



Range size, taxon age and hotspots of neoendemism in the California flora

Nathan J. B. Kraft*†, Bruce G. Baldwin and David D. Ackerly

Department of Integrative Biology and Jepson Herbarium, University of California, Berkeley, CA 94720, USA

ABSTRACT

Aim Sustaining biological diversity requires the protection of the ecological, evolutionary and landscape-level processes that generate it. Here, we identify areas of high neoendemism in a global diversity hotspot, the California flora, using range size data and molecular-based estimates of taxon age.

Location California, USA.

Methods We compiled distribution and range size data for all plant taxa endemic to California and internal transcribed spacer (ITS)-based age estimates for 337 putative neoendemics (15% of the endemic flora). This information was combined to identify areas in the state with high proportions of young and restricted-range taxa. We overlaid the distribution of neoendemic hotspots on maps of currently protected lands and also explored correlations between our diversity measures and climate.

Results The central coast of California, the Sierra Nevada and the San Bernardino Range contained endemics with the most restricted distributions on average, while areas in the Desert and Great Basin provinces found within the state were composed of the youngest neoendemics on average. Diversity measures that took age and range size into account shifted the estimate of highest endemic diversity in the state towards the Desert and Great Basin regions relative to simple counts of endemic species richness. Our diversity measures were poorly correlated with climate and topographic heterogeneity.

Main conclusions Substantial portions of California with high levels of plant neoendemism fall outside of protected lands, indicating that additional action will be needed to preserve the geographic areas apparently associated with high rates of plant diversification. The neoendemic flora of the deserts appears particularly young in our analyses, which may reflect the relatively recent origin of desert environments within the state.

Keywords

Desert flora, diversification, endemism, internal transcribed spacer, Mediterranean-type climate.

*Correspondence: Nathan J. B. Kraft, Department of Integrative Biology and Jepson Herbarium, University of California, Berkeley, CA 94720, USA.
E-mail: nkraft@biodiversity.ubc.ca

†Present address: The Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

INTRODUCTION

Evolutionary perspectives on the assessment of biodiversity have become increasingly important to conservation efforts in recent years (Erwin, 1991; Faith, 1992; Moritz, 2002; Forest *et al.*, 2007; Davis *et al.*, 2008; Vandergast *et al.*, 2008). Perhaps the best-known evolutionary approach is to quantify the phylogenetic diversity (PD, Faith, 1992) of species assemblages within geographic areas, with the goal of selecting areas that

currently contain taxa with a disparate evolutionary composition. However, as Whittaker *et al.* (2005) have emphasized, regional-scale biodiversity conservation plans must strike a balance between a focus on patterns of current-day diversity (the compositionalist approach) with an understanding of the dynamic processes shaping the generation and loss of biodiversity (the functionalist approach). Here we focus on the perspective that sustaining biological diversity in the long term requires protection of the ecological and evolutionary

processes that generate it (Moritz, 2002) along with associated features of the environmental landscape.

California and especially the California Floristic Province (CA-FP) (which includes parts of southern Oregon, a small portion of Nevada on the east shore of Lake Tahoe, and northern Baja California, Mexico) contain one of the world's most biologically rich and threatened terrestrial ecoregions (Myers *et al.*, 2000). The state is home to over 6000 minimum-rank taxa (MRT) (including subspecies and varieties) of native vascular plants, 36% of which are endemic (Hickman, 1993). A large part of California's floristic richness can be attributed to the geologic, topographic and climatic diversity found within the state (Raven & Axelrod, 1978; Hickman, 1993). Climate change may pose a great threat to the state's endemic taxa, particularly if barriers to dispersal prevent plants from shifting their ranges to keep pace with the rate of change (Loarie *et al.*, 2008).

The distribution and underlying causes of endemism in the California flora have attracted considerable interest. A portion of the state endemics are thought to be relict taxa (such as *Sequoiadendron*) persisting in spatially restricted climates that were once more widespread across North America or the Northern Hemisphere (Stebbins & Major, 1965). These palaeoendemics have unusually high conservation value because of their genetic and evolutionary distinctness (Faith, 1992; Moritz, 2002). However, a significant proportion of California's endemics are thought to be recently evolved neoendemics (Stebbins & Major, 1965), whose origins may be tied to the uplift of mountain ranges and the general aridification of the climate over the last 2–5 million years (Axelrod, 1973; Graham, 1999; Chamberlain & Poage, 2000; Calsbeek *et al.*, 2003; Ackerly, 2009). The geologic heterogeneity of the state, reflected in such features as patchily distributed serpentine soils, is also known to contribute to the formation of new plant lineages (e.g. Rajakaruna *et al.*, 2003; Baldwin, 2005).

In this study we present a novel analysis of patterns of endemism within the state flora, with a particular emphasis on the neoendemics. Our goal is to identify hotspots of floristic endemism that may be useful for conservation planning and for future study of the physical and climatic features associated with high rates of diversification. We explore four broad areas related to evolutionary hotspots: (1) patterns of endemic richness across the state; (2) patterns of median endemic taxon age and range size across the state; (3) patterns of richness weighted by the inverse of range size and taxon age, two factors associated with neoendemicism; and (4) patterns of diversity in relation to the protection status of distributional areas. Our efforts build on previous analyses of endemic and native plant diversity of the state (Stebbins & Major, 1965; Richerson & Lum, 1980; Thorne *et al.*, 2009) and complement similar analyses of the endemic mammal fauna (Davis *et al.*, 2008).

METHODS

We used the Jepson Herbarium's list of state endemics, available online at <http://ucjeps.berkeley.edu/interchange/>

endemic.html (accessed October 26th, 2009). For this analysis we worked with the smallest taxonomic units (MRT) recognized as an endemic by the Jepson Manual (Hickman, 1993), including subspecies and varieties, referred to simply as 'taxa' throughout the paper. Refer to the manual for nomenclatural authorities. Note that state endemics are not equivalent to CA-FP endemics, as the floristic province excludes the Desert and Great Basin regions and includes small portions of adjacent Oregon, Nevada and Baja California (Raven & Axelrod, 1978; Loarie *et al.*, 2008). Although none of the major floristic provinces within California are entirely contained within the state, an examination of Californian vascular-plant endemics is valuable for comparison with patterns of endemism for Californian mammals (Davis *et al.*, 2008) and for conservation planning at the statewide level.

We first determined the geographic range of each taxon using distributional data reported in the Jepson Manual (Hickman, 1993; Loarie *et al.*, 2008; Ackerly, 2009). The authors of individual treatments in the Jepson Manual, drawing on examination of specimens and personal knowledge, described geographic ranges based on presence or absence in each of 35 floristic subregions and lower and upper elevational limits. To represent this information geographically, we used a digital elevation model to divide the 35 Jepson floristic subregions into 100-m elevational sections (see Ackerly, 2009), ranging in size from 1.0 to 22,860 km² (median = 171 km²). These sections divide the state into 800 non-overlapping regions (referred to as 'bands' throughout the paper), which form the geographic units of our analyses (Table S1). For example, we grouped all land lying between 1 and 100 m in elevation in the Mojave Desert (DMoj) floristic subregion into one band. The Jepson distribution data were then used to score each taxon as present or absent from each band. Range size was determined as the total area of those bands in which each taxon occurred. This method of range size determination is likely to be an overestimate of the true range of many taxa, particularly for taxa reported to occur in some of the larger bands, such as the bands found in the Great Central Valley (GV), a large area of the state with little topographic variation (GV region in Fig. 1). As an alternative, we also used the number of bands in which each taxon occurs as an index of the geographic range size for each taxon. Note that the varieties and subspecies included in our analyses may have smaller ranges than the species to which they belong, so it may be difficult to compare our results (based on MRT) with other analyses using only species.

To determine the approximate ages of endemic taxa, we searched the recent literature for genera containing California endemic taxa that have been subject to phylogenetic analyses based on 18S–26S nuclear ribosomal sequence data from both internal transcribed spacers, ITS-1 and ITS-2. We restricted our search to analyses with ITS data because ITS-1 and ITS-2 have been shown to undergo approximately clocklike evolution in a number of angiosperm clades (reviewed by Kay *et al.*, 2006) and therefore could be used to obtain roughly comparable estimates of relative divergence time across flowering plant

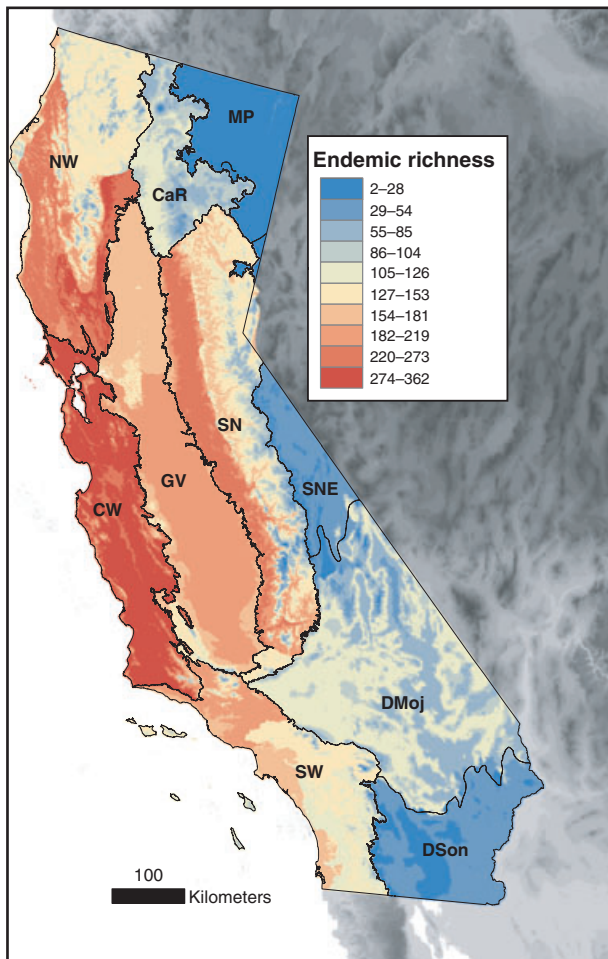


Figure 1 Endemic richness of the California flora, as total number of endemic minimum-rank taxa per regional band. Codes correspond to Jepson bioregions. The north-western California (NW), Cascade Ranges (CaR), Sierra Nevada (SN), Great Central Valley (GV), central-western California (CW), and south-western California (SW) regions fall within the California Floristic Province (CA-FP), while the Modoc Plateau (MP) and East of the Sierra Nevada (SNE) fall within the Great Basin Province. The Mojave (DMoj) and Sonoran (DSon) desert regions are part of the Desert Province.

groups (with some caveats- see discussion below). In addition, ITS sequences are the most widely used nuclear DNA characters for phylogenetic analysis of young plant lineages, notwithstanding the potential for any single line of molecular evidence to provide incomplete or misleading data about evolutionary relationships (see Baldwin *et al.*, 1995; Kay *et al.*, 2006).

We selected studies that included at least one California endemic, published ITS-1 and ITS-2 sequence data for all taxa in the analysis, and a phylogenetic topology based on either parsimony or maximum likelihood analysis of the data. The published topologies were used to determine the sister taxa or taxon for each endemic for this analysis. When studies included multiple phylogenetic analyses based on more data than ITS sequences alone (such as chloroplast sequence data), the topology with the most sources of data was used to

determine sister taxa. ITS-1 and ITS-2 sequence data used in each study were downloaded from GenBank into the software program MEGA version 3.1 (Kumar *et al.*, 2004) and aligned using ClustalW implemented inside MEGA (Gap penalty = 15, gap extension penalty 6.66, transition matrix = IUB, transition weight = 0.5, delay divergent cut-off = 30, negative matrix = OFF). In almost all cases, sister taxa were congeners of the target endemic taxon. Many of the published studies focused on clades with several closely related California endemics, thus, the sister taxon for many endemic MRT in the study is another endemic MRT. Once aligned, the highly conserved 5.8S ribosomal subunit sequence (occurring between ITS-1 and ITS-2) was removed using Baldwin (1992) as a guide for identifying the region. We then calculated the Tamura-Nei (TN) pairwise distance (Tamura & Nei, 1993) between each California endemic and its sister taxon or taxa in MEGA. We calculated net between-group divergence when sequences for multiple individuals per taxon were available.

Tamura-Nei distances measured ITS sequence divergence between endemic taxa and their sister taxon or taxa, and therefore can be used as a provisional estimate of the age of the endemic. There is known to be variation across lineages in the rates of ITS substitutions, some of which is related to life history (Kay *et al.*, 2006; Smith & Donoghue, 2008; see discussion of life history below), although we have not attempted to correct for this in the molecular divergence estimates given the extent of the data synthesized here (c.f. Davis *et al.*, 2008). While this is a limitation of our study, the extremely low divergence levels between many neoendemic clades and close relatives in this analysis lessen the impact of such variation on our findings. Our method also assumes that the phylogenetic study identified the true closest sister group of the endemic taxon; where this is not true, we will have obtained an overestimate of the true taxon age because of extinction or incomplete taxon sampling, which is likely the case for some of the more diverse groups within our study. This bias is unlikely to dramatically alter our analyses given the very low levels of divergence we have measured in many taxa. When subspecies or varieties were not specified in the published phylogenetic studies, TN distances calculated for a species were applied to all endemic MRT associated with the species.

In all, data from 31 phylogenetic studies (Table 1) were used to apply age estimates to 337 California state endemic taxa (Table S1) from 17 families, representing approximately 15% of the endemic state flora. The proportion of taxa sampled was relatively even across the bands in the state (Fig. S1), with the high Sierra Nevada (SN) (eastern portion of the SN region in Fig. S1) and portions of the Deserts [DMoj and Sonoran Desert (DSon)] being slightly under-sampled relative to the rest of the state, while the North Coast Ranges [coastal portions of north-west (NW)], portions of the Modoc Plateau (MP) and the Sacramento Valley (northern portion of GV) were slightly over-sampled relative to the rest of the state.

Maps of endemic diversity were generated using routines in the R statistical language (R Development Core Team 2007) and ArcMap (ESRI 2008). First, simple richness maps were

Table 1 Families and genera sampled for taxon age estimates from previously published studies with internal transcribed spacer-based phylogenies including California state endemics.

Family	Genus	State endemic MRT with ages in this study	Reference
Apiaceae	<i>Sanicula</i>	4	(Vargas <i>et al.</i> , 1999)
Asteraceae	<i>Cirsium</i>	8	(Kelch & Baldwin, 2003)
	<i>Hazardia</i>	3	(Markos & Baldwin, 2001)
	<i>Lasthenia</i>	12	(Chan <i>et al.</i> , 2001)
	<i>Layia</i>	6	(Baldwin, 2005)
	<i>Lessingia</i>	10	(Markos & Baldwin, 2001)
	<i>Stephanomeria</i>	4	(Lee <i>et al.</i> , 2002)
Brassicaceae	<i>Lepidium</i>	4	(Mummenhoff <i>et al.</i> , 2004)
Ericaceae	<i>Arctostaphylos</i>	49	(Boykin <i>et al.</i> , 2005)
Fabaceae	<i>Astragalus</i>	28	(Wojciechowski <i>et al.</i> , 1999)
	<i>Lathyrus</i>	8	(Kenicer <i>et al.</i> , 2005)
	<i>Lupinus</i>	17	(Ainouche <i>et al.</i> , 2004)
	<i>Trifolium</i>	12	(Ellison <i>et al.</i> , 2006)
Fagaceae	<i>Chrysopsis</i>	1	(Manos <i>et al.</i> , 2001)
Grossulariaceae	<i>Ribes</i>	6	(Senters & Soltis, 2003)
Hydrophyllaceae	<i>Phacelia</i>	10	(Gilbert <i>et al.</i> , 2005)
Iridaceae	<i>Iris</i>	10	(Wilson, 2003)
Liliaceae	<i>Lilium</i>	4	(Nishikawa <i>et al.</i> , 1999)
	<i>Smilax</i>	1	(Cameron & Fu, 2006)
	<i>Trillium</i>	2	(Farmer, 2006)
Malvaceae	<i>Sidalcea</i>	21	(Andreasen & Baldwin, 2003)
Onagraceae	<i>Epilobium</i>	2	(Baum <i>et al.</i> , 1994)
Orobanchaceae	<i>Orobanche</i>	1	(Schneeweiss <i>et al.</i> , 2004)
Polemoniaceae	<i>Linanthus</i>	19	(Bell & Patterson, 2000)
	<i>Navarretia</i>	16	(Spencer & Porter, 1997)
Rhamnaceae	<i>Ceanothus</i>	29	(Hardig <i>et al.</i> , 2000)
Saxifragaceae	<i>Boykinia</i>	1	(Soltis <i>et al.</i> , 2003)
	<i>Lithophragma</i>	5	(Kuzoff <i>et al.</i> , 1999)
Scrophulariaceae	<i>Collinsia</i>	15	(Armbruster <i>et al.</i> , 2002)
	<i>Keckiella</i>	4	(Freeman <i>et al.</i> , 2003)
	<i>Mimulus</i>	25	(Beardsley <i>et al.</i> , 2004)

MRT, Minimum-Rank Taxa.

produced by summing the number of endemic taxa occurring in each of the bands within the state. Next, we mapped the median age and the median range size of endemic taxa within each band (median ages were based on the subsample of taxa for which phylogenetic data were available- 337 MRT had age estimates versus 2179 MRT in the endemic flora as a whole). Finally, we explored three diversity metrics (area-weighted diversity, band-weighted diversity and age-weighted diversity) that incorporated age and range size data by mapping richness inversely weighted by ITS TN distance or by range size for all sampled taxon occurring in each band (c.f. Williams *et al.*, 1993; Crisp *et al.*, 2001; Linder, 2001; Davis *et al.*, 2008; Thorne *et al.*, 2009). These metrics give greater weight to areas with younger taxa or smaller-ranged taxa, respectively. We split our diversity indices into 5% quantiles to facilitate comparisons between metrics.

Richness was positively correlated with both the band-weighted and age-weighted diversity indices (band number $r^2 = 0.82$, age $r^2 = 0.61$, $N = 800$), and therefore it is difficult

to parse out the relative role of species richness versus age or range size on the results. As an alternative, we explored using the residuals of the linear fit between richness and age- or range-weighted diversity as a means of removing the influence of endemic richness per se from the age- and range size-weighted measures.

As a final step, we compared our diversity estimates for each band to the distribution of National Parks, California State Parks, Bureau of Land Management holdings and National Forests to estimate the extent to which diversity hotspots fell within areas with some degree of governmental protection.

RESULTS

Total endemic diversity

Endemic diversity ranged from 2 to 362 endemic MRT per band (mean = 107 taxa). In both total number of endemics (Fig. 1) and proportion of taxa that are endemic (Fig. S2), the

highest levels were mapped to the SN foothills (western portion of region marked SN in Fig. 1), portions of the NW region (southern portions of the region marked NW in Fig. 1) and the central-western region (CW region in Fig. 1) of the state, including the San Francisco Bay Area and the South Coast Ranges. The lowest endemic diversity occurred in areas of the state of California outside of the CA-FP, i.e., in the Great Basin (MP and SNE regions in Fig. 1) and Desert provinces (DMoj and DSon regions in Fig. 1). Similar results were recently reported by Thorne *et al.* (2009), based on a partially overlapping plant distribution dataset (Viers *et al.*, 2006).

Median range size

Areas in the SN, central-western California (CW), southwestern California (SW- particularly the San Bernardino Range in the east) and the portion of the Great Basin east of the Sierra Nevada (SNE) contained endemics with the most restricted ranges on average (Fig. 2). Analysis with range size estimated as band number (as opposed to area) revealed similar patterns (Fig. S3). The differences between the two estimates are attributed to the effects of variable band size. The band-based estimate (Fig. S3) showed lower relative median range size in regions such as the GV and the Deserts (DSon and DMoj), which contain large bands.

Median endemic age

A map of median age of endemic taxa across the state (measured as ITS divergence, Fig. 3) reveals that areas in the NW, the Cascade Range (CaR), the MP and the periphery of the GV, among others, harbour older neoendemics than other areas in the state. Endemics of the Desert and portions of the Great Basin provinces (SNE, DMoj, DSon) in particular appear to be quite young in our analysis (Figs 3 & 4). In general, the neoendemic flora appears to be young (see discussion below), with a mean ITS divergence between sampled endemics and their sister taxa of 0.023 substitutions per site (median 0.011). The MRT designated as varieties and subspecies in our analysis showed less divergence (mean TN distance = 0.0165, $N = 168$) than the MRT designated as species (mean TN distance 0.0293, $N = 169$).

Weighted diversity measures

Weighting endemic richness by the inverse of range size (estimated either with area or band number) and taxon age (Fig. 5) reveals a slightly different picture of the state's diversity relative to simple counts of endemic richness. Area- and band-weighted maps place more emphasis on endemic diversity in high-elevation areas in the DMoj and the Transverse Ranges (interior SW) than in an unweighted map (Fig. 1). Age-weighting places more emphasis on diversity in areas of the DSon and portions of the southern GV than in other weighted, or unweighted, maps. All maps still reflect the general pattern of high diversity in CW and west side of the SN and low

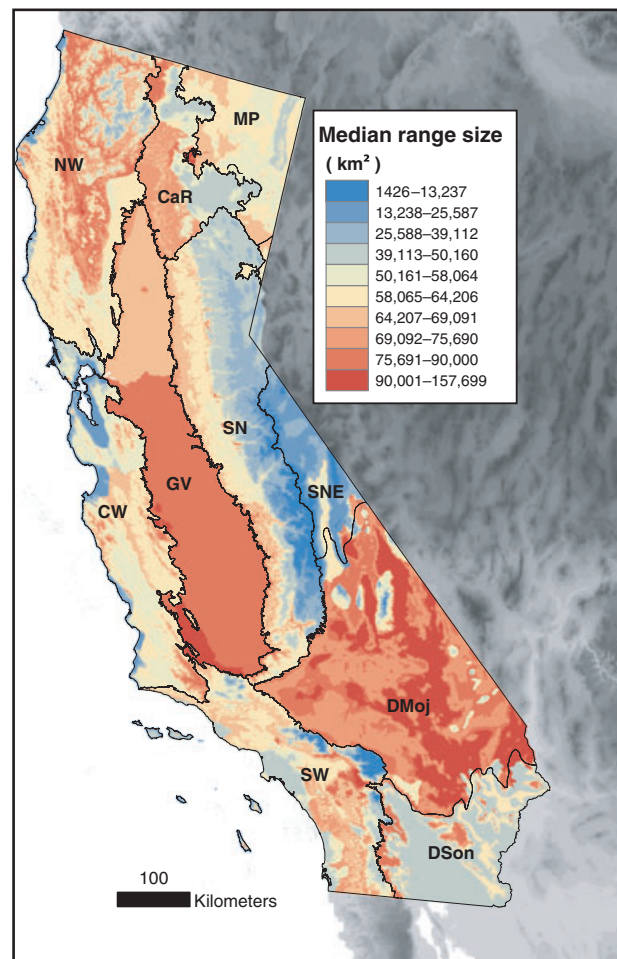


Figure 2 Median range size (km²) of co-occurring endemic taxa. Range size measured as band number is shown in Fig. S3. Region codes follow Fig. 1 legend.

diversity in the Desert (particularly in the Sonoran- DSon) and Great Basin (SNE and MP) provinces of the state. Maps derived from the upper 90th percentile of the age and area-weighted measures—one criterion for defining evolutionary hotspots (Davis *et al.*, 2008)—are shown in Figs 6 & S4 in relation to the distribution of state and federally protected lands.

DISCUSSION

Hotspots of neoendemism

Our analysis of the endemic plant taxa of California reveals that despite taxonomic revisions of a wide diversity of plant groups and decades of growth in herbarium collections, our current understanding of the distribution of angiosperm endemism within the state (Fig. 1) remains quite similar to earlier findings (e.g. Stebbins & Major, 1965). Endemic richness (Fig. 1) in the state is highest in the coastal regions, particularly around the San Francisco Bay Area (northern portion of CW), and, to a lesser extent, in the foothills of the Sierra Nevada (western SN) and the southern periphery of the

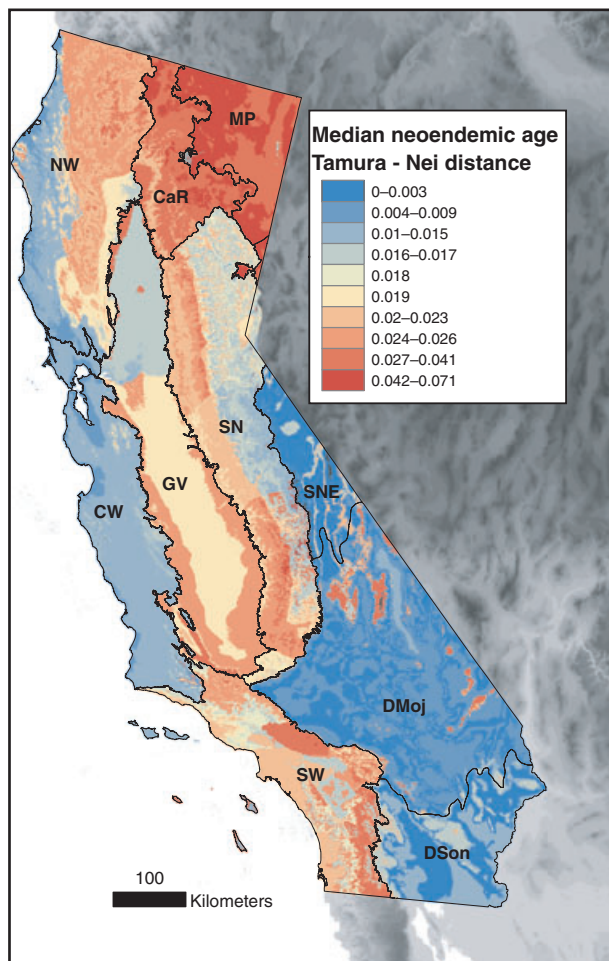


Figure 3 Median age of sampled endemic taxa in each band, as reflected by Tamura-Nei internal transcribed spacer sequence distance between each sampled endemic and its sister taxon. Note that the age-weighted richness map was calculated with a subset of the endemic flora (see Table 1 for list). Region codes follow Fig. 1 legend.

north-west region (NW). Endemics also make up a larger portion of the total floral richness in these areas (Fig. S2) relative to the rest of the state. Peaks in endemic floral richness

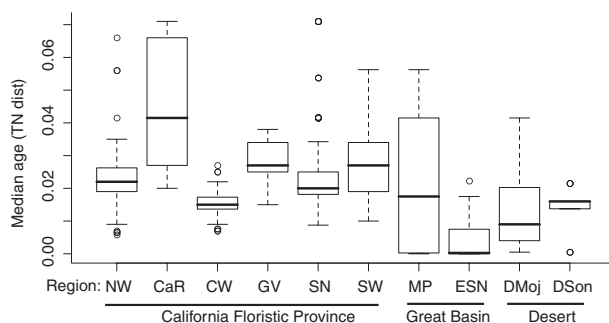


Figure 4 Median age of sampled endemic flora in each Jepson region, as reflected by Tamura-Nei internal transcribed spacer sequence distance between each sampled endemic and its sister taxon. Region codes follow Fig. 1 legend.

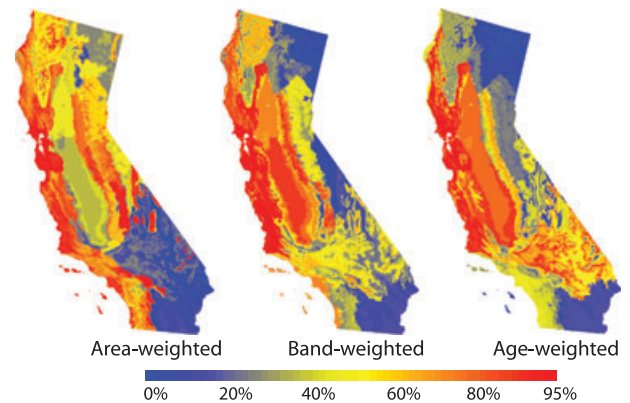


Figure 5 Diversity of the endemic flora, calculated three ways: richness weighted by the inverse of the range size of each taxon (Area-weighted), richness weighted by the inverse of the number of bands occupied by each taxon (Band-weighted) and richness weighted by the inverse of the age of each taxon (Age-weighted). Note that the age-weighted richness map was calculated with a subset of the endemic flora (see Table 1 for list). All metrics have been converted to 5% percentiles to facilitate comparisons between metrics.

(Fig. 1) are somewhat decoupled from the peaks of total plant richness (Fig. 1 in Ackerly, 2009). Endemic diversity is considerably lower in the portions of the state outside of the CA-FP. This disparity may arise in part because most of the CA-FP is contained within the state, and therefore most taxa adapted to conditions unique to the CA-FP will appear as state endemics (unless they range into southern Oregon or northern Baja California). In contrast, the climatic and geologic conditions found within the Desert and Great Basin provinces of the state, and much of the associated flora, extend far beyond the geographic boundaries of California.

Within the endemic flora, our analyses show that low elevations along the Central and North Coast as well as high elevations in the SN and the Transverse Ranges contain taxa with the smallest ranges on average (Fig. 2, see also Thorne *et al.*, 2009). The distinctiveness of the climate within these areas, as well as the reduction in area at high elevations relative to the rest of the state may be responsible for the high concentration of restricted-range endemics found in these regions. These endemics are projected to be among the taxa most severely impacted by climate change (Loarie *et al.*, 2008).

Age- and area-weighted indices of endemic diversity produced similar coarse-scale patterns to unweighted endemic diversity (Fig. 5), with slightly different emphases. Range-weighted maps give more importance to high-elevation bands in the SN, Coast Ranges and Transverse Ranges than do other maps (see Thorne *et al.*, 2009), while age-weighting highlights desert areas more than do other assessment methods.

Age of the neoendemic assemblage

In general, the neoendemic taxa sampled in our study appear to be quite young, with an average TN distance of 0.023

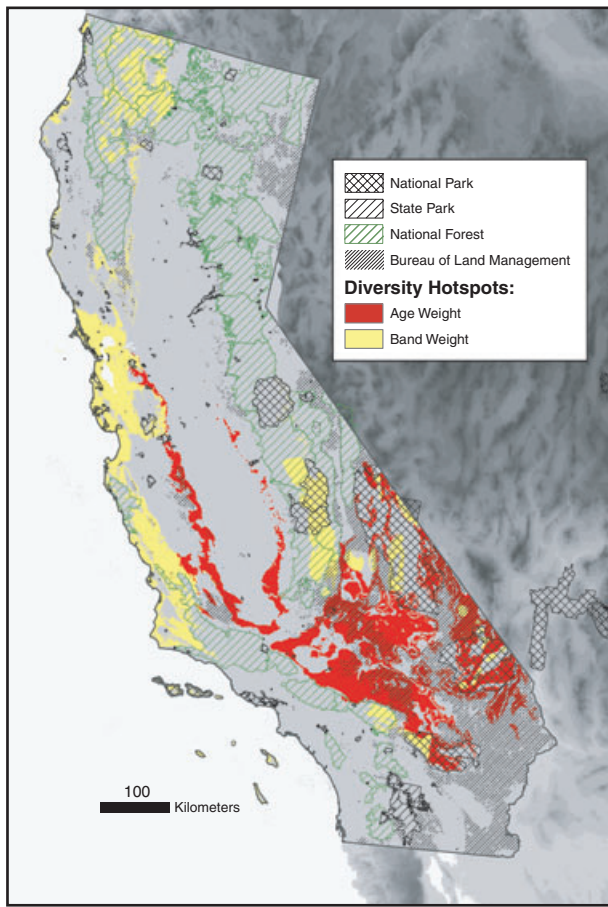


Figure 6 California floral hotspots, identified as the upper 90th percentile of the residuals of our band-weighted (yellow) and age-weighted (red) diversity estimates (Fig. 5) regressed against endemic diversity (Fig. 1). There is no overlap between the Age Weighted and Band Weighted hotspot areas. Hatched areas depict National Forests, State and National parks and Bureau of Land Management lands. Note that age-weighted richness was calculated with a subset of the endemic flora (see Table 1 for list). See Fig. S4 for same map created without using residuals.

between an endemic and its sister taxon. Our primary interest in ITS in this analysis is as a relative dating tool. Nevertheless, we can cautiously use ITS as an approximate molecular clock (Kay *et al.*, 2006) to estimate average endemic age using the formula:

$$\text{Years} = \frac{0.5(\text{TN distance})}{\text{ITS rate}}$$

where TN distance is expressed in substitutions/site and ITS rate is substitutions/site/year. As ITS shows substitution rate variation related in part to life history (Kay *et al.*, 2006; Smith & Donoghue, 2008), calculations should be performed separately for herbaceous and woody neoendemic taxa. Using the data from reported in Kay *et al.* (2006), we calculated the 95% confidence interval for the herbaceous taxa rate ($N = 10$ rates, excluding the same outlier they identified) as $2.93\text{--}4.95 \times 10^{-9}$ substitutions/site/year, which yields an age range of 2.80–

4.73 mybp for the average herbaceous neoendemic plant in our study (mean herbaceous TN distance = 0.028, $N = 236$). Similarly, using Kay *et al.*'s data we calculate the 95% confidence interval on the woody taxa rate ($N = 18$ rates) as $1.07\text{--}2.32 \times 10^{-9}$ substitutions/site/year, which yields an age range of 2.51–5.44 mybp for the woody neoendemism (mean woody TN distance = 0.012, $N = 101$). Note that the lower average TN distance for woody taxa in our sample is matched by lower rates of divergence for woody clades in the Kay *et al.* dataset, so the range of ages we obtain for woody and herbaceous taxa are similar. There are many sources of uncertainty in this calculation (see discussion in methods above and in Kay *et al.*, 2006), and our decision to include subspecies and varieties in the analysis shifts this estimate towards the recent. Palaeoendemism, such as *Sequoiadendron*, which is known from Tertiary fossils outside California (Raven & Axelrod, 1978), are an important part of the overall picture of endemism within the state, but were not included in these estimates and will require a different set of analyses. With current knowledge, it is not known whether there is a truly bimodal distribution of ages, corresponding to the neo- versus palaeoendemic concepts, or if the distribution of taxon ages is continuous.

Of all of the regions in our analyses, the Desert and Great Basin provinces appear to harbour the youngest neoendemism on average (Figs 3 & 4). This could be attributable to the fact that both provinces extend far beyond the state boundaries; state endemics of these regions must have ranges that are much smaller than the total extent of warm and cold desert habitat in western North America. These endemics therefore may include recently diverged species that are still expanding in range. If this is true, one might expect a correlation with age and range size. In our dataset, the two were unrelated in endemic MRT in the Desert and Great Basin provinces [$\log(\text{age}) \sim \log(\text{range size})$ $r^2 = 0.055$, $P > 0.1$, $\log(\text{age}) \sim \log(\text{band number})$ $r^2 = 0.072$, $P > 0.05$; $N = 42$] and were positively related but with very low r^2 values across the state as a whole [$\log(\text{age}) \sim \log(\text{range size})$ $r^2 = 0.028$, $P = 0.002$; $\log(\text{age}) \sim \log(\text{band number})$ $r^2 = 0.043$, $P = 0.0001$; $N = 336$].

A second possible explanation for the recent age estimates of Desert and Great Basin endemics is the young age of modern desert environments within the state. The Mediterranean climate of the CA-FP and the first true desert conditions in western North America are thought to have originated approximately 5 million years ago (Axelrod, 1973). Modern warm-desert vegetation in California was extremely limited during the late Pleistocene, however, and became extensive $< 12,000$ ybp, based on abundant palaeobotanical data from fossil packrat middens (Betancourt *et al.*, 1990).

Most state endemics that are restricted to the deserts occur in habitats that have undergone recent (post-Pleistocene) isolation or that are characterized by extreme climatic conditions of recent origin. The most diverse set of desert endemics is restricted to high-elevation 'sky islands', with wetter, cooler climates than the surrounding desert lowlands, e.g., the high DMOj mountains and the White and Inyo ranges (Baldwin &

Moe, 2002). Divergence of some 'sky island' endemics may have coincided with post-Pleistocene isolation of the once-widespread climatic conditions that characterize modern high-montane desert environments. Endemics at the western edges of the warm desert, where summer monsoons rarely reach, live in essentially an extremely dry Mediterranean climate and include descendants from CA-FP lineages already adapted to (less harsh) Mediterranean conditions (see Baldwin & Moe, 2002).

Physical correlates of neoendemism

We explored the effect of climate on our measures of diversity by regressing our diversity metrics against mean annual, monthly minimum and monthly maximum temperature and precipitation data from the 1 km resolution DAYMET 1980–1998 mean climate database (<http://www.daymet.org>, summarized in Table S1), but found no relationship beyond the effect of younger taxa occurring in the desert (results not shown). Similar analyses of the correlation between rugosity (measured as the standard deviation of elevation within each band) on diversity did not yield any trends (results not shown). These results contrast with the analyses of Richerson & Lum (1980) for the entire flora conducted at larger scales based on earlier species range descriptions (Munz, 1959, 1968). These earlier analyses found that increased diversity was associated with increased precipitation, topographic heterogeneity and decreased temperature seasonality. Harrison & Grace (2007) have also shown relationships between species richness and productivity based on plant diversity of serpentine-containing regions of the state. At this point, we do not know whether the differences between our study and these previous analyses are attributed to the scale of analysis, the distributional datasets or the analysis of the entire flora versus our focus on the state endemics.

Our analysis is based on bioregional and elevational occurrence data from the state flora, a dataset that presents some unique challenges in analysis and interpretation. The Jepson Manual is impressive in the range of taxa that it covers, but it is limited to providing relative coarse-scale spatial distributions of individual taxa, which may translate to false positives when occurrences within bands are spatially mapped. This problem is particularly acute in large subregions with little topographic relief, which lead to relatively large, coarse-scale bands in areas like the GV, whereas we have more confidence in the method in small elevational bands in the SN, for example. We have attempted to account for some of the uncertainty introduced by these distributional data by presenting both band- and area-weighted diversity measures. Nevertheless, it will be valuable to revisit these analyses using distributional models for individual taxa based on georeferenced specimen data; only a small number of endemic taxa currently have enough georeferenced specimens for such an analysis (Loarie *et al.*, 2008).

Another challenge of interpreting analyses drawn from data at this spatial scale is that the maps yield a false sense of

precision. The grain of our data is elevational bands within floristic regions, and this is the scale at which any interpretation should be made.

Conservation

Earlier analyses of Pavlik & Skinner (1994) found that unprotected areas of California tend to harbour the highest numbers of rare plant taxa. Similarly, in our analyses of endemics (many of which are listed as rare), we found that some of the highest levels of diversity within our study, measured as either band-weighted richness or age-weighted richness, were found in areas that fall outside of state and federally protected areas. These areas correspond to much of CW (including the San Francisco Bay Area), areas along the North Coast and portions of the DMoj (Fig. 6). These hotspots of floral neoendemism, particularly portions that overlap mammal evolutionary hotspots in the central western region of California and the western edges of the Desert province (Davis *et al.*, 2008), are prime targets for a conservation plan that favours the protection of areas that may be responsible for both the conservation and ongoing generation of new biodiversity in California.

ACKNOWLEDGEMENTS

We are grateful to Craig Moritz, Edward Davis, Michelle Koo, Scott Loarie, Dick Moe, and Paul Fine for suggestions, comments and help with compiling data. This work was supported by grants from the Resource Legacy Fund Foundation, the California Department of Parks and Recreation and the NSERC CREATE Training Programme in Biodiversity Research.

REFERENCES

- Ackerly, D.D. (2009) Evolution, origin and age of lineages in the Californian and Mediterranean floras. *Journal of Biogeography*, **36**, 1221–1233.
- Ainouche, A., Bayer, R.J. & Misset, M.T. (2004) Molecular phylogeny, diversification and character evolution in *Lupinus* (Fabaceae) with special attention to Mediterranean and African lupines. *Plant Systematics and Evolution*, **246**, 211–222.
- Andreasen, K. & Baldwin, B.G. (2003) Reexamination of relationships, habitat evolution, and phylogeography of checker mallows (*Sidalcea*; Malvaceae) based on molecular phylogenetic data. *American Journal of Botany*, **90**, 436–444.
- Armbruster, W.S., Mulder, C.P.H., Baldwin, B.G., Kalisz, S., Wessa, B. & Nute, H. (2002) Comparative analysis of late floral development and mating-system evolution in tribe Collinsieae (Scrophulariaceae s.l.). *American Journal of Botany*, **89**, 37–49.
- Axelrod, D.I. (1973) History of Mediterranean ecosystems in California. *Mediterranean Type Ecosystems* (ed. by F. di Castri and H.A. Mooney), pp. 225–305, Springer, Berlin.

- Baldwin, B.G. (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution*, **1**, 3–16.
- Baldwin, B.G. (2005) Origin of the serpentine-endemic herb *Layia discoidea* from the widespread *L. glandulosa* (Compositae). *Evolution*, **59**, 2473–2479.
- Baldwin, B.G. & Moe, R.L. (2002) Floristic diversity in the California deserts. *The Jepson Desert Manual: Vascular Plants of Southeastern California* (ed. by B.G. Baldwin, S. Boyd, B.J. Ertter and et al), pp. 40–46, University of California Press, Berkeley.
- Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S. & Donoghue, M.J. (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, **82**, 247–277.
- Baum, D.A., Sytsma, K.J. & Hoch, P.C. (1994) A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA-sequences. *Systematic Botany*, **19**, 363–388.
- Beardsley, P.M., Schoenig, S.E., Whittall, J.B. & Olmstead, R.G. (2004) Patterns of evolution in Western North American *Mimulus* (Phrymaceae). *American Journal of Botany*, **91**, 474–489.
- Bell, C.D. & Patterson, R.W. (2000) Molecular phylogeny and biogeography of *Linanthus* (Polemoniaceae). *American Journal of Botany*, **87**, 1857–1870.
- Betancourt, J.L., Van Devender, T.R. & Martin, P.S. (1990). *Packrat Middens: the Last 40,000 years of Biotic Change*. University of Arizona Press, Tucson.
- Boykin, L.M., Vasey, M.C., Parker, V.T. & Patterson, R. (2005) Two lineages of *Arctostaphylos* (Ericaceae) identified using the internal transcribed spacer (ITS) region of the nuclear genome. *Madrono*, **52**, 139–147.
- Calsbeek, R., Thompson, J.N. & Richardson, J.E. (2003) Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. *Molecular Ecology*, **12**, 1021–1029.
- Cameron, K.M. & Fu, C. (2006) A nuclear rDNA phylogeny of *Smilax* (Smilacaceae). *Aliso*, **22**, 598–605.
- Chamberlain, C.P. & Poage, M.A. (2000) Reconstructing the paleotopography of mountain belts from the isotopic composition of authigenic minerals. *Geology*, **28**, 115–118.
- Chan, R., Baldwin, B.G. & Ornduff, R. (2001) Goldfields revisited: a molecular phylogenetic perspective on the evolution of *Lasthenia* (Compositae: Heliantheae sensu lato). *International Journal of Plant Sciences*, **162**, 1347–1360.
- Crisp, M.D., Laffan, S., Linder, H.P. & Monro, A. (2001) Endemism in the Australian flora. *Journal of Biogeography*, **28**, 183–198.
- Davis, E.B., Koo, M.S., Conroy, C., Patton, J.L. & Moritz, C. (2008) The California Hotspots Project: identifying regions of rapid diversification in mammals. *Molecular Ecology*, **17**, 120–138.
- Ellison, N.W., Liston, A., Steiner, J.J., Williams, W.M. & Taylor, N.L. (2006) Molecular phylogenetics of the clover genus (*Trifolium* – Leguminosae). *Molecular Phylogenetics and Evolution*, **39**, 688–705.
- Erwin, T.L. (1991) An evolutionary basis for conservation strategies. *Science*, **253**, 750–752.
- ESRI (2008) *ArcView*, ESRI, Redlands, CA.
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.
- Farmer, S.B. (2006) Phylogenetic analyses and biogeography of Trilliaceae. *Aliso*, **22**, 579–592.
- Forest, F., Grenyer, R., Rouget, M., Davies, T.J., Cowling, R.M., Faith, D.P., Balmford, A., Manning, J.C., Proches, S., van der Bank, M., Reeves, G., Hedderson, T.A.J. & Savolainen, V. (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature (London)*, **445**, 757–760.
- Freeman, C.E., Harrison, J.S., Janovec, J.P. & Scogin, R. (2003) Inferred phylogeny in *Keckiella* (Scrophulariaceae) based on noncoding chloroplast and nuclear ribosomal DNA sequences. *Systematic Botany*, **28**, 782–790.
- Gilbert, C., Dempcy, J., Ganong, C., Patterson, R. & Spicer, G.S. (2005) Phylogenetic relationships within *Phacelia* subgenus *Phacelia* (Hydrophyllaceae) inferred from nuclear rDNA ITS sequence data. *Systematic Botany*, **30**, 627–634.
- Graham, A. (1999) *Late Cretaceous and Cenozoic history of North American Vegetation north of Mexico*. Oxford University Press, Oxford.
- Hardig, T.M., Soltis, P.S. & Soltis, D.E. (2000) Diversification of the North American shrub genus *Ceanothus* (Rhamnaceae): conflicting phylogenies from nuclear ribosomal DNA and chloroplast DNA. *American Journal of Botany*, **87**, 108–123.
- Harrison, S. & Grace, J.B. (2007) Biogeographic affinity helps explain productivity-richness relationships at regional and local scales. *American Naturalist*, **170**, S5–S15.
- Hickman, J. (1993) *The Jepson Manual: higher plants of California*. University of California Press, Berkeley.
- Kay, K.M., Whittall, J.B. & Hodges, S.A. (2006) A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology*, **6**, 36.
- Kelch, D.G. & Baldwin, B.G. (2003) Phylogeny and ecological radiation of New World thistles (*Cirsium*, Cardueae - Compositae) based on ITS and ETS rDNA sequence data. *Molecular Ecology*, **12**, 141–151.
- Kenicer, G.J., Kajita, T., Pennington, R.T. & Murata, J. (2005) Systematics and biogeography of *Lathyrus* (Leguminosae) based on internal transcribed spacer and cpDNA sequence data. *American Journal of Botany*, **92**, 1199–1209.
- Kumar, S., Tamura, K. & Nei, M. (2004) MEGA3: integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Briefings in Bioinformatics*, **5**, 150–163.
- Kuzoff, R.K., Soltis, D.E., Hufford, L. & Soltis, P.S. (1999) Phylogenetic relationships within *Lithophragma* (Saxifragaceae): hybridization, allopolyploidy, and ovary diversification. *Systematic Botany*, **24**, 598–615.
- Lee, J., Baldwin, B.G. & Gottlieb, L.D. (2002) Phylogeny of *Stephanomeria* and related genera (Compositae-Lactuceae)

- based on analysis of 18S-26S nuclear rDNA ITS and ETS sequences. *American Journal of Botany*, **89**, 160–168.
- Linder, H.P. (2001) Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of Biogeography*, **28**, 169–182.
- Loarie, S.R., Carter, B.E., Hayhoe, K., McMahon, S., Moe, R., Knight, C.A. & Ackerly, D.D. (2008) Climate change and the future of California's endemic flora. *PLoS ONE*, **3**, e2502.
- Manos, P.S., Zhou, Z.K. & Cannon, C.H. (2001) Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. *International Journal of Plant Sciences*, **162**, 1361–1379.
- Markos, S. & Baldwin, B.G. (2001) Higher-level relationships and major lineages of *Lessingia* (Compositae, Astereae) based on nuclear rDNA internal and external transcribed spacer (ITS and ETS) sequences. *Systematic Botany*, **26**, 168–183.
- Moritz, C. (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, **51**, 238–254.
- Mummenhoff, K., Linder, P., Friesen, N., Bowman, J.L., Lee, J.Y. & Franzke, A. (2004) Molecular evidence for bicontinental hybridogenous genomic constitution in *Lepidium sensu stricto* (Brassicaceae) species from Australia and New Zealand. *American Journal of Botany*, **91**, 254–261.
- Munz, P.A. (1959) *A California Flora*. University of California Press, Berkeley.
- Munz, P.A. (1968) *Supplement to a California Flora*. University of California Press, Berkeley.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Nishikawa, T., Okazaki, K., Uchino, T., Arakawa, K. & Nagamine, T. (1999) A molecular phylogeny of *Lilium* in the internal transcribed spacer region of nuclear ribosomal DNA. *Journal of Molecular Evolution*, **49**, 238–249.
- Pavlik, B.M. & Skinner, M.W. (1994) Ecological characteristics of California's rare plants. *California Native Plant Society's Inventory of Rare and Endangered Vascular Plants of California* (ed. by M.W. Skinner and B.M. Pavlik), pp. 4–6, California Native Plant Society, Sacramento.
- R Development Core Team (2007) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Rajakaruna, N., Baldwin, B.G., Chan, R., Desrochers, A.M., Bohm, B.A. & Whitton, J. (2003) Edaphic races and phylogenetic taxa in the *Lasthenia californica* complex (Asteraceae: Heliantheae): an hypothesis of parallel evolution. *Molecular Ecology*, **12**, 1675–1679.
- Raven, P.H. & Axelrod, D.I. (1978) Origin and relationships of the California flora. *University of California Publications in Botany*, **72**, 1–134.
- Richerson, P.J. & Lum, K.-l. (1980) Patterns of plant species diversity in California: relation to weather and topography. *American Naturalist*, **116**, 504–536.
- Schneeweiss, G.M., Colwell, A., Park, J.M., Jang, C.G. & Stuessy, T.F. (2004) Phylogeny of holoparasitic *Orobanchaceae* inferred from nuclear ITS sequences. *Molecular Phylogenetics and Evolution*, **30**, 465–478.
- Senters, A.E. & Soltis, D.E. (2003) Phylogenetic relationships in *Ribes* (Grossulariaceae) inferred from ITS sequence data. *Taxon*, **52**, 51–66.
- Smith, S.A. & Donoghue, M.J. (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science*, **322**, 86–89.
- Soltis, D.E., Kuzoff, R.K., Mort, M.E., Zanis, M., Fishbein, M., Hufford, L., Koontz, J. & Arroyo, M.K. (2003) Elucidating deep-level phylogenetic relationships in Saxifragaceae using sequences for six chloroplastic and nuclear DNA regions. *Annals of the Missouri Botanical Garden*, **88**, 669–693.
- Spencer, S.C. & Porter, J.M. (1997) Evolutionary diversification and adaptation to novel environments in *Navarretia* (Polemoniaceae). *Systematic Botany*, **22**, 649–668.
- Stebbins, G.L. & Major, J. (1965) Endemism and speciation in the California flora. *Ecological Monographs*, **35**, 1–35.
- Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Thorne, J.H., Viers, J.H., Price, J. & Stoms, D.M. (2009) Spatial patterns of endemic plants in California. *Natural Areas Journal*, **29**, 344–366.
- Vandergast, A.G., Bohonak, A.J., Hathaway, S.A., Boys, J. & Fisher, R.N. (2008) Are hotspots of evolutionary potential adequately protected in southern California? *Biological Conservation*, **141**, 1648–1664.
- Vargas, P., Baldwin, B.G. & Constance, L. (1999) A phylogenetic study of *Sanicula* sect. *Sanicoria* and S. sect. *Sandwicensis* (Apiaceae) based on nuclear rDNA and morphological data. *Systematic Botany*, **24**, 228–248.
- Viers, J.H., Thorne, J.H. & Quinn, J.F. (2006). CalJep: a spatial distribution database of CalFlora and Jepson plant species. *San Francisco Estuary and Watershed Science*, **4**, 1–18.
- Whittaker, R.J., Araujo, M.B., Paul, J., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–23.
- Williams, P.H., Vane-Wright, R.I. & Humphries, C.J. (1993) Measuring biodiversity for choosing conservation areas. *Hymenoptera and Biodiversity* (ed. by J. LaSalle and I.D. Gauld), pp. 309–328, Oxford University Press, New York.
- Wilson, C.A. (2003) Phylogenetic relationships in *Iris* series *Californicae* based on ITS sequences of nuclear ribosomal DNA. *Systematic Botany*, **28**, 39–46.
- Wojciechowski, M.F., Sanderson, M.J. & Hu, J.M. (1999) Evidence on the monophyly of *Astragalus* (Fabaceae) and its major subgroups based on nuclear ribosomal DNA ITS and chloroplast DNA trnL intron data. *Systematic Botany*, **24**, 409–437.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 Proportion of the endemic flora (MRT) in each band sampled for age estimation in this study.

Figure S2 Floral endemism, expressed as the proportion of minimum rank taxa (MRT) that are endemic.

Figure S3 Average range size of co-occurring endemic taxa, measured as the number of bands a taxon occurs in.

Figure S4 California floral hotspots, identified as the upper 90th percentile of our band-weighted and age-weighted diversity estimates.

Table S1 Data for the 800 bands used in the analysis, created by dividing the bioregions of California into 100 m elevational slices.

Table S2 Tamura-Nei distance measures for 337 minimum rank taxa (MRT) endemic to the state of California.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

Nathan J. B. Kraft is currently a biodiversity postdoctoral fellow at the Biodiversity Research Centre at the University of British Columbia, specializing in functional trait and phylogenetic-based methods for the study of plant community dynamics. His research focuses on tropical forests in addition to plant communities in California.

Bruce G. Baldwin is a professor in the Department of Integrative Biology and Curator of the Jepson Herbarium at the University of California, Berkeley. His research focuses on the evolution and biogeography of Californian plants and their descendants elsewhere, especially in the Hawaiian Islands. He is also convening editor of the Jepson Flora Project.

David D. Ackerly is an associate professor in the Department of Integrative Biology and Curator of Ecology in the Jepson Herbarium at the University of California, Berkeley. His research examines community ecology, phylogenetics and the impacts of climate change on California's native plants.

Author contributions: D.D.A. and N.J.B.K. conceived the study; N.J.B.K., B.G.B. and D.D.A. compiled the data; N.J.B.K. performed the analyses and led the writing; and D.D.A. and B.G.B. contributed to the writing and interpretation.

Editor: David Richardson