ORIGINAL ARTICLE

doi:10.1111/evo.13276

The phylogeny and biogeography of *Hakea* (Proteaceae) reveals the role of biome shifts in a continental plant radiation

Marcel Cardillo,[1,2] Peter H. Weston,[3] Zoe K. M. Reynolds,[1] Peter M. Olde,[3] Austin R. Mast,[4] Emily M. Lemmon,[4] Alan R. Lemmon,[5] and Lindell Bromham[1]

[1] *Macroevolution and Macroecology Group, Research School of Biology, Australian National University, Canberra 0200, Australia*

[2] *E-mail: marcel.cardillo@anu.edu.au*

[3] *National Herbarium of New South Wales, Royal Botanic Gardens and Domain Trust, Sydney NSW 2000, Australia*

[4] *Department of Biological Science, Florida State University, Tallahassee, Florida 32306*

[5] *Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, Florida 32306*

Received December 8, 2016
Accepted May 4, 2017

The frequency of evolutionary biome shifts during diversification has important implications for our ability to explain geographic patterns of plant diversity. Recent studies present several examples of biome shifts, but whether frequencies of biome shifts closely reflect geographic proximity or environmental similarity of biomes remains poorly known. We explore this question by using phylogenomic methods to estimate the phylogeny of *Hakea*, a diverse Australian genus occupying a wide range of biomes. Model-based estimation of ancestral regions indicates that *Hakea* began diversifying in the Mediterranean biome of southern Australia in the Middle Eocene–Early Oligocene, and dispersed repeatedly into other biomes across the continent. We infer around 47 shifts between biomes. Frequencies of shifts between pairs of biomes are usually similar to those expected from their geographic connectedness or climatic similarity, but in some cases are substantially higher or lower than expected, perhaps reflecting how readily key physiological traits can be modified to adapt lineages to new environments. The history of frequent biome-shifting is reflected in the structure of present-day assemblages, which tend to be more phylogenetically diverse than null-model expectations. The case of *Hakea* demonstrates that the radiation of large plant clades across wide geographic areas need not be constrained by dispersal limitation or conserved adaptations to particular environments.

KEY WORDS: Anchored enrichment phylogenomics, diversification, geographic ranges, niche conservatism, species tree.

Large-scale studies of plant species distribution and phylogeny often point to a strong tendency toward phylogenetic conservatism of environmental niches. According to these studies it is usual for species to inhabit a similar environment to their ancestors, and rare for evolutionary shifts to new habitats or biomes (geographic regions of broadly similar ecoclimatic character) to occur (Prinzing 2001; Qian and Ricklefs 2004; Crisp et al. 2009; Kerkhoff et al. 2014). At the same time, more narrowly focused studies of particular plant clades seem to offer evidence for frequent evolutionary shifts between habitats or biomes (e.g., Holstein and Renner 2011; Töpel et al. 2012; Jara-Arancio et al. 2014; Weeks et al. 2014; Souza-Neto et al. 2016, further references in Donoghue and Edwards 2014). The strength of affinity of plant lineages to particular environments through evolutionary time has important implications for our understanding of the historical events and evolutionary processes that underlie present-day patterns of plant diversity. If environmental niches tend to be conserved, then most diversification is likely to occur within biomes, with limits to diversity mediated by the age and size of the biome. Shifts between biomes will be comparatively rare, and assemblages of species within biomes will tend to be phylogenetically clustered: that is,

This article corresponds to Chloe M. N. (2017), Digest: Shifting biomes: Insight into patterns of plant radiation and dispersal. Evolution. https://doi.org/10.1111/evo.13300.

they will be more closely related to each other than to species from other biomes (Fine et al. 2014; Kerkhoff et al. 2014). So, under prevailing niche conservatism the history of diversification of plant clades is likely to be strongly coupled with the history of the physical environment, with opportunities for diversification determined by the expansion, contraction, appearance, and disappearance of biomes (Donoghue 2008; Donoghue and Edwards 2014). On the other hand, if environmental niches are more labile and evolutionary shifts between biomes common, the diversification of plant clades may be less tightly coupled with the history of changing environments. Instead, much diversification may be stimulated by biome shifts, and thus more strongly linked to the evolution of traits that adapt species to particular climates.

Many previous analyses of biome or habitat shifts are based on informal or parsimony-based reconstructions of ancestral distributions from phylogenies (see Donoghue and Edwards 2014), but an increasing number apply model-based, statistical biogeographic methods (e.g., Koecke et al. 2013; Jara-Arancio et al. 2014; Kerkhoff et al. 2014; Weeks et al. 2014; Duchêne and Cardillo 2015). Regardless of the method of inferring biome shifts, the number of shifts can only be judged to be unusually high or low with reference to some kind of null expectation. It is challenging to devise a meaningful null model for absolute numbers of biome shifts, but it may be more tractable to do so for the relative numbers of shifts expected between different pairs of biomes. Donoghue and Edwards (2014) presented a conceptual framework for interpreting biome shifts, in which the probability that a lineage will undergo a biome shift is a product of (1) the geographic opportunity for movement, (2) the lineage's intrinsic capacity for evolution into new environments, and (3) biotic interactions, for example competitive resistance of the existing assemblage in a biome to new invaders. Of these, the first is perhaps the most straightforward to parameterize because it represents a set of fairly simple sampling effects. The geographic opportunity for biome shifts to occur, and hence the probability of a biome shift, should increase as the sizes of the ancestral and descendant biomes increase, because larger biomes tend to harbor more species, and because they offer a larger "target" for range shifts. Shift probability should also increase with species richness of the ancestral biome independently of geographic area, with the age of the ancestral and descendant biomes, with the connectedness of biomes (e.g., the length of the shared boundary of two biomes) on relevant timescales, and with the environmental similarity of biomes, assuming that a shift into a similar environment is more likely because it requires a smaller adaptive change (Donoghue and Edwards 2014).

A thorough exploration of the role that biome conservatism and biome shifts play in plant radiations requires the integration of species-level data on phylogeny and geographic distributions, ideally for large, widely distributed plant clades that occupy several distinct biomes. The set of analyses we present in this paper are based on *Hakea* Schrad. & J. C.Wendl. (Proteaceae), a genus of 151 sclerophyllous shrub and tree species endemic to Australia and nearby islands. *Hakea* reaches greatest diversity (approximately 60% of species) in the Mediterranean-climate shrublands and heathlands of southern, and particularly southwestern, Australia. However, the genus is distributed across the Australian continent, and occupies most of the major biomes in Australia, from arid deserts to humid coastal forests. Knowledge of the mechanisms of dispersal in *Hakea* is limited, but most species have winged seeds, suggesting wind-dispersal, while insects and vertebrates are likely also involved in seed dispersal (Barker et al. 1999).

We used data from hybrid enrichment targeting hundreds of nuclear loci (Lemmon et al. 2012) to construct the first near-complete species-level phylogeny of *Hakea*, and we used this in combination with species distribution data to carry out statistical, model-based estimates of ancestral biomes, and infer evolutionary shifts between biomes. Our overarching question was whether the radiation of *Hakea* across the Australian continent was driven more by biome conservatism and within-biome diversification, or by evolutionary transitions between biomes (biome shifts). To answer this question, we asked a number of more specific questions:

1. Does present-day biome occupancy show phylogenetic signal that is stronger, weaker, or as expected under a neutral model of geographic range evolution?
2. Which biome was occupied by the most recent common ancestor of all extant *Hakea*, and from where did the ancestral lineage begin its geographic expansion?
3. How many biome shifts occurred overall, and between each pair of biomes, during the diversification of *Hakea*? How do these numbers compare with the relative numbers of shifts expected from the geographic proximity and environmental similarity of biomes?
4. Does the phylogenetic structure of assemblages within biomes show clustering (indicative of biome conservatism) or overdispersion (indicative of multiple biome shifts)?

Material and Methods

SAMPLE COLLECTION AND DNA EXTRACTION

Tissue samples for *Hakea* and outgroup taxa were collected from cultivated plants grown at the Australian National Botanic Gardens (Canberra) the Royal Botanic Garden Sydney, the Australian Botanic Garden Mt Annan, the Blue Mountains Botanic Garden Mt Tomah, and a private collection (Paul Kennedy, Strathmerton), as well as from plants growing in the wild. The list of samples with herbarium accession numbers is given in the Supplementary

Material (Table S1). We extracted total genomic DNA using the Qiagen DNEasy Plant Mini Kit according to the manufacturer's protocol (Qiagen Inc., California, USA).

ANCHORED HYBRID ENRICHMENT

The anchored hybrid enrichment approach uses taxon-specific probes (in this case, for angiosperms) to target highly conserved "anchor" regions of the nuclear genome, flanked by less conserved regions. This results in the capture of sequences for hundreds of loci representing a mix of coding regions, introns, and other sequences. The method is described in detail elsewhere (Lemmon et al. 2012; Prum et al. 2015), and development of the angiosperm enrichment probe kit used for this study is described in a forthcoming paper (Buddenhagen et al. manuscript). Here, we provide only a brief overview of the methods with details specific to this study.

Library preparation

Library preparation and read data processing were carried out at the Center for Anchored Phylogenomics at Florida State University. Genomic DNA was sonicated to a fragment size of ∼200–600 bp before library preparation and indexing following a modified protocol from Meyer and Kircher (2010). Indexed samples were pooled and enriched using the Angiosperm v.1 enrichment kit (Buddenhagen et al. manuscript). Sequencing was done on 4.5 PE150 Illumina HiSeq 2500 lanes (190 Gb total yield) at the Translational Science Laboratory, College of Medicine, Florida State University.

Read assembly

To increase read accuracy and length, paired reads were merged before assembly, following Rokyta et al. (2012). Reads were mapped to the probe regions using *Arabidopsis thaliana*, *Billbergia nutans*, and *Carex lurida* as references, combined with a de novo assembly approach to extend the assembly into flanking regions (Prum et al. 2015; Buddenhagen et al. manuscript). Read files were traversed repeatedly until no additional mapped reads were produced. Following read assembly, consensus bases were called from assembly clusters either as ambiguous or unambiguous bases, depending on the probability of sequencing error. Assembly contigs based on fewer than 100 reads were removed to reduce effects of rare sequencing errors.

Orthology assessment

For each locus, orthology was determined following procedures described in Prum et al. (2015). A pairwise distance matrix among homologs was calculated using an alignment-free approach, and used to cluster sequences with a neighbor-joining algorithm. This allowed the assessment of whether gene duplication occurred prior to or following the basal divergence of the clade. Duplication following basal divergence usually results in two clusters, one of which contains only a subset of the taxa. These were removed from further analysis if they contained fewer than 155 taxa (92%).

Alignment and trimming

Sequences in each orthologous cluster were first aligned using MAFFT v7.023b (Katoh and Standley 2013), then trimmed and masked using the following procedure (Prum et al. 2015). Sites with the same character in $>50\%$ of sequences were considered "conserved." A 20 bp sliding window was then moved across the alignment, and regions with <13 characters matching the common base at the corresponding conserved site were masked. Sites with <152 unmasked bases were removed. Finally, the masked alignments were inspected by eye and regions considered obviously misaligned or paralogous were removed.

PHYLOGENETICS AND DIVERGENCE DATING

To reconstruct the phylogeny of *Hakea* we used a coalescent-based species tree approach. We first estimated phylogenies from 450 orthologous loci with good alignments by maximum likelihood using RAxML (Stamatakis 2014) with the default rapid hill-climbing search algorithm and GTRGAMMA substitution model. We then used these 450 trees to estimate a species tree using ASTRAL-II (Mirarab et al. 2014; Mirarab and Warnow 2015), a method based on the multispecies coalescent. ASTRAL-II achieves a high level of computational efficiency by maximizing the number of common unrooted quartet subtrees among gene trees, and by constraining the tree search space to a restricted set of bipartitions, making it more computationally tractable for large phylogenomic datasets than Bayesian species tree methods and some other non-Bayesian methods such as MP-EST, yet it operates at a high level of accuracy (Mirarab and Warnow 2015). The ASTRAL-II analysis returned a cladogram with nodal bootstrap values calculated from 100 bootstrap trees estimated for each locus by RAxML.

To estimate divergence times and branch lengths for the species tree we first filtered the set of alignments for 450 orthologous loci based on three criteria: (1) taxonomic completeness: we selected only loci represented by 100% of taxa in the dataset; (2) informative sequences: we selected loci with >200 parsimony-informative sites; (3) low substitution rate variation: we selected loci for which the coefficient of variation in root-to-tip branch lengths on the RAxML tree was <0.5 (following Jarvis et al. 2014). After filtering, our dataset was reduced to 154 loci. We estimated divergence times and branch lengths from the concatenated alignment of the 154 loci, using the program mcmctree in the PAML package (Yang 2007), with the ASTRAL-II cladogram supplied as a fixed topology. There are no known macro- or microfossils that can be confidently attributed to *Hakea* (Sauquet et al. 2009; Mast et al. 2015), so we were restricted to the use of secondary calibrations of outgroup divergences to calibrate the timescale of the phylogeny. We specified three point calibrations: (1) 70.6 Mya for the split between *Banksia* + *Lambertia* and the remaining taxa. This corresponds to the minimum age possible for node "I" in a phylogenetic analysis of Proteaceae genera by Sauquet et al. (2009), represented by the fossil taxon *Propylipollis crotonoides*; (2) 55.8 Mya for the split between *Banksia* and *Lambertia*, corresponding to the minimum possible age for Sauquet et al.'s node "A," represented by the fossil taxon *Banksieaeidites elongatus*; (3) 35.4 Mya for the split between *Telopea* and *Alloxylon*, corresponding to the minimum possible age for Sauquet et al.'s node "D," represented by the fossil taxon *Granodiporites nebulosus*. We expressed the uncertainty and low precision of these three calibrations in the way they were specified as priors for the divergence times. The three calibration distributions were specified as: "SN(0.35, 0.1, 50)," "SN(0.55,0.1,50)," and "SN(0.71,0.1,50)," where SN specifies a skewed-normal distribution and the first number is the mean, the second number is a scaling factor in millions of years, and the third is a shape parameter where the value 50 gives a fairly diffuse distribution for the prior. The specified calibration priors may differ from the "effective priors" actually used by the model, which are generated through interactions between calibration distributions, the root constraint, and the birth-death model (Yang 2007). To estimate the effective priors we ran the analysis without the sequence alignments, then adjusted the specified mean values until we obtained an effective mean prior for the root node that approximated the value of calibration 1 (71 Mya). We used the PAML program baseml to obtain an initial maximum-likelihood estimate of substitution rate to use as a starting value in the Bayesian analysis, and specified a GTR substitution model and an autocorrelated-rates clock model. We ran the mcmc chain for a burnin of 50,000 generations followed by 500,000 generations with a sample frequency of 50. We then ran a second chain under the same parameters and confirmed that the chains had reached convergence by finding an almost perfect correlation between the posterior mean divergence times of the two chains.

SPECIES OCCURRENCE AND ENVIRONMENTAL DATA

We obtained 30618 records of herbarium specimens of *Hakea* from the online repository, the Australian Virtual Herbarium (http://chah.gov.au/avh/). We reduced this to a set of 18610 records by removing any that lacked geographic coordinates with a precision $\leq$10 km, appeared to be well outside the species' distribution limits indicated by range maps in the Flora of Australia volume 17B, or were clearly wrong (e.g., in the sea).

There is no generally accepted definition of a biome, nor any "standard" biome scheme currently in use (Donoghue and Edwards 2014). We used the second level of classification of the World Wildlife Fund's Terrestrial Ecoregions (Olson et al. 2001), which groups ecoregions into 14 major biomes, of which seven are represented in Australia. This is a geographic definition of biomes, and each biome is not necessarily uniform in habitat type. For example, the "Tropical and Subtropical Moist Broadleaf Forests" biome in Australia contains areas of open sclerophyll (Eucalypt-dominated) forest as well as closed-canopy tropical rainforest. Nonetheless, each biome has a broadly distinctive climatic and ecological character, and we believe that this biome scheme captures the range of environmental selective regimes relevant to a continental-scale analysis of biogeographic shifts. In the Discussion, we discuss the implications of our use of this biome scheme for inferences about niche or habitat conservatism.

For the purposes of testing phylogenetic signal in biome occupancy, we assigned each species to the single biome in which the majority of its occurrence records are found. In most cases, this could be done unambiguously: of the 151 *Hakea* species, 113 are found in one biome only, and a further 22 have the great majority of their records (>75%) within one biome. For the purposes of modeling range evolution, estimating ancestral biomes, and testing phylogenetic assemblage structure, we assigned each species to all biomes in which its occurrence records are found. There were very few records and species from the "Montane Grasslands and Shrublands" biome, so we collapsed these records into the surrounding "Temperate Broadleaf and Mixed Forests" biome. Because the names of the biomes are quite long, in this paper we use shortened versions of biome names as follows, with the WWF biome names in parentheses: Arid (Deserts and Xeric Shrublands), Mediterranean (Mediterranean Forests, Woodlands and Scrub), Temperate Forest (Temperate Broadleaf and Mixed Forests), Temperate Grasslands (Temperate Grasslands, Savannas, and Shrublands), Tropical Savanna (Tropical and Subtropical Grasslands, Savannas and Shrublands), Tropical Forest (Tropical and Subtropical Moist Broadleaf Forests).

Although we filtered out occurrence records with a coordinate precision of >10 km, imprecise coordinates may still place a species incorrectly in a biome in which it does not naturally occur. It is also possible that individual plants occasionally may be recorded from within the boundary of a biome in which it does not maintain a viable population. For these reasons, we selected only occurrence records found within a 10 km buffer inside the boundary of each biome. This reduced the number of biomes occupied for a few species, so we repeated all analyses using a biome-occupancy matrix constructed without buffering the biome boundaries. These alternative results are presented in the Supplementary Material (Tables S3–S5, Figs. S1–S3).

PHYLOGENETIC SIGNAL TESTS

To test whether the pattern of biome occupancy of *Hakea* species is consistent with the random drift of species' ranges among biomes, we quantified phylogenetic signal in biome occupancy with the maximum likelihood estimate of Pagel's λ (Pagel 1999), using the fitDiscrete function in the R library geiger (Harmon et al. 2012). Phylogenetic signal in biome occupancy does not necessarily indicate conserved adaptations to particular environments: it can also result from the historic signal of speciation and limited dispersal of species away from their ancestral range (Crisp and Cook 2012; Cardillo 2015). To test this, we quantified phylogenetic signal in the positions of species ranges in continuous geographic space, ignoring biome occupancy (Cardillo 2015). To do this we extracted the latitude and longitude of the geographic centroid of each species range, and found the joint maximum likelihood estimate of λ for both of these using the phyl.pca function in the R library phytools (Revell 2012).

ESTIMATING ANCESTRAL BIOMES AND INFERRING BIOME SHIFTS

To estimate ancestral biome occupancy we used the R package BioGeoBEARS (Matzke 2013). BioGeoBEARS provides a flexible likelihood-based framework for modeling range evolution along a phylogeny as a series of shifts among a set of discrete regions. These shifts can take the form of anagenetic events (range expansion or contraction between adjacent regions along a single branch of the phylogeny) or cladogenetic events (shifts that occur at the time of branching, including sympatric, vicariant, or founder-event speciation). For convenience, we bundled different combinations of range-evolution parameters into a set of models that correspond to some of the most widely used biogeographic models in the recent literature: Dispersal-Vicariance Analysis (DIVA; Ronquist 1997), Dispersal-Extinction-Cladogenesis (DEC; Ree and Smith 2008), and BayArea (Landis et al. 2013). We compared each of these models to an alternative model in which we added the parameter *j* to represent founder-event speciation, giving a total of six models. We then selected the model with the lowest AIC score as the best representation of the range evolution process in *Hakea*.

The next step was to test the influence of geographic proximity and environmental similarity of biomes on the range evolution parameters. To do this, we constructed two distance matrices. In the geographic distance matrix (X), the distance from one biome to another was 1- (length of the shared boundary/total (noncoastal) boundary length of the ancestral biome). This meant that the distances between two biomes differed depending on which was the ancestral and which the descendant biome (i.e., the matrix was asymmetric). In the environmental distance matrix (N), distance was the Euclidean distance in the mean values of a set of five climatic variables obtained from BIOCLIM: mean annual temperature, mean annual precipitation, temperature seasonality, precipitation seasonality, and precipitation in the warmest quarter. These variables were chosen because they summarize much of the broad-scale climatic variation that distinguishes biomes. When a distance matrix is included in a BioGeoBEARS model, the matrix elements are used as modifiers on the probability of dispersal between each pair of regions, which in turn influences the estimated values of model parameters, including the jump dispersal parameter *j* (Matzke 2013). We took the best model from the first round of model comparison, and added the X matrix, the N matrix, and both X + N matrices, then compared the fit of models with and without distance matrices using likelihood ratio tests. The X and N matrices are provided in the Supplementary Material (Table S2A, S2B).

For the best-fitting model, we then used biogeographic stochastic mapping (50 runs) in BioGeoBEARS to calculate the probability of ancestral states under the model, for each node in the phylogeny. This allowed us to tabulate the inferred numbers of shifts, both anagenetic and cladogenetic, between each pair of biomes, during the diversification of *Hakea*. We used the X and N matrices to calculate expected values for the number of shifts by first scaling the matrix elements to relative values in the range 0–1 (i.e., dividing each element by the maximum value), then multiplying each relative matrix value by the inferred total number of shifts among all biome pairs. We then plotted the expected values derived from each matrix against the numbers of shifts inferred under the best-fitting BioGeoBEARS model. As a further test, we inferred the number of shifts under a simple dispersal-only "null" model of range evolution in BioGeoBEARS that included only the geographic distance (X) matrix and parameters for range expansion, contraction, and sympatric speciation (range-copying).

PHYLOGENETIC STRUCTURE OF ASSEMBLAGES

To test whether the assemblage of *Hakea* species within each biome is phylogenetically clustered (species more closely related to one another than expected), overdispersed (species less closely related to one another than expected), or neither, we used the Net Relatedness Index (Webb et al 2002) implemented as the ses.mpd function in the R library picante (Kembel et al. 2010). This calculates the mean lengths of the branches connecting each pair of species (mpd), standardized by a distribution of mpd values for random assemblages generated by shuffling the biome occupancy data using an independent-swap algorithm. A value of the test statistic in the lower tail of a null distribution ($P \leq 0.025$) is indicative of phylogenetic clustering; a value in the upper tail ($P \geq 0.975$) indicates phylogenetic overdispersion.

Results

ANCHORED HYBRID ENRICHMENT

We obtained contig assemblies for 176 taxa (151 *Hakea* and 25 outgroups), with a total of 505 loci captured, averaging 876 bp in length across the samples. The mean number of loci >125 bp captured per taxon was 492; the mean number of loci >500 bp was 445; and the mean number of loci >1000 bp was 58. After removing loci with missing data for >92% of samples we were left with 450 loci for which we could obtain good alignments.

PHYLOGENETICS AND DIVERGENCE DATING

The species tree produced by ASTRAL-II shows strong support for a monophyletic *Hakea* (bootstrap support = 1), as well as for the majority of clades throughout the tree: 74.8% of nodes have bootstrap support values ≥0.99, and 81.9% of nodes have bootstrap support values ≥0.90 (Fig. 1). Of the 24 nonmonotypic informal taxonomic groupings within *Hakea* proposed by Barker et al. (1999), nine are supported as monophyletic, but there is no support for the monophyly of the remaining groups. The two clades inferred by Mast et al. (2012, 2015) to have arisen at the basal divergence of *Hakea* are supported by our more complete taxonomic sampling and larger dataset. The branch length estimation and calibration in mcmctree (Fig. 2) placed the estimated range of ages of the crown node of *Hakea* from the Middle Eocene to Early Oligocene.

PHYLOGENETIC SIGNAL TESTS

The maximum likelihood estimate of Pagel's λ for biome occupancy of *Hakea* species was 0.92, and there was no significant difference in the log-likelihoods of the model in which λ was estimated, and a model in which λ was fixed at a value of 1 (Likelihood Ratio Test, $P = 0.2$). This result is consistent with a degree of phylogenetic signal expected under a constant-rates, random drift model. This result was similar for the unbuffered biome occupancy matrix ($\lambda = 0.91$; $P(\lambda = 1) = 0.14$).

In contrast, there was virtually no phylogenetic signal in the positions of species distributions in continuous geographic space, with a maximum likelihood estimate of $\lambda < 0.0001$, jointly estimated for latitude and longitude of range centroids. This result indicates that species distributions are highly dynamic through evolutionary time: a species is no more likely to be found in close geographic proximity to its nearest relatives than to more distant relatives.

ESTIMATING ANCESTRAL BIOMES AND INFERRING BIOME SHIFTS

We compared the fit of six models of geographic range evolution implemented in BioGeoBEARS (Table 1). The model with the lowest AIC score was "BAYAREALIKE + J": that is a likelihood implementation of the BayArea model, with the addition of a *j* parameter for founder-event speciation. The BAYAREALIKE model includes parameters for anagenetic range evolution by range expansion and contraction, and cladogenetic range evolution by sympatry (range-copying), both within one region (narrow sympatry) and across multiple regions (widespread sympatry), but does not include parameters for vicariant speciation across regions. We selected the BAYAREALIKE + J model for further comparison with more complex models that include geographic and environmental distance matrices. Adding either of the two distance matrices improved the fit of the BAYAREALIKE + J model to the data, with the geographic distance matrix (X) providing slightly greater improvement than the environmental distance matrix (N) (Table 2). With the unbuffered biome occupancy matrix, the BAYAREALIKE + J model was also selected over other models (Table S3), but the addition of the geographic or environmental distance matrices did not improve the fit (Table S4).

We used the BAYAREALIKE + J + X model to estimate ancestral biome occupancy for all nodes in the *Hakea* phylogeny. Almost all (48/50) of the stochastic mapping runs estimated the Mediterranean Shrub & Woodlands as the ancestral biome at the crown node of *Hakea* (Fig.3), implying multiple dispersal events away from the Mediterranean biome during the course of *Hakea* diversification. The mean number of inferred shifts between biomes across the 50 stochastic mapping runs are shown in Table 4.

Plots of the numbers of biome shifts expected under a dispersal-only "null" BioGeoBEARS model against the number inferred under the BAYAREALIKE + J + X model are shown in Fig. 4A. Two biome pairs stand out as large outliers in this plot: the BAYAREALIKE + J model reconstructed many more shifts from Forest to Mediterranean biomes, and from Mediterranean to Arid biomes, compared to the dispersal-only model. Because the BAYAREALIKE + J model differs from the dispersal-only model only in the addition of the *j* parameter and the omission of distance matrices, the excess of biome shifts in these two cases reflects the contribution of founder-event dispersal. When compared to expected numbers of biome shifts calculated from X and N matrices, there are biome pairs both with more and fewer inferred shifts than expected (Fig. 4B, C). Results were similar for the unbuffered biome occupancy data (Fig. S4).

PHYLOGENETIC STRUCTURE OF ASSEMBLAGES

Assemblages of *Hakea* species within the two most species-rich biomes (Mediterranean and Temperate Forest) are significantly phylogenetically overdispersed. In other words, values of mean pairwise distance (mpd) are in the upper tails of distributions of random values, indicating that pairs of species within these two biomes are less closely related than expected under a null model in which species are shuffled across all biomes (Table 3). Values of mpd for the other biomes were not significantly different from the null values. Results were similar under all other null models available in the picante package, as well as alternative phylogenetic structure metrics such as phylogenetic species

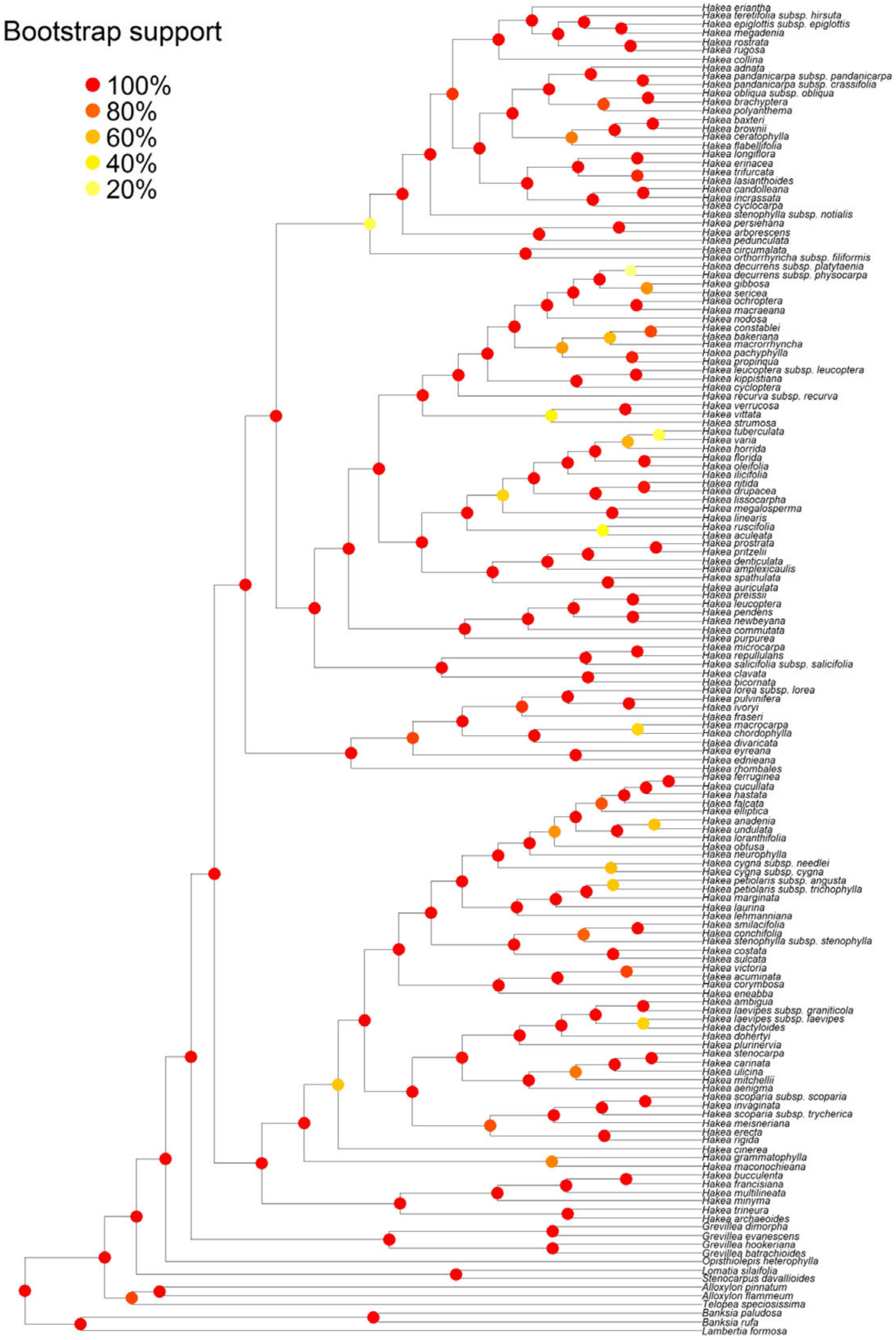

Figure 1. ASTRAL-II species tree showing the branching relationships of *Hakea* species and Proteaceae outgroups. Colors of the node labels are proportional to the degree of support from 100 bootstrap trees.

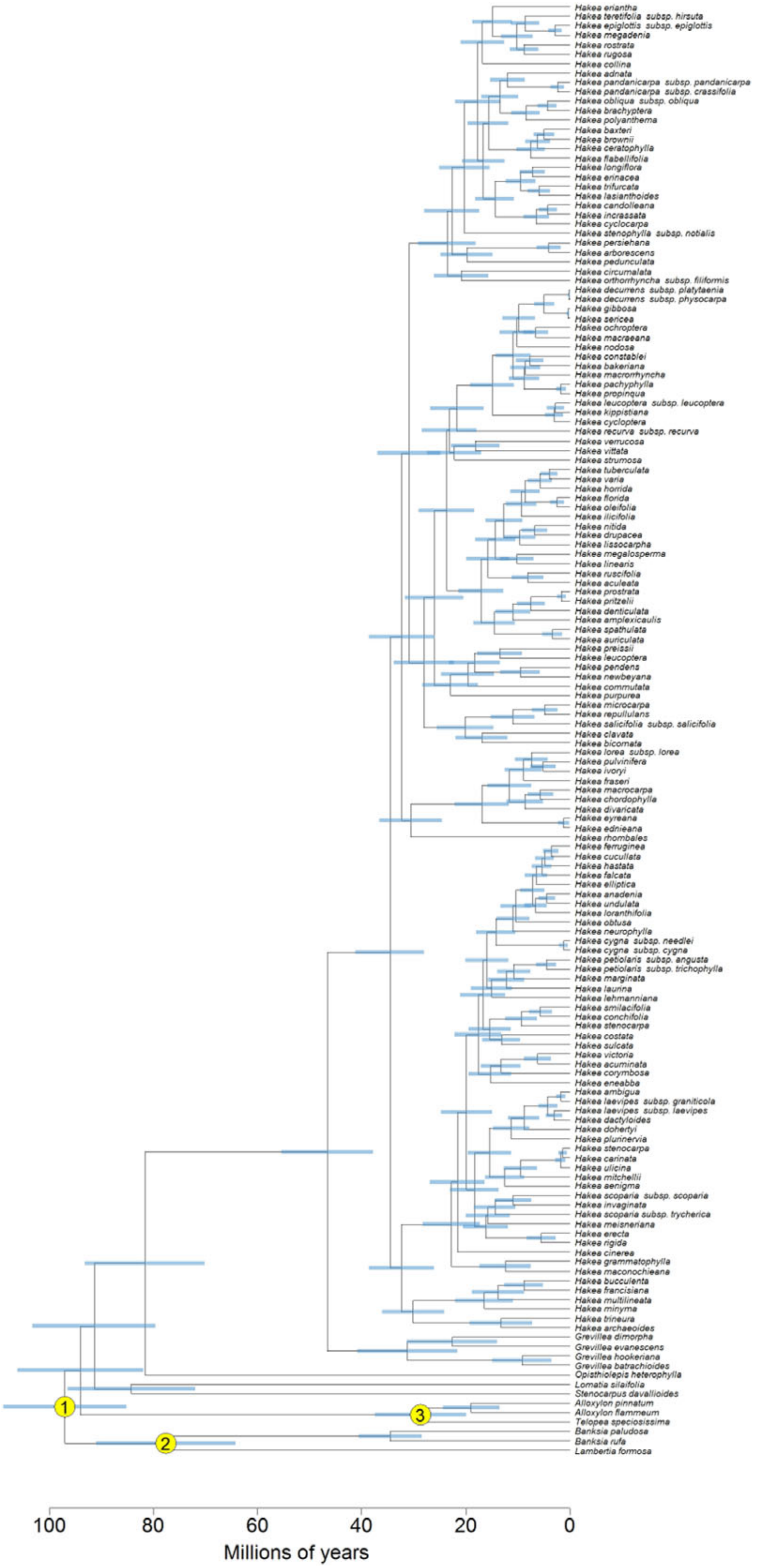

Figure 2. Divergence times among *Hakea* and Proteaceae outgroup species estimated using mcmctree. Yellow symbols indicate three calibration points as described in the text. Node heights are the medians from the posterior distribution, and node bars indicate the 95% highest posterior density limits.

Table 1. Comparison of models for the evolution of *Hakea* geographic ranges fitted in BioGeoBEARS.

	Parameters fitted								
Model	*d*	*e*	*y*	*s*	*v*	*j*	Log-likelihood	*d.f.*	AIC
DEC	*	*	*	*	*		–216.53	2	437.07
DIVALIKE	*	*	*		*		–222.45	2	448.9
BAYAREALIKE	*	*	*				–248.27	2	500.55
DEC + J	*	*	*	*	*	*	–215.99	3	437.98
DIVALIKE + J	*	*	*		*	*	–222.08	3	450.17
BAYAREALIKE + J	*	*	*			*	–212.61	3	431.21

The six models differ in the set of parameters that are allowed to vary freely. Parameters are range expansion (*d*), range contraction (*e*), sympatric speciation: range-copying (*y*), sympatric speciation: range subsetting (*s*), vicariant speciation (*v*), and founder-event speciation (*j*).

variance (PSV). Results were also very similar for the unbuffered biome occupancy matrix (Table S5).

Discussion

Our reconstruction of the phylogeny and divergence times of *Hakea* suggests that the ancestors of present-day species began diversifying in the Mediterranean Shrub and Woodland biome, sometime between the Middle Eocene and Early Oligocene. This does not necessarily imply that the genus had its origins in the Mediterranean biome; it is only an estimate of the likely geographic distribution of the most recent common ancestor of extant *Hakea* species. Hence, our biogeographic result does not necessarily conflict with a previous suggestion that the *Hakea* stem lineage originated in the seasonal tropics of north-eastern Australia (Barker et al. 1999). However, this suggestion was based on a cladistic analysis of morphological characters in which *Grevillea glauca*, a species from tropical north-eastern Australia, was resolved as the sister group of *Hakea* without bootstrap support. More recent molecular analysis (Mast et al. 2015) strongly supports a sister group of *Hakea* made up of a large clade of *Grevillea* species (not including *G. glauca*), that is distributed widely across Australia and neighboring landmasses, providing little evidence for the geographic origin of *Hakea*. Regardless of where the genus originated, however, the Mediterranean biome seems to have been the launching point for a radiation of *Hakea* across the Australian continent and into a wide range of environments in all of the major biomes, a conclusion also reached by Lamont et al. (2016).

Under the best-fitting BioGeoBEARS model, our estimate of ancestral biomes suggests that during the *Hakea* radiation there were around 47 shifts between biomes, including 28 shifts out of or into the Mediterranean biome. Is this unusually high? It is not straightforward to answer this question because it is difficult to conceive of a suitable null model to generate an expected absolute number of biome shifts. One approach used previously is to shuffle species distributions before estimating ancestral biomes (Holstein and Renner 2011), but this is essentially a nonevolutionary null model that erases all historical signal of range evolution, so is predisposed to infer a very large number of biome shifts. However, comparisons with inferred numbers of biome shifts in other large plant genera that occupy Mediterranean-type environments implies that the number of shifts in *Hakea* has indeed been unusually high. For example, the ancestors of extant *Banksia* have dispersed away from the Southwest Australian Floristic Region only twice in the 32–62 million-year history of this clade (Cardillo and Pratt 2013). In *Protea*, there was only one shift away from the Cape Floristic Region in 11–27 million years (Valente et al. 2010), while in *Leucocoryne* there were two shifts from the sclerophyll biome to the arid winter-rainfall biome in Chile during a period of 8–14 million years (Jara-Arancio et al. 2014). In a recent comparative analysis of evolutionary shifts between Mediterranean-climate hotspots and nonhotspot regions (Skeels and Cardillo 2017), *Hakea* stood out with a large number of shifts, compared to very few shifts in three other large hotspot plant genera (*Banksia*, *Protea*, and *Moraea*).

The apparently high frequency of evolutionary transitions between biomes in *Hakea* suggests a high degree of lability in the boundaries of geographic distributions. This conclusion is further supported by the virtual absence of phylogenetic signal in the geographic centroids of species ranges. This suggests that *Hakea* species ranges are highly dynamic and the limitations of time and opportunity for dispersal away from ancestral ranges have played a relatively minor role in shaping the present-day distributions of species (although we must keep in mind that applying the lambda test to geographic centroids is likely to violate some of the assumptions of the test, such as unbounded variance). But if geographic ranges are labile and biome shifts have been frequent, why do we see a high degree of phylogenetic signal in present-day biome occupancy ($\lambda = 0.92$), as expected under a random-drift model of range evolution? We suspect that the high lambda values for biome occupancy are driven by the high proportion of *Hakea* species, and hence large clades of close

Table 2. Comparisons of range evolution models with and without distance matrices.

Model		Log-likelihood		*d.f.*		*D*	AIC		AIC weight	
Null	Alternative	Null	Alternative	Null	Alternative		Null	Alternative	Null	Alternative
BAYAREALIKE+J	BAYAREALIKE+J+X	−212.61	−208.27	3	4	8.67**	431.2	424.55	0.034	0.966
BAYAREALIKE+J	BAYAREALIKE+J+N	−212.61	−209.83	3	4	5.56*	431.2	427.66	0.144	0.856
BAYAREALIKE+J	BAYAREALIKE+J+X+N	−212.61	−210.79	3	5	3.63†	431.2	429.58	0.306	0.694

The "null" model is the best-fitting BioGeoBEARS model. Alternative models include a matrix of geographic distances between biomes (X), a matrix of climatic distances (N), or both X and N. *D* is the *D*-statistic of a likelihood-ratio test, with asterisks indicating significance as follows: ***P* ≤ 0.001; **P* ≤ 0.01; †*P* ≤ 0.1.

relatives, that are found in a single biome, the Mediterranean shrub and woodlands. Further, for the phylogenetic signal tests we assumed each species occupied a single biome, the one in which the majority of its records were found. Given that multiple-biome species are moderately clustered on the phylogeny ($\lambda = 0.46$), recoding them as single-state results in a slight underestimate of the phylogenetic signal in biome occupancy. Removing multiple-biome species from the tree leads to a slightly elevated value of lambda for the remaining species ($\lambda = 0.97$).

Both the large number of inferred biome shifts and the lack of phylogenetic signal in range positions suggest that *Hakea* species have been geographically labile, but does this necessarily argue against phylogenetically conserved environmental niches in this genus? Again, there is no clear answer here, because the definition of phylogenetic niche conservatism, and whether it is represented by measures of phylogenetic signal, is itself the subject of disagreement (Losos 2008; Wiens et al. 2010). Furthermore, it is important to emphasize that biome shifts are not synonymous with habitat shifts, and that our biomes do not necessarily correspond with all key aspects of the environment that may be important in limiting *Hakea* distributions. However, we believe that biomes effectively capture the broad climatic and vegetational differences that reflect fundamentally different selective regimes for plant growth. Two recent studies of niche evolution in Mediterranean-type environments provide evidence that seems to support this notion. Skeels and Cardillo (2017) showed that lineages of *Hakea* and several other genera are evolving toward different climatic optimum values within and outside Mediterranean hotspots, suggesting different selective regimes, while Onstein et al. (2016) came to a very similar conclusion for Proteaceae species in open vegetation types compared to closed forest habitats. In this context, our results probably reflect a high degree of adaptability and lability in the broad-scale climatic niche of *Hakea* species, but they do not allow us to draw conclusions about finer-scale adaptability to particular habitats, topography, or soil types. Other authors have approached this problem by using common methods to infer ancestral geographic regions and habitat types (Weeks et al. 2014).

Our inferences of the relative numbers of shifts between different pairs of biomes show that both geographic proximity and climaticsimilarity play a strong role in facilitating or inhibiting shifts across biome boundaries. In some cases, the inferred number of shifts diverges substantially from the number expected from geographic proximity (Fig. 4B), but appears much closer to the expected number when environmental similarity is taken into account (Fig. 4C). For example, the inferred mean number of shifts from Tropical Forest to Tropical Savanna (0.18) was very low compared to the expectation based on their geographic proximity (2.2). However, given the climatic distinctness of these two biomes (year-round humidity and precipitation in Tropical

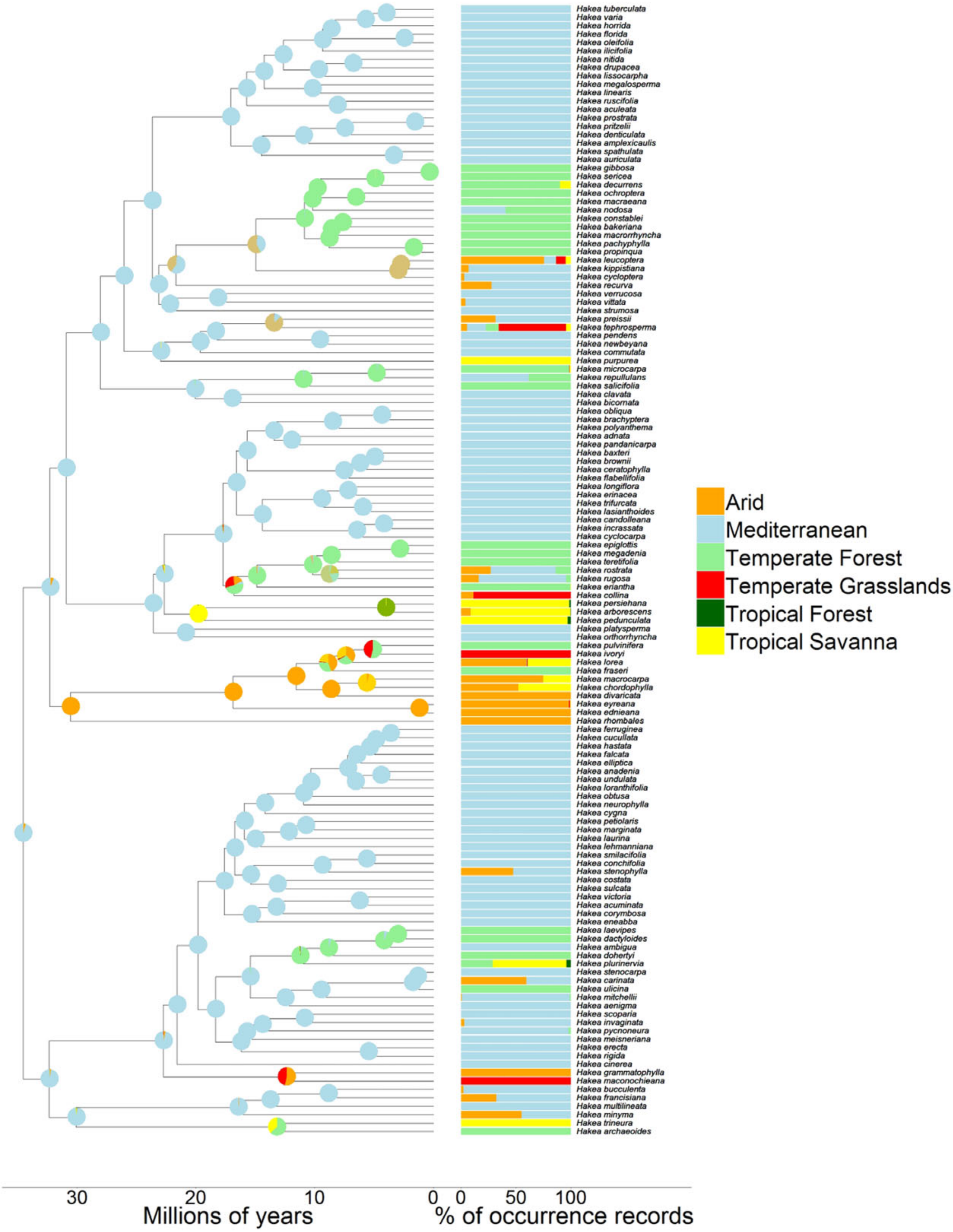

Figure 3. Estimated ancestral biomes of *Hakea* under the BAYAREA + J + X model in BioGeoBEARS. Node symbols indicate the proportions of 50 stochastic mapping runs in which occupancy of each biome was estimated. Colored bars at the tips of the phylogeny indicate the present-day distributions of species: length of each colored section represents the relative number of species occurrence records found in different biomes. Multiple ancestral geographic states are indicated by blended colours; these are not given in the legend. Note that the tips in this tree do not exactly match those in Figs. 1 and 2, because subspecies have been collapsed to species, and outgroups have been removed.

Table 3. Phylogenetic assemblage structure statistics.

Biome	Species	MPD observed	MPD random (mean ± SD)	z-score
Arid	24	4.08	4.63 ± 0.52	–1.07
Mediterranean	96	5.49	4.71 ± 0.13	6.04**
Temperate forest	31	6.83	4.59 ± 0.40	5.54**
Temperate grassland	5	2.27	5.01 ± 2.26	–1.21†
Tropical Savanna	12	4.46	4.78 ± 0.98	–0.33
Tropical forest	3	10.19	5.12 ± 3.89	1.3†

MPD observed is the mean pairwise phylogenetic distance among species found within each biome; MPD random is the mean pairwise distance after shuffling species between biomes. The z-score is the standardized effect size of the observed MPD with respect to the random values. Asterisks indicate significant departures from the null distribution, as follows: $^{**}P \le 0.001$; $^{*}P \le 0.01$; $^{\dagger}P \le 0.1$.

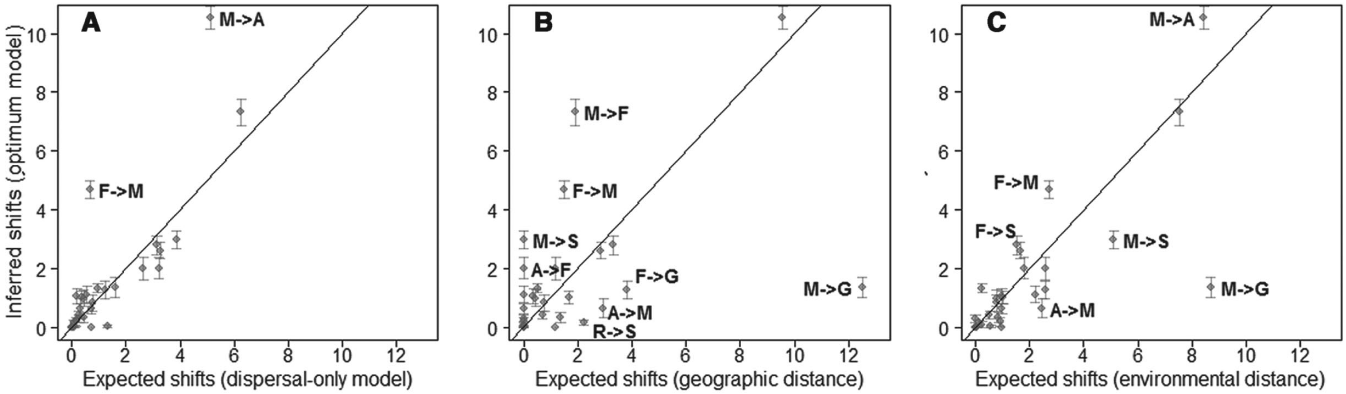

Figure 4. Inferred and expected numbers of shifts between each pair of biomes during the diversification of *Hakea*.
The *y*-axis of each plot shows the numbers of shifts inferred under the BAYAREA + J + X model in BioGeoBEARS (mean and SD of 50 stochastic mapping runs). The *x*-axes show the numbers of shifts expected (A) under a "dispersal-only" model in BioGeoBEARS; (B) from geographic distance between biomes, (C) from environmental distance between biomes. Biome pairs for which the number of shifts deviates substantially from the expected value are labeled. A = Arid; M = Mediterranean; F = Temperate Forest; G = Temperate Grassland; S = Tropical Savanna; R = Tropical Forest.

Forest, strongly seasonal winter drought in Tropical Savanna), the number of shifts expected from the climatic distance matrix (0.13) was far closer to the number inferred. We should also note, however, that few (if any) species of *Hakea* inhabit closed tropical rainforest; those species recorded from the Tropical Forest biome tend to be found in wet sclerophyll forests within this biome.

In other cases, the inferred number of shifts diverged from both the geographic and environmental expectations. There are probably different reasons for this. First, species distributions may be limited by aspects of the environment other than climate. There were far fewer shifts inferred from Mediterranean to Temperate Grassland biomes (1.38) than the large number expected from their close geographic connectedness (12.5) and relatively similar climates (8.7). This may reflect the distinct soil types that characterize these two biomes: often sandy and infertile in the Mediterranean biome, and often more fertile, with higher clay content, in the Temperate Grasslands (Specht 1981). Conversely, we inferred a larger number of shifts from Mediterranean to Arid biomes (10.54) than expected from both geography (2.9) and climate (2.5). There was also an excess of shifts between these two biomes in the BAYAREA + J + X compared to the dispersal-only model (Fig. 4A), which points to a key role for founder-event speciation in the successful colonization of the Arid zone from the Mediterranean biome. This does not necessarily reflect a greater propensity of Mediterranean species to long-distance dispersal. It is more likely that dispersal events had a greater chance of leading to the successful founding of a new lineage, given that these two biomes share climatic features (drought, high maximum temperatures, and high solar radiation) that select for similar plant anatomical features such as scleromorphic structures (Jordan et al. 2005), stomatal encryption (Jordan et al. 2008), or small leaves without toothed margins (Peppe et al. 2011). The adaptive changes required for species of *Hakea* to make the transition between Mediterranean and Arid biomes are likely to be far smaller than those required to move between wetter and drier biomes.

Table 4. Number of inferred shifts between pairs of biomes in *Hakea*.

	Descendant biome					
Ancestral biome	Arid	Mediterranean	Temperate forest	Temperate grassland	Tropical Savanna	Tropical forest
Arid		0.7 ± 0.64	2 ± 0.71	2 ± 0.79	2.6 ± 0.53	0
Mediterranean	10.5 ± 0.76		7.3 ± 0.88	1.4 ± 0.69	3 ± 0.6	0.04 ± 0.1
Temperate Forest	1.1 ± 0.54	4.7 ± 0.59		1.3 ± 0.6	2.8 ± 0.66	0.6 ± 0.32
Temperate Grassland	1.1 ± 0.5	0	1 ± 0.58		0.4 ± 0.27	0
Tropical Savanna	1 ± 0.41	0.2 ± 0.3	0.8 ± 0.56	0.3 ± 0.33		1.3 ± 0.33
Tropical Forest	0.3 ± 0.23	0.02 ± 0.07	0.1 ± 0.21	0	0.2 ± 0.19	

The table gives numbers of shifts (mean and SD of 50 stochastic mapping runs) inferred under the BAYAREA + J + X model in BioGeoBEARS.

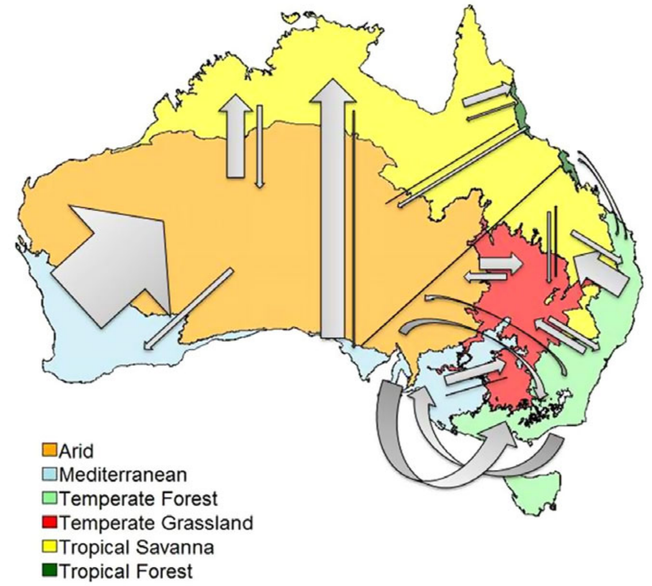

Figure 5. Inferred shifts among Australian biomes during the diversification of *Hakea*.
Width of the arrows is proportional to the relative number of inferred shifts under the BAYAREA + J + X model.

Of course, our reconstructions of ancestral biomes and inference of biome shifts need to be considered with a view to past as well as present-day environments. In the early part of the *Hakea* radiation, the Australian climatic environment and configuration of biomes looked very different to that of today, with the Arid zone yet to develop (Byrne et al. 2008) and the forested biomes far more extensive (Byrne et al. 2011). Yet, some of our inferred shifts into the Arid biome appear to predate the onset of aridification around 15 Mya. There are several possible explanations for this. Adaptations to seasonally dry climates in southwestern Australia may have evolved as early as the late Eocene, long before the Arid zone developed (Carpenter et al. 2014), so that a number of *Hakea* lineages may have been preadapted to arid climates, allowing them to expand their distributions through inland Australia as the Arid zone developed. For some species, their distributions remain spread across multiple biomes to the present day. If some species made a full transition into the Arid biome, and these species are connected to their relatives by long branches, it could lead us to reconstruct deep origins for some of the shifts into the Arid zone. Another possible explanation for the deep origins of shifts into the Arid zone is extinction. The extinction of a lineage pushes the reconstructed divergence time of its sister lineage deeper into the past, so the extinction of close relatives in other biomes could push back the timing of inferred shifts into the Arid zone. This last scenario may be consistent with evidence from the fossil record for widespread extinction in Australian sclerophyll floras since the early Pleistocene, probably associated with severe climatic cycling or the contraction of wet forest habitats to the coastal margins of eastern Australia (Sniderman et al. 2013).

Conclusions

Our analyses of the phylogeny and biogeography of *Hakea* offer strong evidence that evolutionary transitions between major biomes can occur with surprising ease and frequency, facilitating the radiation and dispersal of a large clade across a continent. In the case of *Hakea*, our biogeographic analyses suggest that the primary mechanisms of biome-shifting have been anagenetic range expansion and founder-event speciation (jump dispersal), rather than vicariance (ecological differentiation) across biome boundaries. The fact that many present-day *Hakea* species distributions span several biomes can be seen as a snapshot in time of this process of range expansion. As a result, the assemblages of *Hakea* species within biomes tend to be composed of phylogenetically diverse sets of species derived from multiple invasions from other regions, as opposed to sets of closely related species that arose from *in situ* radiations within biomes. The case of *Hakea* seems to be one of the more dramatic examples in a growing list of studies that demonstrate biome shifts within plant genera, and runs counter to a common view of plant diversification shaped by environmental niche conservatism. Future work should focus on revealing the physiological and functional traits, and the underlying genomic architecture, that may be associated with adaptive flexibility and biome shifts in *Hakea*.

AUTHOR CONTRIBUTIONS

P.H.W., Z.K.M.R., P.M.O. collected and catalogued tissue samples and voucher specimens; Z.K.M.R. carried out labwork for DNA extraction, A.R.M., E.L., A.R.L. developed protocols, and carried out labwork and bioinformatics procedures to produce sequence datasets; M.C. designed the study, carried out phylogenetic and biogeographic analyses, and wrote the manuscript; all authors contributed to writing and revising the manuscript.

ACKNOWLEDGMENTS

This research was funded by Australian Research Council Discovery Grant DP110103168 to M.C. and L.B., and National Science Foundation Grant IIP-1313554 to A.R.L. and E.M.L. Many thanks to Nick Matzke for help with BioGeoBEARS, Alex Skeels for providing the processed species occurrence data, and Russell Dinnage for producing the ancestral biomes figure. Many thanks to Greg Jordan and the anonymous referees for insightful comments and feedback, and to Alyssa Bigelow, Kirby Birch, Ameer Jalal, and Sean Holland at the FSU Center for Anchored Phylogenomics for assistance with molecular data collection and bioinformatics analysis. Paul and Barbara Kennedy kindly allowed us to collect specimens of many species for this project from their living *Hakea* collection.

DATA ARCHIVING

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.j8qv9

LITERATURE CITED

Barker, R. M., L. Haegi, and W. R. Barker. 1999. Hakea. *in* A. Wilson, ed. Flora of Australia. ABRS/CSIRO, Canberra.

Buddenhagen, C., A. R. Lemmon, E. M. Lemmon, J. Bruhl, W. L. Cappa, M. J. Clement, M. J. Donoghue, E. J. Edwards, A. L. Hipp, M. Kortyna, et al. Manuscript. Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. Syst. Biol.

Byrne, M., D. A. Steane, L. Joseph, D. K. Yeates, G. J. Jordan, D. Crayn, K. Aplin, D. J. Cantrill, L. G. Cook, M. D. Crisp, et al. 2011. Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. J. Biogeogr. 38:1635–1656.

Byrne, M., D. K. Yeates, L. Joseph, M. Kearney, J. Bowler, M. A. J. Williams, S. Cooper, S. C. Donnellan, J. S. Keogh, R. Leys, et al. 2008. Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. Mol. Ecol. 17:4398–4417.

Cardillo, M. 2015. Geographic range shifts do not erase the historic signal of speciation in mammals. Am. Nat. 185:343–353.

Cardillo, M., and R. Pratt. 2013. Evolution of a hotspot genus: geographic variation in speciation and extinction rates in Banksia (Proteaceae). BMC Evol. Biol. 13:155.

Carpenter, R. J., S. McLoughlin, R. S. Hill, K. J. McNamara, and G. J. Jordan. 2014. Early evidence of xeromorphy in angiosperms: stomatal encryption in a new eocene species of Banksia (Proteaceae) from Western Australia. Am. J. Bot. 101:1486–1497.

Crisp, M. D., M. T. K. Arroyo, L. G. Cook, M. A. Gandolfo, G. J. Jordan, M. S. McGlone, P. H. Weston, M. Westoby, P. Wilf, and H. P. Linder. 2009. Phylogenetic biome conservatism on a global scale. Nature 458:754–756.

Crisp, M. D., and L. G. Cook. 2012. Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? New Phytol. 196:681–694.

Donoghue, M. J. 2008. A phylogenetic perspective on the distribution of plant diversity. Proc. Natl. Acad. Sci. 105:11549–11555.

Donoghue, M. J., and E. J. Edwards. 2014. Biome shifts and niche evolution in plants. Ann. Rev. Ecol. Evol. Syst. 45:547–572.

Duchêne, D. A., and M. Cardillo. 2015. Phylogenetic patterns in the geographic distributions of birds support the tropical conservatism hypothesis. Global Ecol. Biogeogr. 24:1261–1268.

Fine, P. V. A., F. Zapata, and D. C. Daly. 2014. Investigating processes of Neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protiaee (Burseraceae). Evolution 68:1988–2004.

Harmon, L. J., J. T. Weir, C. T. Brock, R. E. Glor, and W. Challenger. 2012. GEIGER: investigating evolutionary radiations. Bioinformatics 24:129–131.

Holstein, N., and S. S. Renner. 2011. A dated phylogeny and collection records reveal repeated biome shifts in the African genus Coccinia (Cucurbitaceae). BMC Evol. Biol. 11:1–16.

Jara-Arancio, P., M. T. K. Arroyo, P. C. Guerrero, L. F. Hinojosa, G. Arancio, and M. A. Méndez. 2014. Phylogenetic perspectives on biome shifts in Leucocoryne (Alliaceae) in relation to climatic niche evolution in western South America. J. Biogeogr. 41:328–338.

Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. W. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320–1331.

Jordan, G. J., R. A. Dillon, and P. H. Weston. 2005. Solar radiation as a factor in the evolution of scleromorphic leaf anatomy in Proteaceae. Am. J. Bot. 92:789–796.

Jordan, G. J., P. H. Weston, R. J. Carpenter, R. A. Dillon, and T. J. Brodribb. 2008. The evolutionary relations of sunken, covered, and encrypted stomata to dry habitats in Proteaceae. Am. J. Bot. 95:521–530.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kerkhoff, A. J., P. E. Moriarty, and M. D. Weiser. 2014. The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. Proc. Natl. Acad. Sci. 111:8125–8130.

Koecke, A. V., A. N. Muellner-Riehl, T. D. Pennington, G. Schorr, and J. Schnitzler. 2013. Niche evolution through time and across continents: the story of Neotropical Cedrela (Meliaceae). Am. J. Bot. 100:1800–1810.

Lamont, B. B., T. He, and S. L. Lim. 2016. Hakea, the world's most sclerophyllous genus, arose in southwestern Australian heathland and diversified throughout Australia over the past 12 million years. Australian J. Bot. 64:77–88.

Landis, M. J., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck. 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol 62:789–804.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61:727–744.

Losos, J. B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. Ecol. Lett. 11:995–1003.

Mast, A. R., E. F. Milton, E. H. Jones, R. M. Barker, W. R. Barker, and P. H. Weston. 2012. Time-calibrated phylogeny of the woody Australian genus Hakea (Proteaceae) supports multiple origins of insect-pollination among bird-pollinated ancestors. Am. J. Bot. 99:472–487.

Mast, A. R., P. M. Olde, R. O. Makinson, E. Jones, A. Kubes, E. T. Miller, and P. H. Weston. 2015. Paraphyly changes understanding of timing and tempo of diversification in subtribe Hakeinae (Proteaceae), a giant Australian plant radiation. Am. J. Bot. 102:1634–1646.

Matzke, N. 2013. BioGeoBEARS: biogeography with Bayesian (and Likelihood) evolutionary analysis in R scripts. University of California, Berkeley, Berkeley.

Meyer, M., and M. Kircher. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010:pdb.prot5448.

Mirarab, S., R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, and T. Warnow. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548.

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

Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. N. Powell, E. C. Underwood, J. A. D'Amico, I. Itoua, H. E. Strand, J. C. Morrison, et al. 2001. Terrestrial ecoregions of the worlds: a new map of life on Earth. Bioscience 51:933–938.

Pagel, M. D. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Peppe, D. J., D. L. Royer, B. Cariglino, S. Y. Oliver, S. Newman, E. Leight, G. Enikolopov, M. Fernandez-Burgos, F. Herrera, J. M. Adams, et al. 2011. Sensitivity of leaf size and shape to climate: global patterns and paleoclimatic applications. New Phytol. 190:724–739.

Prinzing, A. 2001. The niche of higher plants: evidence for phylogenetic conservatism. Proc. Royal Soc. Lond. B Biol. Sci. 268:2383–2389.

Prum, R. O., J. S. Berv, A. Dornburg, D. J. Field, J. P. Townsend, E. M. Lemmon, and A. R. Lemmon. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526:569–573.

Qian, H., and R. E. Ricklefs. 2004. Geographical distribution and ecological conservatism of disjunct genera of vascular plants in eastern Asia and eastern North America. J. Ecol. 92:253–265.

Ree, R. H., and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57:4–14.

Revell, L. J. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Method Ecol. Evol. 3:217–223.

Rokyta, D. R., A. R. Lemmon, M. J. Margres, and K. Aronow. 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13:312.

Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Syst. Biol. 46:195–203.

Sauquet, H., P. H. Weston, C. L. Anderson, N. P. Barker, D. J. Cantrill, A. R. Mast, and V. Savolainen. 2009. Contrasted patterns of hyperdiversification in Mediterranean hotspots. Proc. Natl. Acad. Sci. USA 106:221–225.

Skeels, A., and Cardillo, M. 2017. Environmental niche conservatism explains the accumulation of species richness in Mediterranean-hotspot plant genera. Evolution. https://doi.org/10.1111/evo.13276

Sniderman, J. M. K., G. J. Jordan, and R. M. Cowling. 2013. Fossil evidence for a hyperdiverse sclerophyll flora under a non–Mediterranean-type climate. Proc. Natl. Acad. Sci. 110:3423–3428.

Souza-Neto, A. C., M. V. Cianciaruso, and R. G. Collevatti. 2016. Habitat shifts shaping the diversity of a biodiversity hotspot through time: insights from the phylogenetic structure of Caesalpinioideae in the Brazilian Cerrado. J. Biogeogr. 43:340–350.

Specht, R. L. 1981. Major vegetation formations in Australia. Pp. 163–197 *in* A. Keast, ed. Ecological biogeography of Australia. Dr. W. Junk bv, The Hague.

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

Töpel, M., A. Antonelli, C. Yesson, and B. Eriksen. 2012. Past climate change and plant evolution in Western North America: a case study in rosaceae. PLoS ONE 7:e50358.

Valente, L. M., G. Reeves, J. Schnitzler, I. P. Mason, M. F. Fay, T. G. Rebelo, M. W. Chase, and T. G. Barraclough. 2010. Diversification of the African genus Protea (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. Evolution 64:745–760.

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. Ann. Rev. Ecol. Syst. 33:475–505.

Weeks, A., F. Zapata, S. K. Pell, D. C. Daly, J. D. Mitchell, and P.V.A. Fine. 2014. To move or to evolve: contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae" (Anacardiaceae and Burseraceae). Front. Genet. 5:409.

Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I. Damschen, et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecol. Lett. 13:1310–1324.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Associate Editor: M. Vallejo-Marin
Handling Editor: M. Noor

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. List of samples and herbarium accession numbers for *Hakea* and outgroup taxa
Table S2. Distance matrices used in the BioGeoBEARS analyses of geographic range evolution.
Table S3. Comparison of models for the evolution of *Hakea* geographic ranges fitted in BioGeoBEARS, based on an unbuffered biome occupancy matrix (see main text).
Table S4. Comparisons of range evolution models with and without distance matrices, based on an unbuffered biome occupancy matrix (see main text).
Table S5. Phylogenetic assemblage structure statistics, based on an unbuffered biome occupancy matrix (see main text).
Figure S1. Estimated ancestral biomes of *Hakea* under the BAYAREA + J model in BioGeoBEARS, based on an unbuffered biome occupancy matrix (see main text).
Figure S2. Inferred and expected numbers of shifts between each pair of biomes during the diversification of *Hakea*, based on an unbuffered biome occupancy matrix (see main text).
Figure S3. Inferred shifts among Australian biomes during the diversification of *Hakea*, based on an unbuffered biome occupancy matrix (see main text).