers, London. ISBN: 0340691514.

**Wilson, P.G.,** 1975. A taxonomic revision of the genus Maireana (Chenopodiaceae). Nuytsia, 2(1), pp.2-83. https://biostor.org/reference/222856.

**Wilson, P.G.,** 1984. Chenopodiaceae. Flora of Australia, 4, Phytolaccaceae to Chenopodiaceae, pp.81-317.

**Whitford, W.G.,** 2002. Ecology of desert systems. Academic Press, San Diego, California. ISBN: 0127472614.

**Wu, S., You, F., Hall, M. and Huang, L.,** 2021. Native plant Maireana brevifolia drives prokaryotic microbial community development in alkaline Fe ore tailings under semi-arid climatic conditions. Science of The Total Environment, 760, p.144019. https://doi.org/10.1016/j.scitotenv.2020.144019.

**Xi, Z., Liu, L., Davis, C.C.,** 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Molecular Phylogenetics and Evolution, 92, 63–71. https://doi.org/10.1016/j.ympev.2015.06.009.

**Xi, Z., Liu, L., Davis, C.C.,** 2016. The impact of missing data on species tree estimation. Moleculra Biology and Evolultion, 33 (3), 838–860. https://doi.org/10.1093/molbev/msv266.

**Xu, B., Yang, Z.,** 2016. Challenges in species tree estimation under the multispecies coalescent model. Genetics, 204 (4), 1353–1368. https://doi.org/10.1534/genetics.116.190173.

**Yang, X. and Goudie, A.,** 2007. Geomorphic processes and palaeoclimatology in deserts. Quaternary International, 1(175), pp.1-2. doi:10.1016/j.quaint.2007.06.021.

**Zerdoner Calasan, A., Hammen, S., Sukhorukov, A.P., McDonald, J.T., Brignone, N.F., Böhnert, T. and Kadereit, G.,** 2022. From continental Asia into the world: Global historical biogeography of the saltbush genus Atriplex (Chenopodieae, Chenopodioideae, Amaranthaceae). Perspectives in Plant Ecology, Evolution and Systematics, 54, p.125660. https://doi.org/10.1016/j.ppees.2022.125660.

**Zheng, Y. and Wiens, J.J.,** 2015. Do missing data influence the accuracy of divergence-time estimation with BEAST?. Molecular Phylogenetics and Evolution, 85, pp.41-49. https://doi.org/10.1016/j.ympev.2015.02.002.

**Zhong, H., Zhou, J., Wong, W.S., Cross, A. and Lambers, H.,** 2021. Exceptional nitrogen-resorption efficiency enables Maireana species (Chenopodiaceae) to function as pioneers at a mine-restoration site. Science of the Total Environment, 779, p.146420. https://doi.org/10.1016/j.scitotenv.2021.146420.

**Zhu, G.L., Mosyakin, S.L. and Clemants, S.E.,** 2003. Chenopodiaceae. Flora of China, 5, pp.351-414. Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis. ISBN: 193072327X.

## *7. LIST OF TABLES*



## *8. LIST OF FIGURES*

## *9. IMAGE SOURCE AND COPYRIGHT INFORMATION ON FIGURES*

**Figure 1**: **A**: Desert Lake, **B**: Open Eucalyptus Woodland, **E**: *Maireana lobiflora*, **I**: Sand dune, **J**: Stony Desert, **M**: *Enchylaena tomentosa*, **O**: *Eriochiton sclerolaenoides*, **P**: *Malacocera tricornis*, **Q**: Murray River, **U**: *Roycea spinescens* and **V**: *Roycea pycnophylloides* by Philipp Hühn, Mainz. **C**: *Maireana atkinsiana*, **D**: *Maireana georgei*, **F**: *Sclerolaena burbidgeae*, **G**: *Sclerolaena eurotioides*, **H**: *Sclerolaena alata*, **K**: *Didymanthus roei* and **N**: *Eremophea aggregata* by Christopher J. French, Perth. **L**: *Dissocarpus paradoxus* by Mark Marathon, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Dissocarpus_fruit.jpg. **R**: Nullarbor Plain by State of Western Australia, Department of Primary Industries and Regional Development, Copyright Act 1968, https://www.agric.wa.gov.au/rangelands/rangeland-inventory-and-condition-survey-nullarbor-region-western-australia. **S**: *Neobassia proceriflora* by Kym Nicolson, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Neobassia_proceriflora.jpg. **T**: *Osteocarpum dipterocarpum* by Kym Nicolson, CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/>, via iNaturalist, https://www.inaturalist.org/observations/108562091. **W**: *Threlkeldia diffusa* by Sue Jaggar, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0/>, via iNaturalist, https://www.inaturalist.org/observations/70989013.

**Figure 2**: **A**: *Sclerolaena decurrens* by D. Whibley, Flora of Australia (1984), Fig. 48 V; **B**: *Sclerolaena alata* by A. Ashby, Flora of Australia (1984), Fig. 48 I; **C**: *Malacocera gracilis* by L. Dutkiewicz, Flora of Australia (1984), Fig. 38 C; **D**: *Dissocarpus biflorus* by E. McBarron, Flora of Australia (1984), Fig. 40 H; **E**: *Eremophea aggregata* by R. Cranfield, Flora of Australia (1984), Fig. 39 B; **F**: *Enchylaena tomentosa* by P. Wilson, Flora of Australia (1984), Fig. 36 B; **G**: *Maireana georgei* by P. Wilson, Flora of Australia (1984), Fig. 35 H; **H**: *Roycea spinescens* by P. Wilson, Flora of Australia (1984), Fig. 37 F. Copyright: Commonwealth of Australia, 1984.

**Figure 3**: **A**: *Osteocarpum dipterocarpum* by F. Hilton, Flora of Australia (1984), Fig. 41 F; **B**: *Neobassia proceriflora* by H. Mohr, Cabrera et al. (2009), Fig. 4 G; **C**: *Sclerolaena fimbriolata* by A. George, Flora of Australia (1984), Fig. 46 C; **D**: *Maireana ciliata* by M. Koch, Flora of Australia (1984), Fig. 33 L; **E**: *Malacocera gracilis* by L. Dutkiewicz, Flora of Australia (1984), Fig. 38 C; **F**: *Maireana dichoptera* by E. Schneider, Flora of Australia (1984), Fig. 33 O; **G**: *Maireana ovata* by P. Wilson, Flora of Australia (1984), Fig. 34 R; **H**: *Maireana turbinata* by P. Wilson, Flora of Australia (1984), Fig. 33 X; **I**: *Enchylaena lanata* by P. Wilson, Flora of Australia (1984), Fig. 36 E; **J**: *Maireana spongiocarpa* by N. Lothian, Flora of Australia (1984), Fig. 34 B; **K**: *Maireana aphylla* by J. Jessop, Flora of Australia (1984), Fig. 34 Z; **L**: *Sclerolaena stelligera* by H. Mohr, Cabrera et al. (2009), Fig. 5 D; **M**: *Maireana pentatropis* by J. Jessop, Flora of Australia (1984), Fig. 34 G; **N**: *Maireana erioclada* by A. George, Flora of Australia (1984), Fig. 34 F; **O**: *Maireana excavata* by R. Wallace, Flora of Australia (1984), Fig. 33 S; **P**: *Maireana brevifolia* by T. Aplin, Flora of Australia (1984), Fig. 33 F; **Q**: *Maireana carnosa* by A. Ashby, Flora of Australia (1984), Fig. 33 Q; **R**: *Maireana lobiflora* by R. Helms, Flora of Australia (1984), Fig. 33 J; **S**: *Roycea spinescens* by P. Wilson, Flora of Australia (1984), Fig. 37 F; **T**: *Maireana oppositifolia* by P. Wilson, Flora of Australia (1984), Fig. 33 H. Copyright: Commonwealth of Australia, 1984

**Figure 4**: **A**: *Sclerolaena burbidgeae* by A. Oliver, Flora of Australia (1984), Fig. 47 B; **B**: *Sclerolaena parallelicuspis* by E. Ising, Flora of Australia (1984), Fig. 47 E; **C**: *Sclerolaena decurrens* by D. Whibley, Flora of Australia (1984), Fig. 48 V; **D**: *Sclerolaena divaricata* by E. Constable, Flora of Australia (1984), Fig. 47 J; **E**: *Sclerolaena cuneata* by A. Oliver, Flora of Australia (1984), Fig. 47 K; **F**: *Sclerolaena glabra* by A. Beauglehole, Flora of Australia (1984), Fig. 47 L; **G**: *Sclerolaena ventricosa* by M. Dodson, Flora of Australia (1984), Fig. 48 O; **H**: *Threlkeldia diffusa* by H. Mohr, Cabrera et al. (2009), Fig. 4 B; **I**: *Sclerolaena fusiformis* by A. Oliver, Flora of Australia (1984), Fig. 46 H; **J**: *Sclerolaena bicornis* by N. Ford, Flora of Australia (1984), Fig. 48 T; **K**: *Sclerolaena alata* by A. Ashby, Flora of Australia (1984), Fig. 48 I; **L**: *Sclerolaena convexula* by E. Constable, Flora of Australia (1984), Fig. 48 B; **M**: *Sclerolaena napiformis* by E. D'Arnay, Flora of Australia (1984), Fig. 48 K; **N**: *Osteocarpum acropterum* by C. Moore, Flora of Australia (1984), Fig. 41 D; **O**: *Osteocarpum salsuginosum* by D. Symon, Flora of Australia (1984), Fig. 41 G; **P**: *Neobassia proceriflora* by H. Mohr, Cabrera et al. (2009), Fig. 4 G; **Q**: *Dissocarpus fontinalis* by A. Williams, Flora of Australia (1984), Fig. 40 B; **R**: *Sclerolaena fimbriolata* by A. George, Flora of Australia (1984), Fig. 46 C. Copyright: Commonwealth of Australia, 1984

**Figure 5 and 6**: Timeline of landscape (habitat) history in Australia since the Middle Miocene based on McDonald (2020), Fig. 3.1.

**Figure 7**: Timeline of landscape (habitat) history in Australia since the Middle Miocene based on McDonald (2020), Fig. 3.1. Tectonic shift and movement of the Sub-Tropical High Pressure system taken from Fujioka and Chappell (2010), Fig. 2, who based their illustration on Bowler (1982). Cartoon Kangaroo vector by Ery Prihananto, free vector graphic available at < https://www.uidownload.com/>, < https://www.uidownload.com/en/vector-jgslh>. Cartoon ant vector by Brian Goff, Published under the "free licence" by Vecteezy: < https://support.vecteezy.com/en_us/new-vecteezy-licensing-ByHivesvt>, < https://de.vecteezy.com/vektorkunst/546797-cartoon-ant-insektenwanze>. Changes in sea level based on Martin (2006), Fig. 3 and Byrne et al. (2008), Fig. 1. Changes in vegetation composition based on Martin (2006) and Fujioka and Chappell (2010), Fig. 5. Changes in climate based on Byrne et al. (2008), Fig. 1 and Fujioka and Chappell (2010), Fig. 5.

# *10. APPENDIX AND SUPPLEMENTARY MATERIAL*

## *Appendix*



## *Electronic supplementary material*

[see USB flash drive attached at the back endpaper of this book]

Supplement 1: Sampling and Assemblies.

Supplement 2: Vegetative and reproductive morphological characteristics, soil type and salinity preferences, colonized habitats and distributions of the Australian Camphorosmeae.

Supplement 3: CT selection and assembly statistics.

Supplement 4: Concatenated supermatrices, ipyrad statistics, tree-files and CA-ML phylogenies showing all BS support values of assemblies 1-4, and the results of the BEAST analysis (Assembly 4).

## *Sequence data*

The demultiplexed and trimmed RADseq sequence data are stored at Dryad:

Hühn, Philipp (2023), Phylogenetics and Evolution of the Australian Camphorosmeae (Amaranthaceae), Dryad, Dataset. DOI: https://doi.org/10.5061/dryad.j6q573njz.

Download link:

https://datadryad.org/stash/share/2cO4w5lse9Y4QNnm4Mu7obNhwVF0xaLUzhjThDkeF00.

Molecular Phylogenetics and Evolution 167 (2022) 107342

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# How challenging RADseq data turned out to favor coalescent-based species tree inference. A case study in Aichryson (Crassulaceae)

Philipp Hühn [a,b], Markus S. Dillenberger [b,c], Michael Gerschwitz-Eidt [b], Elvira Hörandl [d], Jessica A. Los [a,e], Thibaud F.E. Messerschmid [a,b,e], Claudia Paetzold [d,f], Benjamin Rieger [b], Gudrun Kadereit [e,*]

[a] Institute of Molecular Physiology (iMP), Johannes Gutenberg-University Mainz, Germany
[b] Institute of Organismic and Molecular Evolution (iOME), Johannes Gutenberg-University Mainz, Germany
[c] Institut für Biologie, Systematische Botanik und Pflanzengeographie, Freie Universität Berlin, Germany
[d] Department of Systematics, Biodiversity and Evolution of Plants, Georg-August-University Göttingen, Germany
[e] Prinzessin Therese von Bayern Lehrstuhl für Systematik, Biodiversität & Evolution der Pflanzen, Ludwig-Maximilians-Universität München, Germany
[f] Department of Botany and Molecular Evolution, Senckenberg Research Institute and Natural History Museum Frankfurt am Main, Germany

ARTICLE INFO

*Keywords:*
[clustering threshold selection
coalescent-based summary method
data bias
locus filtering
RADseq
species tree inference]

ABSTRACT

Analysing multiple genomic regions while incorporating detection and qualification of discordance among regions has become standard for understanding phylogenetic relationships. In plants, which usually have comparatively large genomes, this is feasible by the combination of reduced-representation library (RRL) methods and high-throughput sequencing enabling the cost effective acquisition of genomic data for thousands of loci from hundreds of samples. One popular RRL method is RADseq. A major disadvantage of established RADseq approaches is the rather short fragment and sequencing range, leading to loci of little individual phylogenetic information. This issue hampers the application of coalescent-based species tree inference. The modified RADseq protocol presented here targets ca. 5,000 loci of 300-600nt length, sequenced with the latest short-read-sequencing (SRS) technology, has the potential to overcome this drawback. To illustrate the advantages of this approach we use the study group *Aichryson* Webb & Berthelott (Crassulaceae), a plant genus that diversified on the Canary Islands. The data analysis approach used here aims at a careful quality control of the long loci dataset. It involves an informed selection of thresholds for accurate clustering, a thorough exploration of locus properties, such as locus length, coverage and variability, to identify potential biased data and a comparative phylogenetic inference of filtered datasets, accompanied by an evaluation of resulting BS support, gene and site concordance factor values, to improve overall resolution of the resulting phylogenetic trees. The final dataset contains variable loci with an average length of 373nt and facilitates species tree estimation using a coalescent-based summary approach. Additional improvements brought by the approach are critically discussed.

## 1. Introduction

Resolving phylogenetic relationships of recently and rapidly radiating species complexes is a challenge because first, standard markers using universal primers are too conserved and fail to provide sufficient information, and second, inferring relationships is often complicated due to incomplete lineage sorting (ILS), hybridization/introgression and gene duplication/loss events (Pamilo and Nei, 1988; Maddison, 1997;

*Abbreviations:* BSC, between-sample-clustering; CA-ML, maximum likelihood analysis of concatenated loci; CB-SM, coalescent-based summary method; CT, clustering threshold; gCF, gene concordance factor; GTEE, gene tree estimation error; HTS, high throughput sequencing; ILS, incomplete lineage sorting; ISC, in-sample-clustering; ML, maximum likelihood; MSC, multi-species coalescent (model); NPL, new polymorphic loci; PE, paired-end; PIC, parsimony informative character; PIS, parsimony informative site; RADseq, restriction site-associated DNA sequencing; REase, restriction endonuclease; RRL, reduced-representation library (methods); sCF, site concordance factor; SNP, single nucleotide polymorphism; SRS, short-read sequencing; SVD, SVDquartets; VAR, variable sites (sequence variation); var, variability (VAR/locus length/number of samples.

* Corresponding author at: Prinzessin Therese von Bayern Lehrstuhl für Systematik, Biodiversität & Evolution der Pflanzen, Ludwig-Maximilians-Universität München, Germany.
*E-mail address:* G.Kadereit@lmu.de (G. Kadereit).

https://doi.org/10.1016/j.ympev.2021.107342
Received 3 July 2020; Received in revised form 5 July 2021; Accepted 29 October 2021
Available online 14 November 2021

Maddison and Knowles, 2006; Kubatko and Degnan, 2007; Whitfield and Lockhart, 2007; Degnan et al., 2006, Degnan and Rosenberg, 2009; Heled and Drummond, 2009; Yang and Rannala, 2010; Rannala et al., 2020). Since different parts of the genome can have different evolutionary backgrounds, approaches analyzing multiple genomic regions have become the baseline for resolving such challenging lineages. The multi-species coalescent (MSC) model provides a natural framework for species tree inference that accounts for gene tree discordance due to ILS. However, full-coalescence approaches under the MSC are computationally very intensive when applied on large-scale genomic data and thus often not feasible (McCormack et al., 2013a; Smith et al., 2014; Zimmermann et al., 2014). Other approaches, such as maximum likelihood analysis of concatenated multi-locus data (de Queiroz et al., 1995; Yang 1996; de Queiroz and Gatesy 2007), coalescent-based summary methods that estimate species trees from independently inferred gene trees (here called "locus trees") (Mirarab et al., 2014a; Mirarab and Warnow, 2015; Rannala et al., 2020) or coalescent-based methods that use site patterns of assembled loci for species tree inference (Bryant et al., 2012; Chifman and Kubatko, 2014; Bryant and Hahn, 2020), became increasingly popular and widely used. Despite their popularity, these methods each have advantages and disadvantages and their correct application to modern high-throughput data, in particular approaches that generate short loci with high amounts of missing data such as RADseq, is highly controversial.

High-throughput sequencing (HTS) technologies and lab workflows for sample preparation improved enormously during the last decade and provide the opportunity to generate extensive datasets for phylogenetic inference (reviewed in Good, 2012; Reuter et al., 2015; Andrews et al., 2016; Mardis, 2017; McKain et al., 2018). Some of the most popular sample preparation protocols are grouped under the term reduced-representation library (RRL) preparation protocols, which are often combined with short-read sequencing (SRS). These methods target only a reduced subset of the studied genome for sequencing, therefore reducing computational complexity during assembly and analysis, facilitating a deeper sequencing depth per locus while increasing the number of samples included. The combination of both HTS and RRL enable simultaneous acquisition of genomic data of hundreds up to thousands of loci from dozens to hundreds of samples for systematic researchers and extend the questions and taxa that can be investigated tremendously. Widely used RRL approaches are hybridization capturing methods, e.g., on-array capture or in-solution capture (Mamanova et al., 2010), Hyb-Seq (Weitemier et al., 2014), targeted sequence capture (Grover et al., 2012) and restriction-site associated DNA sequencing (RADseq; Miller et al., 2007; Baird et al., 2008). The term RADseq comprises several methods that all rely on the enzymatic digestion of genomic DNA for complexity reduction, followed by adapter ligation, further reduction by size selection (either direct or indirect) and high-throughput sequencing (reviewed in Andrews et al., 2016). The cross-over approach hyRAD by Suchan et al. (2016) combines RADseq with capturing using either biotinylated DNA- or RNA-probes (Schmid et al., 2017; Suchan, 2018) obtained from the enzymatically fragmented DNA resources of the target group itself. Yet, the lab workflow is quite complex and time consuming. Thanks to the modular principle of RADseq, the individual wet lab steps, restriction endonucleases (REase/s) and adapters can be modified as required (see also McCormack et al. 2013b; Andrews et al., 2016; McKain et al., 2018; Parchman et al., 2018). This flexible toolbox of cheap, fast and individually scalable wet lab modules, as well as the fact that no prior genomic information is required, paved the way for the success of RADseq methods in various fields of evolutionary research, particularly in non-model organisms (e.g., Eaton and Ree, 2013; Escudero et al., 2014; Harvey et al., 2016; Herrera and Shank, 2016; Razkin et al., 2016; de Oca et al., 2017; Dillenberger and Kadereit, 2017; Hamon et al., 2017; Curto et al., 2018; Wagner et al., 2018; Gerschwitz-Eidt and Kadereit, 2019; Paetzold et al., 2019; Rancilhac et al., 2019; Hipp et al., 2020; Karbstein et al., 2020; Wagner et al., 2020; Buono et al., 2021).

Despite these obvious benefits of RADseq, the approach poses some inherent challenges regarding the wet lab workflow, sequence assembly, data set processing and the application of coalescent-based species tree inference. Characteristically, RADseq datasets comprise relatively short loci (typically 100–250nt) and a high proportion of missing data (Ree and Hipp, 2015; Andrews et al., 2016; Eaton et al., 2017; Lee et al., 2018; McKain et al., 2018). The average fragment length obtained (and locus length assembled) depends on the degree of genomic reduction, which in turn depends on the REase/s chosen, the selected size segregation window and the genome size of the study group. To some extent, missing data (absence of data or missingness) in RADseq data is inherently expected due to mutations of the REase-specific recognition sites (Rubin et al., 2012; Eaton et al., 2017; Lee et al., 2018). Technical causes for missingness include: varying DNA quantity and quality, size selection artifacts, PCR bias or low sequencing depth and quality. All of these factors influence the average information content per locus and the uniformity with which it is distributed across taxa, consequently limiting the applicability of inference methods (Gatesy and Springer, 2014; Xi et al., 2015; Xu and Yang, 2016; Eaton et al., 2017; Sayyari et al., 2017; Lee et al., 2018; Molloy and Warnow, 2018).

RADseq is particularly appealing for studying non-model taxa, as large genome-sized datasets can be generated quickly and cost-effectively and assembled without requiring a reference genome. However, *de novo* assembly and data processing can also be a major challenge. The bioinformatics effort related to RADseq data is often not straightforward and can heavily impact the assembly outcome regarding differentiation of orthologs and paralogs, as well as the quantity of recovered loci, sequence variation (VAR), single nucleotide polymorphisms (SNPs) and parsimony informative sites (PIS), respectively (Rubin et al., 2012; Ilut et al., 2014; Harvey et al., 2015; Shafer et al., 2017; Lee et al., 2018). To facilitate data processing, assembly pipelines such as Stacks (Catchen et al., 2013), dDocent (Puritz et al., 2014) and *ipyrad* (Eaton and Overcast, 2020) have been developed. These pipelines implement several main steps. 1) In-sample-clustering (ISC), in which reads within each sample are grouped by sequence similarity into putative loci. 2) Consensus calling of allele sequences from clustered reads. 3) Between-sample-clustering (BSC) of consensus sequences of all loci across all samples are clustered by sequence similarity to generate putatively homologous loci. 4) Data filtering based on given thresholds such as the number of samples per locus required (locus coverage) or the maximum proportion of shared heterozygous sites in a locus (detection of potential paralogs). To determine which reads represent the same genomic locus, a clustering threshold (CT) based on sequence similarity is used. Yet, genetic variation within the target genomes and across the studied taxa makes it difficult to find an appropriate CT (Rubin et al., 2012; Catchen et al., 2013; Hirsch and Buell, 2013; Ilut et al., 2014; Harvey et al., 2015; Ilut et al., 2014; Paris et al., 2017; Shafer et al., 2017; Lee et al., 2018; McCartney-Melstad et al., 2019). Both over- and undermerging are major issues in RADseq datasets, affecting ISC and BSC and therefore the resulting datasets. To ensure the homology of the assembled loci (Springer and Gatesy, 2018; McCartney-Melstad et al., 2019; Fernández et al., 2020; Simion et al., 2020), detailed evaluations of dataset metrics are used to find balanced dataset-specific CTs for ISC and BSC (e.g. Ilut et al., 2014; Mastretta-Yanes et al., 2015; McKinney et al., 2017; Paris et al., 2017; McCartney-Melstad et al., 2019). Approaches to facilitate this problem aim at the determination of suitable CTs for homology assessment by analyzing trends of several assembly metrics over a wide range of tested CTs (hereafter referred to as "CT selection approach"). This is accomplished by plotting the metrics as a function of the CT range and searching for a region that avoids over- and undermerging areas and that provides an accurate clustering for the majority of loci (hereafter referred to as "transition zone"). This transition zone is assumed to minimize the assembly of paralogs, to maximize the yield of sequence variation, and to form the smallest distance among taxa (Ilut et al., 2014; Mastretta-Yanes et al., 2015; McCartney-Melstad et al., 2019). In other words: an informed selection of dataset-

specific CTs yields maximum phylogenetic information with minimum missingness and least paralogs. Still, such CT selection approaches have to be taken with care because 1) the determined CT (for ISC and BSC) can never represent all taxa equally well and 2) all other chosen assembly parameters affect the outcome (Shafer et al., 2017; McCartney-Melstad et al., 2019).

Phylogenetic inference of assembled RADseq data presents the next challenge because the data properties often limit the choice of methods. Added to this is an ongoing, intense debate on the utilization of phylogenetic inference methods. The focus is mainly on: 1) the statistical consistency under the MSC, 2) the evolutionary framework to which the methods are applied (e.g. hybridization, horizontal gene transfer, ILS), and 3) the estimation accuracy under varying dataset conditions (e.g. linkage, phylogenetic information content, missingness, homology of data), leading to constant re-analyses and comparisons of simulated and empirical data to proof the diverging concepts (e.g. de Queiroz and Gatesy 2007; Edwards et al., 2007, 2016; Kubatko and Degnan, 2007; Degnan and Rosenberg, 2009; Leaché and Rannala, 2011; Song et al., 2012; Bayzid and Warnow, 2013; Wu et al., 2013; Gatesy and Springer, 2013, 2014; Springer and Gatesy 2014, 2016, 2018; Mirarab et al., 2014a, b, 2016; Chou et al., 2015; Roch and Steel 2015; Mendes and Hahn, 2018; Molloy and Warnow, 2018; Bryant and Hahn, 2020; Rannala et al., 2020). This somewhat amusing and abstruse debate, with sometimes remarkably tailored data for proof, complicates the search for appropriate phylogenetic inference methods for RRL-SRS data. Fact is that the locus properties are pivotal for selecting appropriate species tree inference methods. Due to the short fragment length, RADseq loci are generally assumed to lack sufficient phylogenetic information to generate locus trees as input for coalescent-based summary methods (Rubin et al., 2012; Gatesy and Springer, 2014; Xi et al., 2015; Hosner et al., 2016; Molloy and Warnow, 2018).

Gene-tree-based coalescent methods (summary methods; hereafter referred to as CB-SM) are a favorable choice for phylogenetic inference of rather long and informative loci (Mirarab et al., 2014a, 2016; Vachaspati and Warnow, 2015; Xu and Yang, 2016; Molloy and Warnow 2018; Rannala et al., 2020). CB-SM infer species trees by a two-step system: individual gene trees are estimated, and their summary statistics are then used as data input for species tree estimation. While CB-SM are becoming popular for their ability to handle large amounts of data in a short time, they are best known for their sensitivity to gene tree estimation error (GTEE). When applied to datasets composed of short loci of little individual phylogenetic information and a high proportion of missingness, as is characteristic of RADseq datasets, the effect on estimation accuracy can get quite severe (Chou et al., 2015; Roch and Warnow, 2015; Xi et al., 2015; Xu and Yang, 2016; Sayyari et al., 2017; Molloy and Warnow, 2018). Therefore, the focus on the effects of filtering loci for specific properties prior to gene and species tree estimation is becoming increasingly relevant (e.g. Lanier et al., 2014; Chen et al., 2015; Xi et al., 2015; Hosner et al., 2016; Huang and Knowles 2016; Simmons et al., 2016; Sayyari et al., 2017; Molloy and Warnow 2018).

Coalescent-based site-based methods are another option for species tree inference (Bryant et al., 2012; Chifman and Kubatko, 2014; Xu and Yang, 2016). Such approaches bypass the generation of locus trees by generating the species tree directly from all given site patterns, thus avoid the issue of GTEE. The sites are required to have individual histories or at least very little linkage. Violation of this assumption leads to a statistically inconsistent species tree estimate (Bryant et al., 2012; Chifman and Kubatko 2014; Xu and Yang, 2016). Under certain challenging data conditions, site-based methods were found to be more accurate than gene tree-based summary (Chou et al., 2015; Long and Kubatko, 2018; Molloy and Warnow, 2018).

RADseq data are most commonly analyzed using maximum likelihood analysis of a concatenated supermatrix (hereafter referred to as CA-ML) (Yang, 1996; de Queiroz and Gatesy, 2007; Rubin et al., 2012). In case of CA-ML, several thousand loci are treated as one locus that evolved under a single evolutionary history. This is violating the MSC and may theoretically lead to poorly resolved, incomplete, or positively misleading species tree estimates (Degnan et al., 2006, Degnan and Rosenberg, 2009; Kubatko and Degnan, 2007; Knowles, 2009; Roch and Steel, 2015, Xu and Yang, 2016; Mendes and Hahn, 2018; Rannala et al., 2020). In addition, bootstrapping is also commonly performed across the entire supermatrix, potentially resulting in spuriously high support values caused by the sheer dataset size (Kubatko and Degnan, 2007; Kumar et al., 2012; Rubin et al., 2012; Liu et al., 2015; Wang et al., 2017, Minh et al. 2020a). Still, it also has been shown that CA-ML can be comparably or more accurate than coalescent-based methods under various conditions of linkage, locus length, information content, missingness, ILS and GTEE (Mirarab et al., 2014a; Chou et al., 2015; Roch and Warnow, 2015; Mirarab et al., 2016; Springer and Gatesy, 2016; Long and Kubatko, 2018; Molloy and Warnow, 2018).

Despite the ongoing debate about the pros and cons of approaches to sequence generation, data assembly, phylogenetic inference, and, the assumption that RAD data do not favor coalescent-based summary methods, we think there is a need to take advantage of the significant methodical progress made in the last decade and explore their potential for practical use. Our objective is to test whether longer RADseq loci enable coalescent-based species tree inference, and to provide advice on how to handle and analyze challenging data.

We modified several modules of the RADseq toolbox to obtain a library containing a small number of fragments (ca 5,000 assembled loci), with lengths of ca. 300-600nt, sequenced with the latest SRS technology (Illumina MiSeq v3 kit, 300nt PE) and applied this protocol (Fig. 1) to the plant genus *Aichryson* Webb & Berthel. (Crassulaceae), a rapidly radiated yet relatively small genus distributed in Macaronesia, for which standard sanger sequenced markers failed to provide a resolved phylogeny (Fairfield et al., 2004). The data analysis (Fig. 2) included a CT selection approach to facilitate an informed choice of suitable CTs for ISC and BSC during *de novo* assembly (Fig. 3) and an exploratory approach to determine the properties of the assembled loci, with respect to locus coverage (missingness), locus variability (phylogenetic information) and locus length, and thus their suitability as input for CB-SM (Fig. 4). We compared the phylogenetic outcome of this assembly using CA-ML (RAxML by Stamatakis, 2014), CB-SM (ASTRAL III by Zhang et al., 2018) and put it in perspective to the site-based approach SVDquartets by Chifman and Kubatko (2014). To assess the phylogenetic results, we also evaluated the resulting BS support values relative to gene and site concordance factors that were calculated using IQ-TREE (Minh et al. 2020a, b).

## 2. Materials and methods

### 2.1. Study group, sampling and DNA extraction

Together with *Monanthes* Haw. and *Aeonium* Webb & Berthel., *Aichryson* belongs to the Macaronesian tribe Aeonieae of the Crassulaceae family (Eggli, 2008). The genus comprises 15 species with the centre of diversity on the Canary Islands (11 species; Bañares Baudet, 2002, 2015a, Baudet, 2017), three species on Madeira, and one species on the island of Santa Maria in the Azores (Moura et al., 2015). *Aichryson* is divided into two sections, sect. *Aichryson* and sect. *Macrobia* Webb & Berthel. Section *Macrobia* includes only *Aichryson tortuosum* (Aiton) Webb & Berthel., a perennial, small shrub endemic to Lanzarote (subsp. *tortuosum*) and Fuerteventura (subsp. *bethencourtianum* Botte & Bañares). All other species belong to sect. *Aichryson* and are monocarpic, mostly annual herbaceous plants (Bañares, 2015a). Within sect. *Aichryson* several natural hybrids are described (Bañares, 2015b). *Aichryson* proved to be monophyletic and likely sister to *Monanthes icterica* (Webb ex Bolle) Christ in molecular phylogenetic studies on Aeonieae based on cp markers and ITS (Mort et al., 2002; Fairfield et al., 2004). The genus comprises both diploid and tetraploid species (Uhl, 1961; Suda et al., 2005).

We sampled a total of 29 individuals representing 14 species of *Aichryson* (only *A. santa-mariensis* M.Moura, Carine & M.Seq. is missing) and two accessions of *Monanthes icterica* as outgroup (Supplementary Table 1, "sampling"). For 20 samples we were able to assess the ploidy level on a CyFlow cytometer (PARTEC) using the isolation buffer "OTTO I" (2.1 g Citric-acid-1-hydrat, 10 ml 5% Triton X-100, 90 ml dd$H_2O$). FloMax v2.8.2 (QA GmbH, Münster, Germany) was used for the particle analysis and the measurement of the peaks (Table S1, "flow cytometry"). For the remaining samples, published ploidy levels were incorporated (Uhl, 1961; Suda et al., 2005).

DNA-extraction was conducted using the DNeasy Plant Mini-Kit (QIAGEN, Venlo, Netherlands) according to the manufacturer's protocol for "Purification of Total DNA from Plant Tissue (Mini Protocol)" with a number of modifications outlined in the online Appendix 1. The DNA concentration and quality were evaluated using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and gel electrophoresis.

### 2.2. *In silico digestion, restriction enzyme choice and adapter design*

The search for suitable restriction enzymes for our approach was performed *in silico* and based on 1) the desired fragment length (300-600nt), 2) the number of samples per library (up to 50), 3) the expected sequencing output of the MiSeq v3 kit (up to 25 Million), and 4) the targeted sequencing depth (aimed at ~ 10 × per fragment), leading to the required fragment yield of 5,000 within the target length range. Initially we tested commonly used REases individually. However, the REases tested did not meet our requirements, thus we tested combinations of two REases each. For this, we have taken into account a minimum length of 6nt for the recognition site and the simultaneous applicability of two REases in a single reaction. The *in silico* digestion was performed using the software CLC genomics Workbench v9.5.5 (Qiagen) with its included "Restriction Site Analysis" for several genomes of various focal groups including *Beta vulgaris* L., Amaranthaceae (BioProject PRJNA41497) and *Kalanchoe fedtschenkoi* Raym.-Hamet & H.Perrier, Crassulaceae (BioProject PRJNA397334). The resulting restriction maps were evaluated with respect to fragments showing two cut sites within the desired length window of 300-600nt. Among other suitable REase combinations, the REases *BamHI* (G'GATCC) and *KpnI* (GGTAC'C) best met our criteria for a double digest (for excerpts of the REase selection, see also Supplementary Table S2, "*in silico* digest"). In case of *Aichryson*, the *in silico* digest of the distantly (yet closest) related *K. fedtschenkoi* genome (divergence to Aeonieae is roughly 58.60 [44.60–73.62] mya, Messerschmid et al., 2020), resulted in 61,692 fragments, of which 4,429 fragments fell in the targeted length range.

In contrast to widely established strategies (Elshire et al., 2011; Peterson et al., 2012; Andrews et al., 2016), we aimed at sequencing all generated fragment types, including fragments framed by identical restriction motifs. Thus, we designed the barcode and common adapters for both REases motifs (Table S2, "BamHI adapter", "KpnI adapter"). The barcode sequences were obtained from Elshire et al. (2011) and van Gurp (2017). Both barcode and common adapter fit to the overhang of the *BamHI* and *KpnI* cut sites (Fig. 1b). We were able to achieve the set aim with this design, however, we recommend a more flexible adapter/indexing strategy that accounts for technical bias during wet lab and sequencing (e.g. MacConaill et al., 2018; Bayona-Vásquez et al., 2019).

### 2.3. *RADseq*

The major changes compared to other RADseq approaches such as ddRADseq (Peterson et al., 2012) or Genotyping by Sequencing (GBS; Elshire et al., 2011) are: the usage of two rare cutter REases that produce c. 5,000 fragments within a target range of 300-600nt (Fig. 1a), adapters binding to all generated fragments (Fig. 1b), an extended size selection range (Fig. 1d) and an extra size selection step during the final purification (Fig. 1f). In particular the two size selections were important to fully exploit the sequencing range (see also Appendix 1, Fig. A1.6, A1.7). Since the RADseq toolbox includes many modifiable modules, various protocols might be capable of generating libraries/datasets of an extended length range and we encourage an impartial testing of this potential (see also: McCormack et al. 2013b; Andrews et al., 2016; McKain et al., 2018; Parchman et al., 2018). The following is a brief overview of the workflow. For the detailed protocol, see Appendix 1 and Supplementary Table S3.

#### 2.3.1. *RADseq lab workflow*

We used 200 ng genomic DNA as input for the double digest reaction (Fig. 1a), which was followed by adapter ligation (Fig. 1b) in the same reaction tube. For thorough saturation of cut sites, 6 µl adapter working solution (0.5 ng/µl) containing equimolar amounts each motif pair were used. Reactions were incubated for 3 h at 37 °C, respectively. The libraries were multiplexed using 100 ng DNA each (Fig. 1c), followed by a column-based cleaning of the pool. Size selection (Fig. 1d) was performed using Pippin Prep (Sage Science, Beverly, MA, USA) with a segregation range of 350-720nt. The size-selected products were amplified using a low-cycle 2-step PCR protocol (Fig. 1e). Subsequently, PCR products were collected in three pools (Table S3), purified and quantified. Final purification, accompanied by the 2nd size segregation, was done using the NucleoMag NGS kit (Macherey-Nagel, Düren, Germany) with a ratio of 0.8 bead suspension to one part library. The purified library was resuspended in 25 µl Buffer AE for sequencing.

#### 2.3.2. *Library quality assessment and sequencing*

Library quality was validated by measuring the DNA concentration by Qubit Fluorometer and assessing the fragment distribution by Bioanalyzer electropherogram (Appendix 1). Sequencing was performed on an Illumina MiSeq (San Diego, CA, USA; Reagent Kit v3 600-cycle) at StarSEQ (Mainz, Germany) producing 300nt PE reads in three different runs (Supplementary Table S4).

### 2.4. *Data assembly*

#### 2.4.1. *Raw sequence treatment*

Raw data quality was assessed with FastQC 0.11.4 (Andrews, 2010; Fig. 2a; Table S4 "run I-III"). Raw reads were demultiplexed (Table S2 and S3) using *ipyrad* v0.9.52 (Eaton and Overcast, 2020) twice, once for each REase cut site (Fig. 2a). This two-fold demultiplexing was necessary due to the motifs occurring on both read directions. The fastq-files were combined and adapter sequences were removed with Cutadapt 1.18 (Martin, 2011). FastQC reports of the demultiplexed/adapter trimmed samples were combined using MultiQC v1.9 (Ewels et al., 2016; Table S4 "mean quality scores").

#### 2.4.2. *ipyrad*

We used *ipyrad* v0.9.52 (Eaton and Overcast, 2020) for *de novo* RADseq assembly. Several filtering parameters of the *ipyrad* pipeline (v9 or above, Eaton and Overcast, 2020) represent percentages, allowing the application of the selected thresholds to variable read lengths and thus supporting clustering of datasets obtained by a broad sequencing range. We used default parameters, except for the ones outlined below.

#### 2.4.3. *Assembly parameter settings*

The de-multiplexed samples were split into two groups according to ploidy level (di- or tetraploid; Table S1). The diploid dataset contained nine *Aichryson* samples, the tetraploid dataset contained 18 *Aichryson* and two *Monanthes icterica* samples. Parameter #18 (max_alleles_consens) was set to two and four, respectively (Supplementary Table S5). With respect to the extended read length, we allowed up to 24 indels per locus (parameter #23). We assumed increased gene flow and set parameter #24 to 0.7 (Bañares, 2015b; max_Hs_consens). Parameters #11 and #12, which give the minimum depth for statistical and majority

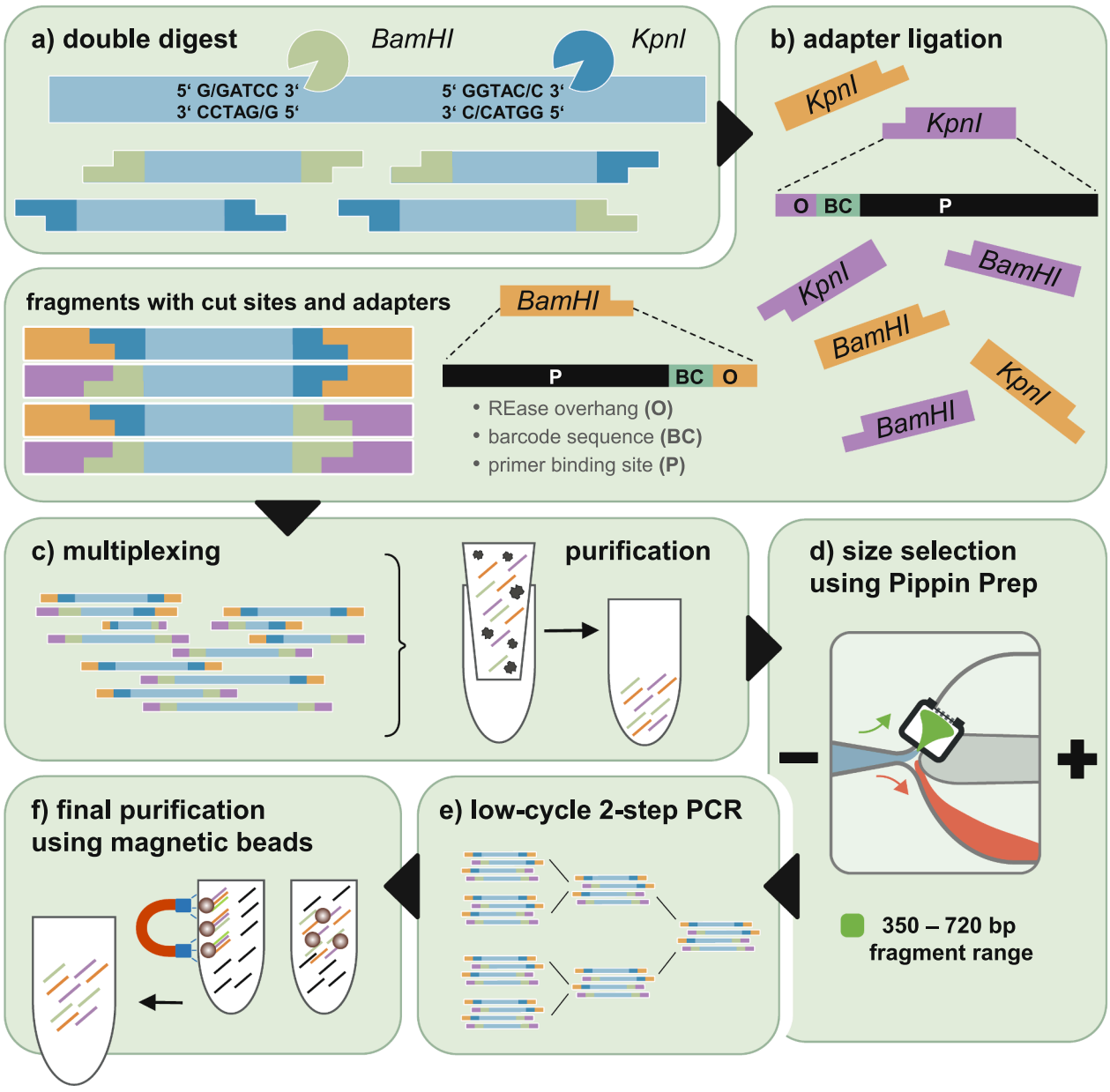


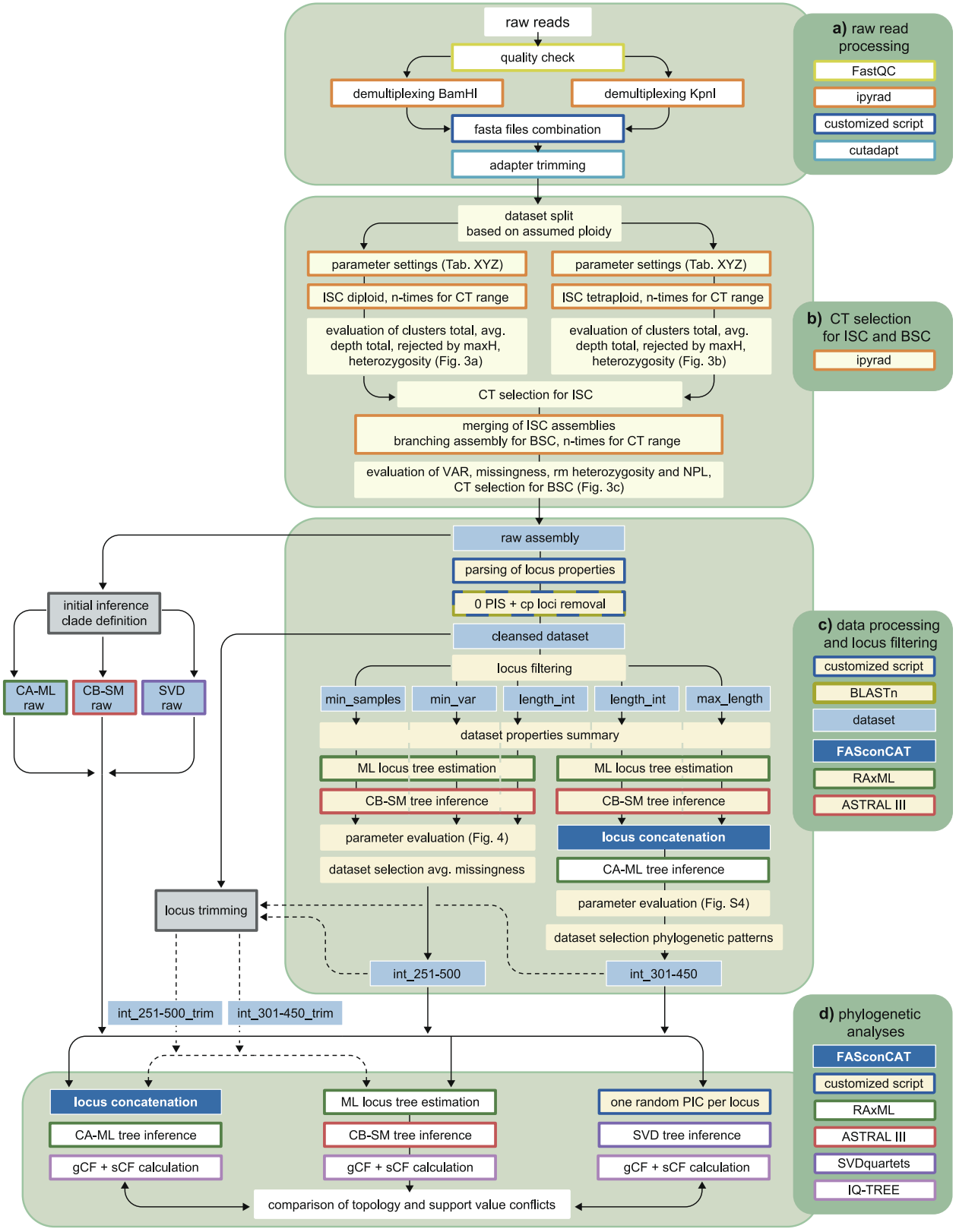

**Fig. 1.** The lab workflow of the modified RADseq protocol consists of six steps (a – f). a) Genomic DNA is digested simultaneously using the REases *BamHI* and *KpnI*. b) Barcode and common adapters are ligated to the fragments. c) The barcoded samples are multiplexed and purified. d) The pool is size selected to a 350 – 720 bp length range using Pippin Prep. e) The size selected pool is amplified using a low-cycle 2-step PCR. f) The final purification using magnetic beads removes PCR and size selection artifacts.

rule base calling, were set to 10. We aimed at an average cluster depth (avg_depth_mj) of > 20 × for statistical base calling (Pamilo et al., 2011; Eaton and Overcast, 2020).

*2.4.4. Selection of suitable clustering thresholds for ISC and BSC*

Avoiding both, over- and undermerging of putative loci is not trivial in high-throughput datasets. If the selected CT is too lax, paralogous reads will be incorrectly clustered and treated as orthologs (overmerging) and if the selected CT is too strict, reads belonging to an actual locus will incorrectly be split into several loci (undermerging) with low variability (Supplementary Figure S1.A). To determine suitable CTs for ISC and BSC, we used several CT selection approaches as guidance (Ilut et al., 2014; Mastretta-Yanes et al., 2015; Paris et al., 2017; McCartney-Melstad et al., 2019) and defined the assumptions to determine suitable CTs. 1) Over- and undermerging ranges have to be identified to avoid merging/splitting effects within these areas. 2) Overmerging is indicated by highly heterozygous clusters/alleles with a high proportion of filtered paralogs (Ilut et al., 2014; McCartney-Melstad et al., 2019). Hence, a suitable CT is expected in an increasing area of heterozygosity and a decreasing area of flagged paralogs, between the maxima of both metrics. 3) Undermerging of orthologs leads to an increased number of loci (and lower locus coverage in ISC, lower sample coverage per locus in BSC) while sequence divergence among taxa decreases (Mastretta-Yanes et al., 2015; McCartney-Melstad et al., 2019). Thus, sequence variation declines while missingness increases. A suitable CT is expected near a steep increase in number of clusters/loci and amount of missingness while heterozygosity is biologically realistic (ISC) and locus variability is high (BSC).

To prevent introducing a potential bias due to ploidy, we split the samples into two groups (di- and tetraploid) for ISC assembly (*ipyrad* assembly steps 1–5, Fig. 2b). Following ISC CT selection, all samples were merged for BSC (*ipyrad* assembly steps 6 and 7). A CT range of




(caption on next page)

**Fig. 2.** The schematic overview of the data analysis is split into four major parts (a-d, boxes on the right side). The boxes in light blue indicate sub-/datasets. Dashed arrows illustrate parameter applications between datasets. Colored box edges show the software used for the work step. During the raw read processing (a) the quality is assessed using FastQC, the reads are demultiplexed two times with respect to the REase cut sites and the sample specific barcodes, combined into sample fasta files, and adapter and cut sites are removed using cutadapt. For the clustering threshold (CT) selection approach (b), the data set is split based on the assumed ploidy and the *ipyrad* parameters are adjusted as required. For in-sample-clustering (ISC) a CT range of 0.81 – 0.99 is tested for both datasets and *ipyrad* outputs are evaluated with respect to the number of total clusters, total average read depth, clusters rejected by maxH (flagged paralogs) and heterozygosity (Fig. 3a and b). The selected ISC assemblies are merged and branched to test the CT range (see above) for between-sample-clustering (BSC). The resulting assemblies are evaluated with respect to the number of retained loci, the retained sequence variation (VAR), missingness and the number of new polymorphic loci (NPL, Fig. 3c). The selected "raw" assembly is used for initial phylogenetic inference and clade definition (c). The locus properties (locus ID, length, number of samples, number of SNPs, PIS and VAR) are parsed using a customized script. Loci showing no variation and chloroplast loci are removed. The loci of the "cleansed" dataset are filtered into several sub-datasets based on their properties. The first locus filtering approach, using a missingness threshold for dataset selection, resulted in the "int_251-500" dataset. The second filtering approach, using sub-dataset properties and resulting phylogenetic patterns for dataset selection, resulted in the "int_301-450" dataset. The truncated loci of the "raw" assembly were re-arranged based on the selected datasets of the locus filtering (locus truncation, dashed arrows). The datasets ("raw", "int_251-500", "int_301-450" and "short") are used for comparative phylogenetic inference (d). Individual loci are either concatenated using FASconCAT for CA-ML inference or used to calculate ML locus trees as input for CB-SM inference. The SVD datasets are created by picking a single randomly selected parsimony informative character (PIC) of each locus. To assess the resulting trees of the tested inference methods across datasets, we compared changes in BS support values and gene (gCF) and site concordance factor (sCF) values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.81–0.99 (in 0.01 increments) was tested. To assess the above-mentioned criteria for CT selection, we plotted a variety of metrics as a function of the tested CT range as box- and scatter plots (see also Figure S1.B and S1.C). For the ISC CT selection, we evaluated the number of clusters (clusters_total), the average read depth (avg_depth_total), the number of filtered paralogs (filtered_by_maxH) and the heterozygosity. For the BSC CT selection, we additionally evaluated the number of retained loci, sequence variation (VAR, SNPs and PIS) and proportion of missingness (sequences_missing). In addition, we calculated the "new polymorphic loci" (NPL) in order to detect the assembly containing most accurately clustered sequence variation, which is indicated by the so-called "hockey stick signal" (Paris et al., 2017). We expected the transition zone from over- to undermerging to be characterized by trend changes, e.g. prominent differences in the medians of adjacent CTs and compressions or expansions of the quartiles (in boxplots) or changes in the slope intensity (in scatter plots). Multiple suitable CTs within a transition zone of a metric and across metrics were averaged to determine a consensus CT.

#### 2.4.5. Processing of the unfiltered ipyrad assembly

The *ipyrad* loci-file of the unfiltered "raw" assembly was parsed with a custom Perl script (available on GitHub https://github.com/philipphuehn/RADseq-locus-filtering) for the specific locus ID, the length, the number of samples, SNPs and PIS (VAR in total) and the proportion of missingness for each locus (Fig. 2c "parsing of locus properties"). We used BLAST + 2.7.1 (Camacho et al., 2009) to identify chloroplast loci by blasting all loci against four reference plastomes from the Crassulaceae (GenBank accessions: *Sedum uniflorum* subsp. *oryzifolium* (Makino) H.Ohba: NC_027837, *Sedum sarmentosum* Bunge: NC_023085, *Phedimus takesimensis* (Nakai) 't Hart: NC_026065, *Phedimus kamtschaticus* (Fisch. & C.A.Mey.) 't Hart: NC_037946). Loci of a plastid origin as well as loci showing no parsimony informative sites were removed (Fig. 2c, "0 PIS + cp loci removal"). In addition, this "raw" assembly was used for initial phylogenetic inference and clade definition to compare potentially different phylogenetic results from subsequently filtered datasets (see 3.4.1).

### 2.5. Locus filtering and dataset selection

In general, phylogenetic inference by CB-SM is very sensitive to GTEE, which most often is caused by loci showing little sequence variation, high missingness or fractional coverage (Chou et al., 2015; Roch and Warnow, 2015; Xi et al., 2015, 2016; Xu and Yang, 2016; Sayyari et al., 2017; Hosner et al., 2016; Lee et al., 2018; Molloy and Warnow, 2018). We filtered the here generated RADseq loci into several sub-datasets to test for a potential influence of locus properties on phylogenetic inference (Fig. 2c, "locus filtering"). First, we determined the impact of the locus properties on CB-SM reconstruction (see 2.5.1). This filtering approach suggested a potential impact of biased phylogenetic signal due to non-randomly distributed partial taxon coverage (Sanderson et al., 2010, 2011, 2015; Simmons, 2012; Xi et al., 2015; 2016; Hosner et al., 2016; Sayyari et al., 2017; Dobrin et al., 2018). This so-called "biased missingness" has been shown to cause high GTEE, thus results in conflicting, unsupported locus trees and consequently in a decline of species tree estimation performance (Xi et al, 2015, 2016; Hosner et al., 2016; Sayyari et al, 2017; Molloy and Warnow, 2018). We therefore performed a second locus filtering with respect to locus length and evaluated phylogenetic patterns of CB-SM and CA-ML reconstructions (see 2.5.2). The locus filtering scripts are available at GitHub (https://github.com/philipphuehn/RADseq-locus-filtering).

#### 2.5.1. Locus filtering by coverage, variability, length intervals and dataset selection based on average missingness

The loci were filtered with respect to the average variability (var = VAR/locus length/number of samples; "min_var"), minimum number of samples per locus (number of samples/locus; "min_samples"), and locus length intervals ("length_int") and rearranged to new sub-datasets (Fig. 2c, "locus filtering", Supplementary Table S6). For the "min_var" sub-datasets, seven thresholds were used (0.01, 0.25, 0.50, 0.75, 1.0, 2.0, 3.0, "min_var_001" – "min_var_300"). Six thresholds by increments of four were used for the "min_samples" sub-datasets (4, 8, 12, 16, 20, 24, "min_samples_4" – "min_samples_24"). The locus length interval datasets were created based on eight intervals starting from the minimum length to 250nt, and then ranging by 50nt steps from 251nt to 550nt, and 551nt to the maximum length ("int_min-250" – "int_551-max"). Properties of these datasets, such as the total number of loci, VAR, SNPs, PIS (average per locus), sample coverage/missingness, and average locus length were recorded (Fig. 2c, "sub-dataset properties summary"). For each rearranged sub-dataset, ML locus trees were estimated and used for CB-SM inference (see 2.6.2). We recorded the bootstrap support values of all branches of each tree and assigned them to three categories: backbone, clade and within clade branch support values. Clade branches contained all samples of the defined clades (see 3.4.1 for clade definition). All support values within the defined clades were assigned to within clade branches. All other support values, spanning from the outgroup to the clade branches, were recorded as backbone support values. Topology changes and conflicts were not accounted for. Based on this and on recommendations by studies investigating the impact of locus filtering for summary methods (Xi et al., 2016; Sayyari et al., 2017; Molloy and Warnow, 2018), we selected an average missingness threshold to filter the locus sets (Fig. 2c, "dataset selection avg. missingness"). The resulting dataset was subsequently used for comparative phylogenetic inference (Fig. 2d).

*2.5.2. Locus filtering by length and dataset selection based on sub-dataset properties and phylogenetic patterns*

In order to narrow down the suspected dataset bias in terms of fractional, non-random locus and/or taxon coverage, we used phylogenetic patterns to assess sub-datasets filtered by length. CA-ML inference of datasets exhibiting this type of bias can result in unsupported or overly high supported polytomies resolved as a terraced topology (Sanderson et al., 2010, 2011, 2015; Simmons, 2012; Dobrin et al., 2018). Dobrin et al. (2018) have reported numerous empirical multilocus datasets to be impacted by this issue (e.g. Springer et al., 2012; Burleigh et al., 2015; Shi and Rabosky, 2015). Since we generated ML locus trees as input for species tree estimation with CB-SM, we assumed this terraced topology to also appear if the bias of the underlying data was strong. Besides, Hosner et al. (2016) and Sayyari et al. (2017) found that a high proportion of fragmentary data (biased incongruence of locus trees) can lead to a sharp drop of the resulting BS support values for CB-SM inference.

In addition to the length interval sub-datasets of the first filtering ("int_min-250" – "int_551-max"), we filtered the loci requiring an increasing, cumulative maximum length (Fig. 2c, "locus filtering", Supplementary Table S7). The eight maximum locus length sub-datasets were generated starting at a threshold of 250nt ("max_250", all loci up to 250nt length were included) increasing by 50nt increments up to the maximum locus length. Each sub-dataset was subjected to phylogenetic inference using CA-ML and CB-SM. The sub-dataset properties and resulting BS support values were recorded as described in 2.5.1.

While bootstrapping across a concatenated matrix almost automatically increases the resulting support values with increasing matrix size (Kubatko and Degnan, 2007; Liu et al., 2015; Minh et al. 2020a), the multi-locus bootstrapping used with CB-SM employs a 2-stage system that accounts for variations among loci by resampling during BS calculation (Seo, 2008) and thus reacts very sensitive to fragmentary data (Xi et al., 2015, 2016; Hosner et al., 2016; Sayyari et al., 2017). We expected the BS support values to collapse as soon as the ratio of biased to unbiased data (respecting a non-randomly distributed partial taxon coverage) became too high. For CA-ML, we expected a similar but less sensitive pattern, in particular for the sub-datasets of an increasing maximum locus length.

For the evaluation of a terrace-like topology pattern, the number of samples resolved on terraced branches was recorded. We defined that a terraced branch must either -originate from a dichotomous branch of the tree's backbone, - the clade's backbone containing that sample, - or must follow an individual branch within a clade, - but must not be included within a dichotomous constellation. For instance, phylogenetic inference of the "raw" dataset using CA-ML, CB-SM and SVD resulted in two, five and three terraced branches for clade 5, respectively (Supplementary Figure S2). The SVD tree contained another terraced branch in clade 4, but the CA-ML and CB-SM trees did not. By increasing the maximum locus length required, we expected the topology to switch from a terraced to a dichotomous tree pattern once the biased area has been passed or compensated (and vice versa). CB-SM was expected to react more sensitive than CA-ML due to the reduced amount of data, with individual gene trees as input (Xu and Yang, 2016). Therefore, the terraced pattern was assumed to be over-expressed once the amount of data became too small (in particular for the length interval sub-datasets), and likewise a larger portion of unbiased data would be needed for compensation (for the maximum length sub-datasets).

The dataset, which was intended to be a reasonable compromise for both methods, had to meet the following criteria: 1) relatively low average missingness, 2) relatively high ratio of PIS to SNPs, 3) relatively high BS support values for all tree sections, 4) relatively low number of samples resolved on terraced branches, 5) and had to avoid over- and under-represented assembly regions. The selected dataset was used for comparative phylogenetic inference (Fig. 2d).

*2.5.3. Generating 'short' loci by locus truncation*

The loci of the *ipyrad* "raw" assembly were truncated to one third of their original length to compare potential performance differences of the here generated loci to a RAD dataset obtained by assembly of 100nt PE reads. These shorter loci were intended to show less sequence variation and thus negatively affect phylogenetic inference. The truncated loci were re-arranged based on the selected datasets of the locus filtering (Table 1, Fig. 2c, "locus truncation").

*2.6. Phylogenetic inference*

We have chosen three commonly used approaches for phylogenetic inference of the generated main- and sub-datasets (Table 1, S6 and S7). CA-ML and CB-SM were used for inference during locus filtering. For the comparative phylogenetic inference, we additionally used SVDquartets as third inference approach. We decided not to test a full-coalescent method that uses co-estimation of locus trees and species trees such as implemented in BEST (Liu, 2008) or BEAST 2 (Bouckaert et al., 2014) because computation time and capacities required increase sharply with the number of loci and samples. Thus, full-coalescent methods are currently not practical for large-scale datasets with thousands of loci (e.g. Bayzid and Warnow, 2013; McCormack et al., 2013a; Zimmermann et al., 2014).

*2.6.1. Phylogenetic inference with RAxML (CA-ML)*

We used RAxML v8.2.12 (Stamatakis, 2014) to infer maximum likelihood phylogenies using GTRGAMMA as substitution model, 20 runs for BestML and 1,000 bootstrap replicates to assess statistical support of relationships. We used the unfiltered *ipyrad* supermatrix for inference of the "raw" assembly. For all other datasets, we concatenated individual loci to a supermatrix using FASconCAT v1.11 (Kück and Meusemann, 2010).

*2.6.2. Species tree inference with ASTRAL-III (CB-SM)*

ASTRAL-III v5.7.4 (Zhang et al., 2018) estimates species relationships based on gene/locus trees. To generate these locus trees, we used RAxML v8.2.12 (Stamatakis, 2014) under the GTRGAMMA model with 20 runs for BestML and 1,000 bootstrap replicates. ASTRAL was run in default mode using unrooted locus trees. We used multilocus bootstrapping (Seo, 2008) to compute branch support for the estimated species trees.

*2.6.3. Svdquartets analysis (SVD)*

SVDquartets (Chifman and Kubatko, 2014) is a quartet-based algorithm to compute species trees from SNP datasets. We used FASconCAT-G (Kück and Longo, 2014) to extract and concatenate the 25,320 parsimony informative characters (polymorphisms that are shared by at least two samples, PICs) of the 3,818 loci constituting the "raw" assembly. To meet the requirement for linkage of the dataset (sites must be unlinked), we randomly selected a single PIC of each informative locus for each dataset (Table 1, "unlinked PICs"). Analyses were run in SVDquartets as implemented in PAUP*4.0a168 (Swofford, 2003) with 1,000 bootstrap replicates under the multilocus bootstrap (Seo, 2008). The scripts for generating PIC datasets are available at GitHub (https://github.com/philipphuehn/RADseq-locus-filtering).

*2.6.4. IQ-TREE analysis*

We used IQ-TREE v2.1.2 (Minh et al. 2020a, b) to calculate the gene (gCF) and site concordance factors (sCF) of the resulting phylogenies, which give the percentage of decisive locus trees and alignment sites containing or supporting a specific branch in a given reference tree, respectively. Locus trees obtained with RAxML were used for gCF calculation. For sCF calculation, 1000 quartets were used to obtain stable estimations. To assess the resulting phylogenies with respect to a potential influence of biased data, we put the resulting topologies and BS support values in context with the gCF and sCF values and value

**Table 1**
The properties of the unfiltered "raw" assembly, the "cleansed" dataset, the datasets selected by locus filtering and their length truncated variants.

| dataset | raw | cleansed | int_251-500 | int_301-450 | int_251-500_short | int_301-450_short |
|---|---|---|---|---|---|---|
| loci | 3,818 | 3,225 | 2,788 | 1,599 | 2,788 | 1,599 |
| VAR total | 71,691 | 68,490 | 56,448 | 33,480 | 18,590 | 10,625 |
| VAR per locus | 18.78 (±16.69) | 21.24 (±16.82) | 20.24 (±15.70) | 20.94 (±16.15) | 6.67 (±5.65) | 6.65 (±5.63) |
| SNPs total | 36,413 | 33,261 | 26,533 | 15,673 | 8,779 | 5,040 |
| SNPs per locus | 9.54 (±5.25) | 10.31 (±9.89) | 9.51 (±8.66) | 9.80 (±8.86) | 3.15 (±3.17) | 3.15 (±3.16) |
| PIS total | 35,278 | 35,229 | 29,915 | 17,807 | 9,811 | 5,585 |
| PIS per locus | 9.24 (±10.73) | 10.92 (±10.86) | 10.73 (±10.67) | 11.14 (±11.05) | 3.52 (±3.93) | 3.49 (±3.91) |
| unlinked PICs total | 2,723 | | 2,220 | 1,287 | | |
| locus coverage | 8.86 (±5.25) | 9.37 (±5.45) | 9.67 (±5.57) | 9.96 (+-5.62) | 9.67 (±5.57) | 9.96 (+-5.62) |
| sample coverage | 1,166 (±467) | | 930 (±333) | 549 (±204) | | |
| missingness avg. [%] | 69.79 | 67.69 | 66.66 | 65.64 | 66.66 | 65.64 |
| locus length avg. [nt] | 376 (±93) | 379 (±93) | 360 (±70) | 373 (±43) | 120 (±23) | 123 (±18) |

Given are the total number of loci (loci), the total and average values per locus (standard deviations in parentheses) for the number of variable sites (VAR), single nucleotide polymorphisms (SNPs), and parsimony informative sites (PIS), the total number of unlinked PICs as input for SVD inference, and the average locus coverage (samples per locus), sample coverage (loci per sample), the average proportion of missingness [%] and the average locus length [nt].

differences. In general, both concordance factors are expected to be similar if the phylogenetic signal is only impacted by discordant signal, e.g. due to ILS (Minh et al. 2020a, b). If other processes affect the dataset, such as limited information or a data bias, the gCF values can be a lot lower than the sCF values, resulting in large factor value differences. A large proportion of conflicting signal or a significant variation of sites in the dataset can lead to a completely random resolution, which is indicated by sCF values ~ 33%. The reasons are either true phylogenetic signal caused by ILS or biased signal caused by uneven coverage. Distinct factor value differences of alternative topologies may indicate non-phylogenetic signal.

## 3. Results

### 3.1. Final library and MiSeq output

The fragment distribution of the final library ranged from ca. 370-770nt. The majority of fragments outside the target range were successfully removed (Appendix 1, Fig. A1.4, A 1.5). The MiSeq runs generated a total of 6,870,208 paired raw reads for the 29 samples (Table S4, "samples"). Sequence quality decreased with increasing read length (Table S4, "run I-III"). The quality of the R2 reads started to decline below a Phred quality score of 20 from ca 260nt read length (Table S4, "mean quality scores"). The number of reads per sample ranged from 98,754 for *A. laxum* var. *latipetalum* Bañares & M.Marrero to 587,377 for *M. icterica* BG Bonn with an average of 236,903 reads per sample. Demultiplexed raw data is available at the NCBI Sequence Read Archive in BioProject PRJNA642981.

### 3.2. ISC and BSC threshold selection

In general, the plots of the selected metrics showed the expected trends and met the requirements (Fig. 3 and S1.B and S1.C). For the ISC metrics, however, the indicators were not as distinct as expected. The transition zones of the metrics were averaged to consensus CTs for the diploid and tetraploid samples, respectively (Supplementary Table S8).

For the ISC of diploid samples (Fig. 3a and S1.B, "ISC 2n"), the onset of the undermerging area was initiated by an abrupt increase in the number of clusters at CT 0.95, which was indicated by a compression of the third quartile (Q3) for the CTs 0.93 and 0.94 and a simultaneous increasing slope intensity in the scatter plots (Fig. 3a and S1.B, "clusters total", transition zone: 0.93–0.94). Allelic variation was highest in the transition zone of 0.92–0.95 and started to decrease strongly with increasing sample coverage (Fig. 3a and S1B, "heterozygosity"). The peak CT for heterozygosity was 0.92 (transition zone: 0.92–0.95) while the paralog peak was 0.88 (transition zone: 0.88–0.95). These maxima were preceded by irregular jumps of adjacent medians and an intensity change of the slopes (Fig. S1.B). This area was enclosed by the transition zone of the average read depth per cluster trend, which was indicated by an increasing Q3 compression and a steady slope shift (Fig. 3a, Fig. S1.B, "avg. depth total", transition zone: 0.92–0.95). The CTs within the described transition zones were averaged to a consensus CT of 0.93 (Table S8, "ISC consensus CT").

For the ISC of tetraploid samples (Fig. 3b and S1.C, "ISC 4n"), undermerging was initiated by a Q3 compression within the transition zone of the number of clusters and increased in slope from CT 0.94 on (Fig. 3b and S1C, "clusters total", transition zone: 0.92–0.93), while allelic variation also started to decline steeply with increasing CTs (Fig. 3b and S1.C, "heterozygosity", peak at 0.94, transition zone: 0.89–0.94). The transition zone of the average depth per cluster showed a steadily declining trend, a few slight median jumps and an increasing Q2 compression (Fig. 3b and S1.C, "avg. depth total", transition zone: 0.89–0.92). The transition zone of filtered paralogs showed a prominent median jump and a moderate slope decline towards the undermerging area (Fig. 3b and S1.C, "filtered by maxH", peak at 0.90, transition zone: 0.90–0.92). The averaged consensus CT was 0.91 (Table S8, "ISC consensus CT").

The scatter plots of the ISC metrics showed that some samples can have a larger effect on the overall trend of a metric than others. For instance, the sample "A_tort_RIII_A36_J49" (*A. tortuosum* subsp. *tortuosum*) showed one of the lowest average cluster depths ("avg. depth total") while a high number of clusters ("clusters total") was found (Fig. S1.B). It also showed the highest amount of filtered paralogs ("filtered by maxH") and a two times higher heterozygosity than the other diploid samples, although flow cytometry confirmed its diploid status (Table S1). The tetraploids also showed some samples that were clearly different from the others (Figure S1.C).

For the BSC threshold selection (Fig. 3c), the undermerging area was indicated by the steady increase in retained loci while the sequence variation (VAR) started to decrease at CT 0.92. At this point, the missingness of the assembly matrix was still low before it increased abruptly starting at CT 0.92. According to McCartney-Melstad et al. (2019) and Mastretta-Yanes et al., (2015), a suitable CT is right before the decrease in sequence variation and the steep increase in missingness while the sample coverage (retained loci) still increases, at CT 0.91. The hockey-stick signal was identified by the first positive swing of the "blade" following the NPL minimum (Fig. 3c, "new polymorphic loci", Paris et al., 2017). This upward swing was in the transition of the CTs 91/90 that corresponds to a CT of 0.91 (Table S8, "NPL") and thus supports the other requirements. We selected 0.91 as BSC threshold.

### 3.3. ipyrad assembly output

The average total read depth (avg_depth_total) for the diploid and tetraploid samples was 6.21 (±2.17) at CT 0.93 and 5.55 (±1.80) at CT 0.91, respectively (Supplementary Table S9, "ISC 2n", "ISC 4n"). After

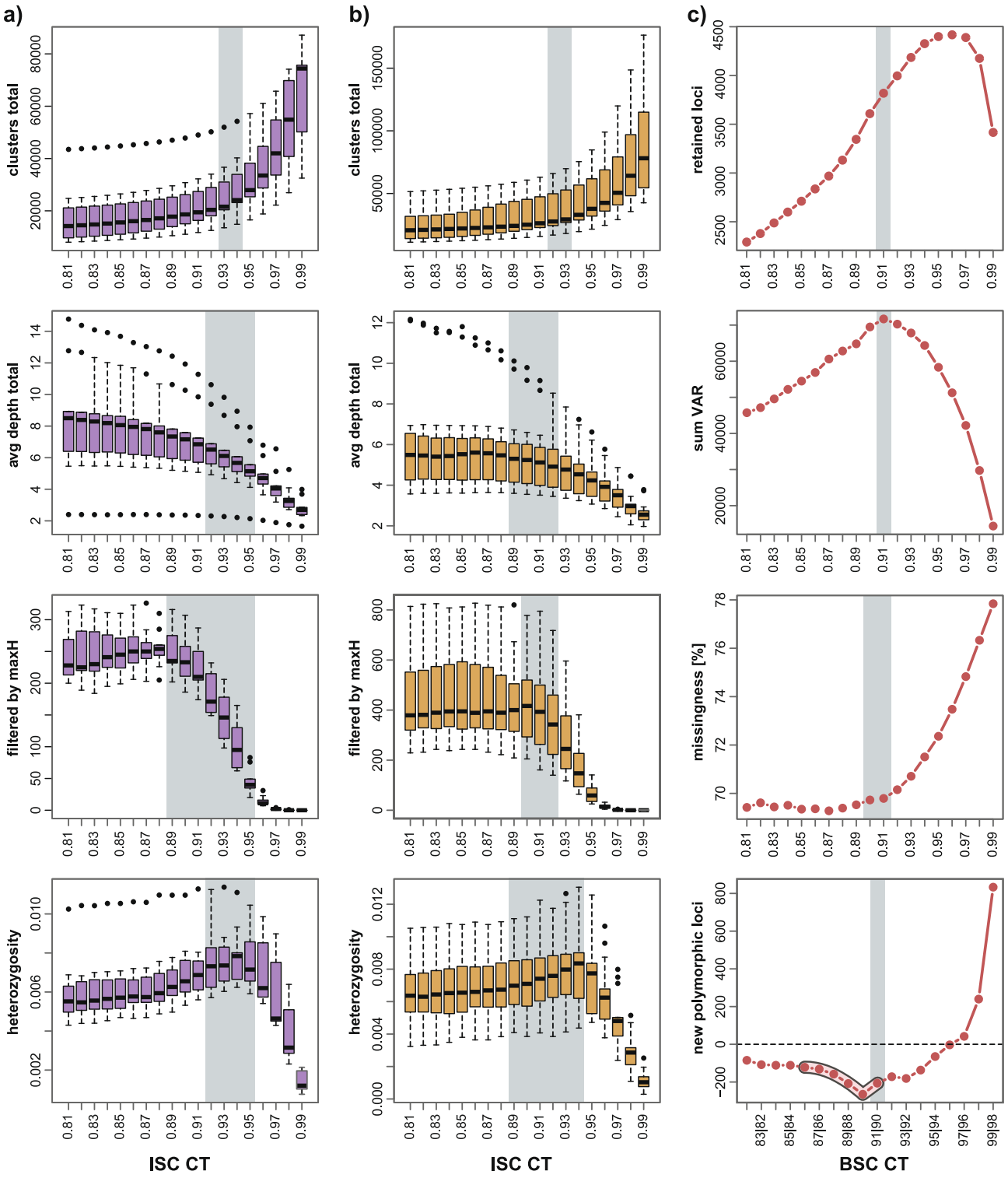



**Fig. 3.** To determine suitable thresholds for in-sample-clustering (ISC) and between-sample-clustering (BSC), trends of several metrics tested across a CT range of 0.81–0.99 were evaluated. For ISC threshold selection of the diploid (a) and tetraploid (b) samples, the number of clusters, the average read depth, flagged paralogs (filtered by maxH) and the allelic variation (heterozygosity) were recorded and plotted. Transition zones from the over- to the undermerging area containing several suitable CTs are shaded in grey. CTs within these zones were averaged to a consensus CT. To select a suitable threshold for clustering between samples of the merged ISC assemblies, the number of retained loci, the retained sequence variation (VAR), the missingness and the number of new polymorphic loci (NPL) were recorded (c). The "hockey stick signal" in the NPL plot, which indicates the assembly containing most accurately clustered sequence variation, is in line with the requirements for the other metrics.

applying the min_depth threshold of 10 for clustering, the majority read depth (avg_depth_mj) rose to 40.24 (±7.52) for the diploid and 39.22 (±17.10) for the tetraploid samples. On average, 26,280 (±11,873) clusters per individual were found for the diploids and 34,436 (±15,023) clusters per sample for the tetraploids. The average count of consensus reads was 2,635 (±692) for the diploid and 2,633 (±645) for the tetraploid samples.

The unfiltered assembly using a BSC threshold of 0.91 comprised 3,818 loci and 71,691 variable sites (Table 1, Fig. 2, "raw" assembly). Of these variable sites, 36,413 were unique SNPs and 35,278 were PIS. 92 loci showed no variation and 581 loci contained no PIS. The dataset included 69.79% missingness, on average 10 unique SNPs and 9 PIS per locus. The retained loci had an average length of 376nt (±93) with a maximum locus length of 618nt (including uncalled bases and gaps). The majority of retained loci ranged in length from 250 to 550nt (Table S9, "locus coverage"). The assembly length range > 500nt showed a prominent gap at ca. 540-580nt, after which a denser region with some samples of comparatively high coverage followed, at ca 590nt. Locus coverage per sample was fairly heterogeneous with an average of 1,242 (±385) and ranged from a minimum of 640 loci for *A. laxum* A29_J41 to a maximum of 2,092 loci for *A. roseum* A01_J02 (Table S9, "sample coverage"). The two outgroup samples contained 127 (*M. icterica* M30_N36) and 155 loci (*M. icterica* BG Bonn) in the final assembly. The BLAST results showed that our dataset contained 21 loci (118 SNPs and 66 PIS) with identities of 78.5–100% with the reference plastomes. After removing non-parsimony-informative loci and cp loci, the dataset contained 3,225 loci with an average of 67.69% missing data. Each locus contained on average 10 SNPs and 11 PIS and had an average length of 379nt (±93) (Table 1, "cleansed", Fig. 2c, "cp loci + 0-PIS loci removal").

### 3.4. Initial inference of the raw dataset and clade definition

Phylogenetic inference of the *ipyrad* "raw" assembly resulted in incongruent topologies (Table 2, Fig. S2). CA-ML (Fig. S2.A) and CB-SM (Fig. S2.B) yielded unsupported backbones, while the SVD reconstruction was fully supported (Fig. S2.C). All trees showed five well supported main clades: clade 1 comprised *A. laxum*, *A. pachycaulon* subsp. *parviflorum* and *A. palmense*, clade 2 included two subspecies of *A. pachycaulon,* subsp. *immaculatum* and subsp. *pachycaulon*, clade 3 was formed by three species from Madeira (*A. villosum*, *A. dumosum* and *A. divaricatum*), clade 4 comprised both subspecies of *A. tortuosum* and clade 5 comprised all remaining taxa (*A. roseum*, *A. punctatum*, *A. bituminosum*, *A. porphyrogenetos*, *A. brevipetalum*, *A. bollei* and *A. parlatorei*) as well as two subspecies of *A. pachycaulon,* i.e., *A. pachycaulon* subsp. *praetermissum* and subsp. *gonzalezhernandezii*. Relationships among clades was not resolved due to a lack of reliable BS support among reconstructions.

### 3.5. Locus filtering

The 3,225 loci of the "cleansed" dataset (Table 1, Fig. 2c) were first filtered respecting the locus coverage (minimum number of samples required), the locus variability (VAR/locus length/number of samples) and locus length intervals by 50nt steps. The properties of the resulting sub-datasets were recorded and phylogenies were inferred using CB-SM (see 3.5.1, Table S6, Supplementary Figure S3, all tree files available at Mendeley, https://doi.org/10.17632/yb6fd93dbw.1). For the second locus filtering, in addition to the length interval sub-datasets, the loci were filtered requiring an increasing, cumulative maximum length ("max length") and subjected to phylogenetic inference using CA-ML and CB-SM (see 3.5.2, Table S7, Supplementary Figure S4, all tree files available at Mendeley, https://doi.org/10.17632/yb6fd93dbw.1).

#### 3.5.1. Locus filtering by coverage, variability, length intervals and dataset selection based on average missingness

We created six "min samples" sub-datasets by increments of four (Fig. 4a, Table S6, Fig. S3). The locus count and sequence variation (total) decreased as the number of samples increased (Fig. S3.A1 and S3.A4). The average number of SNPs per locus remained nearly constant across datasets, whereas the number of PIS per locus increased proportionately with VAR/locus until the "min_samples_16" dataset and then remained constant when increasing the parameter (Fig. S3.A4 and S3.B1). As expected, missingness declined with increasing number of samples (Fig. S3.B1 and S3.C1). The average locus length was constant across the datasets (Fig. S3.B1 and S3.C4). The branch support values of the CB-SM phylogenies showed a steady, slightly decreasing pattern across the datasets (Fig. S3.D1 and S3.D2). The backbone and within clade support values were around 80 and dropped by ca ten points with the "min_samples_24 dataset". The average clade branch support was close to 100 in all datasets.

Seven sub-datasets were filtered for the "min var" parameter (Fig. 4b, Table S6, Fig. S3). The number of loci and sequence variation (total) decreased with increasing minimum variability (Fig. S3.A2 and S3.A5).

**Table 2**
Bootstrap support values and concordance factor values and differences of the inferred datasets using CA-ML, CB-SM and SVD.

| inference method | CA-ML | | | CB-SM | | | SVD | | |
|---|---|---|---|---|---|---|---|---|---|
| dataset | raw | int_251-500 | int_301-450 | raw | int_251-500 | int_301-450 | raw | int_251-500 | int_301-450 |
| BS backbone branches | 86.80 | 99.20 | 99.20 | 83.06 | 90.14 | 96.70 | 100 | 100 | 100 |
| BS clade branches | 94.80 | 100 | 99.60 | 99.92 | 99.98 | 99.08 | 100 | 100 | 100 |
| BS within clade branches | 93.71 | 95.29 | 94.41 | 80.25 | 83.92 | 83.94 | 100 | 100 | 100 |
| BS all branches | 92.63 | 96.89 | 96.26 | 84.41 | 88.05 | 89.10 | 100 | 100 | 100 |
| CF clade 1 | 44.4; 69.2; 24.8 | 44.5; 68.8; 24.4 | 46.7; 69.4; 22.7 | 45.4; 69.1; 23.8 | 45.0; 68.0; 23.0 | 47.9; 68.7; 20.7 | 45.6; 70.0; 24.4 | 44.5; 69.0; 24.6 | 45.1; 61.3; 16.2 |
| CF clade 2 + 3 | 48.1; 62.3; 14.2 | 48.7; 62.3; 13.6 | 50.0; 58.5; 8.5 | 42.6; 72.4; 29.8 | 43.1; 72.8; 29.7 | 44.6; 70.0; 25.4 | 43.2; 72.1; 28.8 | 41.8; 71.7; 29.9 | 43.1; 73.7; 30.6 |
| CF clade 4 | 40.1; 64.1; 24.0 | 40.5; 64.5; 24.1 | 42.5; 66.1; 23.6 | 36.4; 60.1; 23.7 | 40.9; 65.4; 24.5 | 42.5; 65.8; 23.3 | 37.6; 54.4; 16.8 | 40.9; 63.8; 22.9 | 36.6; 59.1; 22.6 |
| CF clade 5 | 17.4; 57.8; 40.5 | 17.5; 57.8; 40.3 | 17.4; 58.4; 41.0 | 16.1; 59.8; 43.7 | 18.9; 61.0; 42.1 | 19.9; 61.5; 41.6 | 18.7; 60.8; 42.1 | 18.4; 61.7; 43.3 | 17.3; 60.4; 43.1 |
| CF clade branches | 56.2; 83.3; 27.1 | 55.9; 83.2; 27.3 | 58.7; 81.7; 22.9 | 51.5; 80.8; 29.3 | 55.8; 81.5; 25.7 | 59.3; 80.2; 20.9 | 57.6; 83.2; 25.6 | 54.9; 81.7; 26.8 | 53.2; 79.2; 26.1 |
| CF backbone branches | 58.6; 75.9; 17.3 | 55.9; 69.5; 13.6 | 57.9; 71.2; 13.3 | 55.0; 68.1; 13.1 | 55.9; 69.6; 13.7 | 57.9; 71.6; 13.6 | 49.1; 62.7; 13.6 | 49.9; 66.7; 16.7 | 48.9; 64.8; 15.9 |

Given are the average BS support values (sectional and total) and the average gene (gCF) and site concordance factor (sCF) values (of the within clade branches, the clade branches and backbone branches) of the inferred datasets ("raw", "int_251-500″, "int_301-450″, "int_251-500_short", "int_301-450_short") using CA-ML, CB-SM and SVD. The average concordance factor (CF) values are shown in this order: gCF; sCF; gCF-sCF-difference.

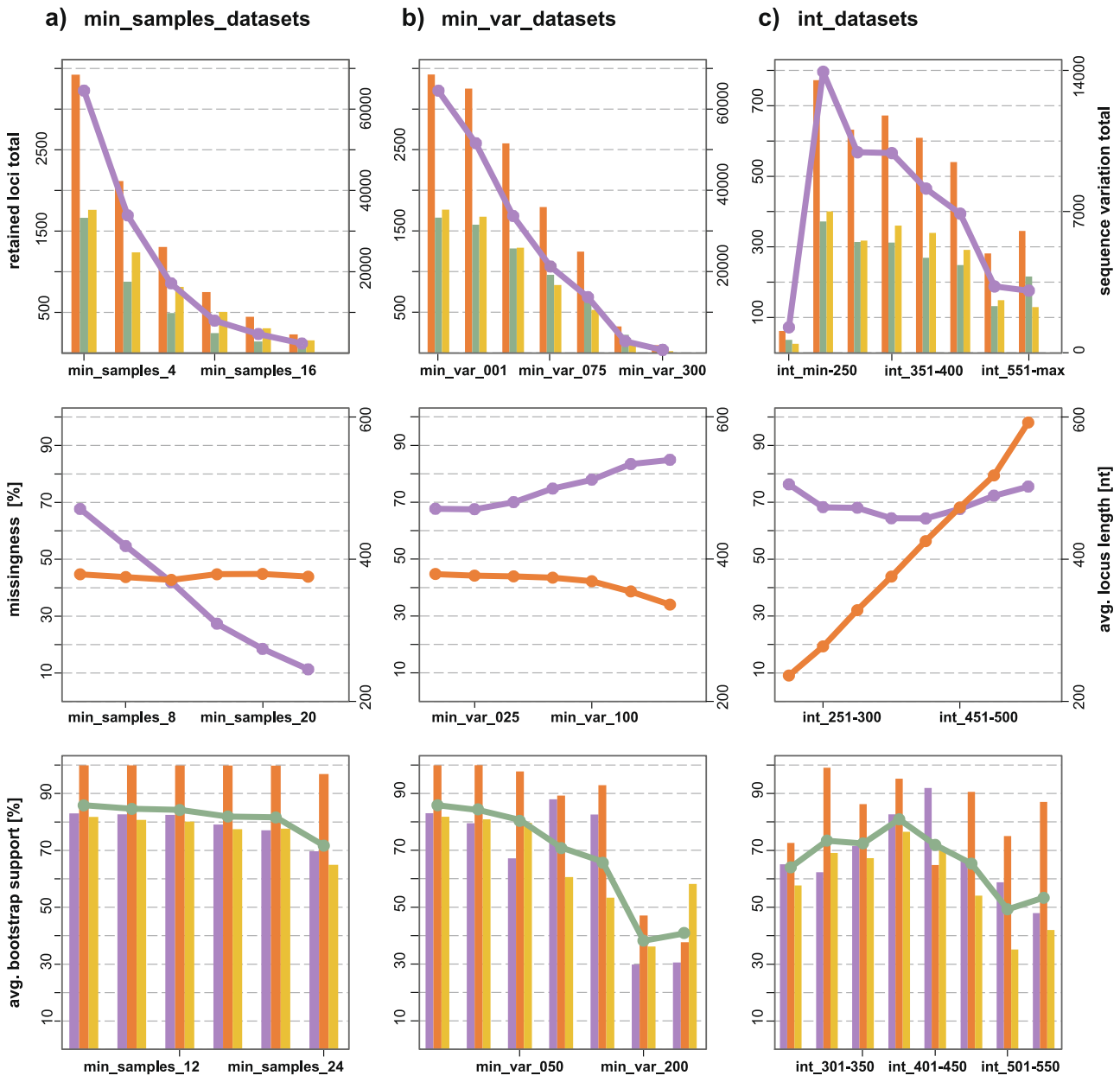



**Fig. 4.** The loci of the „cleansed" assembly were rearranged into sub-datasets based on the minimum number of samples required (a), the minimum variability required (b) and locus length intervals (c). For each sub-dataset, properties such as the number of retained loci (upper plots, purple line with data points), sequence variation (orange = VAR, green = SNPs, yellow = PIS), the average missingness (middle plots, purple line with data points) and average locus length (orange line) were recorded. The average BS support values of the resulting CB-SM trees are given in total (bottom plots, green line with data points) and for the three sections (purple = backbone branches, orange = clade branches, yellow = within clade branches). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In terms of the sequence variation (VAR) total and per locus, the ratio of SNPs to PIS shifted towards a higher SNPs proportion with increasing required minimum variability (Fig. S3.A5 and S3.B2) and missingness increased as well (Fig. S3.B2 and S3.C2). The average locus length decreased slightly with increasing variability required, with the "min_var_300" sub-dataset showing a clear shift towards shorter loci (Fig. S3.B5 and S3.C5). The BS support values showed a decreasing trend (Fig. S3.D2). The tree topologies received varying support across the sub-datasets. The backbone branches were supported highest for the "min_var_075" and "min_var_100" datasets, while the clade and within clade branches had highest support values in the "min_var_001, 025, 050" datasets. The average branch support decreased with increasing missingness (Fig. S3.D5).

The properties and resulting support values of the eight length interval datasets showed irregular trends (Fig. 4c, Table S6, Fig. S3). The amount of loci and sequence variation total (excluding the first sub-dataset containing only 72 loci) dropped from the highest value at "int_251-300" to the adjacent dataset, then rose and declined moderately until the next sharp decline from "int_451-500" to "int_501-550" (Fig. S3.A3 and S3.A6). The average sequence variation per locus was rising with increasing locus length. The proportions of SNPs and PIS in sequence variation (VAR) shifted towards a higher proportion of parsimony-uninformative sequence variation for the datasets "int_min-250" and "int_551-max", respectively (Fig. S3.B3). The missingness had a slightly convex trend with maxima for the flanking datasets (Fig. S3.B6 and S3.C3). The steadily increasing trend of the locus length showed unexpected averages for the two datasets containing the longest loci (Table S6, Fig. S3.B6 and S3.C6), matching the uneven locus length

distribution of the "raw" assembly (Table S9, "locus length distribution" and "locus coverage"). The resulting branch support values showed contrasting patterns (Fig. S3.D3 and S3.D6). The overall trend was shaped concavely. The backbone support initially increased to a maximum at "int_401-450" and then decreased with increasing locus length. The average clade support values were highest at "int_251-300", "int_351-400" and "int_451-500". The within clade branches were supported highest by the "int_351-400" sub-dataset, embedded in a descending trend towards the dataset edges.

Regarding the "min_samples" and the "min_var" datasets, the results were as expected and consistent with findings of previous studies (e.g. Chen et al., 2015; Huang and Knowles, 2016; Eaton et al., 2017; Molloy and Warnow, 2018,). For both parameters, the overall support decreased with increasing requirements, likely due to the simultaneous decline in number of loci and sequence variation. The irregular trends of the locus length interval datasets provided useful clues for subsequent dataset selection and further filtering (see 3.5.2). The trends observed here, together with the declining read quality (Table S4), the heterogeneous coverage of samples and loci, and the irregular assembly coverage respecting the over- and under-represented locus length ranges from ca. 250-280nt and ca. 540-580nt (Table S9), fit the definition of so-called "biased missingness" (Xi et al, 2015, 2016; Hosner et al., 2016; Sayyari et al, 2017; Molloy and Warnow, 2018). To reduce this impact, we selected the average proportion of missingness (69.58% for the length interval datasets) as threshold and discarded all datasets above this cut-off. The retained "int_251-500" dataset (Table 1, "int_251-500") consisted of 2,788 loci, containing in total 56,448 (20.24 ± 15.7 on average) VAR, 26,533 (9.51 ± 8.66) SNPs, 29,915 (10.73 ± 10.76) PIS, 66.66% missingness (9.67 samples/locus) and the average locus length was 360 (±70nt). The locus truncation to one third of the original length lead to a 2/3 reduction of sequence variation and locus length (Table 1, "int_251-500_short").

### 3.5.2. *Locus filtering by length intervals and increasing maximum length and dataset selection based on data qualities and phylogenetic patterns*

The conspicuous trends of the length interval datasets in terms of SNPs/PIS ratio and missingness/locus coverage relative to the resulting BS support values of the species tree sections motivated further filtering to narrow down the extent of potential biased missingness (Table S7 and Fig. S4).

For the locus length interval datasets, CA-ML showed the lowest and highest average BS support values for the "int_0-250" and "int_251-300" datasets, respectively (Fig. S4.C2 and S4.D1). The average branch support decreased steadily with increasing locus length. The three branch sections were irregularly supported by different sub-datasets. The highest count of terraced branches was found for both CA-ML and CB-SM for the "int_0-250" dataset (Fig. S4.D2). The second highest counts were recorded for the sub-datasets "int_501-550" and "int_551-max", respectively. CA-ML resolved the fewest terraced branches for the "int_301-350" and "int_351-400" sub-datasets. The CB-SM trees showed the smallest counts for the "int_251-300" and "int_351-400" datasets, with the latter having the highest average BS support value (Fig. S4.D1 and S4.D3).

For the maximum length datasets, CA-ML showed the lowest sectional and total average BS support values for the first two sub-datasets (Fig. S4.C4 and S4.D2). Then, the BS support raised sharply for the "max_350" sub-dataset and increased steadily up to the maximum for the "max_500" sub-dataset. Beyond this point, there was no gain in branch support. The CB-SM branch support values were lowest for the "max_250" sub-dataset, increased slightly until the "max_350" sub-dataset, showed a strong gain for the "max_400" and a maximum value for the "max_450" sub-dataset (Fig. S4.C2 and S4.D2). Then, the average BS support decreased with increasing maximum locus length, in particular the backbone section lost support. CA-ML and CB-SM resolved the highest terraced branch count for the first sub-dataset (Fig. S4.D4). The number of terraced branches decreased to a minimum of two for the CA-ML trees with increasing maximum length required. CB-SM resolved the fewest terraced branches for the "max_400" sub-dataset. With the addition of loci up to 500nt length ("max_500") the terraced branch count increased strongly and remained high up to the maximum locus length ("max_length").

For the final dataset selection, we classified all recorded locus properties of the sub-datasets and the phylogenetic patterns of the resulting trees into three categories, respectively (Fig. S4.E). The two extreme datasets of both assembly edges were either over- or under-represented (Table S9, "locus length distribution" and "locus coverage"). Those sub-datasets showed also a higher or almost equal ratio of SNPs to PIS relative to the average VAR per locus (Fig. 4, "int_datasets", Table S7, Fig. S4.A1-A6). The average missingness was highest for the filtering parameter edges and decreased towards the inner medium parameters (Fig. S4.B1-B4). The expected average locus lengths were met by the inner filtering parameters, while the values of the sub-datasets increasingly diverged towards the assembly edges (Fig. 4, "int_datasets", Table S7, Fig. S4.B5 and S4.B6). Both CA-ML and CB-SM showed the highest sectional and total BS support values for the inner filtered sub-datasets, with the highest gain for the "max_350" and "max_450" sub-datasets (Fig. S4.C1-C4 and S4.D1-D2). The BS support values of the backbone section profited most within this locus length range. Both approaches resolved the highest number of terraced branches for the filtering parameter edges (Fig. S4.D3-4). The terraced branch count decreased with increasing maximum locus length and increased again strongly beyond a locus length of 450nt for the CB-SM trees. With this locus length also the BS support values started decreasing steadily (Fig. S4.D2 and S4.D4). In summary, the locus properties and phylogenetic patterns associated with non-randomly distributed missingness or biased data were strongest at the filtering parameter edges, while the length ranges from 300 to 450nt appeared to be less affected (Fig. S4.E). The selected "int_301-450" dataset (Table 1) consisted of 1,599 loci of an average length of 373nt (±43nt), containing 15,673 SNPs (avg. 9.82), 17,808 PIS (avg. 11.24) and 65.56% missingness. Truncation resulted in a 2/3 reduction of locus properties (Table 1, "int_301-450_short").

## 3.6. *Phylogenetic inference*

We used three datasets for comparative phylogenetic inference (Table 1, Fig. 2c and 2d). The 3,818 loci of the "raw" assembly were used for initial inference and clade definition (see 3.4). We removed both cp and non-informative loci from this dataset. The retained 3,225 loci of the "cleansed" dataset were the input for the locus filtering approach (see 2.5 and 3.5). The first locus filtering by coverage, variability and length intervals resulted in the "int_251-500" dataset (see 2.5.1 and 3.5.1). The second locus filtering was intended to reduce the presumed biased phylogenetic signal by using phylogenetic patterns relative to the underlying sub-dataset qualities to detect impacted assembly areas (see 2.5.2 and 3.5.2). This approach yielded the "int_301-450" dataset. The filtering steps reduced the number of loci by 58% and the amount of PIS by 50% (Table 1, "raw" compared to "int_301-450"). Sequence variation and locus coverage increased slightly while the average missingness decreased by 4%. Loci-per-sample coverage decreased from an average of 1,166 to 549 loci while sample-per-locus coverage became more homogenous (Table S9, "sample coverage"). Hence, we assumed the "raw" assembly to contain the most, the "int_251-500" dataset to contain less, and the "int_301-450" dataset to contain the least biased phylogenetic signal. Filtering parsimony informative characters (unlinked PICs) resulted in three datasets for the SVD analyses (Table 1). The loci of the "raw" assembly were truncated to one third of their original length, rearranged respecting the locus filtering results and species relationships were inferred with CA-ML and CB-SM to compare potential performance differences in terms of locus length (Table 1, "short", Fig. 2c and 2d).

*3.6.1. Comparative phylogenetic inference of the un-/filtered datasets*

For the "raw" datasets, CA-ML (Fig. S2.A) and CB-SM (Fig. S2.B) resolved incongruent and weakly supported backbone topologies. The CA-ML tree showed an unresolved relationship between the clades 2, 3 and 4. CB-SM inference resulted in an unresolved relationship of clade 1 to clades 2, 3 and 4, with low support and low concordance factor values. The SVD tree (Fig. S2.C) showed full support for a third topology. However, the concordance factor values for the relationship of clade 1 to clade 5 were low. The within clade topology differed among all reconstructions.

For the "int_251-500" dataset, CA-ML (Supplementary Figure S5.A) and CB-SM inferences (Fig. S5.B) resolved congruent backbone topologies, however, for CB-SM the relationships of clades 2 + 3 + 4 to clade 5 lacked support. The concordance factor values increased compared to the "raw" dataset. The SVD tree (Fig. S5.C) showed a maximally supported conflicting topology with low concordance factor values for the relationship of clade 2 to clades 1 + 3 + 4. The within clade topology differed among all reconstructions.

For the "int_301-450" dataset, CA-ML (Supplementary Figure S6.A) and CB-SM (Fig. S6.B) inference resulted in a well-supported, congruent backbone topology (Fig. 5). Concordance factor values for the backbone and clade branches were similar. Again, the SVD tree (Fig. S6.C) showed a maximally supported conflicting topology but low concordance factor values for the relationship of clade 2 to clades 1 + 3 + 4.

*3.6.2. gCF and sCF values obtained with IQ-Tree*

Dataset reduction with respect to the exclusion of potentially biased assembly areas, clearly showed an improvement regarding the concordance factor values and differences for the CB-SM reconstructions (Table 2). The factor difference decreased for all within clade branches and clade branches. The factor values of the clades 1, 2, 3, and 5 decreased stronger compared to clade 4. The gCF value of the clade branches increased by more than 8% compared to the unreduced dataset, while the sCF value decreased slightly. Interestingly, the factor values of the backbone branches increased slightly while the difference increased slightly as well.

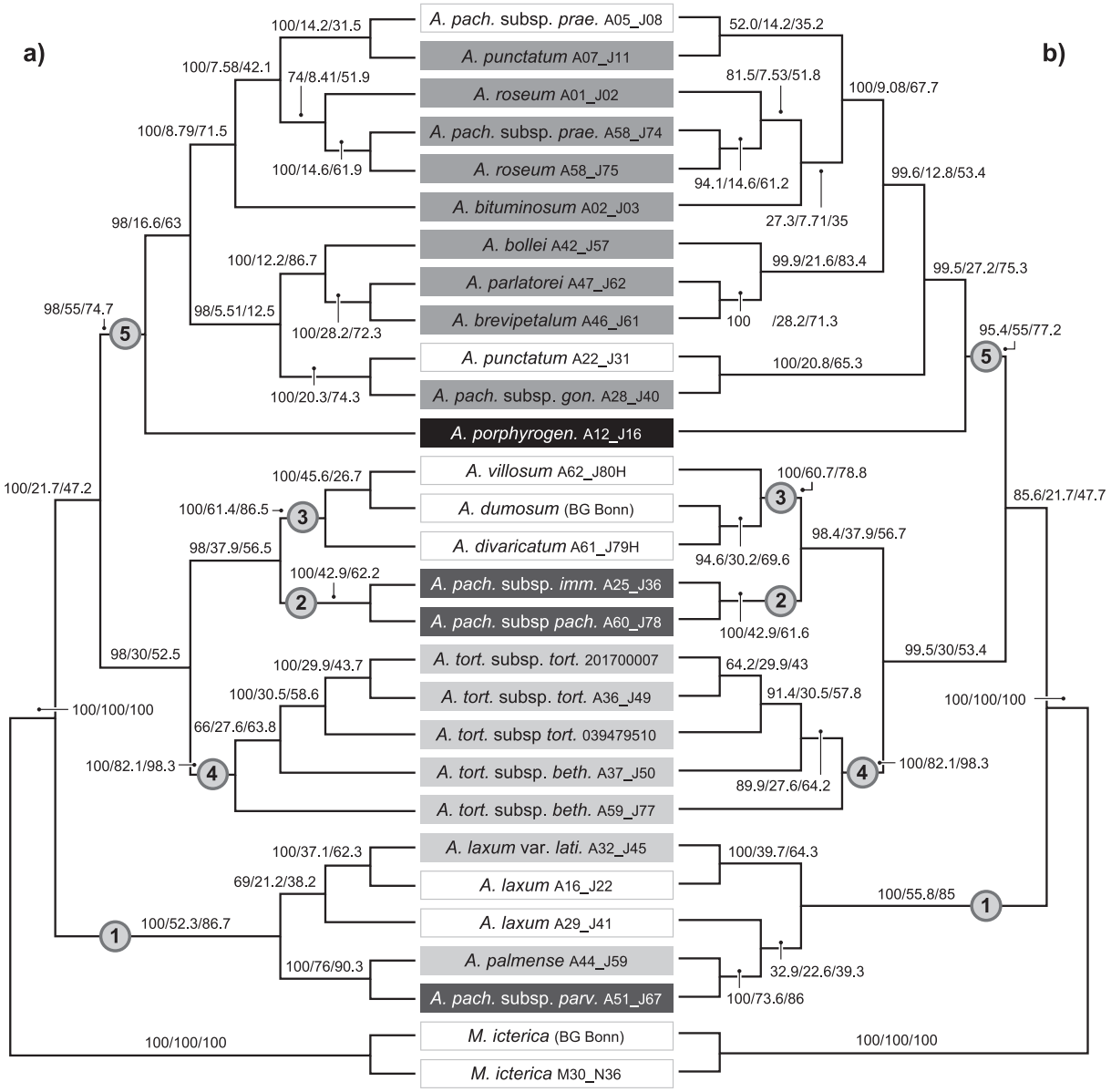



**Fig. 5.** The CA-ML (a) and CB-SM (b) phylogenies of the "int_301-450" dataset. Bootstrap support, gene and site concordance factor values are given above branches. Clades are indicated by the encircled numbers 1–5. Boxes shaded in light and dark gray indicate diploid and tetraploid samples, respectively. The sample *A. porphyrogennetos* A12_J16 showed an intermediate genome size and was treated as tetraploid (black box).

Concordance factors of CA-ML inference showed a similar pattern compared to CB-SM. Overall, the factor values increased with increasing dataset reduction. However, the effect was less pronounced compared to CB-SM and clade 5 even showed an increased factor difference. Notably, the effect of the factor differences for the clade branches was smaller while for the backbone branches it was larger, compared to the CB-SM reconstruction.

In general, the factor effects of the SVD reconstructions were in strong contrast to CB-SM and CA-ML. The SVD factor values were lower compared to CB-SM and CA-ML, and the factor differences raised for the clade and the backbone branches. For the within ancestral branches of clade 2 + 3 and all descendant relationships, the factor difference decreased strongly.

In terms of the resulting BS support values, data reduction had the strongest effect on the backbone branches with an increase in support by ~ 13% for CA-ML and ~ 16% for CB-SM (Table 2). Still, the gCF and sCF values suggest alternative topologies for the relationship of clades 2 + 3 + 4 to clade 5.

#### 3.6.3. Phylogenetic inference of the truncated locus datasets

Inference of the truncated datasets using CA-ML (Supplementary Figure S7.A and S7.B) and CB-SM (Fig. S7.C and S7.D) resulted in alternative topologies compared to the full-length datasets, while also exhibiting distinctly lower concordance factor values and larger factor differences (Supplementary Table S10) or insufficient BS support for the backbone section. The BS support values decreased with decreasing locus length and the decrease was strongest in the backbone branches. The concordance factor values were mostly lower compared to the full-length datasets and the factor difference for the clade and backbone branches increased clearly for all reconstructions.

## 4. Discussion

Modification of several modules of the RADseq toolbox, inspired by GBS (Elshire et al., 2011) and ddRADseq (Peterson et al., 2012), has enabled a strong reduction of the number of targeted fragments. In addition, employing the maximum capacity for sequencing resulted in an extended locus length of up to 618nt. The CT selection approach enabled an informed selection of ISC/BSC thresholds for homology assessment of assembled loci. The locus filtering approach, based on properties known to affect phylogenetic inference, provided the opportunity to observe dataset-specific trends and identify potential adverse properties of the sub-datasets. Additional filtering using phylogenetic patterns for bias detection turned out to improve overall resolution, in particular for CB-SM inference. Besides these positive outcomes, there were also many challenges whose critical consideration led to suggestions for further improvements.

### 4.1. Lab workflow

Compared to other studies employing a RADseq approaches for sample preparation (e.g. Escudero et al., 2014; de Oca et al., 2017; Dillenberger and Kadereit, 2017; Hamon et al., 2017; Wagner et al., 2018; Gerschwitz-Eidt and Kadereit, 2019; Paetzold et al., 2019; Rancilhac et al., 2019; Hipp et al., 2020; Karbstein et al., 2020) we increased the fragment length range and thus the length of assembled loci clearly by shifting the size selection window and fully exploiting the sequencing range of 300nt PE. However, the raw reads varied strongly both in quantity and quality across the samples, which led to a loss of locus and sample coverage, in particular within the higher length range targeted (Supplementary Table S9). This biased distribution of phylogenetic information represented a substantial challenge to data evaluation.

Our lab workflow aims at long RAD loci and has been modified in three aspects: First, we included a specific size selection window ranging from 300 to 600nt for the resulting fragments of the utilized REases BamHI and KpnI. Second, barcode and common adapters were designed for both REase motifs to sequence all generated fragment types in contrast to the classic ddRADseq approach (compare Peterson et al. 2012). Third, the lab protocol contained two size selection steps to ensure complete removal of fragments outside the target range.

#### 4.1.1. Employed REases

The flexible RADseq toolbox allows the use of various REases of a wide range of qualities for complexity reduction (see also: Andrews et al., 2016; McKain et al., 2018; Parchman et al., 2018). Testing and comparing single and dual enzyme strategies with respect to the desired degree of reduction, or in particular a reduced fragment number and an extended length range, either *in silico* or by sequencing a trial library when there is no reference available, can certainly reduce mutation-based locus dropout and ease library prep and adapter design (see also: Lepais and Weir, 2014; Mora-Márquez et al. 2017; Rivera-Colón et al., 2021). Double-digest approaches, using two REases for digestion (e.g. Peterson et al., 2012), are more prone to restriction site mutation disruption than single-digest protocols (e.g. Elshire et al., 2011). Hence, they tend to yield fewer fragments than single-digest approaches which are therefore more easily sequenced to sufficient depth (Andrews et al., 2016; Harvey et al., 2016; Eaton et al., 2017; McKain et al., 2018). Using the *K. fedtschenkoi* genome for *in silico* double-digest using *BamHI* and *KpnI*, we calculated about 4,400 fragments (see 2.2) and received about 3,800 assembled loci (Table 1, "raw"). The difference of ca. 600 fragments may be due to the loss of loci in the assembly range above 500nt (Table S9). Compared to capturing approaches, which typically produce loci of up to thousands of base pairs in length (e.g., McCormack et al. 2013a; Nicholls et al., 2015) the herein obtained locus length of Ø 376nt and 618nt at most may seem short. Still, the resulting loci showed sufficient sequence variation per locus as input for species tree estimation using CB-SM and were in line with approaches targeting similar length ranges (e.g. Hosner et al., 2016; Blom et al., 2017).

#### 4.1.2. Adapter design

The design of adapters herein was based on the original GBS protocol to include and sequence all generated fragments (see Elshire et al., 2011). However, this approach proved not satisfactory. It did not account for potential chimera formation and index hopping (see also: Van der Valk et al., 2020) and the identical flow cell binding motifs meant a potential reduction in sequencing yield. While in general the sequencing output was not influenced, the second sequencing run, containing the majority of samples, yielded only 50% of the maximum sequencing output of the MiSeq v3 kit (Table S4, "run III"). In addition, the reads flanked by identical cut sites introduced a further step in data processing and locus assembly that could be avoided as the raw data had to be demultiplexed twice. Considering these hurdles, we recommend to design each adapter type for one cutsite motif only and to use an indexing approach that accounts for technical bias (e.g. MacConaill et al., 2018; Bayona-Vásquez et al., 2019).

#### 4.1.3. Size selection window and fragment/locus length distribution

The use of coalescent-based summary methods for phylogenetic inference requires a relatively high quality content of sequence variation per locus to reduce GTEE (Chou et al., 2015; Liu et al., 2015; Mirarab et al., 2016; Xu and Yang, 2016; Molloy and Warnow, 2018). Because the average amount of phylogenetic information in a neutrally evolving locus generally correlates with its length (Blom et al., 2017; Mirarab et al., 2016; Chou et al., 2015; Molloy and Warnow, 2018), we chose a size selection window of approximately 300-600nt (ca. 380-720nt segregation range including the adapter and primer length) to obtain longer fragments and thus more informative loci (Fig. 1, Appendix 1). The 2nd size selection using a ratio of 0.8 parts magnetic bead suspension to one part library suspension is particularly important as it removes fragment artifacts from automated fragment segregation and PCR (Fig. 1f, Appendix 1). Compared to a library prepared with the same protocol but without final purification, the precision of the fragment

length segregation was clearly improved (Appendix 1). The length distribution of the final assembly was overall in the range targeted by the lab protocol. However, the strong decline in sequencing quality of R2 reads (Table S4, "run I-III", "mean quality scores") has resulted in a large degree of missingness in the length range of 500-600nt of assembled loci (Table S9, "locus length distribution"). Moreover, the quality filtering thresholds were set quite strictly (Table S5; Eaton and Overcast, 2020). This prevents assembly of erroneous sequences by discarding reads below a specified threshold for base and overall quality. In our dataset this applied especially to the R2 reads, starting at ca 260nt. Thus, a lot of information was lost by excluding high quality partners of low quality mates. Tan et al. (2019) found that declining base quality and higher error rates of fragments above 500nt are a general issue with multiple Illumina sequencing platforms and kits.

The descriptive analysis of the filtered sub-datasets showed that phylogenetic information across the length intervals provided varying support for different sections of the resulting species trees (see 2.5.2 and 3.5.2, Fig. 4c, Fig. S4, "length interval datasets"). Maximum support for all sections was covered by a locus length range of 300-450nt. Considering this and the decreasing quality of R2 reads, we recommend a size selection window of 300-500nt (ca. 380-620nt segregation range including the adapter and primer length). This might avoid locus loss due to the decreasing sequencing quality of the R2 reads and thereby achieve a more uniform assembly and evenly distributed phylogenetic information. However, other focal groups than *Aichryson* might require longer loci, as the retained variation per locus depends on the taxonomic level of interest and is very group specific.

### 4.2. Data analysis

Assembly and analysis of RADseq data is often challenged by various factors depending on the selected library prep and bioinformatics approach, and, of course, the study group itself. The *Aichryson* data shown here united just about every conceivable challenge known to RADseq data. The samples had varying DNA qualities and were sequenced in three different libraries. The output of the three sequencing runs differed in terms of quantity and quality. The R2 reads showed an unevenly distributed drop in quality starting at about 260nt sequencing length (Table S4). And it turned out that this dataset had not only a high proportion of missing data, but also of biased missingness across the assembly length range, impacting sample and locus coverage (Table S9). Despite these unfavorable circumstances, or maybe because of them, the detailed analyses (Fig. 2), including a CT selection and a locus filtering approach, provided detailed insights into the data properties and their impact on phylogenetic inference.

#### 4.2.1. CT selection approach

Clustering threshold selection approaches aim at determining balanced CTs to establish homology while avoiding clustering of paralogous RADseq loci (e.g., Ilut et al., 2014; Mastretta-Yanes et al., 2015; McKinney et al., 2017; Paris et al., 2017; McCartney-Melstad et al., 2019). For this purpose, assembly metrics are compared across a range of CTs to identify values that meet specified requirements. Application of such methods is becoming increasingly popular (e.g. Herrera and Shank, 2016; Razkin et al., 2016; Paetzold et al., 2019; Rancilhac et al., 2019; Karbstein et al., 2020; Wagner et al., 2020) to ensure the assembly of homologous loci (Shafer et al., 2017; Springer and Gatesy, 2018; McCartney-Melstad et al., 2019; Fernández et al., 2020; Simion et al., 2020). Following these previously proposed criteris, we were able to identify areas that met the requirements in terms of 1) the onset of the undermerging area, in which true orthologs are separated into paralogs (McCartney-Melstad et al., 2019), 2) an area of high heterozygosity with decreased clustering of paralogs (Ilut et al., 2014), 3) a maximized sequence variation count while missingness is minimized (Mastretta-Yanes et al., 2015), and 4) an increasing number of new polymorphic loci (NPL) indicated by the hockey stick signal (Paris et al., 2017). This procedure resulted in an assembly comprising 3,818 loci, of which ~ 84% contained parsimony informative sites (Table 1). The loci showed on average ~ 19–21 variable sites, of which ~ 9–11 were parsimony informative. Since these loci were found to be useful for CB-SM inference, we consider the here selected metrics and CT selection approaches in general as promising tools for an informed selection of thresholds during *de novo* assembly. Still, there are some issues that need to be considered: 1) The results shown herein and assumptions arising from them provide more empirical evidence on previous studies, however, are highly specific to our study group and do not constitute proof in general. Hence, simulation studies with known characteristics and focusing on each of these aspects are urgently required. 2) We selected only a few out of many more possible metrics that can be utilized to evaluate dataset-specific trends, such as the pairwise data missingness and genetic dissimilarity (McCartney-Melstad et al., 2019), the proportion of heterozygous loci in a sample and allelic ratios at each locus (McKinney et al., 2017) or the fraction of sequence variation shared by specific proportions of all individuals (Paris et al., 2017). 3) The selected CTs for ISC and BSC are an adequate representation of a majority of loci but one CT cannot appropriately characterize the entire sequence divergence within and across samples. Various causes of sequence divergence among genomic regions (e.g., coding or non-coding regions, thus degree of sequence conservation, and biological processes such as hybridization, horizontal gene-transfer and ILS) lead to a normalization within a range of suitable CTs, which we here referred to as the "transition zone". 4) Polyploid loci composed of greater allele numbers can show greater heterozygosity than loci composed of lower number of alleles presumably containing less sequence variation across orthologous alleles (Hirsch and Buell, 2013; Karbstein et al., 2021), and thus require different CTs for accurate clustering. Hence, merging of ISC samples of varying ploidy for BSC across all taxa leads to a clustering bias. 5) The resulting data, whether used for metric evaluation or inferences of population structure or species relationships, are heavily impacted by all other parameters chosen, depend on numerous properties of the study system (e.g.: taxonomic level, genomic variation, utilized lab protocols, quality and quantity of data) and will affect downstream analysis (e.g. Huang and Knowles 2016; Eaton et al., 2017; Shafer et al., 2017; Crotti et al., 2019; McCartney-Melstad et al., 2019). 6) Metric trends can be affected by heterogeneous read quality and quantity, as well as biological factors, such as genome size or repetitive regions. This presumably leads to different metric trends of individual samples, as seen in the scatter plots for the ISC threshold selection (paragraph 3.2, Supplementary Figure S1). As a consequence, the selection of potential CTs gets less precise. This problem may be improved by re-splitting samples into groups that show similar trend intensities and using specific CTs for each group. Simulation studies focusing on potential impacts of heterogeneous sample qualities on the CT selection and the resulting assembly are required. Nevertheless, we consider a thorough evaluation of assembly metrics, as shown in this and other studies (e.g. Paris et al., 2017; Paetzold et al., 2019; Rancilhac et al., 2019; McCartney-Melstad et al., 2019; Karbstein et al., 2020; Wagner et al., 2020), to be an improvement over simply using default settings.

#### 4.2.2. Locus filtering

The impact of filtering loci regarding specific properties, such as length, sequence variation or missingness, prior to phylogenetic inference has been investigated by numerous studies (e.g. Chou et al., 2015; Liu et al., 2015; Xi et al., 2015, 2016; Hosner et al., 2016; Mirarab et al., 2016; Huang and Knowles 2016; Sayyari et al., 2017; Molloy and Warnow, 2018). We confirm general trends previously observed regarding locus coverage and sequence variation (see 2.5.1 and 3.5.1, Table S6, Fig. S3). As the minimum requirements increased, the number of loci and sequence variation decreased (Huang and Knowles, 2016; Eaton et al., 2017). This information loss resulted in sharply decreasing BS support values of the resulting species tree estimates. This is likely a result of higher locus dropout in more rapidly evolving loci (for the "min

var" datasets). The more conserved loci are less variable but also less prone to mutation-induced cut-site disruption and thus show a higher sample coverage (for the "min samples" datasets). An interesting point is that the two datasets with the highest minimum variability required (Table S6, "min_var_200" and "min_var_300") also showed a trend toward biased locus lengths. In addition, these loci contained on average more missing data and a higher portion of variable sites was parsimony un-informative. The negative impact of this constellation of locus properties on the accuracy of species tree estimation has been demonstrated by Xi et al. (2015), Hosner et al. (2016) and Lee et al. (2018). This constellation was also evident for the length interval datasets containing the shortest and longest loci at the assembly edges (Table S6 and S7, Fig. S3 and S4). For these assembly regions, we assume that the declining sequencing quality of R2 reads led to biased sample and locus coverage, which was reflected by the prominent gap between 500 and 600nt as well as the high number of loci in the 250-300nt length range of the assembly (Table S9). This kind of data bias causes high GTEE and artificial phylogenetic conflicts among taxa and clades, which negatively affects the species tree estimation performance (Sanderson et al., 2010, 2011, 2015; Simmons, 2012; Hosner et al., 2016; Xi et al., 2016; Sayyari et al., 2017; Dobrin et al., 2018).

To reduce this effect, we first chose a controversial approach and filtered the loci based on average missingness, which resulted in the "int_251-500" dataset. Locus filtering based on missingness is generally not recommended because it can lead to a significant loss of information and thus to a performance decline of phylogenetic inference (Huang and Knowles 2016; Eaton et al., 2017; Molloy and Warnow, 2018; Crotti et al., 2019). However, it can lead to an improvement in estimation accuracy if the extent of biased, non-randomly distributed phylogenetic signal is also reduced (Xi et al., 2015, 2016; Sayyari et al., 2017; Molloy and Warnow, 2018). Although this first filtering and dataset selection resulted in a slight improvement of the data quality and the resulting BS support and concordance factor values, it did not yield the required data quality for a successful CB-SM inference. Simply choosing the average missingness as a cutoff value may improve the quality of loci containing evenly distributed phylogenetic information, but not if the bias is unevenly distributed across the assembly.

To further reduce the extent of the biased assembly area, we binned the loci based on length, inferred CA-ML and CB-SM phylogenies for each sub-dataset and put resulting phylogenetic patterns in relation to sub-dataset properties to detect biased locus length ranges (see 2.5.2 and 3.5.2, Table S7, Fig. S4). This approach turned out beneficial with regard to the selection of less biased assembly areas, suitable for CB-SM inference. The typical responses of BS support values and reconstruction of terraced branches confirmed the assembly's edge regions as particularly biased. In these locus length regions of the assembly, either the BS support values collapsed or the number of terraced branches of the resulting topology was high. Consequently, we selected the remaining, presumably less biased, assembly range of 301-450nt length served as third dataset for comparative phylogenetic inference. While this second filtering and dataset selection procedure represented a drastic reduction of overall data quantity, it also increased data quality as indicated by the average sequence variation per locus, locus coverage/missingness and sample coverage (Table 1, Table S9).

The second filtering approach used here to examine the influence of locus properties on the resulting phylogenetic reconstructions resulted in a dataset favorable for CB-SM inference. However, the process was quite tedious, and at times somewhat crude, which indicates a number of opportunities for further refinement in the future. 1) Loci of certain properties within the excluded assembly ranges are likely to be also well suited for CB-SM inference. We filtered the loci by their relative sequence variation including SNPs and PIS (see 2.5.1). However, the notable PIS/SNPs ratio along with the average locus coverage evident in the locus length filtering (Fig. S3 and S4) may be a clue to filter loci by information quality (Xi et al., 2015; Hosner et al., 2016; Lee et al., 2018). 2) The bin sizes chosen for filtering locus properties might be smaller to enable a more accurate detection of potential trend changes respecting phylogenetic outcomes. 3) We calculated only one reconstruction per inference approach for each sub-dataset. Multiple replicates may be generated to identify and statistically assess potential variations. 4) We found overall matching trends of locus properties relative to the resulting phylogenetic patterns of CA-ML and CB-SM used for bias detection. Considering the presumably strongly biased signal scattered across taxa, the relative influence of technical errors and true biological conditions (e.g. ILS) remain difficult to assess. 5) Instead of multi-locus bootstrapping (Seo, 2008), the branch support might be assessed using Local Posterior Probability, which was shown to perform more accurate on locus trees with relatively high error (Sayyari and Mirarab, 2016) or quartet based methods to identify non-informativeness (Pease et al., 2018). 6) Counting the terrace-like branches in the resulting trees helped to identify biased assembly areas but did not provide insight into the actual underlying conflicts among taxa and clades. Besides, terraced branches can also represent the true topology (Sanderson et al., 2011). To account for artificial conflicts in the data, terrace-aware phylogenetic inference tools can be used (Sanderson et al., 2011, 2015; Chernomor et al., 2016; Dobrin et al., 2018). 7) Further approaches may be tested comparatively to allow for a more accurate data quality assessment, such as filtering for fragmentary data to achieve uniform taxon coverage (Xi et al., 2016; Sayyari et al., 2017) or subsampling specific loci to establish congruence across the dataset (Chen et al., 2015; Simmons et al., 2016). For future projects, an automated pipeline that filters loci based on multiple criteria, records the properties of these bins, and evaluates the resulting phylogenetic patterns, thus simplifying the tedious filtering process, would be of great value.

### 4.3. Phylogenetic inference

Previous attempts at resolving phylogenetic relationships in *Aichryson* were mainly hampered by lack of variability in the employed regions (Mort et al., 2002; Fairfield et al., 2004 which failed to resolve relationships at shallow taxonomic levels (e.g., Miller et al., 2003; Abeysinghe et al., 2009; Duan et al., 2015). The application of a modified RADseq approach together with detailed data processing, analysis of filtered sub-datasets and comparative phylogenetic inference resulted in the first well-supported phylogeny for *Aichryson*. Moreover, we gained further insight into the performance of the tested inference methods with respect to underlying data properties.

#### 4.3.1. General trends of the CA-ML and CB-SM inference during locus filtering

During locus filtering, we initially filtered the loci by variability, locus coverage and length intervals (see 2.5.1 and 3.5.1). Contrary to our expectation, we were not able to reconstruct a well-supported CB-SM phylogeny using this approach. Instead, we found that the BS support values of the three species tree sections responded differently to the underlying locus length interval datasets (Fig. 4, Table S6, Fig. S3). The related locus properties in terms of sequence variation and missingness, as well as the distribution of data across the assembly, loci, and samples (Table S9), indicated a data bias (Sanderson et al., 2010; Hosner et al., 2016; Xi et al., 2016; Sayyari et al., 2017; Lee et al., 2018; Molloy and Warnow, 2018).

Subsequently, we used phylogenetic patterns yielded by CA-ML and CB-SM inference of locus length sub-datasets to detect potentially biased assembly areas (see 2.5.2 and 3.5.2, Table S7, Fig. S4). CB-SM resolved more terraced branches than CA-ML across the tested sub-datasets, in particular when the datasets were small (Xi et al., 2016; Fig. S4, "length interval" datasets). This is likely due to the information loss inherent to the method, using only summary statistics of the inferred gene trees as input for species tree estimation (Xu and Yang, 2016). Along with this come the clearly lower resulting support values of the multi-locus bootstrapping (Seo, 2008) when applied to fragmentary data (Xi et al., 2015, 2016; Hosner et al., 2016; Sayyari et al., 2017). The overall higher

and steadily increasing BS support values with increasing dataset size confirm prior observations regarding CA-ML inference (Kubatko and Degnan, 2007; Liu et al., 2015; Minh et al. 2020a). CA-ML inference of the length sub-datasets seemed less sensitive or more robust to data bias (Xi et al., 2016; Molloy and Warnow, 2018). Still, bootstrapping over the concatenated matrix showed quite similar trends compared to the multi-locus bootstrapping employed with CB-SM.

#### 4.3.2. Comparative phylogenetic inference of the un-/filtered datasets

The filtering steps meant a maximum reduction of 58% for the number of loci and 50% for the number of PIS, while the average sequence variation and coverage per locus raised, average missingness declined and sample coverage became more evenly distributed (Table 1, "raw" compared to "int_301-450", Table S9, "sample coverage").

For CA-ML and CB-SM, the exclusion of presumably biased assembly areas, resulted in increasing statistical support while the concordance factor value differences decreased (Table 2, Table S10, Fig. S2, S5 and S6). These trends were stronger for the CB-SM inferences. The concordance factor values and differences of the within clade branches benefited slightly while those of the clade branches benefited most from reduction. This was accompanied by improved factor values and differences of the backbone branches. We suggest that the overall higher locus coverage and the more evenly distributed information across taxa (sample coverage) of the retained assembly area caused less artificial conflicts among clades and thus favored resolution and support of the backbone section (Sanderson et al., 2010, 2011; Xi et al., 2015, 2016; Hosner et al., 2016; Sayyari et al., 2017; Dobrin et al., 2018; Molloy and Warnow, 2018; Minh et al. 2020a, b). This increasing statistical support coincides with an increase in the number of terraced branches. For instance, the CA-ML and CB-SM inferences of the "raw" dataset reconstructed a dichotomous topology for the taxa of clade 4, but there was insufficient statistical support for the backbone sections (Fig. S2). The backbone topology of the strongly reduced "int_301-450" dataset was well supported, but in exchange the taxa of clade 4 were reconstructed on terraced branches (Fig. 5 and S6).

Phylogenetic inference of the datasets using SVD showed some contradictions. The lower factor values of the backbone branches for the alternative topologies and compared to the CA-ML and CB-SM inferences (Fig. S2, S5 and S6), increasing concordance factor value differences with increasing extent of reduction (Table 2), as well as the consistent maximum BS support values, suggest a random resolution due to limited and unevenly distributed information (Long and Kubatko, 2018; Minh et al. 2020a, b). This is certainly in part due to the selection of individual PICs per locus, which we performed to meet the methods requirements in terms of linkage (Bryant et al., 2012; Chiffman and Kubatko 2014; Xu and Yang, 2016). In addition, studies comparing the performance of inference methods under challenging data conditions showed that SVD is often less accurate than CA-ML and CB-SM (Chou et al., 2015; Molloy and Warnow, 2018). Still, the SVD inferences illustrated potentially conflicting topological alternatives.

In summary, phylogenetic inference of the three datasets ("raw", "int_251-500", and "int_301-450") showed positive trends in terms of the resulting BS support values and concordance factor values with increasing degree of dataset reduction for CA-ML and CB-SM. The resulting SVD reconstructions, however, appeared to be impeded by information limitation and data bias.

#### 4.3.3. Phylogenetic inference of the truncated locus datasets

In general, increasing locus length is associated with increasing phylogenetic information, lower GTEE and thus an increased accuracy of species tree estimation (e.g. Mirarab et al., 2014, 2016; Xi et al., 2015; Chou et al., 2015; Hosner et al., 2016; Xu and Yang, 2016; Blom et al., 2017; Molloy and Warnow, 2018). We expected a decrease in locus length to decrease the total and average phylogenetic information per locus, and consequently to negatively affect performance. To test this, the "raw" assembly loci were truncated and used as input for CA-ML (Supplementary Figure S7 A and B) and CB-SM inference (Supplementary Figure S7 C and D).

The truncated datasets showed a 2/3 reduction in phylogenetic information (Table 1, "int_251-500_short" and "int_301-450_short"), resulted incongruently resolved tree topologies (Fig. S7), and yielded decreased estimated BS support and concordance factor values, while the factor value differences of the clade and backbone branches increased strongly compared to the original datasets (Table S10). Therefore, we conclude that the locus length reduction had a substantially negative impact on the phylogenetic inference. This is in line with findings by studies comparing the inference performance over varying locus lengths and information contents (e.g. Mirarab et al., 2014a, b, 2016; Xi et al., 2015; Chou et al., 2015; Xu and Yang, 2016; Molloy and Warnow, 2018).

However, we performed a drastic locus length reduction by 2/3, which resulted in an average locus length of 120/123nt (Table 1). As we found during locus filtering (see 2.5) and phylogenetic inference of the resulting datasets, an average locus length of 373nt (±43nt) in an assembly range of 300-450nt yielded sufficient phylogenetic information per locus and in total for successful CB-SM inference. Other empirical studies using similar or even shorter length ranges also achieved a successful CB-SM inference of the assembled data (e.g. Curto et al., 2018; Rancilhac et al., 2019). Based on our results, and as found by numerous studies (e.g., Gatesy and Springer, 2014; Lanier et al., 2014; Liu et al., 2015; Xi et al., 2015; Hosner et al., 2016; Huang and Knowles, 2016; Blom et al., 2017; Sayyari et al., 2017; Xu and Yang, 2016; Lee et al., 2018), we suggest that locus quality in terms of the information content and its distribution across the assembly and taxa is of greater importance than mere locus length. Yet, this also strongly depends on the taxonomic level, i.e. sequence divergence, of the study group.

#### 4.3.4. On the accuracy of the Aichryson phylogeny

The accuracy of the phylogenetic outcome is the suggested by the emerging congruence of the CA-ML and CB-SM reconstructions with increasing data quality. Inference of the "int_301-450" dataset yielded overall congruent, similarly well-supported topologies as well as similar concordance factor values and differences. In addition, the phylogenetic pattern matches the species distributions. For instance, the species occurring on Madeira (*A. divaricatum*, *A. dumosum*, *A. villosum*) and the two *A. tortuosum* subspecies occurring on the eastern Canary Islands, Lanzarote (subsp. *tortuosum*) and Fuerteventura (subsp. *bethencourtianum*), each form a monophyletic group. The polyphyletic status of the *A. pachycaulon* subspecies is also consistent with previous studies (Mort et al., 2002; Fairfield et al., 2004).

However, as Goethe put it: „We know accurately only when we know little; with knowledge, doubt increases" (von Goethe, 2012, published postum). 1) *Aichryson* is not a model group and lacks comparable studies in terms of data properties (locus length, sequence variation, missingness), data analysis (data assembly, locus filtering) and phylogenetic inference. 2) We did not statistically assess potential variation in phylogenetic inference of the filtered datasets using multiple replicates. 3) The extent to which phylogenetic inference may be impacted by terraces due to artificial conflicts among clades arising from the data structure herein is unclear (Sanderson et al., 2010, 2011, 2015; Simmons, 2012; Dobrin et al., 2018). 4) Although locus properties gained quality and sample coverage became more even, the low concordance factor values of some backbone branches representing the relationships of clades 2 + 3 + 4 to clade 5 and high concordance factor value differences of the within clade branches of clade 5 suggest a strong conflict among clades and taxa, respectively (Minh et al. 2020a, b). However, we cannot assess whether this incongruence of information among locus trees is a true biological signal due to reticulate evolution or an artifact of the data structure. 5) In addition, the ongoing, sometimes heated debate over the most accurate application, analysis, and inference of a variety of RRL/SRS-based approaches, along with a series of comparisons of divergent concepts and opinions, further complicate the

interpretation of the results (e.g. de Queiroz and Gatesy 2007; Edwards et al., 2007, 2016; Kubatko and Degnan, 2007; Degnan and Rosenberg, 2009; Knowles, 2009; Leaché and Rannala, 2011; Song et al., 2012; Gatesy and Springer, 2013, 2014; Springer and Gatesy 2014, 2016, 2018; Mirarab et al., 2014b, 2016; Chou et al., 2015; Roch and Steel 2015; Mirarab and Warnow 2015; Solís-Lemus et al., 2016; Mendes and Hahn, 2018; Molloy and Warnow, 2018; Bryant and Hahn, 2020; Rannala et al., 2020). In particular, the inference accuracy of CA-ML in the presence of gene tree-species tree discordance (Degnan et al., 2006, Degnan and Rosenberg, 2009; Kubatko and Degnan, 2007; Knowles, 2009; Roch and Steel, 2015; Solís-Lemus et al., 2016; Mendes and Hahn, 2018; Bryant and Hahn, 2020) and the performance of CB-SM under conditions of GTEE (Springer and Gatesy, 2014, 2016; Roch and Warnow, 2015; Xi et al., 2015, 2016; Solís-Lemus et al., 2016; Xu and Yang, 2016; Sayyari et al., 2017; Molloy and Warnow, 2018) raise concerns.

In general, CA-ML and CB-SM are expected to yield congruent results under less challenging conditions of gene tree-species tree discordance (Edwards et al., 2007; Kubatko and Degnan, 2007; Leaché and Rannala, 2011). Comparative studies showed that CA-ML and CB-SM performed equally under various levels of ILS, with CA-ML performing more accurate under challenging GTEE conditions (Chou et al., 2015; Xi et al., 2015, 2016; Mirarab et al., 2016; Sayyari et al., 2017; Molloy and Warnow, 2018). Moreover, inference of empirical data using both approaches generally yielded congruent results (e.g. Chiari et al., 2012; Hosner et al., 2016; Blom et al., 2017; Sayyari et al., 2017; Curto et al., 2018; Rancilhac et al., 2019). The bottom line is that we cannot ultimately assess the accuracy of the species tree for *Aichryson*, still, we construe the overall congruence as supporting the accuracy of the phylogenetic outcome.

### 4.4. Conclusion

The methodology presented in this study successfully led to a coalescent-based inference of our focal group *Aichryson*. For some, however, the series of approaches tested by us may be equivalent to a butcher making "phylogenetic sausage" (for the definition of a "phylogenetic sausage" see: Gatesy and Springer, 2014; see further: Springer and Gatesy, 2016, 2018; Bryant and Hahn, 2020; Fernández et al., 2020; Rannala et al., 2020). Admittedly, all methodological components could be modified and improved in many ways. The resulting data were also quite demanding to analyze. Still, particularly the challenging data structure provided the opportunity to gain further valuable insights to drive the development of fast and reliable RRL-SRS approaches. 1) Minor modifications of the RADseq toolbox regarding fragment size selection and sequencing range yielded a strongly reduced locus set of extended length. 2) Evaluation of a few metrics enabled an informed selection of clustering thresholds for data assembly within and across samples. 3) Simple descriptive statistics of the resulting assembly were useful for an initial assessment of the data structure. 4) Locus filtering greatly assisted to identify assembly areas of presumably biased locus and taxon coverage. 5) Comparative evaluation of phylogenetic patterns, such as terrace-like branches, BS support values and concordance factor values highlighted the importance of data quality over mere quantity, in particular for the coalescent-based summary method.

We are convinced that the combination of highly flexible RRL-SRS laboratory, data analysis, and inference approaches is crucial for a fast and reliable biodiversity exploration. Hence, we highly encourage the community to: 1) modify the extensive RADseq toolbox regarding an extended fragment length and sequencing range, 2) reduce the data quantity in favor of data quality, 3) utilize approaches guiding an informed threshold selection for accurate clustering, 4) thoroughly analyze and test the resulting assembly and locus properties for potential biases, 5) and to compare and evaluate the resulting phylogenetic trends using multiple inference approaches.

## Funding

This work benefitted from the sharing of expertise on RADseq data in various groups of organisms at various taxonomic levels within the Deutsche Forschungsgemeinschaft DFG priority program SPP 1991 TaxonOMICS and from financial support from DFG KA 1816/11–1.

## CRediT authorship contribution statement

**Philipp Hühn:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration, Supervision. **Markus S. Dillenberger:** Methodology, Software, Formal analysis, Writing – review & editing. **Michael Gerschwitz-Eidt:** Methodology, Software, Formal analysis, Writing – review & editing. **Elvira Hörandl:** Resources, Writing – review & editing. **Jessica A. Los:** Investigation, Resources. **Thibaud F.E. Messerschmid:** Investigation, Resources. **Claudia Paetzold:** Methodology, Writing – review & editing. **Benjamin Rieger:** Methodology, Software. **Gudrun Kadereit:** Conceptualization, Resources, Writing – review & editing, Supervision, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Acknowledgements*

We thank the following people and institutions for providing materials for this study: Ángel Bañares Baudet (Tenerife), Stephan Scholz (Lanzarote), Stefan Abrahamczyk (Bonn) and Nadine Bobon (Mainz). We thank Hans Zischler (Mainz) and Dirk Albach (Oldenburg) for providing access to their lab facilities. We thank Ursula Martiné (Mainz) and Silvia Wienken (Munich) for lab assistance. We thank Oliver Hawlitschek and three further reviewers for helpful comments on the manuscript. We are grateful to Doris Franke and Maria Geyer (Mainz) for their assistance with the figure design, and to Christopher Wild (Mainz) for taking care of the living collection of Crassulaceae at the Botanical Garden Mainz.

This work used computing infrastructure of: -the Scientific Compute Cluster at the Göttingen Society for Scientific Data Processing (GWDG), as part of the joint data center of Max Planck Society for the Advancement of Science (MPG) and University of Göttingen, -the supercomputer Mogon at Johannes Gutenberg University Mainz, which is a member of the Alliance for High Performance Computing in Rhineland Palatinate (AHRP) and the Gauss Alliance e.V., and -the Center for Genome Research and Biocomputing at the Oregon State University. We gratefully acknowledge the computing time granted. This research was funded by the German Science Foundation (DFG grant KA1816/11-1) within the priority programme 1991 Taxon-OMICS.

*Sequence data*

Demultiplexed raw data is available at the NCBI Sequence Read Archive (www.ncbi.nlm.nih.gov/sra/PRJNA642981), BioProject PRJNA642981 (www.ncbi.nlm.nih.gov/bioproject/PRJNA642981).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107342.

## References

Abeysinghe, P.D., Wijesinghe, K.G.G., Tachida, H., Yoshda, T., 2009. Molecular characterization of Cinnamon (*Cinnamomum verum* Presl) accessions and evaluation of genetic relatedness of Cinnamon species in Sri Lanka based on *trn*L intron region, intergenic spacers between *trn*T-*trn*L, *trn*L-*trn*F, *trn*H-*psb*A and nuclear ITS. Res. J. Agric. Biol. Sci. 5 (6), 1079–1088.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

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 (2), 81–92. https://doi.org/10.1038/nrg.2015.28.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., Fay, J.C., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3 (10), e3376. https://doi.org/10.1371/journal.pone.0003376.

Bañares Baudet, Á., 2002. On some poorly known taxa of *Aichryson* sect. *Aichryson* and *A. bituminosum* sp. nova (Crassulaceae). Willdenowia 32 (2), 221–230. https://doi.org/10.3372/wi.32.32204.

Bañares Baudet, Á., 2015. Las plantas suculentas (Crassulaceae) endémicas de las Islas Canarias. Publicaciones Turquesa, Santa Cruz de Tenerife.

Bañares Baudet, Á., 2015b. Híbridos de la familia Crassulaceae en las islas Canarias. V. Vieraea 43, 189–206.

Bañares Baudet, Á., 2017. Typification of Aichryson pachycaulon subsp. praetermissum and description of *A. roseum* sp. nov. (Crassulaceae) from Gran Canaria, Canary Islands, Spain. Willdenowia 47 (2), 127–134. https://doi.org/10.3372/wi.47.47204.

Bayona-Vásquez, N.J., Glenn, T.C., Kieran, T.J., Pierson, T.W., Hoffberg, S.L., Scott, P.A., Bentley, K.E., Finger, J.W., Louha, S., Troendle, N. and Díaz-Jaimes, P., Mauricio, R., Faircloth, B.C., 2019. Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). PeerJ 7:e7724. doi: 10.7717/peerj.7724.

Bayzid, M.S., Warnow, T., 2013. Naive binning improves phylogenomic analyses. Bioinformatics 29, 2277–2284. https://doi.org/10.1093/bioinformatics/btt394.

Blom, M.P.K., Bragg, J.G., Potter, S., Moritz, C., 2017. Accounting for uncertainty in gene tree estimation: summary-coalescent species tree inference in a challenging radiation of Australian lizards. Syst. Biol. 66, 352–366. https://doi.org/10.1093/sysbio/syw089.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., Prlic, A., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comp. Biol. 10 (4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

Bryant, D., Hahn, M.W., 2020. The Concatenation Question. In: Scornavacca, C., Delsuc, F., Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 3.4, pp. 3.4:1–3.4:23. No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE. HAL Id: hal-02535651.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molec. Biol. Evol. 29 (8), 1917–1932. https://doi.org/10.1093/molbev/mss086.

Buono, D., Khan, G., von Hagen, K.B., Kosachev, P.A., Mayland-Quellhorst, E., Mosyakin, S.L., Albach, D.C., 2021. Comparative Phylogeography of Veronica spicata and V. longifolia (Plantaginaceae) Across Europe: Integrating Hybridization and Polyploidy in Phylogeography. Front. Plant. Sci. 11 https://doi.org/10.3389/fpls.2020.588354.

Burleigh, J.G., Kimball, R.T., Braun, E.L., 2015. Building the avian tree of life using a large-scale, sparse supermatrix. Mol. Phylogenet. Evol. 84, 53–63. https://doi.org/10.1016/j.ympev.2014.12.003.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Molec. Ecol. 22 (11), 3124–3140. https://doi.org/10.1111/mec.12354.

Chen, M.Y., Liang, D., Zhang, P., 2015. Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. Syst. Biol. 64 (6), 1104–1120. https://doi.org/10.1093/sysbio/syv059.

Chernomor, O., Von Haeseler, A., Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. 65 (6), 997–1008. https://doi.org/10.1093/sysbio/syw037.

Chiari, Y., Cahais, V., Galtier, N., Delsuc, F., 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). BMC Biol. 10 (1), 1–15. https://doi.org/10.1186/1741-7007-10-65.

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30 (23), 3317–3324. https://doi.org/10.1093/bioinformatics/btu530.

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T., 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. BMC Genomics 16, S2. https://doi.org/10.1186/1471-2164-16-S10-S2.

Crotti, M., Barratt, C.D., Loader, S.P., Gower, D.J., Streicher, J.W., 2019. Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (*Afrixalus*). Zool. Scripta 48 (2), 157–167. https://doi.org/10.1111/zsc.2019.48.issue-210.1111/zsc.12335.

Curto, M., Schachtler, C., Puppo, P., Meimberg, H., 2018. Using a new RAD-sequencing approach to study the evolution of *Micromeria* in the Canary islands. Molec. Phylogen. Evol. 119, 160–169. https://doi.org/10.1016/j.ympev.2017.11.005.

de Oca, A.N.M., Barley, A.J., Meza-Lázaro, R.N., García-Vázquez, U.O., Zamora-Abrego, J.G., Thomson, R.C., Leaché, A.D., 2017. Phylogenomics and species delimitation in the knob-scaled lizards of the genus Xenosaurus (Squamata: Xenosauridae) using ddRADseq data reveal a substantial underestimation of diversity. Molec. Phylogen. Evol. 106, 241–253. https://doi.org/10.1016/j.ympev.2016.09.001.

de Queiroz, A., Donoghue, M.J., Kim, J., 1995. Separate versus combined analysis of phylogenetic evidence. Stable URL: Annu. Rev. Ecol. Syst. 26 (1), 657–681 https://www.jstor.org/stable/2097223.

de Queiroz, A., Gatesy, J., 2007. The supermatrix approach to systematics. Trends Ecol. Evol 22 (1), 34–41. https://doi.org/10.1016/j.tree.2006.10.002.

Degnan, J.H., Rosenberg, N.A., Wakeley, J., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2 (5), e68. https://doi.org/10.1371/journal.pgen.0020068.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24 (6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009.

Dillenberger, M.S., Kadereit, J.W., 2017. Simultaneous speciation in the European high mountain flowering plant genus *Facchinia (Minuartia* s.l., Caryophyllaceae) revealed by genotyping-by-sequencing. Molec. Phylogen. Evol. 112, 23–35. https://doi.org/10.1016/j.ympev.2017.04.016.

Dobrin, B.H., Zwickl, D.J., Sanderson, M.J., 2018. The prevalence of terraced treescapes in analyses of phylogenetic data sets. BMC Evol. Biol. 18 (1), 1–16. https://doi.org/10.1186/s12862-018-1162-9.

Duan, L., Wen, J., Yang, X., Liu, P.L., Arslan, E., Ertuğrul, K., Chang, Z.Y., 2015. Phylogeny of *Hedysarum* and tribe Hedysareae (Leguminosae: Papilionoideae) inferred from sequence data of ITS, *matK*, *trnL-F* and *psbA-trnH*. Taxon 64 (1), 49–64. https://doi.org/10.12705/641.26.

Eaton, D.A., Overcast, I., 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics, 36(8): 2592-2594. doi: 10.1093/bioinformatics/btz966.

Eaton, D.A.R., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 62 (5), 689–706. https://doi.org/10.1093/sysbio/syt032.

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. Syst. Biol. 66 (3), 399–412. https://doi.org/10.1093/sysbio/syw092.

Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. P. Nati. A. Sci. USA 104 (14), 5936–5941. https://doi.org/10.1073/pnas.0607004104.

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94, 447–462. https://doi.org/10.1016/j.ympev.2015.10.027.

Eggli, U., 2008. Sukkulenten, 2nd Edition. Eugen Ulmer KG, Stuttgart.

Elshire, R.J., Glaubitz, J.C., Sun, Q.i., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., Orban, L., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6 (5), e19379. https://doi.org/10.1371/journal.pone.0019379.

Escudero, M., Eaton, D.A.R., Hahn, M., Hipp, A.L., 2014. Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). Molec. Phylogen. Evol. 79, 359–367. https://doi.org/10.1016/j.ympev.2014.06.026.

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32 (19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

Fairfield, K.N., Mort, M.E., Santos-Guerra, A., 2004. Phylogenetics and evolution of the Macaronesian members of the genus *Aichryson* (Crassulaceae) inferred from nuclear and chloroplast sequence data. Pl. Syst. Evol. 248, 71–83. https://doi.org/10.1007/s00606-004-0190-7.

Fernández, R., Gabaldón, T., Dessimoz, C., 2020. Orthology: definitions, inference, and impact on species phylogeny inference. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 2.4, pp. 2.4:1–2.4:14. No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE. HAL Id: hal-02535414.

Gatesy, J., Springer, M.S., 2013. Concatenation versus coalescence versus "concatalescence". P. Natl. Acad. Sci. USA 110 (13). https://doi.org/10.1073/pnas.1221121110.

Gatesy, J., Springer, M.S., 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80, 231–266. https://doi.org/10.1016/j.ympev.2014.08.013.

Gerschwitz-Eidt, M.A., Kadereit, J.W., 2019. Genotyping-by-sequencing (GBS), ITS and cpDNA phylogenies reveal the existence of a distinct Pyrenean/Cantabrian lineage in the European high mountain genus *Homogyne* (Asteraceae) and imply dual westward migration of the genus. Alp. Botany 129 (1), 21–31. https://doi.org/10.1007/s00035-018-0212-7.

Good, J.M., 2012. Reduced representation methods for subgenomic enrichment and next-generation sequencing. In: Orgogozo, V., Rockman, M.V. (Eds.), Methods in Molecular Biology Vol. 772: Molecular Methods for Evolutionary Genetics. Humana Press, New York, pp. 85–103. https://doi.org/10.1007/978-1-61779-228-1_5.

Grover, C.E., Salmon, A., Wendel, J.F., 2012. Targeted sequence capture as a powerful tool for evolutionary analysis1. Am. J. Botany. 99 (2), 312–319. https://doi.org/10.3732/ajb.1100323.

Hamon, P., Grover, C.E., Davis, A.P., Rakotomalala, J.-J., Raharimalala, N.E., Albert, V. A., Sreenath, H.L., Stoffelen, P., Mitchell, S.E., Couturon, E., Hamon, S., de

Kochko, A., Crouzillat, D., Rigoreau, M., Sumirat, U., Akaffou, S., Guyot, R., 2017. Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (Coffea) and insights into the evolution of caffeine content in its species: GBS coffee phylogeny and the evolution of caffeine content. Molec. Phylogen. Evol. 109, 351–361. https://doi.org/10.1016/j.ympev.2017.02.009.

Harvey, M.G., Judy, C.D., Seeholzer, G.F., Maley, J.M., Graves, G.R., Brumfield, R.T., 2015. Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. PeerJ 3:e895. doi: 10.7717/peerj.895.

Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., Brumfield, R.T., 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Syst. Biol. 65 (5), 910–924. https://doi.org/10.1093/sysbio/syw036.

Heled, J., Drummond, A.J., 2009. Bayesian inference of species trees from multilocus data. Molec. Bio. Evol. 27 (3), 570–580. https://doi.org/10.1093/molbev/msp274.

Herrera, S., Shank, T.M., 2016. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Molec. Phylogen. Evol. 100, 70–79. https://doi.org/10.1016/j.ympev.2016.03.010.

Hipp, A.L., Manos, P.S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., Crowl, A. A., Deng, M., Denk, T., Fitz-Gibbon, S., Gailing, O., González-Elizondo, M.S., González-Rodríguez, A., Grimm, G.W., Jiang, X.-L., Kremer, A., Lesur, I., McVay, J. D., Plomion, C., Rodríguez-Correa, H., Schulze, E.-D., Simeone, M.C., Sork, V.L., Valencia-Avalos, S., 2020. Genomic landscape of the global oak phylogeny. New Phytol. 226 (4), 1198–1212. https://doi.org/10.1111/nph.v226.410.1111/nph.16162.

Hirsch, C.N., Robin Buell, C., 2013. Tapping the promise of genomics in species with complex, nonmodel genomes. Annual Rev. Pl. Biol. 64 (1), 89–110. https://doi.org/10.1146/arplant.2013.64.issue-110.1146/annurev-arplant-050312-120237.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. 33 (4), 1110–1125. https://doi.org/10.1093/molbev/msv347.

Huang, H., Knowles, L.L., 2016. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Syst. Biol. 65 (3), 357–365. https://doi.org/10.1093/sysbio/syu046.

Ilut, D.C., Nydam, M.L., Hare, M.P., 2014. Defining loci in restriction-based reduced representation genomic data from nonmodel species: sources of bias and diagnostics for optimal clustering. BioMed Res. Int. 2014, 1–9. https://doi.org/10.1155/2014/675158.

Karbstein, K., Tomasello, S., Hodač, L., Dunkel, F.G., Daubert, M., Hörandl, E., 2020. Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic Ranunculus auricomus complex (Ranunculaceae). Taxon 69 (6), 1191–1220. https://doi.org/10.1002/tax.v69.610.1002/tax.12365.

Karbstein, K., Tomasello, S., Hodač, L., Lorberg, E., Daubert, M., Hörandl, E., 2021. Moving beyond assumptions: Polyploidy and environmental effects explain a geographical parthenogenesis scenario in European plants. Mol. Ecol. 30 (11), 2659–2675. https://doi.org/10.1111/mec.v30.1110.1111/mec.15919.

Knowles, L.L., 2009. Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. Syst. Biol. 58 (5), 463–467. https://doi.org/10.1093/sysbio/syp061.

Kubatko, L.S., Degnan J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56(1):17–24. doi: 10.1080/10635150601146041.

Kumar, S., Filipski, A.J., Battistuzzi, F.U., Kosakovsky Pond, S.L., Tamura, K., 2012. Statistics and truth in phylogenomics. Molec. Biol. Evol. 29 (2), 457–472. https://doi.org/10.1093/molbev/msr202.

Kück, P., Meusemann, K., 2010. FASconCAT: convenient handling of data matrices. Molec. Phylogenet. Evol. 56 (3), 1115–1118. https://doi.org/10.1016/j.ympev.2010.04.024.

Kück, P., Longo, G.C., 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front. Zool. 11, 81. https://doi.org/10.1186/s12983-014-0081-x.

Lanier, H.C., Huang, H., Knowles, L.L., 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Mol. Phylogenet. Evol. 70, 112–119. https://doi.org/10.1016/j.ympev.2013.09.006.

Leaché, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60 (2), 126–137. https://doi.org/10.1093/sysbio/syq073.

Lee, K.M., Kivelä, S.M., Ivanov, V., Hausmann, A., Kaila, L., Wahlberg, N., Mutanen, M., 2018. Information dropout patterns in restriction site associated DNA phylogenomics and a comparison with multilocus Sanger data in a species-rich moth genus. Syst. Biol. 67(6):925–939. doi: 10.1093/sysbio/syy029.

Lepais, O., Weir, J.T., 2014. Sim RAD: an R package for simulation-based prediction of the number of loci expected in RAD seq and similar genotyping by sequencing approaches. Mol. Ecol. Resour. 14 (6), 1314–1321. https://doi.org/10.1111/men.2014.14.issue-610.1111/1755-0998.12273.

Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24 (21), 2542–2543. https://doi.org/10.1093/bioinformatics/btn484.

Liu, L., Xi, Z., Wu, S., Davis, C., Edwards, S.V., 2015. Estimating phylogenetic trees from genome-scale data. Ann. N. Y. Acad. Sci. 1360, 36–53. https://doi.org/10.1111/nyas.12747.

Long, C., Kubatko, L., 2018. The effect of gene flow on coalescent-based species-tree inference. Syst. Biol. 67(5):770–785. doi: 10.1093/sysbio/syy020.

MacConaill, L.E., Burns, R.T., Nag, A., Coleman, H.A., Slevin, M.K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M.S., Ducar, M.D., Meyerson, M., Thorner, A.R., 2018. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics 19 (1). https://doi.org/10.1186/s12864-017-4428-5.

Maddison, W. P., 1997. Gene trees in species trees. Syst, Biol. 46(3): 523-536. doi: 10.1093/sysbio/46.3.523.

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55(1):21-30. doi: 10.1080/10635150500354928.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., Turner, D.J., 2010. Target-enrichment strategies for next-generation sequencing. Nat. Methods. 7 (2), 111–118. https://doi.org/10.1038/nmeth.1419.

Mardis, E.R., 2017. DNA sequencing technologies: 2006–2016. Nat. Protoc. 12 (2), 213–218. https://doi.org/10.1038/nprot.2016.182.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17 (1), 10. https://doi.org/10.14806/ej.17.110.14806/ej.17.1.200.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D., Emerson, B.C., 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Mol. Ecol. Resour. 15 (1), 28–41. https://doi.org/10.1111/men.2014.15.issue-110.1111/1755-0998.12291.

McCartney-Melstad, E., Gidiş, M., Shaffer, H.B., 2019. An empirical pipeline for choosing the optimal clustering threshold in RADseq studies. Molec. Ecol. Resour. 19 (5), 1195–1204. https://doi.org/10.5281/zenodo.2540263.

McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield, R.T., Alvarez, N., 2013a. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS ONE 8 (1), e54848. https://doi.org/10.1371/journal.pone.0054848.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013b. Applications of next-generation sequencing to phylogeography and phylogenetics. Molec. Phylogen. Evol. 66 (2), 526–538. https://doi.org/10.1016/j.ympev.2011.12.007.

McKain, M.R., Johnson, M.G., Uribe-Convers, S., Eaton, D., Yang, Y.a., 2018. Practical considerations for plant phylogenomics. Appl. Plant Sci. 6 (3), e1038. https://doi.org/10.1002/aps3.1038.

McKinney, G.J., Waples, R.K., Seeb, L.W., Seeb, J.E., 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Mol. Ecol. Resour. 17 (4), 656–669. https://doi.org/10.1111/men.2017.17.issue-410.1111/1755-0998.12613.

Mendes, F.K., Hahn, M.W., 2018. Why concatenation fails near the anomaly zone. Syst. Biol. 67 (1), 158–169. https://doi.org/10.1093/sysbio/syx063.

Messerschmid, T.F.E., Klein, J.T., Kadereit, G., Kadereit, J.W., 2020. Linnaeus' folly – phylogeny, evolution and classification of Sedum (Crassulaceae) and Crassulaceae subfamily Sempervivoideae. Taxon 69 (5), 892–926. https://doi.org/10.1002/tax.12316.

Miller, J.T., Grimes, J.W., Murphy, D.J., Bayer, R.J., Ladiges, P.Y., 2003. A phylogenetic analysis of the Acacieae and Ingeae (Mimosoideae: Fabaceae) based on *trnK*, *matK*, *psbA-trnH*, and *trnL/trnF* sequence data. Syst. Bot. 28 (3), 558–566. https://doi.org/10.1043/02-48.1.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17 (2), 240–248.

Minh, B.Q., Hahn, M.W., Lanfear, R., 2020a. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37(9): 2727-2733. doi: 10.1093/molbev/msaa106.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., 2020b. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37(5): 1530-1534. doi: 10.1093/molbev/msaa015.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014a. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30 (17), i541–i548. https://doi.org/10.1093/bioinformatics/btu462.

Mirarab, S., Bayzid, M.S., Boussau, B., Warnow, T., 2014b. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science 346, 1250463. https://doi.org/10.1126/science.1250463.

Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31 (12), i44–i52. https://doi.org/10.1093/bioinformatics/btv234.

Mirarab, S., Bayzid, M.S., Warnow, T., 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65 (3), 366–380. https://doi.org/10.1093/sysbio/syu063.

Molloy, E.K., Warnow, T., 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. Syst. Biol. 67 (2), 285–303. https://doi.org/10.1093/sysbio/syx077.

Mora-Márquez, F., García-Olivares, V., Emerson, B.C., López de Heredia, U., 2017. ddradseqtools: a software package for in silico simulation and testing of double-digest RAD seq experiments. Mol. Ecol. Resour. 17 (2), 230–246. https://doi.org/10.1111/1755-0998.12550.

Mort, M.E., Soltis, D.E., Soltis, P.S., Francisco-Ortega, J., Santos-Guerra, A., 2002. Phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. Syst. Bot. 27 (2), 271–288. https://doi.org/10.1043/0363-6445-27.2.271.

Moura, M., Carine, M., De Sequeira, M.M., 2015. *Aichryson santamariensis* (Crassulaceae): a new species endemic to Santa Maria in the Azores. Phytotaxa 234 (1), 37–50. https://doi.org/10.11646/phytotaxa.234.1.2.

Nicholls, J.A., Pennington, R.T., Koenen, E.J., Hughes, C.E., Hearn, J., Bunnefeld, L., Dexter, K.G., Stone, G.N., Kidner, C.A., 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus Inga (Leguminosae: Mimosoideae). Front. Plant Sci. 6, 710. https://doi.org/10.3389/fpls.2015.00710.

Paetzold, C., Wood, K.R., Eaton, D.A.R., Wagner, W.L., Appelhans, M.S., 2019. Phylogeny of Hawaiian Melicope (Rutaceae): RAD-seq resolves species relationships and reveals ancient introgression. Front. Plant. Sci. 10 https://doi.org/10.3389/fpls.2019.01074.

Parchman, T.L., Jahner, J.P., Uckele, K.A., Galland, L.M., Eckert, A.J., 2018. RADseq approaches and applications for forest tree genetics. Tree Genet. Genomes. 14 (3), 39. https://doi.org/10.1007/s11295-018-1251-3.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5 (5), 568–583. https://doi.org/10.1093/oxfordjournals.molbev.a040517.

Paris, J.R., Stevens, J.R., Catchen, J.M., 2017. Lost in parameter space: a road map for stacks. Methods Ecol. Evol. 8 (10), 1360–1373. https://doi.org/10.1111/2041-210X.12775.

Pease, J.B., Brown, J.W., Walker, J.F., Hinchliff, C.E., Smith, S.A., 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. Am. J. Bot. 105 (3), 385–403. https://doi.org/10.1002/ajb2.2018.105.issue-310.1002/ajb2.1016.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., Orlando, L., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7 (5), e37135. https://doi.org/10.1371/journal.pone.0037135.

Puritz, J.B., Hollenbeck, C.M., Gold, J.R., 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431. doi: 10.7717/peerj.431.

Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M.R., Elmer, K.R., Vences, M., Steinfarz, S., 2019. Phylogeny and species delimitation of near Eastern Neurergus newts (Salamandridae) based on genome-wide RADseq data analysis. Mol. Phylogenet. Evol. 133, 189–197. https://doi.org/10.1016/j.ympev.2019.01.003.

Rannala, B., Edwards, S.V., Leaché, A., Yang, Z., 2020. The Multispecies Coalescent Model and Species Tree Inference. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 3.3, pp. 3.3:1–3.3:21. No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE. HAL Id: hal-02535622.

Razkin, O., Sonet, G., Breugelmans, K., Madeira, M.J., Gómez-Moliner, B.J., Backeljau, T., 2016. Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex Pyramidula: the power of RADseq data. Mol. Phylogenet. Evol. 101, 267–278. https://doi.org/10.1016/j.ympev.2016.05.002.

Ree, R.H., Hipp, A.L., 2015. Inferring phylogenetic history from restriction site associated DNA (RADseq). In: Hörandl, E., Appelhans, M.S. (Eds.), Regnum Vegetabile Vol. 158: Next-Generation Sequencing in Plant Systematics. Koeltz Scientific Books, Oberreifenberg, pp. 181–204.

Reuter, J.A., Spacek, D.V., Snyder, M.P., 2015. High-throughput sequencing technologies. Molec. Cell 58 (4), 586–597. https://doi.org/10.1016/j.molcel.2015.05.004.

Rivera-Colón, A.G., Rochette, N.C., Catchen, J.M., 2021. Simulation with RADinitio improves RADseq experimental design and sheds light on sources of missing data. Mol. Ecol. Resour. 21 (2), 363–378. https://doi.org/10.1111/men.v21.210.1111/1755-0998.13163.

Roch, S., Steel, M., 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Populat. Biol. 100, 56–62. https://doi.org/10.1016/j.tpb.2014.12.005.

Roch, S., Warnow, T., 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Syst. Biol. 64 (4), 663–676. https://doi.org/10.1093/sysbio/syv016.

Rubin, B.E.R., Ree, R.H., Moreau, C.S., Kolokotronis, S.-O., 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7 (4), e33394. https://doi.org/10.1371/journal.pone.0033394.

Sanderson, M.J., McMahon, M.M., Steel, M., 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. BMC Evol. Biol. 10 (1), 1–13. https://doi.org/10.1186/1471-2148-10-155.

Sanderson, M.J., McMahon, M.M., Steel, M., 2011. Terraces in phylogenetic tree space. Science 333 (6041), 448–450. https://doi.org/10.1126/science.1206357.

Sanderson, M.J., McMahon, M.M., Stamatakis, A., Zwickl, D.J., Steel, M., 2015. Impacts of terraces on phylogenetic inference. Syst. Biol. 64 (5), 709–726. https://doi.org/10.1093/sysbio/syv024.

Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Molec. Biol. Evol. 33 (7), 1654–1668. https://doi.org/10.1093/molbev/msw079.

Sayyari, E., Whitfield, J.B., Mirarab, S., 2017. Fragmentary gene sequences negatively impact gene tree and species tree reconstruction. Molec. Biol. Evol. 34 (12), 3279–3291. https://doi.org/10.1093/molbev/msx261.

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., Alvarez, N., 2017. Hy RAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. Methods Ecol. Evol. 8 (10), 1374–1388. https://doi.org/10.1111/2041-210X.12785.

Seo, T.K., 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Molec. Biol. Evol. 25 (5), 960–971. https://doi.org/10.1093/molbev/msn043.

Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., Wolf, J.B.W., Gilbert, M., 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol. Evol. 8 (8), 907–917. https://doi.org/10.1111/mee3.2017.8.issue-810.1111/2041-210X.12700.

Shi, J.J., Rabosky, D.L., 2015. Speciation dynamics during the global radiation of extant bats. Evolution 69 (6), 1528–1545. https://doi.org/10.1111/evo.12681.

Simion, P., Delsuc, F., Philippe, H., 2020. To What Extent Current Limits of Phylogenomics Can Be Overcome? In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 2.1, pp. 2.1:1–2.1:34. No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE. HAL Id: hal-02535366.

Simmons, M.P., 2012. Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. Cladistics 28 (2), 208–222. https://doi.org/10.1111/j.1096-0031.2011.00375.x.

Simmons, M.P., Sloan, D.B., Gatesy, J., 2016. The effects of subsampling gene trees on coalescent methods applied to ancient divergences. Mol. Phylogenet. Evol. 97, 76–89. https://doi.org/10.1016/j.ympev.2015.12.013.

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Brumfield, R.T., 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63 (1), 83–95. https://doi.org/10.1093/sysbio/syt061.

Solís-Lemus, C., Yang, M., Ané, C., 2016. Inconsistency of species tree methods under gene flow. Syst. Biol. 65 (5), 843–851. https://doi.org/10.1093/sysbio/syw030.

Song, S., Liu, L., Edwards, S.V., Wu, S., 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. P. Natl. Acad. Sci. USA 109 (37), 14942–14947. https://doi.org/10.1073/pnas.1211733109.

Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J., Rabosky, D.L., Stadler, T., Steiner, C., Ryder, O.A., Janečka, J.E., Fisher, C.A., Murphy, W.J., Stanyon, R., 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PLoS ONE 7 (11), e49521. https://doi.org/10.1371/journal.pone.0049521.

Springer, M.S., Gatesy, J., 2014. Land plant origins and coalescence confusion. Trends Plant. Sci. 19 (5), 267–269. https://doi.org/10.1016/j.tplants.2014.02.012.

Springer, M.S., Gatesy, J., 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94, 1–33. https://doi.org/10.1016/j.ympev.2015.07.018.

Springer, M.S., Gatesy, J., 2018. On the importance of homology in the age of phylogenomics. Syst. Biodivers. 16 (3), 210–228. https://doi.org/10.1080/14772000.2017.1401016.

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.

Suchan, T., Pitteloud, C., Gerasimova, N.S., Kostikova, A., Schmid, S., Arrigo, N., Pajkovic, M., Ronikier, M., Alvarez, N., Orlando, L., 2016. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. PLoS ONE 11 (3), e0151651. https://doi.org/10.1371/journal.pone.0151651.

Suchan, T., 2018. hyRAD RNA probes preparation and capture. Lab protocol available at: protocols.io, ID 14096, https://protocols.io/view/hyrad-rna-probes-preparation-and-capture-rzqd75w.

Suda, J., Kyncl, T., Jarolímová, V., 2005. Genome size variation in Macaronesian angiosperms: forty percent of the Canarian endemic flora completed. Pl. Syst. Evol. 252 (3–4), 215–238. https://doi.org/10.1007/s00606-004-0280-6.

Swofford, D.L., 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.0a168. Sinauer Associates, Sunderland, Massachusetts, USA.

Tan, G., Opitz, L., Schlapbach, R., Rehrauer, H., 2019. Long fragments achieve lower base quality in Illumina paired-end sequencing. Sci. Rep. 9 (1), 1–7. https://doi.org/10.1038/s41598-019-39076-7.

Uhl, C.H., 1961. The chromosomes of the Sempervivoideae (Crassulaceae). Amer. J. Bot. 48 (2), 114–123. https://doi.org/10.1002/ajb2.1961.48.issue-210.1002/j.1537-2197.1961.tb11612.x.

Vachaspati, P., Warnow, T., 2015. ASTRID: accurate species trees from internode distances. BMC Genomics 16 (10), 1–13. http://www.biomedcentral.com/1471-2164/16/S10/S3.

van der Valk, T., Vezzi, F., Ormestad, M., Dalén, L., Guschanski, K., 2020. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. Mol. Ecol. Resour. 20 (5), 1171–1181. https://doi.org/10.1111/men.v20.510.1111/1755-0998.13009.

van Gurp, T.P., 2017. GBS Barcode Generator. http://www.deenabio.com/services/gbs-adapter (accessed January 2017).

von Goethe, J.W., 2012. Maximen und reflexionen. Jazzybee Verlag Jürgen Beck, Altenmüster, Germany.

Wagner, N.D., Gramlich, S., Hörandl, E., 2018. RAD sequencing resolved phylogenetic relationships in European shrub willows (*Salix* L. subg. *Chamaetia* and subg. *Vetrix*) and revealed multiple evolution of dwarf shrubs. Ecol Evol. 8 (16), 8243–8255. https://doi.org/10.1002/ece3.2018.8.issue-1610.1002/ece3.4360.

Wagner, F., Ott, T., Schall, M., Lautenschlager, U., Vogt, R., Oberprieler, C., 2020. Taming the Red Bastards: Hybridisation and species delimitation in the Rhodanthemum arundanum-group (Compositae, Anthemideae). Mol. Phylogenet. Evol. 144, 106702. https://doi.org/10.1016/j.ympev.2019.106702.

Wang, X., Ye, X., Zhao, L., Li, D., Guo, Z., Zhuang, H., 2017. Genome-wide RAD sequencing data provide unprecedented resolution of the phylogeny of temperate bamboos (Poaceae: Bambusoideae). Sci. Rep. 7, 11546. https://doi.org/10.1038/s41598-017-11367-x.

Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., Liston, A., 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. Appl. Plant Sci. 2 (9), 1400042. https://doi.org/10.3732/apps.140004210.3732/apps.1400042.s1.

Whitfield, J.B., Lockhart, P.J., 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22 (5), 258–265. https://doi.org/10.1016/j.tree.2007.01.012.

Wu, S., Song, S., Liu, L., Edwards, S.V., 2013. Reply to Gatesy and Springer: the multispecies coalescent model can effectively handle recombination and gene tree heterogeneity. P. Natl. Acad. Sci. USA 110 (13). https://doi.org/10.1073/pnas.1300129110.

Xi, Z., Liu, L., Davis, C.C., 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol. Phylogenet. Evol. 92, 63–71. https://doi.org/10.1016/j.ympev.2015.06.009.

Xi, Z., Liu, L., Davis, C.C., 2016. The impact of missing data on species tree estimation. Molec. Biol. Evol. 33 (3), 838–860. https://doi.org/10.1093/molbev/msv266.

Xu, B., Yang, Z., 2016. Challenges in species tree estimation under the multispecies coalescent model. Genetics 204 (4), 1353–1368. https://doi.org/10.1534/genetics.116.190173.

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. U.S.A. 107 (20), 9264–9269. https://doi.org/10.1073/pnas.0913022107.

Yang, Z., 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42 (5), 587–596. https://doi.org/10.1007/BF02352289.

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19, 153. https://doi.org/10.1186/s12859-018-2129-y.

Zimmermann, T., Mirarab, S., Warnow, T., 2014. BBCA: Improving the scalability of* BEAST using random binning. BMC Genomics 15 (6), S11. https://doi.org/10.1186/1471-2164-15-S6-S11.

*RADseq lab protocol by Hühn et al. (2022)*

**DNA extraction**

DNA-extraction was conducted using the DNeasy Plant Mini-Kit (QIAGEN, Venlo, Netherlands) according the manufacturer's protocol for "Purification of Total DNA from Plant Tissue (Mini Protocol)" with following changes: For cell lysis, 600 µl of AP1 buffer was added to 20 mg of disrupted plant tissue, incubated at 65 °C in a heating block and mixed gently every 10 minutes. After 50 minutes, 4 µl RNAse A was added, mixed and incubated for additional 10 minutes. For precipitation of proteins and polysaccharides, 195 µl of P3 buffer was added, incubated on ice for 5 min, and then centrifuged at 13000 rpm. DNA was eluted in 50 or 25 µl Buffer AE (10 mM Tris·Cl, 0.5 mM EDTA, pH 9.0) after 20 min incubation time at room temperature. The eluates were stored at -20 °C until usage. The DNA concentration and quality was evaluated using NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and gel electrophoresis.

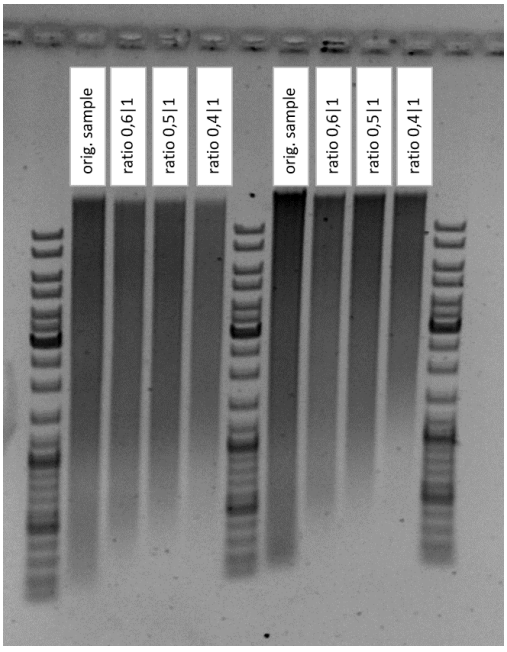



**Figure S1**. Result of the DNA size selection to remove degraded fragments within the size selection window. The degraded DNA samples are size selected using ratios of 0.6, 0.5 and 0.4 MB suspension to 1 part DNA eluate. Using a ratio of 0.4/1 removes fragments smaller than 1 kbp.

As for other RAD protocols, highly degraded DNA is problematic since it may contain a considerable number of fragments that unintendedly fall within the desired size selection window of the library and do not or just partly contain the recognition site of the used REases. These "non-target fragments" can hamper sequencing capacity since they might result in an erroneous library quantification and subsequently lead to under-loading of the flow cell. In addition, degraded samples lead to uneven library composition and partly ligated fragments bind to the flow cell without being sequenced. This results in artificial missing data, caused by DNA quality and technical issues. If the DNA sample shows a thick band in the high molecular range but also a wide smear of degraded fragments reaching the size selection window, a size selection of the genomic DNA sample by magnetic beads (MB) prior to the actual lab workflow can reduce the negative impacts on downstream sequencing. By using a ratio of 0.4 MB suspension to 1 part DNA eluate, fragments

smaller than 1 kbp are removed from the sample (Fig. S1). However, care should be taken with low quantity samples because a loss of ~ 50 % can be expected. DNA samples showing low concentrations may be normalized to an even library composition by pooling each sample using a fixed DNA pooling amount (e.g. 100 ng each sample) after the ligation reaction (see also Table S3).

**Sequencing adapter preparation**

The lyophilized single stranded adapter oligonucleotides were diluted in 1x TE to a concentration of 200 µM and resuspended over night at 7°C. 25 µl of each, top and bottom strand, were mixed and filled up to a total volume of 100 µl with 1x TE buffer. The strands were annealed in a thermocycler (95°C, 2 min; decreased to 25°C by 0.1°C/s; 25°C, 30 min; 4°C hold). Barcode and common adapters were quantified by Qubit Fluorometer. The barcode and common adapters were mixed together (ratio 1:1) and diluted in 1x TE buffer to a concentration of 3.0 ng/µl stock. This stock was used to prepare the working solution (WS) of 0.5 ng/µl diluted in $ddH_2O$. All stocks were kept at -20 °C for long-time storage.

The amount of adapter working solution required to equip all fragments with adapters must be estimated prior to ligation by adapter titration using 200 ng genomic DNA and a range of adapter amounts (2–12 ng by 2 ng steps). The ligation products were amplified by PCR and peak distribution was analyzed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The resulting electropherograms were then evaluated with respect to adapter peaks. We found 6ng adapter adequate for the ligation reaction in Aichryson.

**DNA digestion and adapter ligation**

Enzymatic digestion and adapter ligation were performed consecutively in the same reaction tube. The required volume DNA isolate (min. conc. of 40 ng/µl) needed for 200 ng genomic DNA for each sample was first filled up to a maximum volume of 5 µl with $ddH_2O$ and then topped with 15 µl master mix composed of 1 µl BamHI-HF (NEB R3136), 1 µl KpnI-HF (NEB R3142), 2 µl 10x CutSmart Buffer (NEB B7204) and 11 µl $ddH_2O$ per sample. The digestion reactions were incubated for three hours at 37°C in a thermocycler. The ligation reactions contained 20 µl digestion reaction, 6 µl BamHI adapter working solution (0.5 ng/µl), 6 µl KpnI adapter working solution (0.5 ng/µl), 1 µl T4 DNA Ligase (NEB M0202), 5 µl T4 Reaction Buffer (NEB B0202) and 12 µl $ddH_2O$ for each sample and were incubated at 22°C for three hours in a thermocycler. In case of a sample's DNA concentration was lower than the minimum required 40 ng/µl for optimal results, 5 µl of the highest available concentration was used as input for the digestion.

**Multiplexing, purification and size selection**

The ligation products of all samples were multiplexed by mixing 100 ng of each sample. Remaining ligation products were stored short-time at 7°C. The pooled samples were purified using a slightly modified protocol of the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany): two times washing using 650 µl Buffer NT3, two incubations for 15 min and two times elution in 16.5 µl Buffer AE. The eluate was quantified using the Qubit Fluorometer (dsDNA HS Assay Kit, Invitrogen by Thermo Fisher Scientific) and an aliquot of 1 µl was used for the Bioanalyzer plot. Size selection was performed by automated DNA fragment segregation using Pippin Prep (Sage Science, Beverly, MA, USA) with a 1.5% Agarose Gel Cassette (300–1500bp) and Marker K, according the manufacturer's protocol and a broad segregation range of 350–720bp. The size selected library was again quantified by Qubit Fluorometer and the size selection range was checked using a Bioanalyzer.

**Two-step PCR, pooling and final purification**

To reduce PCR bias, 15 PCR reactions (or even more) containing 1 µl size selected pool, 25 µl Q5 Hot Start High-Fidelity 2X Master Mix (NEB M0494), 19 µl ddH2O and 2.5 µl (10 pmol) of each primer (P5 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'; P7 5'-CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3') were set up and a low-cycle 2-step PCR was performed. The P7 primer contains an additional index for further multiplexing. The reactions were amplified with the following conditions: Initial denaturation at 98 °C for 30 sec; 12 cycles of denaturation at 98 °C for 10 sec, annealing and extension at 72 °C for 30 sec each cycle; final extension at 72 °C for 5 min, hold at 4 °C. The reactions were pooled and purified using the NucleoSpin Gel and PCR Clean-up kit: two times washing using 650 µl Buffer NT3, two incubations for 15 min and two times elution in 25 µl Buffer AE. Quantification was done by Qubit Fluorometer and quality was assessed using a Bioanalyzer using 1 µl pooled PCR product (Fig. S2). Final purification (Fig. 1f) was done using the NucleoMag NGS Clean-up and Size Select kit (Macherey-Nagel, Düren, Germany), according to the manufacturer's protocol for DNA clean-up and single size selection, using a ratio of 0.8 parts magnetic bead suspension to one part library. The purified library was resuspended in 25 µl Buffer AE (5 mM Tris/HCl, pH 8.5) for sequencing.

## Taxa pooling

The amplification output of the taxa pools can vary. In case the taxa pools show different concentrations, normalization of pools can be done during the final purification. Assuming the taxa pool A contains 10 samples, taxa pool B contains 20 samples, and taxa pool C contains 30 samples and the pools show the concentrations in Table 1, mixing the volumes in column "Input [µl] will ensure equal loading of the flow cell with respect to the number of included samples per taxa pool (20 ng per sample).

**Table A1**. Mixing taxa pools with respect to the concentration and the number of samples included leads to a sample-number-specific normalization for sequencing.

| Taxa pool | samples | Conc. [ng/µl] | Volume [µl] | DNA [ng] | Input [µl] | DNA library [ng] |
|---|---|---|---|---|---|---|
| A | 10 | 30 | 25 | 600 | 6.67 | 200 |
| B | 20 | 20 | 25 | 400 | 20 | 400 |
| C | 30 | 25 | 25 | 700 | 24 | 600 |

## Library quality assessment

Library quality was assessed by Qubit Fluorometer quantification and evaluation of the fragment length distribution using Bioanalyzer (Fig. S3). The ready-to-load library was sequenced on an Illumina MiSeq (San Diego, CA, USA; Reagent Kit v3 600-cycle), producing 300bp paired-end reads.

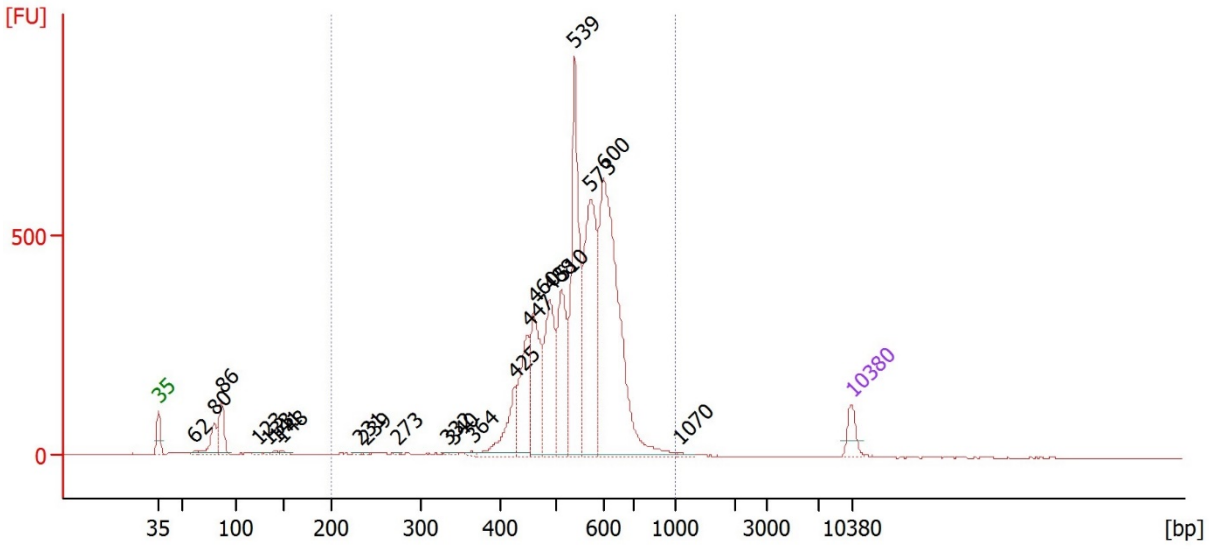



**Figure S2**. The Bioanalyzer electropherogram of the size-selected and amplified library shows the fragment length distribution (x-axis) with corresponding fluorescence (y-axis). The primer peak is at 80bp. The area below 400bp still contains size selection artefacts.

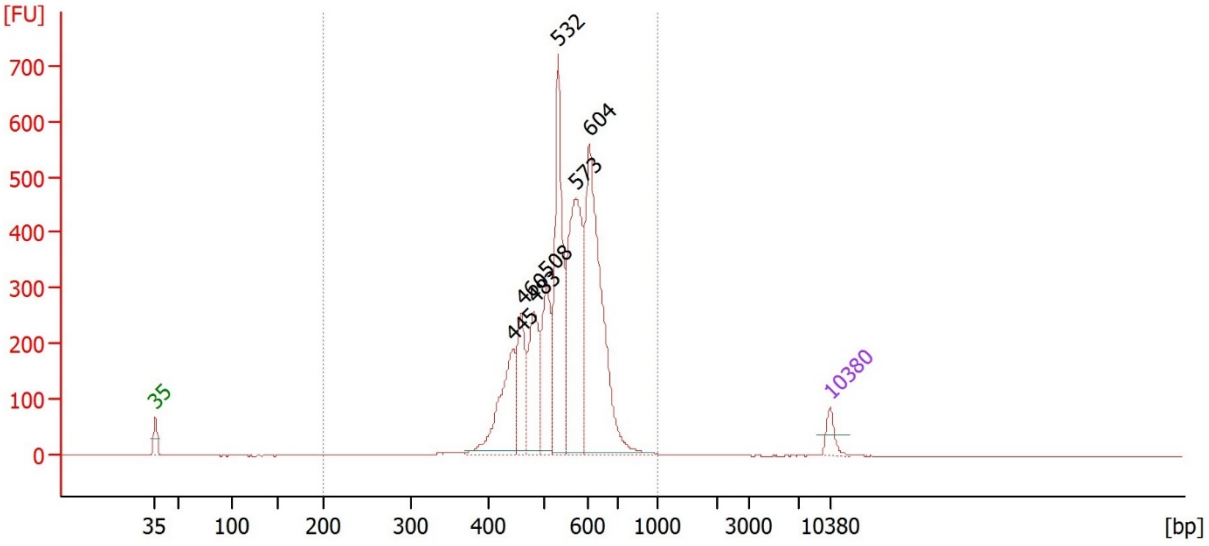



**Figure S3**. The Bioanalyzer electropherogram of the purified (2nd size selection using magnetic beads) ready-to-load library shows the fragment length distribution (x-axis) with corresponding fluorescence (y-axis). Primers and size selection artifacts are removed.

## Impact of the 2nd size selection using magnetic beads

After size selection using Pippin Prep, the area below the target area (ca. 400–700bp) contains some minor fragment peaks (Fig. S4). After amplification, the extent of these artifacts becomes clear (Fig. S5). Without a second purification, these artifacts reduce the output of raw reads of the actual target range. A large proportion of sequences (adapter + insert) was smaller than the lower edge of the targeted fragment range without a second size selection of the library (Fig. S6). Using magnetic beads for a final clean-up, the adapter sequences were detected clearly closer to the target range of the inserts (Fig. S7).

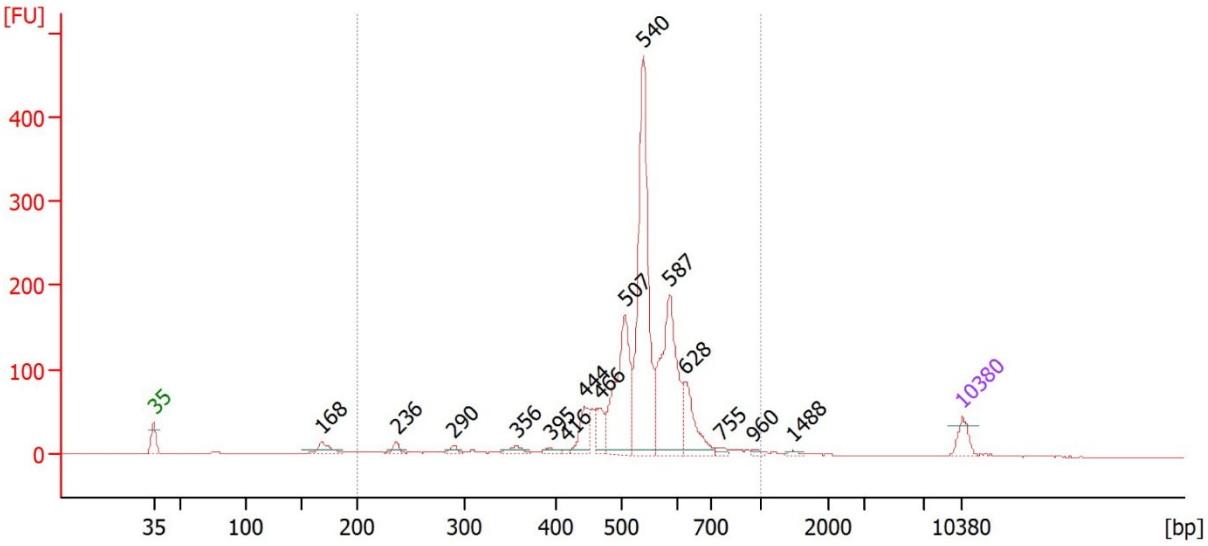



**Figure A1.4** The Bioanalyzer electropherogram of a size-selected library shows the fragment length distribution (x-axis) with corresponding fluorescence (y-axis). The area below the target range still contains a small amount of size selection artefacts.

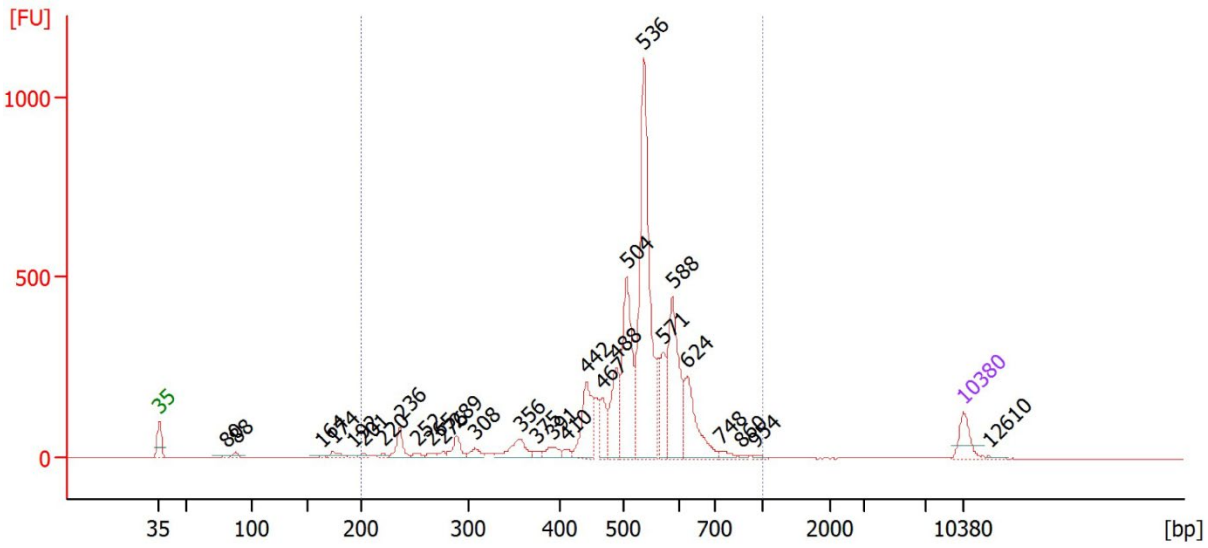



**Figure S5** The Bioanalyzer electropherogram of a size-selected and amplified library shows the fragment length distribution (x-axis) with corresponding fluorescence (y-axis).The size selection artifacts are now clearly visible.

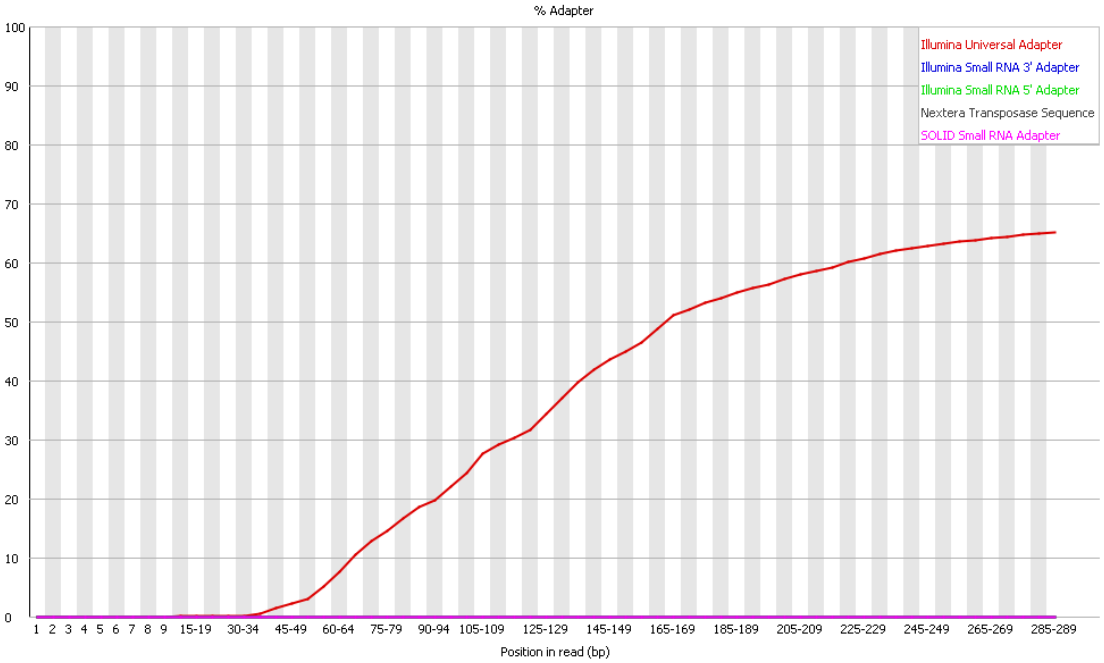



**Figure S6**: The FastQC adapter content plot (Andrews, 2010) shows the distribution of adapter sequences detected across a library that was not purified and size selected by the final clean-up using magnetic beads. The adapter sequences were detected starting at read position 30-34bp.

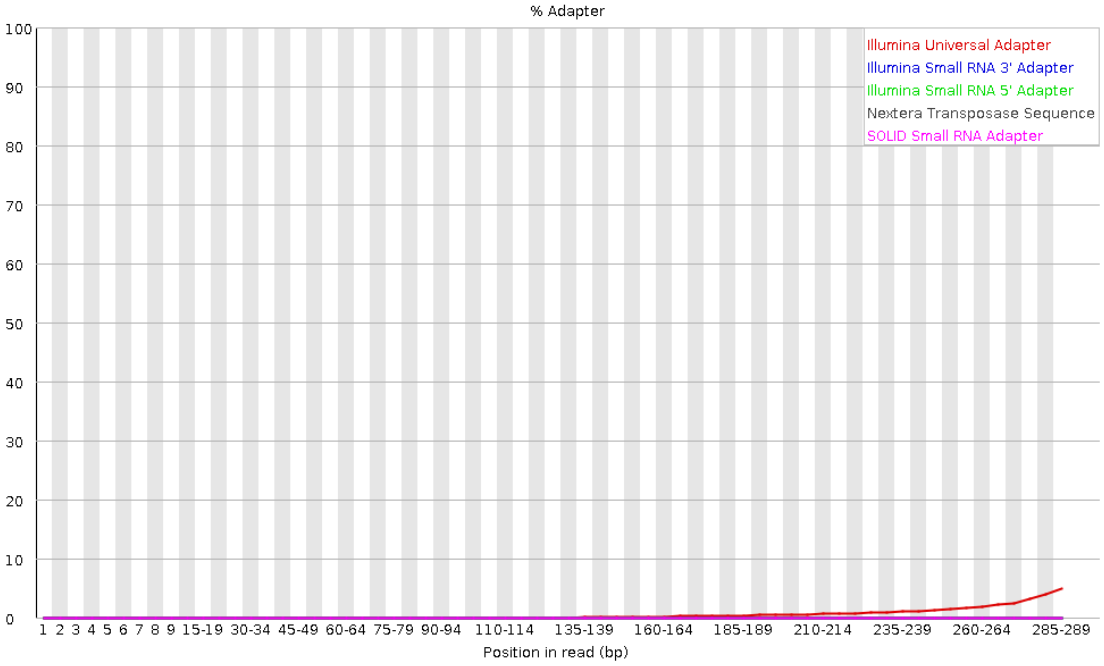



**Figure S7**: The FastQC adapter content plot (Andrews, 2010) shows the distribution of adapter sequences detected across a library that has been purified and size selected by the final clean-up using magnetic beads. The adapter sequences were detected starting at read position 210-214bp.

*Habitat descriptions*

Information on the habitat types occupied by Australian Camphorosmeae (see also: Table 3). Habitat descriptions were largely adopted from Mabbutt (1988) and McDonald (2020).

**Sand Desert** (SD) habitat in Australia takes two forms of comparable extent but different settings; dunefield and sandplain (Mabbutt, 1988). Dunefield sands have been derived from sedimentary rocks, carried by rivers over vast distances and subjected to wind action. Thus, the main dunefields occur in lower parts of basins. Sandplain sands are primary sands derived from crystalline rocks and associated with local Shield Plain desert formations from which they derived. Salinity may occur locally where thin sand sheets cover older landscapes. The surfaces of the sandy desert are largely stabilized by vegetation. Mobile sands are limited to dune ridges and local areas of exceptional sand supply, such as the northern margins of Lake Eyre. The stability of Sand Desert depends on an interplay of wind regimes and rainfall patterns, impacting the protective vegetation cover (Ash and Wasson, 1983; Mabbutt, 1988; Fujioka and Chappell, 2010). The characteristic vegetation form of arid sand deserts is hummock grassland, but open *Eucalyptus* woodland and tall *Acacia* shrubland are also common. The formation of dunefields is associated with substantial denudation of landscapes. The first dunes formed in the center of Australia during the late Pleistocene (ca. 1 Mya) and were active during Pleistocene glacial stages (Nanson et al., 1991; Fujioka and Chappell, 2010). The orientation of parallel dune-ridges represents a continental-scale anticlockwise whorl (Bowler 1976; Mabbutt, 1988; Fujioka and Chappell 2010), corresponding the anticlockwise Sub-Tropical High Pressure system positioned at the center of the continent, which causes an arid climate across a broad area of the continent during glacial phases (Fujioka and Chappell, 2010; Bowler, 1982).

**Desert Upland** (DU) habitats include a variety of forms, such as rocky ranges, hillslopes and tablelands of sedimentary rock, such as sandstone and quartzite (Mabbutt, 1988). Upland Desert

landscapes rise abruptly above wide desert plains but their elevations are modest. Only small areas of the arid zone rise above 1000 m. Thus, they have little impact on the desert climate (slight enhancement of rainfall, faint rain shadow), but are hydrologically important, since their slopes provide the main runoff to ephemeral desert stream networks. The thin stony soils are commonly non-saline and non-alkaline, except on limestone substrates. Saline sections are most commonly found downslope within the ephemeral stream systems. The characteristic vegetation forms of Upland Desert are sclerophyllous woodland, *Acacia* shrubland and hummock grassland.

**Stony Desert** (SP) habitats are characterized by stone pavements of siliceous pebbles together with remnants of dissected upland tablelands (Mabbutt, 1988; McKenzie et al., 2004). The stone pavements are underlain by deeply weathered bedrock, representing complex drainage networks that cause high surface run-off. The pavements accumulate alluvial deposits of interior drainage systems. These clayey soils contain variable amounts of lime and gypsum and are moderately to strongly saline. This hostile habitat is the driest in the hot deserts due to its stony surface and soil texture (Laity, 2009; McDonald, 2020). Salt toxicity, alkalinity, and the lack of ground moisture over long time periods severely limit plant growth and diversity (Stace et al., 1968; McKenzie et al., 2004). The characteristic vegetation forms are low grassland and low chenopod shrubland (Mabbutt, 1988; McDonald, 2020). Gibber pavements formed 2-4 Mya in areas of least rainfall during a global dry phase, such as the area north-west of the Lake Eyre Basin (Fujioka et al. 2005; 2009; Fujioka and Chappell 2010).

**Shield Plain** (SH) habitats are subdued, rocky landscapes of igneous and metamorphic rocks and are associated with granitic shield blocks in western, central and south Australia (Mabbutt, 1988). Characteristic landforms include granitic domes, tors and hills, metamorphic inliers, extensive erosional plains and smooth alluvial washes. The rocky layers are thinly mantled with sand and grit (sandplain mantles). The soils are commonly saline or calcareous with varying

amounts of sand or clay. The characteristic vegetation forms are sclerophyllous woodland, shrubland and grassland. The onset of arid conditions in Australia is indicated by extensive fluvial and lacustrine limestone in the shield plains, resulting from the progressive ceasing of formerly out-going river systems and the accumulation of sediments in seasonal lakes. The age estimations date to the middle Miocene (Bowler, 1976; Benbow et al., 1995; Fujioka and Chappell, 2010).

**Riverine Desert** (RL) habitats are associated with rivers rising on better-wettered uplands at the margins of the arid and semi-arid zones or have been formed by drainage from desert uplands (Mabbutt, 1988). The rivers originating within the desert zone are ephemeral but the rivers entering the arid zone on its periphery flood in most summers. The largest desert floodplains occur on the slopes of the Lake Eyre basin, in the Murray-Darling plains in the south-east, and in the Fitzroy-Lennard plains in the north-west. The desert rivers form disconnected systems defeated by the flatness and the aridity of the continent. Most of the rivers are broad and shallow sand-bed channels. The alluvial soils of the riverine desert plains are prone to salting and wind erosion. Characteristic soil types are sand and slit adjoining the main channels, and heavy swelling clay on the vast backplains. The soils are slightly to moderately saline. The characteristic vegetation forms are open woodlands along the channels and floodplains, grass and chenopod communities in the broad flats, and extensive low chenopod shrubland on the dry alluvial plains. Riverine Deserts emerged in western Australia in the middle to late Miocene as the palaeodrainage systems ceased to flow (van de Graaff et al., 1977). Under declining rainfall Riverine Desert landscapes became Desert Lake landscapes.

**Desert Lake** (DL) habitats are disconnected, saline complexes in pans and flats. They occur in the lower parts of the Australian arid zone near drainage terminals in alluvial plains (Mabbutt, 1988). Most of them are small and all of them are dry for most of the time. In inland Western Australia, river-lake systems are aligned along shallow valleys. The smallest desert lakes occur

as shallow deflation pans of ephemeral river floodplains. The largest desert lakes (Lake Eyre and Lake Torrens) occupy structural groundwater basins and are of tectonic origin (Mabbutt, 1988). The salts of the saline groundwater accumulate on their surface to form thick crusts. Because of periodic flooding that occurs for less a quarter of the time, the high salinity that inhibits a protective vegetation layer, and the associated high erosion by wind, most desert lakes have thin sedimentary layers and aeolian sand and clay lunettes along the margins. Thus, most desert lakes are playas. The soil features appear as a gradient from highly saline salt crust in the lake center, over less saline sandy and clayey lake margins with sand and clay lunettes to gypseous dunes and calcareous lake shores adjoining the lake systems. The salinity gradient and the sediment composition of the Desert Lakes and adjacent areas determine the characteristic vegetation forms. The highly saline lake center crusts are mostly free of vegetation. Low-growing halophytic samphire establish on the less saline margins. Halophytic shrubs, such as *Atriplex* and *Maireana*, occupy the higher sandy and clayey lunettes. Eucalypts are found on claypan margins and lake shores. Desert Lakes emerged from gradually dissected Riverine Desert landscapes. The ephemeral inland desert lake systems in the Western Australian shield plains are relicts of the paleodrainage systems ceased in the late Miocene (van de Graaff et al., 1977). The desert lake systems of Lake Eyre and Lake Torrens are relicts of once extensive perennial lakes of the early and middle Miocene and formed as aridity intensified during the Pliocene (Fujioka and Chappell, 2010).

**Desert Clay Plains** (CP) are broad, alluvial plains of heavy soil (Mabbutt, 1988). They formed on or from alluvium of fine-textured rocks such as shale and limestone. In north-eastern Australia, the heavy, alluvial clay deposits of the Barkley Tableland mark former swamps and shallow lakes of inland drainage systems. The strongly alkaline or moderately saline clayey soils are occupied by open grassland and chenopods in the groundcover.

**Karst Plain** (KP) landscapes are undulating plains of pedogenic sheet calcrete across southern Australia (Mabbutt, 1988). This landscape form is represented by the Nullarbor Plain in southern Australia at the Great Australian Bight. Most of the Nullarbor Plain is covered with a thin calcareous sheet. This sheet is covered by shallow, alkaline loams. Soil salinity increases towards the coast. The characteristic vegetation forms are chenopod shrubland (*Atriplex*, *Maireana*) on rock outcrops and tall shrubland with *Acacia* and *Eucalyptus* towards the better-watered coastal parts. The initial karst development probably occurred during the warm, seasonally wet climate of the Oligocene, when the retreat of the sea exposed the deposited limestone (Webb and James, 2006). These were inundated by the return of the sea, which finally retreated in the late Miocene due to regional uplift.

**Mesic Plain and Range** (MP) habitats are temperate to subtropical ranges and plains of eastern, southern and western continental margins, and represent a transition from coastal to arid inland areas (Crisp et al., 2004; Byrne et al., 2011; McDonald, 2020). This land type grades into Desert Upland, Shield Plain and Riverine Desert. The igneous peneplains, alluvial sand plains and dune landscapes receive more precipitation than the arid land types. Thus, the soils are generally less saline but can be alkaline in drier regions. Characteristic vegetation forms of mesic habitats adjoining the arid zone are sclerophyllous forest, open woodland and heath. Chenopod shrubland is found on heavy soils in drier areas or in areas of strong seasonality with a prominent dry season. Mesic Plain and Range habitats with typical sclerophyllous taxa occurred along the gradual expansion of the arid zone since the early Miocene (Hill, 1994; Byrne et al., 2008).

**Coast** (CS) habitats comprise beach-dune complexes, coastal lagoons and estuaries, rocky shores and cliffs (Chapman, 1976; McDonald, 2020). Coast dunes have deep sands prone to surface disruption, wind erosion and dry subsurface soil. The salinity depends on the frequency of salt-water inundation and freshwater immission from rainfall or rivers. Coastal foredunes are

nutrient-poor and have no or only sparse vegetation. The soils of coastal lagoons, estuaries and salty and brackish marshes are generally nutrient-rich and favor a stable vegetation cover compared to arid zone landscapes. Characteristic vegetation forms are low shrubland, heath and saltmarsh. Some saltmarsh communities are floristically similar to Desert Lake and Riverine Desert communities (Saintilan, 2009; McDonald, 2020). These communities include a number of halophytic and salt-resistant Amaranthaceae species. Among them are also Camphorosmeae: *Dissocarpus biflorus*, *Enchylaena tomentosa*, *Maireana brevifolia*, *M. oppositifolia*, *Neobassia astrocarpa*.

## *References*

**Ash, J.E. and Wasson, R.J.,** 1983. Vegetation and sand mobility in the Australian desert dunefield. Zeitschrift für Geomorphologie, Supplementband, 45, 7–25.

**Benbow, M.C., Lindsay, J.M. and Alley, N.F.,** 1995. Eucla Basin and palaeodrainage. The geology of South Australia, 2, pp.178-186. In: Drexel, J.H., Preiss, W.V., 'The geology of South Australia, Volume 2. The Phanerozoic. South Australia. Geological Survey. Bulletin, 54. ISBN: 0730806219.

**Bowler, J.M.,** 1976. Aridity in Australia: age, origins and expression in aeolian landforms and sediments. Earth-science reviews, 12(2-3), 279-310. https://doi.org/10.1016/0012-8252(76)90008-8.

**Bowler, J.M.,** 1982. Aridity in the late Tertiary and Quaternary of Australia. In: Barker, W.R., and Greenslade, P.J.M., Evolution of the flora and fauna of arid Australia. Peacock Publications, South Australia. ISBN: 0909209626.

**Byrne, M., Yeates, D.K., Joseph, L., Kearney, M., Bowler, J., Williams, M.A.J., Cooper, S., Donnellan, S.C., Keogh, J.S., Leys, R. and Melville, J.,** 2008. Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. Molecular ecology, 17(20), p.4398. https://doi.org/10.1111/j.1365-294X.2008.03899.x.

**Byrne, M., Steane, D. A., Joseph, L., Yeates, D. K., Jordan, G. J., Crayn, D., Aplin, K., Crisp, M.D., Porch, N., Kale Sniderman, J.M., Sunnucks, P. and Weston, P.H.,** 2011. Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. Journal of biogeography, 38(9), 1635-1656. https://doi.org/10.1111/j.1365-2699.2011.02535.x.

**Chapman, V.J.,** 1976. Coastal vegetation. Elsevier. ISBN: 1483279588

**Crisp, M., Cook, L. and Steane, D.,** 2004. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of

diversity in present–day communities? Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359(1450), pp.1551-1571. https://doi.org/10.1098/rstb.2004.1528.

**Fujioka, T., Chappell, J., Honda, M., Yatsevich, I., Fifield, K. and Fabel, D.,** 2005. Global cooling initiated stony deserts in central Australia 2–4 Ma, dated by cosmogenic 21Ne-10Be. Geology, 33(12), pp.993-996. https://doi.org/10.1130/G21746.1.

**Fujioka, T., Chappell, J., Fifield, L.K. and Rhodes, E.J.,** 2009. Australian desert dune fields initiated with Pliocene–Pleistocene global climatic shift. Geology, 37(1), pp.51-54. https://doi.org/10.1130/G25042A.1.

**Fujioka, T. and Chappell, J.,** 2010. History of Australian aridity: chronology in the evolution of arid landscapes. Geological Society, London, Special Publications, 346(1), pp.121-139. https://doi.org/10.1144/SP346.8.

**Hill, R.S.,** 1994. History of the Australian vegetation: Cretaceous to Recent. Cambridge University Press. http://library.oapen.org/handle/20.500.12657/31444. ISBN: 9781925261479.

**Laity, J.J.,** 2009. Deserts and desert environments (Vol. 3). John Wiley & Sons. ISBN: 9781577180333.

**Mabbutt, J.A.,** 1988. Australian desert landscapes. GeoJournal, 16, pp.355-369. https://doi.org/10.1007/BF00214394.

**McDonald, J.T.,** 2020. Biogeography of Australian chenopods: landscape in the evolution of an arid flora (Doctoral dissertation). https://hdl.handle.net/2440/126029.

**McKenzie, N., Jacquier, D., Isbell, R. and Brown, K.,** 2004. Australian soils and landscapes: an illustrated compendium. CSIRO publishing. ISBN: 0643069585.

**Nanson, G.C., Price, D.M., Short, S.A., Page, K.J., and Nott, J.F.,** 1991. Major episodes of climate change in Australia over the last 300 000 years. In: Gillespie, R. (ed.) Quaternary Dating Workshop 1990. Department of Biogeography and Geomorphology, Australian National University, Canberra, 45–50.

**Saintilan, N.,** 2009. Biogeography of Australian saltmarsh plants. Austral Ecology, 34(8), 929-937. https://doi.org/10.1111/j.1442-9993.2009.02001.x.

**Stace, H.C.T., Hubble, G.D., Brewer, R., Northcote, K.H., Sleeman, J.R., Mulcahy, M.J. and Hallsworth, E.G.,** 1968. A handbook of Australian soils, Rellim Technical Publications, Glenside, South Australia.

**Van de Graaff, W.E., Crowe, R.W.A., Bunting, J.A., Jackson, M.J.,** 1977. Relict early Cainozoic drainages in arid Western Australia. Zeitschrift für Geomorphologie, 21(4), pp.379-400.

**Webb, J.A. and James, J.M.,** 2006. Karst evolution of the Nullarbor plain, Australia. https://doi.org/10.1130/2006.2404(07).

# *11. SUMMARY*

The Australian Camphorosmeae are the most species-rich tribe of Amaranthaceae Juss. *sensu lato* in Australia, with about 150 species in currently 12 described genera. Their closest relatives, the genus *Grubovia* with three described species, are distributed in Central Asia.

This study re-examined hypotheses regarding phylogeny, taxonomy, trait evolution, as well as diversification, radiation and migration in the course of Australia's aridification since the Miocene. A total of 104 species representing all 12 recognized genera of the Australian Camphorosmeae were analyzed using a modified RADseq laboratory and data analysis approach with subsequent Maximum Likelihood phylogenetic inference and Bayesian inference of species divergence times. In total, 25 vegetative, reproductive and ecological characteristics were investigated for their systematic relevance.

The phylogenetic results do not support the current taxonomy and confirm concerns regarding an artificial generic delimitation. The species-rich genera *Sclerolaena* and *Maireana* turned out polyphyletic. The included *Maireana* species formed several species-poor clades, while the majority of *Sclerolaena* species formed a species-rich clade. Species-poor genera were resolved within larger genera or formed clades with other small genera. The phylogenetic clades may be circumscribed using combinations of morphological traits. In particular, the features of the multifariously developed fruiting perianth showed systematic relevance.

The phylogenetic results, mapping of ecological traits and the outcome of the divergence time estimation shed new light on biogeographical hypotheses. The Australian Camphorosmeae presumably arrived in Australia during the Middle to Late Miocene. Colonization was likely promoted by pre-adaptation to arid and saline conditions. A littoral connection may have taken place by species adapted to coastal conditions that migrated along declining palaeodrainage systems to hypersaline inland habitats. Two rapid radiations during the Late Miocene overlap

in time. Major climatic and geological changes during this period may have promoted initial diversification. As aridity intensified, this group diversified further during the Pliocene. Riverine Desert and Desert Lake habitats may have been the main migration corridors. Migration may have been multidirectional from south-west Australia to north-west Australia with subsequent west to east migration across the continents interior and diversification into more diverse habitats.

The multifariously developed diaspores relate to different dispersal syndromes. The wind-dispersed wing-like perianths of *Maireana* favor fast, non-directional dispersal over long distances, across Australia's relatively flat interior. This likely promoted rapid colonization of vacant habitats during the Late Miocene and Pliocene. Increased mobility and consequent increased gene flow probably resulted in lower speciation rates, which is now apparent in the species-poor clades of the *Maireana* grade. The emergence of spine-like appendages and the onset of *Sclerolaena* diversification coincide with radiations of large herbivores, termites and ants. The spine-like appendages may have been a key innovation as protection against large herbivores, to reduce transpiration in hot environments, and with respect to dispersal by ants. In contrast to the wind-dispersed *Maireana* species, colonization of new habitats and migration between habitats may have been slower and on a smaller scale in *Sclerolaena* species. This reduced mobility may have resulted in lower gene flow and consequently may have favored a steady speciation rate, which is now apparent in the continuous cladogenesis of the species-rich *Sclerolaena* clade.

## *11. ZUSAMMENFASSUNG*

Die Australischen Camphorosmeae sind mit ca. 150 Arten in derzeit 12 Gattungen die artenreichste Tribus der Amaranthaceae Juss. *sensu lato* in Australien. Ihre engsten Verwandten, die Gattung *Grubovia* mit drei beschriebenen Arten, sind in Zentralasien beheimatet.

In dieser Studie wurden Hypothesen zur Phylogenie, Taxonomie, Merkmalsevolution und Diversifizierung im Zuge der Aridifizierung Australiens seit dem Miozän, überprüft. Insgesamt 104 Arten wurden mit Hilfe einer modifizierten RADseq-Labor- und Datenanalyse-Methodik, mit anschließender phylogenetischer Maximum-Likelihood Inferenz und Bayes'scher Inferenz der Divergenzzeiten, untersucht. Insgesamt wurden 25 vegetative, reproduktive und ökologische Merkmale auf ihre systematische Relevanz hin untersucht.

Die phylogenetischen Ergebnisse stützen die derzeitige Taxonomie nicht und bestätigen Bedenken hinsichtlich einer artifiziellen Gattungsabgrenzung. Die artenreichen Gattungen *Sclerolaena* und *Maireana* erwiesen sich als polyphyletisch. Die untersuchten *Maireana*-Arten bildeten mehrere artenarme Kladen, während die Mehrheit der *Sclerolaena*-Arten eine artenreiche Klade bildete. Artenarme Gattungen wurden innerhalb größerer Gattungen aufgelöst oder bildeten Kladen mit anderen kleinen Gattungen. Die Kladen können mittels Kombinationen morphologischer Merkmale beschrieben werden. Insbesondere die Merkmale des vielgestaltig ausgebildeten Perigons stellten sich als systematisch relevant heraus.

Die phylogenetischen Ergebnisse, die Zuordnung ökologischer Präferenzen und das Ergebnis der Divergenzzeiten-Analyse geben neue Aufschlüsse über biogeographische Hypothesen. Diese Gruppe besiedelte Australien vermutlich im Mittleren bis Oberen Miozän. Die Besiedlung war wahrscheinlich durch eine Voranpassung an aride Bedingungen und saline Böden begünstigt. Vermutlich erfolgte eine Migration von Küstengebieten ins Landesinnere durch an Küstenbedingungen angepasste Arten, die entlang der austrocknenden Paläodrainage-

Systeme in hypersaline Habitate im Landesinneren einwanderten. Zwei rapide Radiationen während des Oberen Miozäns überschnitten zeitlich. Massive klimatische und geologische Veränderungen während dieses Zeitraums könnten die initiale Diversifizierung gefördert haben. Mit der zunehmenden Intensivierung der Aridität diversifizierten diese Gruppe im Pliozän weiter. Ephemere Wüsten-Flusshabitate und Wüsten-Salzsee-Systeme könnten wichtige Migrationskorridore zur Besiedlung des Kontinentinneren gewesen sein. Die Migration erfolgte vermutlich von Südwest- nach Nordwestaustralien mit anschließender West-Ost-Ausbreitung.

Die vielgestaltigen Diasporen stehen in Bezug zu verschiedenen Ausbreitungssyndromen. Die Fruchtflügel der wind-verbreiteten *Maireana*-Arten begünstigen eine schnelle, ungerichtete Ausbreitung über große Entfernungen. Dies förderte wahrscheinlich eine rasche Besiedlung freier Lebensräume im Oberen Miozän und Pliozän. Die erhöhte Mobilität und der daraus resultierende verstärkte Genfluss führten vermutlich zu niedrigeren Speziationsraten, somit zu artenarmen *Maireana*-Kladen. Das Aufkommen der dornigen Fruchthüllen und der Beginn der Diversifizierung der *Sclerolaena*-Klade fällt zeitlich mit der Ausbreitung großer Herbivore, Termiten und Ameisen zusammen. Die dornartigen Anhängsel könnten eine wichtige Innovation zum Schutz vor Herbivoren, zur Verringerung der Transpiration in ariden Umgebungen und im Hinblick auf die Ausbreitung durch Ameisen gewesen sein. Im Gegensatz zu den wind-verbreiteten *Maireana*-Arten war die Besiedlung neuer Lebensräume und die Migration zwischen Lebensräumen bei *Sclerolaena*-Arten möglicherweise langsamer. Diese geringere Mobilität könnte einen geringeren Genfluss zur Folge gehabt haben und folglich eine stetige Speziationsrate begünstigt haben, was sich nun in der kontinuierlichen Kladogenese der artenreichen *Sclerolaena*-Klade zeigt.

**FACHBEREICH BIOLOGIE DER JOHANNES GUTENBERG-UNIVERSITÄT MAINZ**

Archiv-Nr. / 20

**Zusammenfassung der Dissertation von Philipp Hühn**
Thema: Phylogenetics and Evolution of the Australian Camphorosmeae (Amaranthaceae)

Die Australischen Camphorosmeae sind mit ca. 150 Arten in derzeit 12 Gattungen die artenreichste Tribus der Amaranthaceae Juss. *sensu lato* in Australien. Ihre engsten Verwandten, die Gattung *Grubovia* mit drei beschriebenen Arten, sind in Zentralasien beheimatet. In dieser Studie wurden Hypothesen zur Phylogenie, Taxonomie, Merkmalsevolution und Diversifizierung im Zuge der Aridifizierung Australiens seit dem Miozän, überprüft. Insgesamt 104 Arten wurden mit Hilfe einer modifizierten RADseq-Labor- und Datenanalyse-Methodik, mit anschließender phylogenetischer Maximum-Likelihood Inferenz und Bayes'scher Inferenz der Divergenzzeiten, untersucht. Insgesamt wurden 25 vegetative, reproduktive und ökologische Merkmale auf ihre systematische Relevanz hin untersucht.
Die phylogenetischen Ergebnisse stützen die derzeitige Taxonomie nicht und bestätigen Bedenken hinsichtlich einer artifiziellen Gattungsabgrenzung. Die artenreichen Gattungen *Sclerolaena* und *Maireana* erwiesen sich als polyphyletisch. Die untersuchten *Maireana*-Arten bildeten mehrere artenarme Kladen, während die Mehrheit der *Sclerolaena*-Arten eine artenreiche Klade bildete. Artenarme Gattungen wurden innerhalb größerer Gattungen aufgelöst oder bildeten Kladen mit anderen kleinen Gattungen. Die Kladen können mittels Kombinationen morphologischer Merkmale beschrieben werden. Insbesondere die Merkmale des vielgestaltig ausgebildeten Perigons stellten sich als systematisch relevant heraus. Die phylogenetischen Ergebnisse, die Zuordnung ökologischer Präferenzen und das Ergebnis der Divergenzzeiten-Analyse geben neue Aufschlüsse über biogeographische Hypothesen. Diese Gruppe besiedelte Australien vermutlich im Mittleren bis Oberen Miozän. Die Besiedlung war wahrscheinlich durch eine Voranpassung an aride Bedingungen und saline Böden begünstigt. Vermutlich erfolgte eine Migration von Küstengebieten ins Landesinnere durch an Küstenbedingungen angepasste Arten, die entlang der austrocknenden Paläodrainage-Systeme in hypersaline Habitate im Landesinneren einwanderten. Zwei rapide Radiationen während des Oberen Miozäns überschnitten zeitlich. Massive klimatische und geologische Veränderungen während dieses Zeitraums könnten die initiale Diversifizierung gefördert haben. Mit der zunehmenden Intensivierung der Aridität diversifizierten diese Gruppe im Pliozän weiter. Ephemere Wüsten-Flusshabitate und Wüsten-Salzsee-Systeme könnten wichtige Migrationskorridore zur Besiedlung des Kontinentinneren gewesen sein. Die Migration erfolgte vermutlich von Südwest- nach Nordwestaustralien mit anschließender West-Ost-Ausbreitung. Die vielgestaltigen Diasporen stehen in Bezug zu verschiedenen Ausbreitungssyndromen. Die Fruchtflügel der wind-verbreiteten *Maireana*-Arten begünstigen eine schnelle, ungerichtete Ausbreitung über große Entfernungen. Dies förderte wahrscheinlich eine rasche Besiedlung freier Lebensräume im Oberen Miozän und Pliozän. Die erhöhte Mobilität und der daraus resultierende verstärkte Genfluss führten vermutlich zu niedrigeren Speziationsraten, somit zu artenarmen *Maireana*-Kladen. Das Aufkommen der dornigen Früchte und der Beginn der Diversifizierung der *Sclerolaena*-Klade fällt zeitlich mit der Ausbreitung großer Herbivore, Termiten und Ameisen zusammen. Die dornartigen Anhängsel könnten eine wichtige Innovation zum Schutz vor Herbivoren, zur Verringerung der Transpiration in ariden Umgebungen und im Hinblick auf die Ausbreitung durch Ameisen gewesen sein. Im Gegensatz zu den wind-verbreiteten *Maireana*-Arten war die Besiedlung neuer Lebensräume und die Migration zwischen Lebensräumen bei *Sclerolaena*-Arten möglicherweise langsamer. Diese geringere Mobilität könnte einen geringeren Genfluss zur Folge gehabt haben und folglich eine stetige Speziationsrate begünstigt haben, was sich nun in der kontinuierlichen Kladogenese der artenreichen *Sclerolaena*-Klade zeigt.

Genehmigt vom 1. Gutachter / von der 1. Gutachterin _____
(Unterschrift)

FACHBEREICH BIOLOGIE DER JOHANNES GUTENBERG-UNIVERSITÄT MAINZ

Archiv-Nr. / 20

**Zusammenfassung der Dissertation von Philipp Hühn**
Thema: Phylogenetics and Evolution of the Australian Camphorosmeae (Amaranthaceae)

The Australian Camphorosmeae are the most species-rich tribe of Amaranthaceae Juss. *sensu lato* in Australia, with about 150 species in currently 12 described genera. Their closest relatives, the genus *Grubovia* with three described species, are distributed in Central Asia.
This study re-examined hypotheses regarding phylogeny, taxonomy, trait evolution, as well as diversification, radiation and migration in the course of Australia's aridification since the Miocene. A total of 104 species representing all 12 recognized genera of the Australian Camphorosmeae were analyzed using a modified RADseq laboratory and data analysis approach with subsequent Maximum Likelihood phylogenetic inference and Bayesian inference of species divergence times. In total, 25 vegetative, reproductive and ecological characteristics were investigated for their systematic relevance.

The phylogenetic results do not support the current taxonomy and confirm concerns regarding an artificial generic delimitation. The species-rich genera *Sclerolaena* and *Maireana* turned out polyphyletic. The included *Maireana* species formed several species-poor clades, while the majority of *Sclerolaena* species formed a species-rich clade. Species-poor genera were resolved within larger genera or formed clades with other small genera. The phylogenetic clades may be circumscribed using combinations of morphological traits. In particular, the features of the multifariously developed fruiting perianth showed systematic relevance.
The phylogenetic results, mapping of ecological traits and the outcome of the divergence time estimation shed new light on biogeographical hypotheses. The Australian Camphorosmeae presumably arrived in Australia during the Middle to Late Miocene. Colonization was likely promoted by pre-adaptation to arid and saline conditions. A littoral connection may have taken place by species adapted to coastal conditions that migrated along declining palaeodrainage systems to hypersaline inland habitats. Two rapid radiations during the Late Miocene overlap in time. Major climatic and geological changes during this period may have promoted initial diversification. As aridity intensified, this group diversified further during the Pliocene. Riverine Desert and Desert Lake habitats may have been the main migration corridors. Migration may have been multidirectional from south-west Australia to north-west Australia with subsequent west to east migration across the continents interior and diversification into more diverse habitats.
The multifariously developed diaspores relate to different dispersal syndromes. The wind-dispersed winged perianths of *Maireana* favor fast, non-directional dispersal over long distances, across Australia's relatively flat interior. This likely promoted rapid colonization of vacant habitats during the Late Miocene and Pliocene. Increased mobility and consequent increased gene flow probably resulted in lower speciation rates, which is now apparent in the species-poor clades of the *Maireana* grade. The emergence of spine-like appendages and the onset of *Sclerolaena* diversification coincide with radiations of large herbivores, termites and ants. The spine-like appendages may have been a key innovation as protection against large herbivores, to reduce transpiration in hot environments, and with respect to dispersal by ants. In contrast to the wind-dispersed *Maireana* species, colonization of new habitats and migration between habitats may have been slower and on a smaller scale in *Sclerolaena* species. This reduced mobility may have resulted in lower gene flow and consequently may have favored a steady speciation rate, which is now apparent in the continuous cladogenesis of the species-rich *Sclerolaena* clade.

Genehmigt vom 1. Gutachter / von der 1. Gutachterin _____
(Unterschrift)

Anlage zum Antrag auf Zulassung zur Promotionsprüfung
gem. § 12 der Promotionsordnung des Fachbereichs Biologie
der Johannes Gutenberg-Universität Mainz vom 01.04.2018

## VERSICHERUNG

Name: Philipp Hühn

Hiermit versichere ich gemäß § 12, (2) der Promotionsordnung vom 01.04.2018:
(zutreffendes ist angekreuzt.)

☐ Ich habe die heute als Dissertation vorgelegte Arbeit selbst angefertigt und ausschließlich die angegebenen Quellen und Hilfsmitte verwendet.

☐ Ich habe oder hatte die jetzt als Dissertation vorgelegte Arbeit noch an keiner anderen deutschen oder ausländischen Hochschule oder vergleichbaren Einrichtung zur Erlangung eines akademischen Grades eingereicht.

☐ Ich habe noch kein Promotions- PhD,- oder ein vergleichbares Graduierungsverfahren im Promotionsfach erfolglos beendet.

☐ Ich habe noch kein Promotions- PhD,- oder ein vergleichbares Graduierungsverfahren im Promotionsfach erfolgreich beendet.

☐ Für die Anfertigung der vorgelegten Arbeit wurde keine entgeltliche Hilfe Dritter, insbesondere eine Promotionsberatung oder –vermittlung in Anspruch genommen.

Mainz, den _____ _____
(Unterschrift)