WILEY | Journal of Biogeography

Big data for a large clade: Bioregionalization and ancestral range estimation in the daisy family (Asteraceae)

Christiana McDonald-Spicer[1,2] | Nunzio J. Knerr[1] | Francisco Encinas-Viso[1] |
Alexander N. Schmidt-Lebuhn[1] (iD)

[1]CSIRO, Centre for Australian National Biodiversity Research, Canberra, Australian Capital Territory, Australia

[2]The Australian National University, Canberra, Australian Capital Territory, Australia

Correspondence
Alexander N. Schmidt-Lebuhn, CSIRO, Australian National Herbarium, Canberra, ACT, Australia.
Email: alexander.s-l@csiro.au

Editor: Felix Forest

Abstract

Aim: In recent years biogeography has been transformed by the increased availability of large-scale distributional data, phylogenies, and novel quantitative analysis methods and models. More case studies, however, are needed to test the performance of various approaches, in particular at global scales and in species-rich groups. In this study, we inferred bioregionalization and estimated ancestral areas for the largest plant family, the Asteraceae.

Location: Global.

Methods: We used the Global Compositae Checklist data to infer Asteraceae bioregions with cluster and modularity analysis. We reconstructed a phylogeny of genus-terminals for the Asteraceae family from a supermatrix of nuclear ribosomal internal transcribed spacer and chloroplast data. Combining areas based on the bioregions from modularity analysis and the phylogeny, we then estimated ancestral ranges across the Asteraceae phylogeny under 12 biogeographic models.

Results: Cluster analysis resulted in several small bioregions from areas with low taxon numbers and linear and disjunct bioregions between Eurasia and Africa. Modularity analysis produced larger and compact bioregions, and we based downstream analysis on its results. The favoured model for ancestral area estimation was BAYAREALIKE+j+x, demonstrating the importance of long distance dispersal in the biogeographic history of the Asteraceae and a strong distance-dependence of dispersal.

Main conclusions: Differences between cluster and modularity analysis suggest that the latter may be more robust to incomplete data and produces less disjunct and thus presumably biologically more realistic bioregions. With few exceptions, results of ancestral area estimation confirmed the results of previous studies, in particular South America as the ancestral area of the family, subsequent dispersal to and a secondary radiation from Africa, and the ancestral areas of individual tribes of the family.

KEYWORDS
ancestral ranges, Asteraceae, bioregions, Compositae, modularity analysis

1 | INTRODUCTION

The daisy or sunflower family Asteraceae is the largest family of flowering plants, rivalled only by the orchids (Funk, Susanna, Stuessy, & Bayer, 2009). Depending on the estimate, the family comprises *c.* 22,500–30,000 species, accounting for *c.* 10% of all flowering plants (Funk et al., 2009). It is also one of the fastest diversifying families of angiosperms (Barreda et al., 2012). There is large ecological and morphological variation across the family (Figure 1), with plants occurring from open grasslands to forests across all climate zones, and habits varying from herbs to shrubs, vines or trees (Funk et al., 2009).

One of the key morphological features of the Asteraceae is the flowering head. While there is great diversity within the family, all species have their flowers arranged into a condensed inflorescence known as the capitulum. The fruits are single-seeded, indehiscent, nearly always non-fleshy, and usually co-dispersed with the pappus, which is homologous to the calyx of other flowering plants and often developed as hairs or scales serving as a flight organ or as spines for exozoochorous dispersal.

The monophyly of Asteraceae is well established (Funk et al., 2005), although their internal classification continues to be updated. The family is currently divided into 12 subfamilies and 43 or more tribes (Funk et al., 2009), with exact numbers varying between authors.

1.1 | Biogeography and age of the Asteraceae

The first formal, family level and to date most comprehensive study of the biogeographic history of Asteraceae was published by Funk et al. (2005). It resolved South America as the ancestral area of the family, as had been suggested previously (Bremer, 1994), and inferred a large secondary radiation that produced several of the most diverse tribes (Anthemideae, Astereae, Calenduleae, Gnaphalieae, Senecioneae) to have taken place in Africa.

Subsequent biogeographic studies have focused either on the most probable dispersal pathways during the earliest stages of the evolution of the Asteraceae or examined individual tribes in more detail. Katinas, Crisci, Hoch, Tellería, and Apodaca (2013) argued that dispersal from South America to Africa would most likely have taken place via island hopping, as many early-diverging African lineages exhibited a growth habit typical of island floras. At the tribal level, the Cardueae have been inferred to be ancestrally West Asian (Barres et al., 2013), the Gnaphalieae as ancestrally South African (Bergh & Linder, 2009; Nie et al., 2016), and the Vernonieae as having originated in the Old World (Keeley, Forsman, & Chan, 2007).

The earliest pollen fossils of Asteraceae have been found in areas previously part of Gondwana, dated to 76–38 Ma (Barreda et al., 2015; Zavada & de Villiers, 2000), although the supercontinent most likely separated before the Asteraceae evolved. The oldest well-dated macrofossil was found in southern South America, dated to approximately 47.5 Ma (Barreda et al., 2012). Fossils found in Antarctica, however, complicate the historical biogeography of the Asteraceae (Barreda et al., 2015).

The first time-calibrated phylogeny of Asterids suggested an age of the split between Asteraceae and their sister family Calyceraceae of *c.* 51 Ma (Bremer, Friis, Bremer, & Linder, 2004). Recently, Barreda et al. (2015) inferred a considerably older age, dating the Asteraceae crown to *c.* 85.9 Ma based on a chloroplast phylogeny calibrated with new Antarctic pollen fossils dated to *c.* 76–66 Ma. Their interpretation was rejected by Panero (2016), followed by a rebuttal from Barreda et al. (2016). In another study, Panero and Crozier (2016) inferred a stem age of *c.* 69.5 Ma and a crown age of *c.* 64.75 Ma for the Asteraceae from a time-calibrated chloroplast phylogeny of selected Asterales and Apiales.

1.2 | Recent developments in biogeography

In recent years, biogeographic research has been transformed by the increased availability of occurrence records, phylogenies, and novel quantitative methods of analysis.

Large-scale distributional data are increasingly available in publicly accessible biodiversity databases such as GBIF (gbif.org), which aggregate geocoded specimen data from natural history collections. Their coverage, however, is uneven across the globe, with large areas remaining poorly represented (Engemann et al., 2015). In addition, the underlying data are curated to varying degrees of reliability, resulting in inconsistent taxonomy and geocoding errors (Maldonado et al., 2015). An alternative source of distributional data are checklists, such as the Global Compositae Checklist (GCC; compositae.landcareresearch.co.nz) for the daisy family Asteraceae. In contrast to point occurrence data they generally use unequal-size areas such as political units but provide better coverage and are based on expert opinion. At the same time, increasing amounts of sequence data are available, as are data matrices and phylogenies in databases such as TreeBase (treebase.org) and Dryad (datadryad.com).

The estimation of ancestral ranges and the inference of biogeographic events has benefited from the development of several explicit models and software tools. DIVA, or DIspersal-Vicariance Analysis, was originally developed as a parsimony algorithm allowing for both dispersal events and vicariance scenarios, although it favours vicariance (Ronquist, 1997). The popular Dispersal-Extinction-Cladogenesis model (DEC) models the processes of range expansion, local extinction and cladogenesis (Ree & Smith, 2008). The BayArea model was designed to remove the practical limitations on the number of areas that occur when using likelihood methods. It uses data augmentation and the Bayesian statistical framework to calculate likely ancestral ranges (Landis, Matzke, Moore, & Huelsenbeck, 2013). Most recently, variants of these models have been introduced that include additional parameters to model jump dispersal (j) and distance-dependence of dispersal (x) (Matzke, 2013; Van Dam & Matzke, 2016).

Freely available software such as RASP (Yu, Harris, Blair, & He, 2015) and the R package "BioGeoBEARS" (Matzke, 2014) allows the direct comparison of results from the most popular models. The latter has been developed in particular to enable model testing, in direct analogy to model comparison in other fields of evolutionary biology.

FIGURE 1 Representative members of major tribes of the Asteraceae, illustrating the diversity of the family. (a) *Barnadesia*, Barnadesieae; (b) *Mutisia acuminata*, Mutisieae; (c) *Cynara cardunculus*, globe artichoke, Cardueae; (d) *Cyanthillium cinereum*, Vernonieae; (e) *Tragopogon porrifolius*, purple salsify, Cichorieae, (f) *Arctotheca calendula*, cape weed, Arctotideae; (g) *Senecio vulgaris*, groundsel, Senecioneae; (h) *Aster alpinus*, Astereae; (i) *Matricaria chamomilla*, chamomile, Anthemideae; (j) *Xerochrysum subundulatum*, Gnaphalieae; (k) *Espeletia uribei*, Millerieae; (l) *Echinacea pallida*, Heliantheae. All photos by A.N.S.-L

This practice and the *j* parameter have recently been criticized by Ree and Sanmartín (2018), who argued that the node-based *j* parameter could favour "degenerate" solutions in which all biogeographic history is explained with founder events, and that models including the parameter cannot be compared statistically with those that do not. Maximization of the *j* parameter would, however, require biologically unrealistic scenarios in which all taxa are restricted to one area each and sister taxa never share an area, or a two-area problem with one area containing only one maximally nested taxon, and are consequently of little concern for studies using empirical data sets.

Objective delimitation of geographical areas for analyses such as the above remains a problem (Ferrari, 2018; Morrone, 2018). In the present context we are concerned with bioregions, understood as regions defined by the distribution of taxonomic groups as opposed to climate zones and vegetation zones (González-Orozco et al., 2014; Morrone, 2018). The general principle is to group smaller areas into larger bioregions based on shared taxa, thus maximizing the taxonomic homogeneity of the resulting bioregions and the differences between them (Stoddart, 1992), but numerous approaches have been suggested to achieve this. Early quantitative methods were developed by area cladists and used parsimony analysis of absence/presence matrices of taxa, for example, Cladistic Analysis of Distributions and Endemism (Porzecanski & Cracraft, 2005) and Parsimony Analysis of Endemicity (Nihei, 2006).

In recent years cluster analysis has been used increasingly in bioregionalization (Bradshaw, Colville, & Linder, 2015; Kreft & Jetz, 2010). It builds a hierarchical dendrogram using the dissimilarity of smaller areas in their taxon content, grouping the areas into clusters that are interpreted as bioregions (Mackey, Berry, & Brown, 2008), but it requires the user to make a ranking decision on the number of regions to accept. Most recently, network based approaches, for example, the map equation (Rosvall, Axelsson, & Bergstrom, 2009) and Modularity Analysis (Newman, 2006), have been explored. In a biogeographic context, bipartite networks in which areas are connected to, and thus through, all taxa occurring in them are divided into modules that are interpreted as bioregions. The analyses subdivide networks into non-hierarchical modules, thus providing an algorithmic solution for the number of bioregions to accept. To date, few studies have explicitly compared the performance of cluster and network approaches, and more case studies are needed (Bloomfield, Knerr, & Encinas-Viso, 2017; Morrone, 2018; Vilhena & Antonelli, 2015).

1.3 | Premise of the study

Because of at that time limited coverage of sequence data, Funk et al. (2005) constructed an Asteraceae supertree without branch lengths from the topologies of previously published phylogenies. (Branch lengths were estimated by a later publication that did not focus on biogeography [Torices, 2010].) Funk et al. inferred ancestral ranges using parsimony character tracing, which if interpreted as a model of biogeographic processes models only range shifting and sympatric speciation in a single area, and generally limits ancestral ranges to a single area (Ronquist, 1997).

In this study, we aimed to:

1. Delimit global Asteraceae bioregions using GCC data and quantitative approaches, and to compare the results against previous, less formal regionalizations.
2. Compare the bioregionalization results of cluster and modularity analysis, for the first time using a global scale data set.
3. Infer ancestral areas across the phylogeny of the Asteraceae using a supermatrix of available sequence data and biogeographic model testing, and compare the results against those of Funk et al.'s (2005) seminal study and other previous studies at the tribal level.
4. Estimate biogeographic events and the directionality of dispersal between regions to elucidate the global assembly of the present distribution of Asteraceae.

2 | MATERIALS AND METHODS

Additional information on the methodology is available as Appendix S1. Data matrices and phylogenetic trees are available in the CSIRO Data Access Portal, https://doi.org/10.25919/5bf781632e8ab.

2.1 | Spatial data

The spatial data set was based on data extracted from the GCC (compositae.landcareresearch.co.nz, accessed 15 Aug 2014), a database of distribution information for the Asteraceae family. We used OpenRefine (Huynh & Mazzocchi, 2014) to clean the data set, correcting spelling of taxon names, collapsing varieties and subspecies to species level, and removing hybrids and taxa with distribution listed as "null". Additional distribution information was added for New Zealand, the Cordoba Province in Argentina, Mongolia, South Africa, and Mexico. We removed species from regions where they are non-native. The final spatial data set used in this study included 27,019 species representing 1,636 genera. All analyses were conducted using the TDWG level 3 of spatial resolution.

2.2 | Bioregionalization

We tested two different approaches for grouping areas into larger bioregions, using genera as the taxonomic level. Cluster analysis was conducted in BIODIVERSE 1.1 (Laffan, Lubarsky, & Rosauer, 2010) with S2 dissimilarity, link average, and tie breaking by maximizing first corrected-weighted endemism and then weighted endemism. Modularity analysis was conducted in NETCARTO (Guimerà & Amaral, 2005).

2.3 | Phylogeny

Sequences from the nuclear ribosomal internal transcribed spacer region (ITS) and three chloroplast regions (*mat*K, *rbc*L, and *trn*L-*trn*F) were obtained from GenBank and BOLD (Ratnasingham & Hebert,
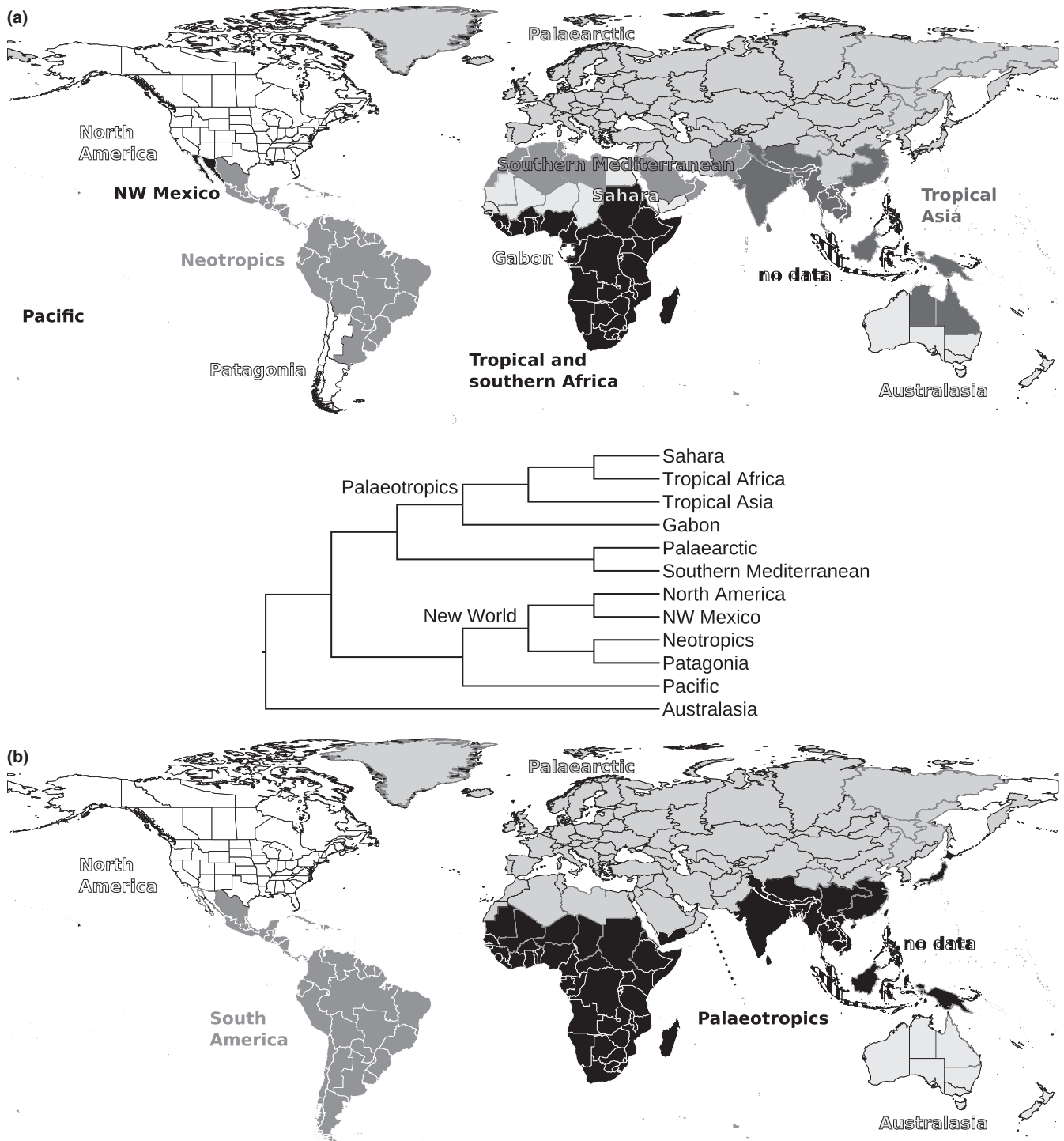
FIGURE 2 (a) Asteraceae bioregions inferred from WPGMA clustering and S2 dissimilarity in BIODIVERSE at the level of 12 clusters. The dendrogram shows the relationships of the clusters and accordingly what larger regions would be inferred for smaller numbers of clusters. (b) Asteraceae bioregions inferred from modularity analysis in NETCARTO. These regions were used for ancestral area inference except that the Palaeotropics were divided into an African and Asian area to increase resolution and make results more comparable with previous studies. The dotted line indicates the border

2007). We used genera as operational taxonomic units and selected a representative sequence for each genus and locus. GenBank accession numbers for all sequences used in this study are listed in Appendix S2.

Gene regions were individually aligned using MAFFT 7 (Katoh & Standley, 2013) and manually edited in BIOEDIT 7.0.5 (Hall, 1999). The four sequence regions were combined into a supermatrix of 1,273 genera and 9,030 characters. We tested ITS versus chloroplast

data using the incongruence length difference test (Farris, Källersjö, Kluge, & Bult, 1994) in TNT (Goloboff, Farris, & Nixon, 2008) with 99 replicates and did not detect significant incongruence. The final phylogeny was inferred using RAxML (Stamatakis, 2014) under the GTRCAT model and partitioning by sequence region. The tree was rooted on the Barnadesieae, which are sister to the rest of the family (Funk et al., 2005).

2.4 | Time calibration

Because Bayesian analysis is not computationally feasible for very large phylogenies, we time calibrated our phylogeny using penalized likelihood as implemented in the chronos function of the R package "APE" (Paradis, Claude, & Strimmer, 2004; R Core Team, 2016; Sanderson, 2002). We set nine calibration points (Appendix S3). We tested all three implemented clock models (relaxed, correlated, and discrete) and found discrete to be favoured.

2.5 | Ancestral area inference

Geographical units for ancestral area inference were based on the results of modularity analysis, but the Palaeotropics were divided into an African and an Asian part to provide a higher degree of resolution and make the analysis more comparable with that of Funk et al. (2005), resulting in a total of six areas (Figure 2b). Model comparisons and ancestral state estimates were conducted using the R package "BioGeoBEARS" (Matzke, 2013), limiting ranges to a size of two areas and removing all 135 genera occurring in more areas from the phylogeny, as well as 125 genera lacking spatial data.

We tested the models DIVALIKE (Ronquist, 1997), BayAreaLIKE (Landis et al., 2013) and DEC (Ree & Smith, 2008) each with and without the jump dispersal parameter j and with and without the x parameter, for a total of 12 models. To produce a matrix of geographical distances for estimation of the x parameter, we measured the shortest contemporary distance between any two areas in Google Earth (google.com/earth), disregarding oceanic islands. For use in "BioGeoBEARS," we scaled these distances to a maximum of 100 and considered all other area pairs adjacent, setting distances to 1 (Appendix S4).

We used biogeographic stochastic mapping (BSM) (Dupin et al., 2017) as implemented in "BioGeoBEARS" to estimate the number of different kinds of biogeographic events from 50 discrete historical scenarios fitting the optimized model.

3 | RESULTS

3.1 | Bioregionalization

At the level of six clusters, cluster analysis resolved the following Asteraceae bioregions: (a) North and South America; (b) the Palaearctic including the Mediterranean and Japan; (c) Gabon; (d) the Palaeotropics including New Guinea and northern Australia; (e)

southern Australia and New Zealand; and (f) a group of four Pacific island nations (Figure 2a). At a finer resolution of 12 clusters, the bioregions are further subdivided into: (a) North America; (b) northwestern Mexico; (c) the Neotropics; (d) Patagonia; (e) the Palaearctic; (f) a disjunct area including the southern Mediterranean coast except Egypt, Saudi Arabia and several of its neighbours, Afghanistan and Pakistan; (g) a disjunct area including the central Sarah region, Egypt and Yemen; (h) Gabon; (i) tropical and southern Africa and Madagascar; (j) tropical Asia, New Guinea and northern Australia; (k) southern Australia and New Zealand; and (l) a group of four Pacific island nations (Figure 2a).

The co-occurrence network was significantly modular ($M = 0.53$, $p < 0.001$) and detected five Asteraceae bioregions (Figure 2b): (a) North America and Magadan, Russia; (b) Central and South America; (c) the Palaearctic including the Mediterranean; (d) the Palaeotropics including also Japan, New Guinea, and the island nations that formed a separate region in cluster analysis; and (e) Australia and New Zealand (subsequently Australasia).

3.2 | Phylogeny

Following the divergence of the Barnadesieae, the phylogeny (Figure 3; Appendix S5) showed the genus *Hecastocleis* as sister to the remaining Asteraceae, followed by a clade of Gochnatieae and Wunderlichieae. They were followed by a clade of Hyalideae, Mutisieae, Nassauvieae, Onoserideae and Stifftieae. Pertyeae and *Erythrocephalum* were recovered as sister to the remaining Asteraceae, which included all large tribes. In this remainder, Carduoideae including Dicomeae, Oldenburgieae, Tarchonantheae, and the large thistle tribe Cardueae, Cichorioideae including Arctotideae, Cichorieae, Liabeae, Moquinieae, *Platycarphella*, and Vernonieae, *Gymnarrhena*, and *Corymbium* were arranged on a grade leading up to the large subfamily Asteroideae. Calenduleae were the sister clade of the other Asteroideae, followed by a clade of Anthemideae, Astereae and Gnaphalieae. Among the remaining Asteroideae, *Doronicum* and Senecioneae were sister to a large clade including Athroismeae, Inuleae, Plucheae, and a group of tribes sometimes referred to as the "Heliantheae alliance", including Bahieae, Coreopsideae, Eupatorieae, Helenieae, Heliantheae, Madieae, Millerieae, and Tageteae. Most of the deeper relationships in the phylogeny did not receive high bootstrap support, but the monophyly of many tribes such as Anthemideae, Astereae, Calenduleae, Gnaphalieae, Helenieae, Inuleae, Nassauvieae, Senecioneae, and Vernonieae was well supported (BS ≥75).

3.3 | Ancestral area inference

The favoured model in model comparison was BAYAREALIKE+j+x with a log likelihood of LnL = −1,889 and AIC weight of 0.9959 (Appendix S5). The second best model, BAYAREALIKE+j, had LnL = −1,982 and AIC weight of 0.0041. Estimated model parameters for BAYAREALIKE+j+x were $d = 0.019$ (dispersal, i.e., range expansion), $e = 0.013$ (local extinction), $j = 0.027$ (jump dispersal), and $x = -0.466$ (distance-dependence of dispersal).
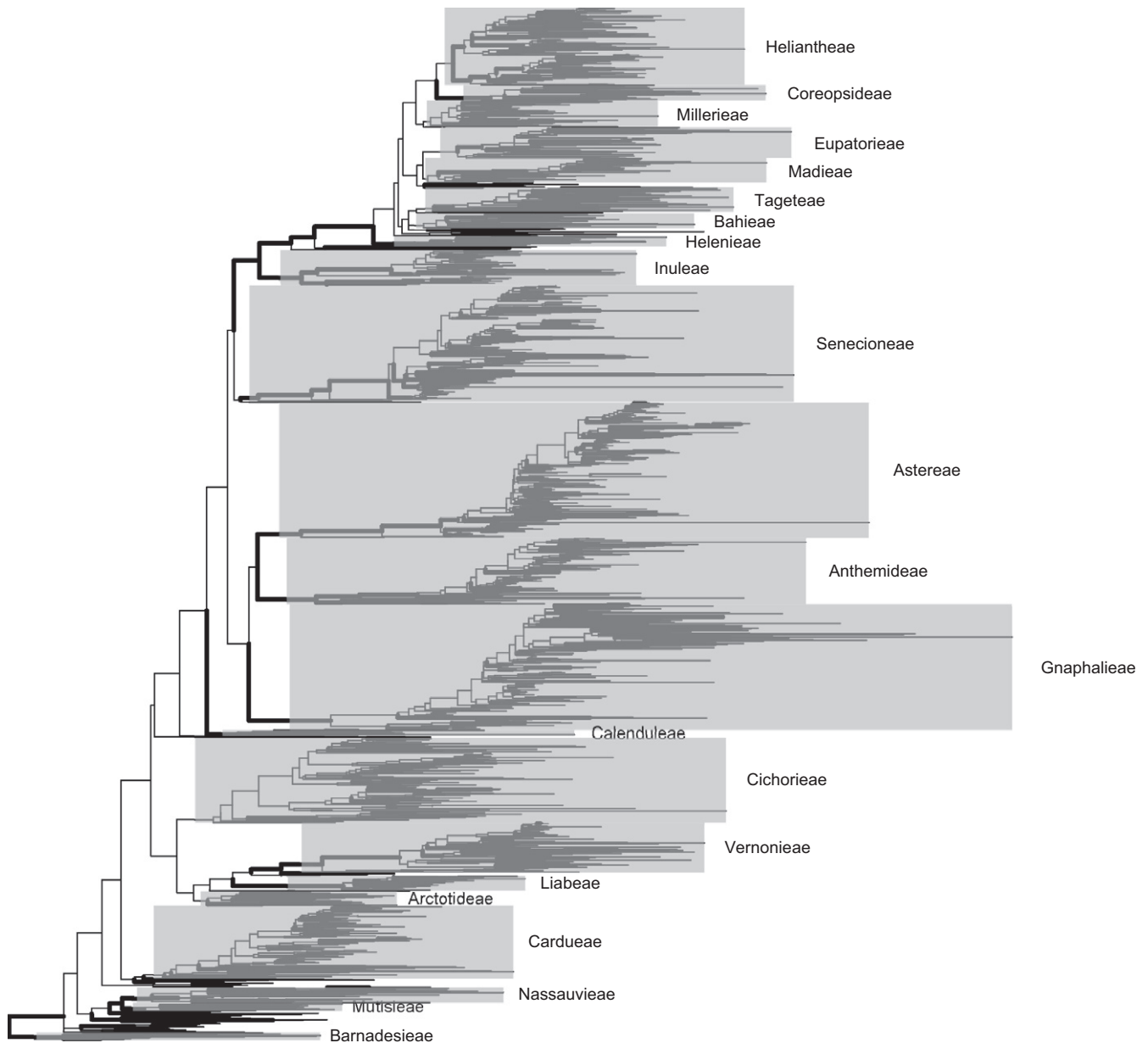
FIGURE 3 Phylogeny of the Asteraceae inferred from a sequence supermatrix analysed in RAxML and using genera as terminals. Major tribes are marked on the tree, and thick internodes indicate bootstrap support of 75 or above

Inference of the ancestral range of all Asteraceae was relatively uncertain, which is unsurprising because it represents the root state (Figure 4; Appendix S6). The highest probabilities were assigned to South America and Tropical Asia (0.31), South America (0.26) and South America and the Palaearctic (0.21). All six possible ranges including South America collectively accounted for 0.99 of the probability distribution. Early divergences were inferred to have occurred in South America with a higher likelihood. After the divergence of the Hyalideae, Mutisieae, Nassauvieae, Onoserideae and Stifftieae the ancestor of the remaining Asteraceae most likely dispersed to Africa (0.38; ≥0.79 after the divergence of Pertyeae). Carduoideae, Cichorioideae and Asteroideae were accordingly inferred to be ancestrally African (0.53, 0.79, and 0.92 respectively). The clade of

the tribes of the "Heliantheae alliance" was inferred as ancestrally North American (0.8).

Of individual tribes, Barnadesieae, Onoserideae, Mutisieae, Nassauvieae, Stifftieae, Wunderlichieae, Gochnatieae, Liabeae, Millerieae, Moquinieae, Neuroleneae, Perityleae, Eupatorieae, Coreopsideae, Vernonieae, and Heliantheae were inferred to be ancestrally South American; Bahieae, Helenieae, Chaenactideae, Madieae, and Tageteae as ancestrally North American; Cardueae as ancestrally Palaearctic; and Arctotideae, Cichorieae, Calenduleae, Astereae, Anthemideae, Gnaphalieae, and Senecioneae as ancestrally African. The most probable ancestral range of Inulae included two areas, Africa and Tropical Asia.

Biogeographic stochastic mapping estimated 224.7 (±9.8, standard deviation) events of range expansion, 908.7 (±4.1) of sympatric
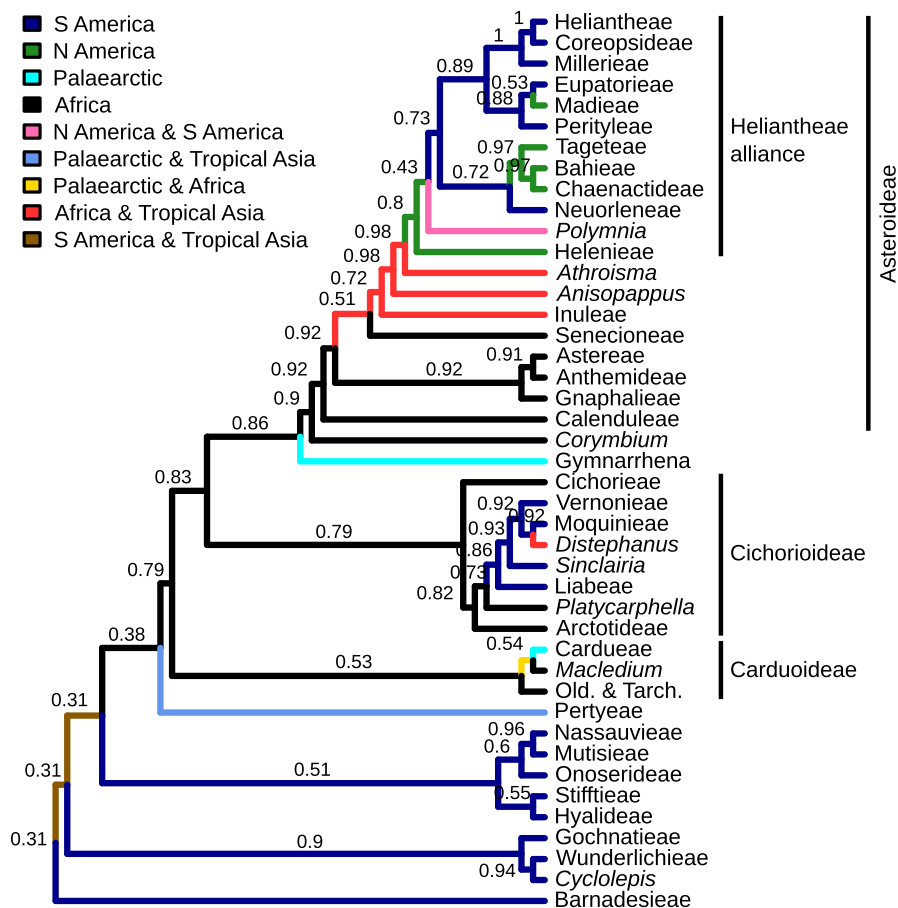
FIGURE 4 Results of ancestral area inference with the R package "BioGeoBEARS." Tree branches are coloured according to the most probable range (combination of areas as defined in Figure 2b) inferred for the ancestral lineage. Numbers above branches indicate the probability share of the relevant range. The phylogeny was here collapsed to the tribal level, and the complete phylogeny is available as Appendix S3

speciation, zero of local extinction, and 103.3 (±4.1) founder events. The highest number of estimated dispersal events between areas (including both range expansion and founder events) were from North America to South America with 58.92 (±5.46), South America to North America with 51.42 (±7.37), Palaearctic to Tropical Asia with 43.34 (±4.53), and Africa to Palaearctic with 29.52 (±6.55) (Figure 5a). All other combinations of areas accounted for less than twenty estimated dispersal events. Some rates of exchange were very uneven, for example, considerably fewer dispersal events were estimated for Tropical Asia to South America, the Palaearctic and Africa, or from the Palaearctic to Africa, than for the respective opposite directions.

4 | DISCUSSION

4.1 | Asteraceae bioregions

The number of studies using network analysis to define bioregions is still limited, although growing (Droissart et al., 2018). Additional studies exploring the performance of cluster and network approaches in direct comparison are needed (Bloomfield et al., 2017; Morrone, 2018).

At the level of six clusters our cluster analysis produced two very small and biologically unrealistic bioregions with very deep phenogram splits (Gabon, Pacific), and another as the number of clusters

was increased (NW Mexico). This may partly have resulted from incomplete data: Only eight genera were registered for Gabon, and three, one, and three, respectively, for the Pacific Island cells forming their own cluster, whereas the average number of genera in an area was 61.4 (±48.1). It is less clear what drove the recognition of the Mexican cluster, as the relevant cells had eleven and one hundred genera respectively. Modularity Analysis did not retrieve similarly small modules, suggesting it is more robust to incomplete data.

Another phenomenon restricted to the cluster analysis was the resolution of approximately linear but geographically disjunct clusters across northern Africa and the Middle East. These shapes and their positions suggest that the relevant clusters may represent interzones or transition zones between two larger bioregions, the Palaearctic and the Palaeotropics. Interzones remain a challenge for many if not all bioregionalization approaches (Morrone, 2018). A previous comparison between cluster analysis, map equation and modularity analysis suggested that the latter was the least likely of the three to recognize interzones as distinct bioregions (Bloomfield et al., 2017). In our case, modularity analysis likewise produced more compact and less disjunct regions than cluster analysis.

Of the differences between the two analyses, perhaps the most difficult to reconcile is the boundary between the Palaeotropics and Australasia. Cluster analysis assigned northern Australia to the former, presumably reflecting the presence of several predominantly extra-Australian, tropical genera in that area and the fact that the
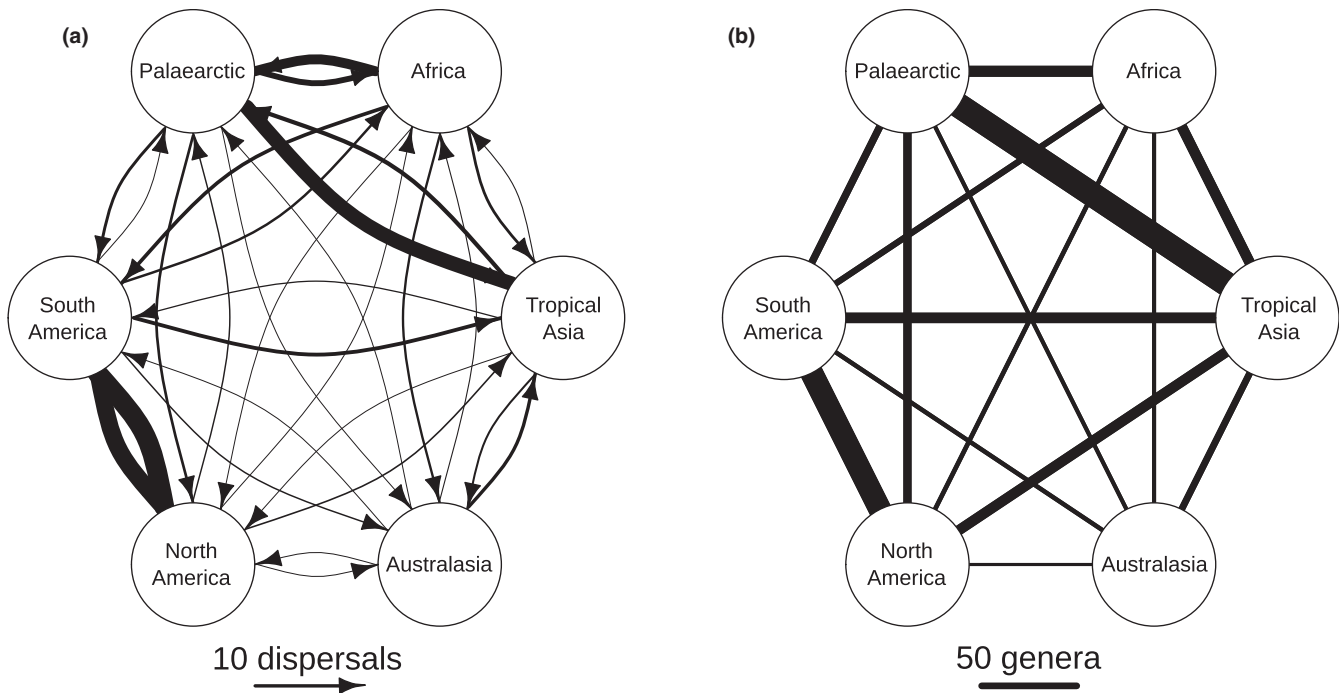
FIGURE 5 (a) Estimated dispersal events, including both range expansions and long distance dispersal, between areas as defined in Figure 2b estimated under the BAYAREALIKE+j+x model. (b) Illustration of the number of genera shared between areas

most diverse tribes in Australia (Gnaphalieae, Astereae and Senecioneae) have their diversity concentrated in the southern half of the continent (Schmidt-Lebuhn, Knerr, & Gonzalez-Orozco, 2012). Modularity analysis, however, assigned northern Australia to Australasia, producing a geographically more compact solution. Although missing data in South East Asia could be considered a potential cause for the discrepancy, New Guinea was consistently assigned to the Palaeotropics despite its closer biogeographic connections to Australasia. Both regionalizations are biologically realistic, and use of a finer spatial scale may have produced consistent results.

There are, however, also some interesting commonalities between cluster and network analyses. They include a clear separation of North America from the Palaearctic, as opposed to the recognition of a Holarctic region, and the absence of a separate Capensis region. The former presumably reflects the presence of many endemic genera of the Heliantheae alliance in North America (Funk et al., 2005). The latter could partly be a result of the taxonomic and spatial scale at which the analysis was conducted. At the genus level, endemicity in the Cape region may have been too low to resolve the region as distinct, and at the same time the TDWG level 3 Cape region may be large enough to include numerous genera more characteristic of the African tropics.

Funk et al. (2005, 2009) used an informal two-tiered Asteraceae bioregionalization to plot ranges on phylogenies, albeit without supplying precise boundaries. At the broader scale, they distinguished North America, South America, Eurasia, Africa and Australia-Pacific. This division differs from our results in particular in the assignment of Central America and the Caribbean, Northern Africa, and the Pacific, suggesting that it was intended to be geographic rather

than biogeographic. At the finer scale, Funk et al. recognized 17 regions with obviously biogeographic circumscription, dividing, for example, South America into Brazil, Guiana Shield, Northern and Central Andes, and southern South America. Some of our results are congruent with this regionalization, in particular the recognition of southern South America (Patagonia) as a finer scale cluster. There are, however, also significant differences, as our results consistently placed the northern Mediterranean with the Palaearctic (vs. with Northern Africa) and New Guinea outside of Australasia (vs. with).

Although it enabled much better geographical coverage than presently available point occurrence data, a drawback of the checklist approach used in this study is the use of biogeographically meaningless political boundaries. At a sufficiently large scale, here the entire globe, the TDWG level 3 areas can be expected to be small enough to nonetheless produce meaningful results. To further test the regionalization of Funk et al. (2005, 2009) and other previous suggestions in particular at finer scales and regarding the precise placement of boundaries, it would, however, be preferable to use either smaller areas or point distribution data (where available at sufficient coverage).

4.2 | Biogeographic history

With the inclusion of 1,013 terminals, our ancestral area inference represents the to date most comprehensive analysis for the Asteraceae. The preferred model, BAYAREA+j+x, models the biogeographic processes of range expansion ("dispersal"), local extinction, sympatric speciation, peripatric speciation ("founder" or "jump

dispersal"), and distance-dependent probability of dispersal, but does not allow vicariance and sympatric speciation in only a subset of the ancestral range.

As in all biogeographic analyses these processes apply at the geographical scale of the study, meaning for example that an inferred event of sympatric speciation relative to the entire Palaearctic would very likely have been allopatric at a finer geographical scale. At the large scale of the areas used in this study it was consequently unsurprising that the biogeographic event estimated as most important was sympatry, that is, the same range for ancestor and both descendants, and that a model without vicariance events was preferred. Model optimization and BSM also indicated, however, that peripatric speciation was an important process in the biogeographic history of the Asteraceae, as suggested by previous studies including evidence for many radiations on oceanic islands (Baldwin, Kyhos, Dvorak, & Carr, 1991; Francisco-Ortega, Santos-Guerra, Hines, & Jansen, 1997; Sancho, de Lange, Donato, Barkla, & Wagstaff, 2015; Swenson, Nylinder, & Wagstaff, 2012; Vijverberg, Mes, & Bachmann, 1999), the aggressive colonizing ability of many Asteraceae, and the most common dispersal syndrome of the family (Katinas et al., 2013).

Previous studies inferring ancestral ranges used simpler inference methods or were restricted to individual tribes, which has the disadvantage of increasing uncertainty for ranges close to the root. In many cases, however, our results confirmed those of previous studies. A South American origin of the Asteraceae had been suggested by several authors (Bentham, 1873; Bremer, 1994) and formally inferred in the first family scale analysis (Funk et al., 2005). Several early Asteraceae fossils have been found on the continent (Barreda, Palazzesi, & Tellería, 2008; Barreda et al., 2010, 2012), but the recent controversy around an Antarctic pollen fossil (Barreda et al., 2015, 2016; Panero, 2016) also draws attention to the obvious problem that ancestral range estimation with the most commonly used models and using only extant ranges as input is by necessity unable to take into account wholesale local extinction. It is consequently probable that the ancestral range included (then adjacent) Antarctica.

A large secondary radiation took place after dispersal of an ancestral species to Africa, producing with Carduoideae, Cichorioideae, and Asteroideae the largest subfamilies of Asteraceae, which all independently colonized all continents except Antarctica (Funk et al., 2005; Katinas et al., 2013). The large assemblage of tribes known as the Heliantheae alliance is most diverse in the Americas, and their inferred ancestral range was North America, matching the results of Funk et al. (2005).

For individual tribes, our results are in accord with previous studies showing Anthemideae and Gnaphalieae as ancestrally African (Bergh & Linder, 2009; Nie et al., 2016; Watson, Evans, & Boluarte, 2000), Cardueae as ancestrally Asian (Barres et al., 2013) but Carduoideae as a whole as ancestrally African (Funk et al., 2005).

The Vernonieae were inferred to have had a most probable ancestral range in South America. This contrasts with previous studies suggesting an origin in Africa and subsequent dispersal to South America (Funk et al., 2005; Keeley et al., 2007), based in particular on two African clades forming a grade leading up to a South American clade. The results of our analysis are clearly influenced by outgroup states, where Moquinieae and Liabeae are ancestrally South American.

Cichorieae have previously been inferred to be ancestrally Mediterranean or North African (Tremetsberger et al., 2013), areas that would be part of the Palaearctic region in our study. Our results suggest an origin in subsaharan Africa, presumably likewise based on outgroup states. In this case, however, the discrepancy is less significant, because both regions would be situated on the same landmass, and changing climates would have blurred historical boundaries.

4.3 | Floristic exchange

Our historical analysis excluded terminals with ranges of more than two areas, so that many anagenetic dispersal events at the tips of the tree are missing from the BSM estimates. They are captured by plotting the number of genera shared between areas, which is missing the historical dimension. Despite this difference, both plots show a surprisingly similar pattern of strong connections between North and South America and between the Palaearctic and Africa as well as Tropical Asia (Figure 5b).

This may in most cases simply reflect geographical distance as, conversely, similar climate does not appear to be a predictor of high dispersal or taxonomic overlap, which are low for Africa—Tropical Asia but high for (temperate) North America and (largely tropical) South America. Australasia is the least connected area, reflecting its long floristic isolation (Byrne et al., 2008).

Of particular interest are unbalanced floristic exchanges. Considerably more dispersal events were estimated to flow from Africa to all other regions except North America (exchange with which is negligible), potentially reflecting the role of the continent as the setting of a large secondary radiation (Funk et al., 2005; Katinas et al., 2013). A potential explanation for a much larger number of movements from the Palaearctic to Tropical Asia than vice versa could be the availability of cold mountain habitats in the tropics and the absence of comparable warm "pockets" in temperate to Arctic regions (Gehrke & Linder, 2009).

This study collated a larger amount of data than was available for previous studies at the level of the entire Asteraceae family, providing insights into its global floristic assembly and biogeographic history. Given the size of the family and analytic constraints, further study is required to elucidate finer scale patterns. As computational power increases and more sequence data become available, a species-level phylogeny would provide more robust estimates of biogeographic events. Conversely, smaller phylogenies with more precise time calibrations will increase our understanding of the timing of events (e.g., Sancho et al., 2015; Wagstaff, Breitwieser, & Swenson, 2006). Similarly, more detailed spatial data than presently available will be required to test the boundaries of floral assemblages at regional scales.

DATA ACCESSIBILITY

Data used in this study are included as appendixes or available in publicly accessible databases.

ORCID

Alexander N. Schmidt-Lebuhn http://orcid.org/0000-0002-7402-8941

REFERENCES

Baldwin, B. G., Kyhos, D. W., Dvorak, J., & Carr, G. D. (1991). Chloroplast DNA evidence for a North American origin of the Hawaiian silversword alliance (Asteraceae). *Proceedings of the National Academy of Sciences*, 88, 1840–1843. https://doi.org/10.1073/pnas.88.5.1840

Barreda, V. D., Palazzesi, L., Katinas, L., Crisci, J. V., Tellería, M. C., Bremer, K., … Corsolini, R. (2012). An extinct Eocene taxon of the daisy family (Asteraceae): Evolutionary, ecological and biogeographical implications. *Annals of Botany*, 109, 127–134. https://doi.org/10.1093/aob/mcr240

Barreda, V., Palazzesi, L., & Tellería, M. C. (2008). Fossil pollen grains of Asteraceae from the Miocene of Patagonia: Nassauviinae affinity. *Review of Palaeobotany and Palynology*, 151, 51–58. https://doi.org/10.1016/j.revpalbo.2008.02.002

Barreda, V. D., Palazzesi, L., Tellería, M. C., Katinas, L., Crisci, J. V., Bremer, K., … Bechis, F. (2010). Eocene Patagonia fossils of the daisy family. *Science*, 329, 1621–1621. https://doi.org/10.1126/science.1193108

Barreda, V. D., Palazzesi, L., Tellería, M. C., Olivero, E. B., Raine, J. I., & Forest, F. (2015). Early evolution of the angiosperm clade Asteraceae in the Cretaceous of Antarctica. *Proceedings of the National Academy of Sciences*, 112, 10989–10994. https://doi.org/10.1073/pnas.1423653112

Barreda, V. D., Palazzesi, L., Tellería, M. C., Olivero, E. B., Raine, J. I., & Forest, F. (2016). Reply to Panero: Robust phylogenetic placement of fossil pollen grains: The case of Asteraceae. *Proceedings of the National Academy of Sciences*, 113, E412–E412. https://doi.org/10.1073/pnas.1521642113

Barres, L., Sanmartín, I., Anderson, C. L., Susanna, A., Buerki, S., Galbany-Casals, M., & Vilatersana, R. (2013). Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *American Journal of Botany*, 100, 867–882. https://doi.org/10.3732/ajb.1200058

Bentham, G. (1873). Notes on the classification, history, and geographical distribution of Compositæ. *Journal of the Linnean Society of London, Botany*, 13, 335–577. https://doi.org/10.1111/j.1095-8339.1873.tb02575.x

Bergh, N. G., & Linder, P. H. (2009). Cape diversification and repeated out-of-southern-Africa dispersal in paper daisies (Asteraceae–Gnaphalieae). *Molecular Phylogenetics and Evolution*, 51, 5–18. https://doi.org/10.1016/j.ympev.2008.09.001

Bloomfield, N. J., Knerr, N., & Encinas-Viso, F. (2017). A comparison of network and clustering methods to detect biogeographical regions. *Ecography*, 41, 1–10.

Bradshaw, P. L., Colville, J. F., & Linder, H. P. (2015). Optimising regionalisation techniques: Identifying centres of endemism in the extraordinarily endemic-rich cape floristic region. *PLoS ONE*, 10, e0132538. https://doi.org/10.1371/journal.pone.0132538

Bremer, K. (1994). *Asteraceae: Cladistics and classification*. Portland: Timber Press.

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

Byrne, M., Yeates, D. K., Joseph, L., Kearney, M., Bowler, J., Williams, M. A. J., … Wyrwoll, K.-H. (2008). Birth of a biome: Insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology*, 17, 4398–4417. https://doi.org/10.1111/j.1365-294X.2008.03899.x

Droissart, V., Dauby, G., Hardy, O. J., Deblauwe, V., Harris, D. J., Janssens, S., … Couvreur, T. L. P. (2018). Beyond trees: Biogeographical regionalization of tropical Africa. *Journal of Biogeography*, 45, 1153–1167. https://doi.org/10.1111/jbi.13190

Dupin, J., Matzke, N. J., Särkinen, T., Knapp, S., Olmstead, R. G., Bohs, L., & Smith, S. D. (2017). Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography*, 44, 887–899. https://doi.org/10.1111/jbi.12898

Engemann, K., Enquist, B. J., Sandel, B., Boyle, B., Jørgensen, P. M., Morueta-Holme, N., … Svenning, J. (2015). Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution*, 5, 807–820. https://doi.org/10.1002/ece3.1405

Farris, J. S., Källersjö, M., Kluge, A. G., & Bult, C. (1994). Testing significance of incongruence. *Cladistics*, 10, 315–319. https://doi.org/10.1111/j.1096-0031.1994.tb00181.x

Ferrari, A. (2018). Biogeographical units matter. *Australian Systematic Botany*, 30, 391–402.

Francisco-Ortega, J., Santos-Guerra, A., Hines, A., & Jansen, R. K. (1997). Molecular evidence for a Mediterranean origin of the Macaronesian endemic genus *Argyranthemum* (Asteraceae). *American Journal of Botany*, 84, 1595–1613. https://doi.org/10.2307/2446622

Funk, V. A., Bayer, B. J., Keeley, S., Chan, R., Watson, L., Gemeinholzer, B., … Jansen, R. K. (2005). Everywhere but Antarctica: Using a supertree to understand the diversity and distribution of the Compositae. *Biologiske Skrifter*, 55, 343–374.

Funk, V. A., Susanna, A., Stuessy, T. F., & Bayer, R. J. (2009). *Systematics, evolution, and biogeography of Compositae*. Vienna: International Association for Plant Taxonomy.

Gehrke, B., & Linder, H. P. (2009). The scramble for Africa: Pan-temperate elements on the African high mountains. *Proceedings of the Royal Society of London B: Biological Sciences*, 276, 2657–2665. https://doi.org/10.1098/rspb.2009.0334

Goloboff, P. A., Farris, J. S., & Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics*, 24, 774–786. https://doi.org/10.1111/j.1096-0031.2008.00217.x

González-Orozco, C. E., Ebach, M. C., Laffan, S., Thornhill, A. H., Knerr, N. J., Schmidt-Lebuhn, A. N., … Miller, J. T. (2014). Quantifying phytogeographical regions of Australia using geospatial turnover in species composition. *PLoS ONE*, 9, e92558. https://doi.org/10.1371/journal.pone.0092558

Guimerà, R., & Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature*, 433, 895. https://doi.org/10.1038/nature03288

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.

Huynh, D., & Mazzocchi, S. (2014). *OpenRefine*. Retrieved from http://openrefine.org

Katinas, L., Crisci, J. V., Hoch, P., Tellería, M. C., & Apodaca, M. J. (2013). Trans-oceanic dispersal and evolution of early composites

(Asteraceae). *Perspectives in Plant Ecology, Evolution and Systematics*, *15*, 269–280. https://doi.org/10.1016/j.ppees.2013.07.003

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Keeley, S. C., Forsman, Z. H., & Chan, R. (2007). A phylogeny of the "evil tribe" (Vernonieae: Compositae) reveals Old/New World long distance dispersal: Support from separate and combined congruent datasets (trnL-F, ndhF, ITS). *Molecular Phylogenetics and Evolution*, *44*, 89–103. https://doi.org/10.1016/j.ympev.2006.12.024

Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, *37*, 2029–2053. https://doi.org/10.1111/j.1365-2699.2010.02375.x

Laffan, S. W., Lubarsky, E., & Rosauer, D. F. (2010). Biodiverse, a tool for the spatial analysis of biological and related diversity. *Ecography*, *33*, 643–647. https://doi.org/10.1111/j.1600-0587.2010.06237.x

Landis, M. J., Matzke, N. J., Moore, B. R., & Huelsenbeck, J. P. (2013). Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, *62*, 789–804. https://doi.org/10.1093/sysbio/syt040

Mackey, B. G., Berry, S. L., & Brown, T. (2008). Reconciling approaches to biogeographical regionalization: A systematic and generic framework examined with a case study of the Australian continent. *Journal of Biogeography*, *35*, 213–229.

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., … Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? *Global Ecology and Biogeography*, *24*, 973–984. https://doi.org/10.1111/geb.12326

Matzke, N. J. (2013). Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, *5*, 242–248.

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, *63*, 951–970. https://doi.org/10.1093/sysbio/syu056

Morrone, J. J. (2018). The spectre of biogeographical regionalization. *Journal of Biogeography*, *45*, 282–288. https://doi.org/10.1111/jbi.13135

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*, 8577–8582. https://doi.org/10.1073/pnas.0601602103

Nie, Z.-L., Funk, V. A., Meng, Y., Deng, T., Sun, H., & Wen, J. (2016). Recent assembly of the global herbaceous flora: Evidence from the paper daisies (Asteraceae: Gnaphalieae). *New Phytologist*, *209*, 1795–1806. https://doi.org/10.1111/nph.13740

Nihei, S. S. (2006). Misconceptions about parsimony analysis of endemicity. *Journal of Biogeography*, *33*, 2099–2106. https://doi.org/10.1111/j.1365-2699.2006.01619.x

Panero, J. L. (2016). Phylogenetic uncertainty and fossil calibration of Asteraceae chronograms. *Proceedings of the National Academy of Sciences*, *113*, E411–E411. https://doi.org/10.1073/pnas.1517649113

Panero, J. L., & Crozier, B. S. (2016). Macroevolutionary dynamics in the early diversification of Asteraceae. *Molecular Phylogenetics and Evolution*, *99*, 116–132. https://doi.org/10.1016/j.ympev.2016.03.007

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Porzecanski, A. L., & Cracraft, J. (2005). Cladistic analysis of distributions and endemism (CADE): Using raw distributions of birds to unravel the biogeography of the South American aridlands. *Journal of Biogeography*, *32*, 261–275. https://doi.org/10.1111/j.1365-2699.2004.01138.x

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from http://www.r-project.org.

Ratnasingham, S., & Hebert, P. D. N.. (2007). BOLD: The barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography*, *45*, 741–749. https://doi.org/10.1111/jbi.13173

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

Ronquist, F. (1997). Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, *46*, 195–203. https://doi.org/10.1093/sysbio/46.1.195

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *The European Physical Journal Special Topics*, *178*, 13–23. https://doi.org/10.1140/epjst/e2010-01179-1

Sancho, G., de Lange, J. P., Donato, M., Barkla, J., & Wagstaff, S. J. (2015). Late Cenozoic diversification of the austral genus *Lagenophora* (Astereae, Asteraceae). *Botanical Journal of the Linnean Society*, *177*, 78–95. https://doi.org/10.1111/boj.12224

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution*, *19*, 101–109. https://doi.org/10.1093/oxfordjournals.molbev.a003974

Schmidt-Lebuhn, A. N., Knerr, N. J., & Gonzalez-Orozco, C. E. (2012). Distorted perception of the spatial distribution of plant diversity through uneven collecting efforts: The example of Asteraceae in Australia. *Journal of Biogeography*, *39*, 2072–2080. https://doi.org/10.1111/j.1365-2699.2012.02756.x

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

Stoddart, D. R. (1992). Biogeography of the Tropical Pacific. *Pacific Science*, *46*, 276–293.

Swenson, U., Nylinder, S., & Wagstaff, S. J. (2012). Are Asteraceae 1.5 billion years old? A reply to heads. *Systematic Biology*, *61*, 522–532. https://doi.org/10.1093/sysbio/syr121

Torices, R. (2010). Adding time-calibrated branch lengths to the Asteraceae supertree. *Journal of Systematics and Evolution*, *48*, 271–278. https://doi.org/10.1111/j.1759-6831.2010.00088.x

Tremetsberger, K., Gemeinholzer, B., Zetzsche, H., Blackmore, S., Kilian, N., & Talavera, S. (2013). Divergence time estimation in Cichorieae (Asteraceae) using a fossil-calibrated relaxed molecular clock. *Organisms Diversity & Evolution*, *13*, 1–13. https://doi.org/10.1007/s13127-012-0094-2

Van Dam, M. H., & Matzke, N. J. (2016). Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *Journal of Biogeography*, *43*, 1514–1532.

Vijverberg, K., Mes, T. H. M., & Bachmann, K. (1999). Chloroplast DNA evidence for the evolution of *Microseris* (Asteraceae) in Australia and New Zealand after long-distance dispersal from western North America. *American Journal of Botany*, *86*, 1448–1463. https://doi.org/10.2307/2656926

Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, *6*, 6848. https://doi.org/10.1038/ncomms7848

Wagstaff, S. J., Breitwieser, I., & Swenson, U. (2006). Retrieved from https://www.ingentaconnect.com/content/iapt/tax/2006/00000055/00000001/art00012.

Watson, L. E., Evans, T. M., & Boluarte, T. (2000). Molecular phylogeny and biogeography of tribe Anthemideae (Asteraceae), based on

chloroplast gene *ndh*F. *Molecular Phylogenetics and Evolution*, *15*, 59–69. https://doi.org/10.1006/mpev.1999.0714

Yu, Y., Harris, A. J., Blair, C., & He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. *Molecular Phylogenetics and Evolution*, *87*, 46–49. https://doi.org/10.1016/j.ympev.2015.03.008

Zavada, M., & de Villiers, S. (2000). Pollen of the Asteraceae from the Paleocene-Eocene of South Africa. *Grana*, *39*, 39–45. https://doi.org/10.1080/00173130150503795

BIOSKETCH

The research team at the Centre for Australian National Biodiversity Research including the Australian National Herbarium conducts research on systematics, evolution, biogeography, ecology and conservation of plants with a focus on the Australian flora (https://www.anbg.gov.au/cpbr/herbarium/).

Author contributions: A.N.S.L. conceived the study; C.Mc.D.S. prepared the data; C.Mc.D.S., N.J.K., F.E.V. and A.N.S.L. analysed the data; C.Mc.D.S. and A.N.S.L. led the writing.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.