PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Kling MM, Mishler BD, Thornhill AH, Baldwin BG, Ackerly DD. 2018 Facets of phylodiversity: evolutionary diversification, divergence and survival as conservation targets. *Phil. Trans. R. Soc. B* **374**: 20170397.
http://dx.doi.org/10.1098/rstb.2017.0397

Accepted: 29 August 2018

One contribution of 16 to a theme issue 'Biological collections for understanding biodiversity in the Anthropocene'.

Subject Areas:
evolution, environmental science

Keywords:
phylogenetic diversity, biodiversity conservation, reserve design, California flora, biogeography, macroevolution

Author for correspondence:
Matthew M. Kling
e-mail: mattkling@berkeley.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4258193.

THE ROYAL SOCIETY PUBLISHING

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

Facets of phylodiversity: evolutionary diversification, divergence and survival as conservation targets

Matthew M. Kling, Brent D. Mishler, Andrew H. Thornhill, Bruce G. Baldwin and David D. Ackerly

Department of Integrative Biology, University and Jepson Herbaria/University of California, Berkeley, CA 94720-3200, USA

MMK, 0000-0001-9073-4240

Biodiversity is often described as having multiple facets, including species richness, functional diversity and phylogenetic diversity. In this paper, we argue that phylogenetic diversity itself has three distinct facets—lineage diversification, character divergence and survival time—that can be quantified using distinct branch length metrics on an evolutionary tree. Each dimension is related to different processes of macroevolution, has different spatial patterns and is tied to distinct goals for conserving biodiversity and protecting its future resilience and evolutionary potential. We compared the landscapes identified as top conservation priorities by each of these three metrics in a conservation gap analysis for California, a world biodiversity hotspot, using herbarium data on the biogeography and evolutionary relationships of more than 5000 native plant species. Our analysis incorporated a novel continuous metric of current land protection status, fine-scale data on landscape intactness and an optimization algorithm used to identify complementary priority sites containing concentrations of taxa that are evolutionarily unique, vulnerable due to small range size and/or poorly protected across their ranges. Top conservation priorities included pockets of coastal and northern California that ranked highly for all three phylodiversity dimensions and for species richness, as well as sites uniquely identified by each metric whose value may depend on whether properties such as genetic divergence, high net diversification or independent survival experience are most desirable in an Anthropocene flora.

This article is part of the theme issue 'Biological collections for understanding biodiversity in the Anthropocene'.

1. Introduction

As we move toward the bottleneck of what may become the planet's sixth mass extinction [1], conservationists face the challenge of preserving as much biodiversity as possible given imperfect knowledge and limited resources. Difficult choices will have to be made about which taxa and landscapes to protect [2]. Finding efficient solutions to this problem is critical, and quantitative approaches to conservation prioritization have accordingly received much attention in the academic and applied literature, with particular emphasis on optimally locating future reserves to maximize overall biodiversity protection [3,4].

One aspect of this issue that has attracted recent focus is which of biodiversity's multiple facets to target for conservation [5–7]. It has been increasingly argued that instead of species richness, phylogenetic diversity (phylodiversity, PD) offers a more conservation-relevant metric for biodiversity [8–10]. PD is defined as the total length of all branches connecting a given set of terminal taxa to the root of an evolutionary tree [8]; a set of species that are distantly related or connected by long branches will thus have a higher collective PD measure than a set of closely related species spanning short branches. Because relatedness corresponds to shared evolutionary history and often shared functional traits, focusing on PD rather than species richness during conservation

© 2018 The Author(s) Published by the Royal Society. All rights reserved.

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

prioritization is a basis for protecting non-redundant species with complementary biological characteristics [10–12].

Numerous variants of PD have been introduced, including alternative formulae as well as alternative variables used to represent branch lengths on the tree [8,11–14]—a proliferation of approaches that has led to calls for more conceptual and empirical clarity about the conservation relevance of PD metrics [15,16]. In this paper, we argue that phylogenetic diversity has three distinct facets, each related to different macroevolutionary processes and different conservation goals. Macroevolution comprises three distinct processes: branching events (i.e. diversification), molecular and phenotypic changes happening along the branches (i.e. divergence) and persistence of branches through time (i.e. survival). Each of these three theoretical processes generates empirical patterns that can be quantified using different measures of branch length when calculating PD (figure 1; electronic supplementary material, figure S1).

First, lineages experience episodic divergence, accumulating changes in genetic, morphological or functional characters. By default, most molecular phylogenies are 'phylograms' with branch lengths representing the inferred number of mutations in the genes used to construct the tree. (While branch lengths can also represent phenotypic characters, genetic characters are more common.) Assuming that these genes reflect broader genotypic or phenotypic patterns of evolutionary change, PD measured on a phylogram will thus represent the total cumulative divergence of a given set of extant taxa. The validity of this assumption is uncertain for present phylogenetic methods that rely on relatively few genes [13,17], but will improve with continuing advances in whole-genome phylogenetics and functional genomics. For now it is clear that although rates of molecular evolution vary among genes, rates for individual genes are often persistent within clades due to conserved characteristics such as genetic architecture and life-history traits [18–21], which suggests that molecular branch lengths may have utility in broadly predicting diversity. Divergence across a clade in genomic characteristics and associated functional traits is relevant to conservation in two ways: it represents diversity in biological attributes that may be critical to the goal of conserving ecological function and character, and it could be used to identify taxa with high evolutionary rates that could be more resilient to future environmental change.

Secondly, evolutionary lineages either survive or go extinct over time. Phylogenetic diversity is often assessed on 'chronograms' with evolutionary branches measured in units of time, generally estimated using fossil calibration and molecular clocks. Chronogram PD represents the total number of years that independent ancestral lineages survived to give rise to a given set of taxa, which can be interpreted as their combined ecological experience, or their ancestors' cumulative demonstrated success in persisting through past environmental change. Targeting this survival experience in conservation prioritization could be a strategy for selecting a resilient biota more likely to endure future anthropogenic environmental change. This time-based strategy assumes either that survival rates are somewhat evolutionarily conserved (as commonly associated with certain life history traits [22]) or that properties of the geographical area itself increase lineage survival rates (as commonly associated with climatic refugia [23–25]).

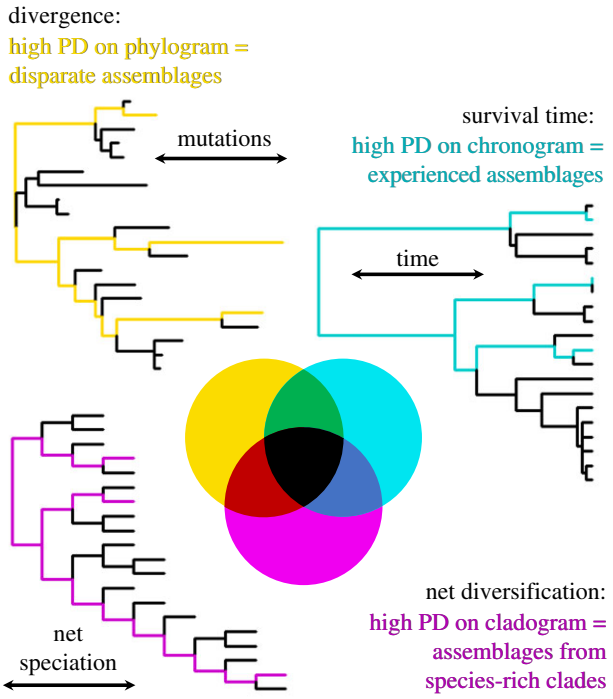

Figure 1. Three facets of phylogenetic diversity: divergence, diversification and survival time. The example phylogenies shown have identical topologies and differ only in branch length metric. For each tree, the set of three terminal taxa with the maximum collective PD on that tree is highlighted, where PD is the total length of the coloured branches connecting them (for the chronogram and cladogram, multiple sets are tied for the maximum and only one is shown). The assemblage of terminal taxa in a given geographical location could be high in none, a subset, or all three of the dimensions, placing it in a given sector of the Venn diagram (electronic supplementary material, figure S1). This same colour scheme is used on the maps in figure 3. Note that, in practice, the number of taxa varies among geographical areas and range size varies among taxa, both of which affect conservation rankings; these are held constant in this figure. (Online version in colour.)

Finally, lineages undergo occasional diversification—lineage splitting events known as cladogenesis or speciation. PD measured on a 'cladogram' phylogeny in which all branch segments are of equal length represents the net number of diversification events that gave rise to a set of taxa, and is similar to early measures of taxonomic distinctiveness [26]. Importantly, diversification rates based on phylogenies of extant taxa do not account for extinct lineages and undetected past speciation events, and so must be viewed as *net* diversification rates; while this obscures insights that might come from disentangling instantaneous rates of speciation and extinction, it still generates a metric of longer-term net speciation that is relevant to evolution and conservation. Like rates of divergence or survival, rates of net diversification can be properties of clades due to life-history characteristics [22,27], or can be driven by landscape features that influence metapopulation dynamics [28]. Prioritizing the protection of such lineages and landscapes could be a long-range strategy for conserving the underlying engines of diversification [29,30], which may be crucial in regenerating biodiversity following an anthropogenic mass extinction.

Here, we consider these alternative evolutionary metrics in the context of a conservation gap analysis for the vascular flora of California. California's biota is both highly diverse and highly threatened, making it a global conservation priority under a wide variety of prioritization schemes [4].

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

Compared with other global biodiversity hotspots, California also has a biota whose biogeography and systematics are uncommonly well documented [31,32], allowing for more robust assessments and making it a good methodological study system. The diversity of California's flora has been extensively catalogued in herbaria, which house millions of specimens representing more than 5000 native vascular plant species; these records are largely digitized and publicly available [33] and form the basis of our analysis.

DNA and geolocations from these herbarium specimens were used in prior studies by our group to model both the phylogenetic and biogeographical relationships of the native California flora at unprecedented levels of detail [34,35]. In this study, we leverage those same datasets, in combination with high-resolution data on land protection status and landscape intactness, to identify optimal future conservation priorities for native California plants. We compare conservation priorities based on the three phylogenetic metrics as well as traditional species richness.

While these PD metrics can be used to assign conservation value to a given taxon, it is also important to consider a taxon's vulnerability when setting conservation priorities [4]. In the absence of sufficiently detailed data on IUCN threat status necessary to calculate metrics such as EDGE scores [14], geographical range size has been found to be a strong predictor of species extinction risk [36]. By weighting clades at every level by the inverse of their range sizes [37], PD calculations can be adjusted to give range-restricted branches extra weight, deriving a synthetic prioritization metric called phylogenetic endemism (PE [38]) that underpins our analysis. (We use the term 'endemism' in reference to a continuous measure of the inverse of range size, not absolute restriction to a particular area as it has traditionally been used.) Geographies with high concentrations of PE represent opportunities to efficiently conserve large fractions of the ranges of many vulnerable taxa while protecting relatively small areas of land.

This efficiency is critical in optimizing biodiversity conservation given limited resources. As redundancies among geographical areas with similar taxonomic composition make the value of preserving any given site dependent on which other sites are also preserved, reserve planning efforts typically use optimization algorithms to identify sets of priority sites with complementary taxa [39,40]. We used forward stepwise selection, an algorithm similar to the reverse stepwise method implemented in the widely used Zonation software [39]. The algorithm generates a nested set of conservation priorities that considers complementarity while identifying immediate priorities and discounting land protection choices more heavily the farther into an uncertain future they would need to be made. Highly ranked sites are those that are (1) currently poorly protected and (2) contain concentrations of taxa that are (a) small-ranged, (b) poorly protected across their ranges and/or (c) high in whatever evolutionary attribute is represented by the phylogenetic or non-phylogenetic metric that was used.

2. Material and methods

(i) Species distributions

Our spatial analysis is based on the California native vascular plant species occurrence dataset described in Baldwin *et al.* [34], comprising more than 1.2 million quality-controlled occurrence records from herbarium specimens and available in an online repository [41]. The 5221 species in the dataset represent nearly every described native plant species recognized in California as of 2015, with an average of 176 occurrence records per species.

Thornhill *et al.* [35] detail the methods we used to fit distribution models for all 5221 species. In brief, modelled ranges incorporated Maxent niche models [42] representing climatic suitability, as well as distances to observed occurrence records representing non-climatic range constraints such as dispersal limitation, edaphic specialization and source–sink dynamics. Here, we reduced the width of the Gaussian distance kernel standard deviation from 50 to 25 km, a trade-off that may result in underpredicting the ranges of some widespread species but limits overpredicting the distributions of highly range-restricted taxa, which are the primary concern for conservation planning and exert greater leverage in our analysis.

The model outputs are suitability scores for each species ranging from 0 to 1 for every 810 m grid cell, with high values for grid cells that are both environmentally suitable and geographically close to recorded occurrences. We use these continuous values throughout the subsequent analysis, avoiding the arbitrary and information-degrading step of thresholding the predicted suitabilities into binary presence–absence values. Because relative occurrence probabilities from species distribution models are positively correlated with species abundances [43], conservation prioritizations based on these values can be viewed as giving higher weight to either locations where species are more likely to occur or where population sizes may be larger.

(ii) Landscape intactness

California's landscapes have been highly impacted by human activities, with less than 25% of primary vegetation in the California Floristic Province remaining intact [44]; we incorporated habitat integrity into distribution models to account for the lower likelihood that species persist in urbanized and agricultural landscapes. (Note that while this approach emphasizes conservation of pristine habitats, there are also critical roles for conservation in urban and working landscapes that are beyond the scope of this study.) We used a California landscape intactness dataset created for conservation planning applications by Degagne *et al.* [45] that incorporates urbanization, agriculture, pollution, roads, resource extraction, fragmentation and invasive species, among other factors. This index has values ranging from −1 to 1 for 1 km grid cells, which we rescaled from 0 to 1 and converted to our 810 m modelling grid (figure 2*a*) using nearest-neighbour resampling. Modelled species presence values in each grid cell were multiplied by these scalars to reflect expected reductions in occurrence and abundance in heavily impacted landscapes; this was done *post hoc* rather than by incorporating intactness as a predictor in the distribution models, because many of the herbarium records predate recent anthropogenic land-use changes. Species richness can be estimated by summing these final suitabilities across all species in a cell (figure 2*d*).

After incorporating landscape intactness, we upscaled model outputs from 810 m to 15 km using spatial averaging of the fine-scale suitability values. All subsequent analysis was performed at the 15 km grid scale, a resolution chosen to emphasize landscape-scale conservation opportunities and allow comparability with prior studies [34,35].

(iii) Phylogeny

Thornhill *et al.* [35] created a phylogenetic dataset [46] representing every California plant species with DNA sequences from GenBank as well as newly generated from herbarium specimens or fresh leaf tissue. In the absence of DNA sequence data for every species, they were grouped into 1083 monophyletic 'operational taxonomic units' (OTUs) based on published molecular

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

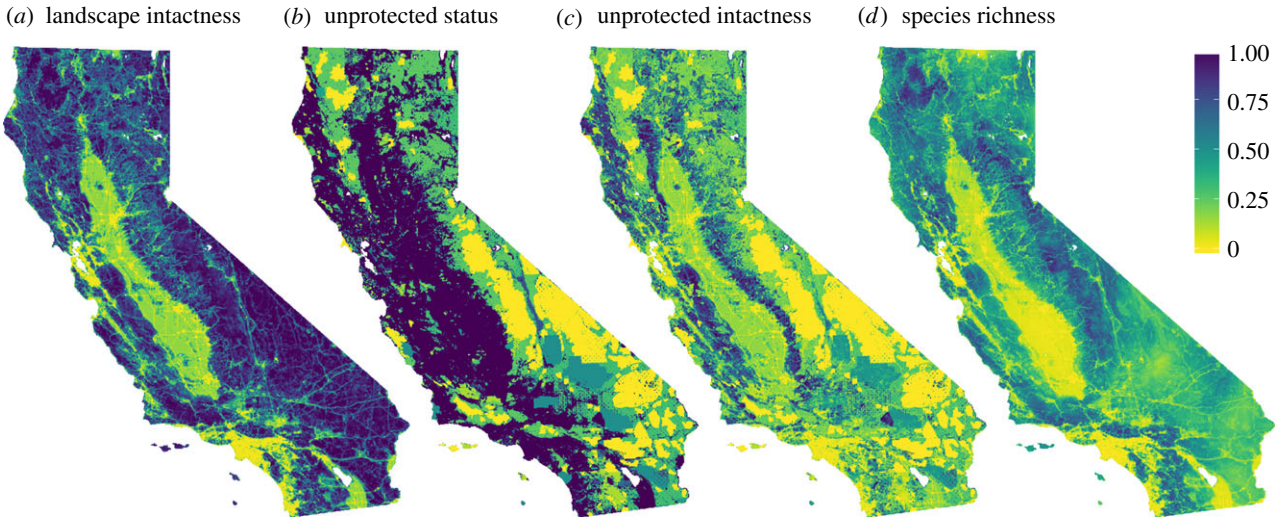

Figure 2. Factors contributing to baseline landscape conservation opportunity, mapped at 1 km grain size for the state of California. (*a*) Landscape intactness [45], with scores ranging from 0 where native ecosystems have been entirely destroyed to 1 where human impacts are minimal. (*b*) Land unprotected status, one minus current protection score, ranging from 0 for cells composed entirely of highly protected public lands to 1 for cells composed entirely of unprotected private lands. (*c*) Unprotected intactness, the product of *a* and *b*, with high scores representing high potential for conservation gains assuming biodiversity is evenly distributed. (*d*) Plant species richness, modelled using herbarium records, climate data and landscape intactness, rescaled from 0 to 1. (Online version in colour.)

studies. A phylogeny of the OTUs was then constructed from nine genetic markers using RAxML (electronic supplementary material, figure S2).

We used species-level distributions to calculate distribution maps for every lineage in this phylogenetic tree, including OTUs and all parent groups. This was done following established methods for phylogenetic aggregation of occurrence probabilities [47] using the formula $p_{ij} = 1 - \prod_{m=1}^{M} (1 - p_{im})$, where p_{ij} is the presence weight of clade j in cell i and p_{im} is the presence weight of member species m in cell i.

(iv) Land conservation status

While most prior phylodiversity conservation studies have treated land conservation status as a binary variable, levels of biodiversity protection in fact vary on a gradient from unprotected private land to highly protected public land, with many intermediate designations offering partial protection. We therefore developed a continuous score varying from 0 to 1, which aims to represent how the long-term security of resident biodiversity varies by land management class. These conservation status scores were assigned at the parcel level using spatial datasets on protected areas [48], conservation easements [49] and wilderness areas [50], as well as military and tribal lands. Parcels were grouped into categories based on ownership and management, ranked independently by the authors using a 0–4 scale, averaged across author rankings and rescaled from 0 to 1. Parcel scores were then averaged within 15 km grid cells, weighted by parcel area (figure 2*b*).

(v) Branch protection

We calculate the proportion of each branch's geographical range that is protected using the equation $C_j = \sum_{i=1}^{n} (p_{ij} s_i / R_j)$, where 'branch' here refers to a lineage segment (or simply a terminal taxon for the non-phylogenetic measures). C_j is the proportion of the range of branch j that is protected, p_{ij} is the occurrence probability of branch j in grid cell i, s_i is the conservation status score of grid cell i and R_j is the California range size of branch j, defined as the sum of p_{ij} across all cells.

The protection level for each lineage is then converted into a conservation 'benefit' value that assigns higher importance to protecting populations of poorly protected lineages. This is done using the equation $\mathcal{B}_j(C, \lambda) = (1 - (1 - C_j)^{2^\lambda})^{1/2^\lambda}$, where λ is a free parameter defining the rate of diminishing returns. The use of this nonlinear benefit function is analogous to the 'additive benefit function' in the Zonation software [39]. Electronic supplementary material, figure S3 shows how λ affects the function shape; except as otherwise noted we use $\lambda = 1$.

(vi) Conservation optimization

We used forward stepwise selection to identify future conservation priorities. We also tried backward stepwise elimination and it yielded nearly identical results. Our algorithm begins with the current conservation status landscape (one minus the scores shown in figure 2*b*), and calculates the hypothetical marginal value of increasing each cell's conservation status from its current level to 1 (full protection, e.g. a national park), using the following equation:

$$MV_i = \sum_{j=1}^{n} v_j \left(\mathcal{B}\left(C_j + \frac{p_{ij}(1 - s_i)}{R_j}, \lambda \right) - \mathcal{B}(C_j, \lambda) \right), \tag{2.1}$$

where v_j is the length of branch j (or 1, in the case of non-phylogenetic measures), and all other variables are as described above. This change is implemented for the cell with the highest marginal value, and the process is repeated, recalculating the conservation status of every lineage and every grid cell at each step, until all cells are fully protected.

We generated conservation prioritizations using five variants of the input dataset: species (no phylogeny), OTUs (no phylogeny), OTUs on the cladogram, OTUs on the chronogram and OTUs on the phylogram. We also generated alternative prioritizations for three types of sensitivity analysis: a 'California-restricted' comparison using a subset of the species dataset containing only species whose ranges are entirely restricted to California (to test for edge effects resulting from the study area boundary), a 'blank slate' comparison using a version of the dataset in which both landscape intactness and the current distribution of protected lands were ignored (to gain insight into how these factors constrained the differences among methodological choices), and finally a 'jackknife' approach comparing four variants of the marginal value function each ignoring variation in taxon range size, taxon range protection status, total site diversity or phylogeny (to assess the relative importance of these factors in shaping priorities).

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

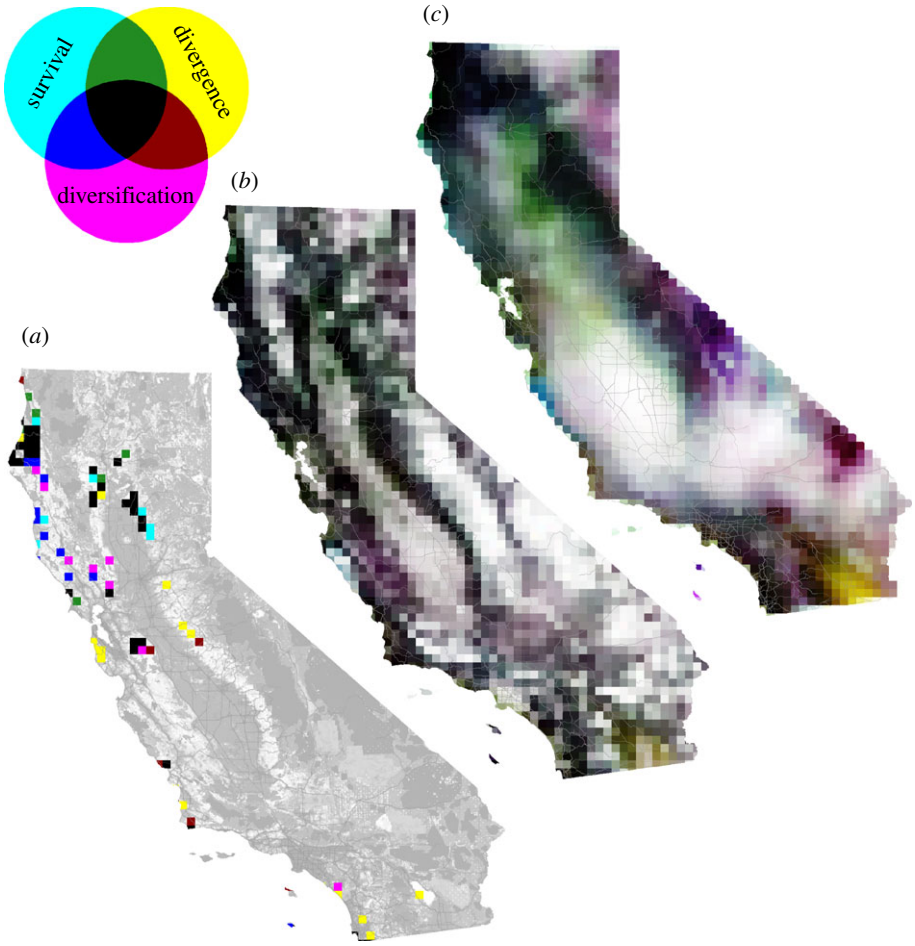

Figure 3. Variation in spatial conservation priorities among the three phylogeny-based metrics. Colour indicates the scheme or schemes under which cells were selected as high priorities, with black cells ranked highly under all schemes, white/grey cells ranked low under all schemes and coloured cells ranked highly under a subset of schemes—see figure 1 and electronic supplementary material, figure S1 for details. (*a*) The top 50 highest priority sites for the three metrics; the background shows unprotected intactness as in figure 2*c*, with darker grey for unlikely priorities that are already protected or degraded. (*b*) Continuous threshold-free ranks for all sites for the metrics. (*c*) Same as *b*, but for the 'blank slate' analysis that ignores current protected land and landscape intactness and uses $\lambda = 0$. *b* and *c* are visualized using a continuous three-dimensional version of the discrete colour palette shown in the legend, with intermediate colour shades indicating intermediate priority. Major roads are shown for reference.

While these conservation prioritizations do not involve directly calculating spatial phylodiversity statistics such as PD or PE, we also calculated several such metrics for reference (electronic supplementary material, table S1). All analyses were done in R [51].

3. Results

(i) Spatial patterns

Conservation priorities differed among methods, though rankings were correlated overall, with shared priorities concentrated in pockets of the immediate coast, Coast Ranges and Sierra Nevada foothills. Some sites differed in ranking by nearly two orders of magnitude across metrics (electronic supplementary material, figure S6). We interpret the five metrics in two groups of three: a chronogram-phylogram-cladogram comparison of the three facets of phylodiversity, and a species-OTU-chronogram comparison of richness versus phylodiversity.

Sites on the North Coast near Eureka, in the Diablo Range east of San Jose, and in the northern Sierra Nevada foothills near Chico were among the fifty highest ranked priorities for all three phylodiversity facets (figure 3*a*)—but each facet also identified top priorities that were unique, with evolutionary time-focused priorities (cyan, blue, green) concentrated in northern California, divergence-focused priorities (yellow, green, red) concentrated in the Central Coast and central Sierra Nevada foothills, and diversification-focused priorities (magenta, red, blue) found mostly near the coast with higher prevalence in the south.

Results for the full rankings rather than just the top priorities (figure 3*b*), and for the blank slate analysis (figure 3*c*) move successively farther away from representing conservation priorities and toward representing the evolutionary characteristics of the underlying endemic biodiversity. These results indicate, for example, that the Sacramento Valley is home to concentrations of taxa from lineages with greater survival and divergence than diversification (green), areas along the eastern California border are home to taxa from young but species-rich clades (magenta, red) and the Sonoran Desert is home to taxa that have undergone particularly rapid rates of divergence (yellow). Endemism weighting means these patterns are driven mostly by lineages with relatively small California ranges.

Priority rankings based on the species, OTU and chronogram metrics differed even more starkly in some areas than did the three phylogeny-based metrics (electronic supplementary material, figure S5). Species-based priorities were concentrated along the Central Coast of California,

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

while phylogeny-based priorities were more concentrated in coastal and interior reaches of far northern California. The OTU metric, an intermediate method that shares attributes of both the species and phylogenetic datasets, yielded very few priorities that were not shared with one of these other datasets; OTU priorities overlapped more with the chronogram priorities than with the species priorities. Differences notwithstanding, all three methods also identified shared priorities as mentioned above.

(ii) Reserve efficiency

The performance efficiency of the stepwise reserve optimization algorithm can be visualized as a curve tracing the proportion of total biodiversity benefit protected across the region as an additional site is protected at each step of the algorithm. As designed, poorly protected taxa accumulated additional protection more rapidly than well-protected taxa (electronic supplementary material, figure S7a). Incorporating phylogenetic information in the prioritization improved the efficiency of the reserve network in protecting evolutionary diversity—for example, using the chronogram to select the 25 highest-priority sites would protect 20% more of the currently unprotected chronogram PD than using species to select sites (electronic supplementary material, figure S7b).

(iii) Drivers of conservation value

Underlying the final conservation prioritizations are strong spatial patterns in different characteristics of biodiversity (electronic supplementary material, figure S4 and table S1). The jackknife analysis quantified the importance of four major variables (phylogeny, range protection status, endemism and diversity) in shaping spatial patterns of conservation priority, and found that they influenced chronogram-based rankings in that order of increasing importance (figure 4*a*); an alternative that ignored all four variables, 'unprotected intactness' (a combination of intactness and conservation status as in figure 2*c*) explained half of the spatial variation in the final chronogram priorities. Conservation priorities were not strictly limited to sites with low current protection—the top 100 priorities included sites with current protection scores as high as 0.6 out of 1 (electronic supplementary material, figure S10).

These four drivers also operate on taxa within individual grid cells to help determine a cell's marginal value. For the top-ranked cell for the chronogram dataset, located in northern coastal California near Eureka, 73% of the cell's value comes from just 10 lineages (figure 4*b*). These select taxa are most exceptional in their endemism scores, which decline more rapidly than branch length, presence probability or conservation status outside the top-ranked taxa. Owing to the weight given to endemism in our optimization, threatened range-restricted taxa such as *Aphanisma blitoides*, *Bergerocactus emoryi* and *Dicranostegia oppositifolia* contributed strongly to priority cells for all facets. Examples of notable taxa contributing disproportionately to priorities for individual facets included *Chamaecyparis lawsoniana* and *Sequoia sempervirens* (for the chronogram), *Stemodia durantifolia* (for the phylogram) and *Constancea nevinii* and *Lyonothamnus floribundus* (for the cladogram).

The California-restricted analysis yielded conservation priorities that were mostly similar to the version with all species, but without the priority areas on the extreme

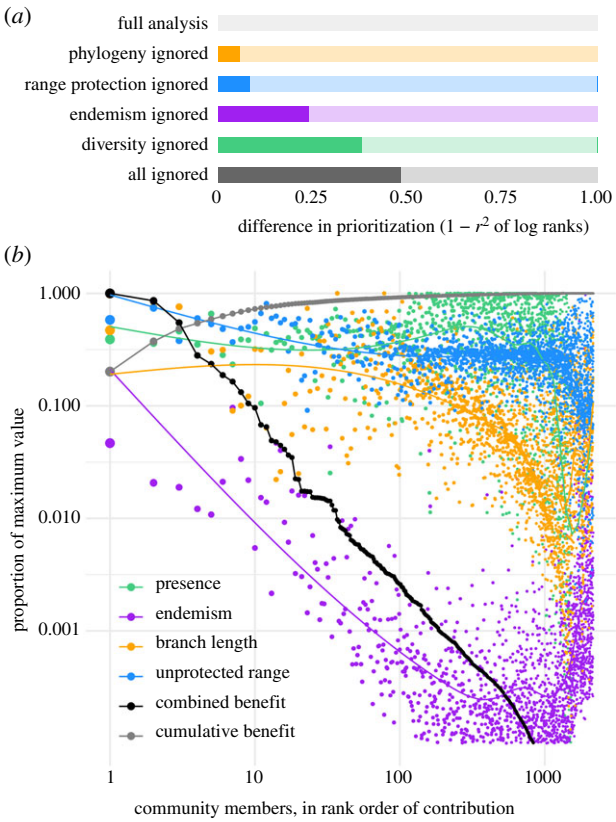

Figure 4. The importance of four variables—presence probability (or its sum, diversity), endemism, taxon range-wide protection status and phylogeny—in driving conservation priorities for the chronogram-based analysis. (*a*) Results of jackknife sensitivity analysis comparing the final region-wide rankings to versions ignoring different components, and to ranked 'unprotected intactness' that ignores all biodiversity. (*b*) Relative contributions of lineages and their attributes to the conservation value of the top-ranked grid cell. (Online version in colour.)

northern (and to a lesser extent southern) end of the California coast (electronic supplementary material, figure S8). The blank slate analysis yielded priorities that differed far more among the five phylogenies than in the full formulation constrained by existing preserves and landscape intactness (figures 3*b*,*c*; electronic supplementary material, figures S5*b*,*c* and S9).

In addition to the final results presented here, the underlying data including modelled ranges of every taxon, spatial biodiversity patterns (electronic supplementary material, figure S4), and phylogenies (electronic supplementary material, figure S2) are incorporated into an interactive web application that allows users to explore how patterns of occurrence, endemism, branch length and conservation status vary across geography, within local communities, and across the phylogenetic tree. The tool is available online at http://ucjeps.berkeley.edu/phylodiversity/.

4. Discussion

This study represents one of the most comprehensive biodiversity gap analyses, to date, for the native flora of California, incorporating extensive herbarium-based data on both the spatial distributions and evolutionary relationships of its full native vascular plant flora of more than 5000 species, a new high-resolution index of current land

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

conservation status, and detailed data on landscape intactness across the region. Our results identify priority areas for increased protection of evolutionarily distinct and currently vulnerable taxa in one of the world's biodiversity hotspots, findings that can serve as a guide for future conservation applications.

One novel aspect of our prioritization analysis is the use of a continuous scoring system, rather than a binary variable, to model current land conservation status. This approach has two advantages. First, it increases the realism of taxon protection scores, reflecting, for example, that populations in national parks are likely more secure than those in national forests due to land management differences. Second, it recognizes a role in conservation strategies for increased protection of sites that are currently moderately protected. For example, an important fraction of high priority sites in our analysis already have intermediate protection, but were still prioritized for increases due to their exceptional biodiversity value (electronic supplementary material, figure S10). These priorities are thus relevant not only for establishing new protected areas, but just as importantly for allocating management zones within moderately protected areas such as national forests that must balance multiple uses including recreation, resource extraction and biodiversity conservation.

It is important to consider potential sampling bias in the herbarium data used in our analysis [52]. While California has been extensively botanized relative to most other regions [31], the number of herbarium specimens varied across species and grid cells in our dataset [34]. Two factors help to mitigate the effect of this nonrandom sampling on our results. First, species distribution models, in general, can help by projecting occurrences into poorly sampled areas [53], and in our models we further explicitly account for sampling bias in our selection of background training data [35]. Second, it is known that herbarium collections are biased toward documenting novel and endemic diversity [34,54], which makes them a far more complete representation of local floristic composition than a random sample would be. Still, it remains possible that our results omit undocumented biodiversity hotspots in remote areas.

Phylogenetic diversity measures have emerged as an increasingly recommended approach to biodiversity quantification for conservation prioritization [8,38,55], but there have been calls for more clarity about the conservation relevance of different kinds of phylodiversity metrics [15]. While phylodiversity is often considered a single dimension of biodiversity, we assessed the conservation relevance of conceptual and empirical differences among three dimensions of phylodiversity, as well as comparing them to two non-phylogenetic biodiversity metrics. The five metrics identified largely similar regions of California as high conservation priorities for native plants. These portions of the immediate coast, Coast Ranges and Sierra Nevada foothills can be seen as good candidates for future conservation regardless of which aspects of biodiversity are most valued.

Our conservation priorities were driven primarily by factors shaping patterns of native plant occurrence (taxonomic richness and landscape intactness) and vulnerability (current land protection and taxon range sizes). These variables narrow the range of reasonably attractive conservation sites in the same way for all five biodiversity metrics, and so unsurprisingly, differences among the five metrics emerged as second-order patterns, with greater differences between traditional richness and phylogenetic approaches than among the three phylogenetic approaches. Endemism played a particularly strong role in defining priorities shared among diversity facets, as weighting taxa by the inverse of range size places an extreme emphasis on prioritizing the most highly restricted lineages; for studies that might reasonably choose a milder range-weighting function, priorities would likely diverge more among metrics. Landscape intactness and current protection status also shaped shared priorities—when we removed those variables, prioritization differences among diversity facets increased roughly fivefold (figure 3; electronic supplementary material, figures S5 and S9). These observations caution that the choice among biodiversity metrics may make an even greater difference in other regions than it did in California, where most landscapes are already either degraded or protected.

Subsets of the five diversity metrics address different aspects of how evolution is quantified in biodiversity studies. Comparing the species, OTU and chronogram metrics, which represent progressive aggregations of species into larger clades and then nested aggregation of those clades on a phylogeny, highlights the importance of taxonomic resolution and evolutionary relatedness in conservation prioritization. It has been argued that even if phylogenetic diversity is conceptually preferable to species richness for conservation purposes, the latter serves as an adequate surrogate as the two metrics are correlated [4]; while we indeed found positive correlations, the top 25 sites selected using chronogram-based PD contained 20% more PD than did the top 25 sites selected using species richness (electronic supplementary material, figure S7b), suggesting that species richness is an inefficient surrogate for PD. The schemes also differed in the broad geographical areas they identified as top priorities, with northern California prized more highly for phylodiversity and central California for species diversity (electronic supplementary material, figure S5). This implies that species and OTUs in central California are more closely related than those in the north, a pattern that is corroborated by the relative values for mean phylogenetic diversity in these two regions (electronic supplementary material, figure S4) and consistent with prior studies [35]. Such differences between taxonomic richness and chronogram diversity are indeed expected for biodiverse regions with imbalanced phylogenetic trees and high environmental heterogeneity [56].

Comparisons among the chronogram-, phylogram- and cladogram-based results address how valuing evolutionary lineages based respectively on their cumulative time since divergence, ancestral mutation rates or net cladogenesis affects conservation priorities. While there was more overlap among these three rankings than among the three discussed above, differences between them highlight key characteristics of certain priority sites. For example, comparisons between cladogram and chronogram metrics have been used to identify areas of neo- and paleoendemism [55], concentrations of recently or anciently diverged taxa that correspond, respectively, to the magenta–red and cyan–green cells in figure 3. From a conservation perspective, centres of neoendemism represent opportunities to conserve potentially ongoing evolutionary radiations that could drive future diversification [28,30], while areas of paleoendemism represent opportunities to conserve collections of lineages that have persisted independently across exceptionally long stretches of time.

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

The latter are found predominantly in areas of low historic climate velocity that may have played a role in reducing the extinction rates of ancient lineages [23,25,57], suggesting these landscapes could also be important to conserve as future climate refugia as well.

Contrasts between phylogram- and chronogram-based PD indicate differential rates of historic genetic change. Sites harbouring assemblages that rank highly on the phylogram but not the chronogram, or vice versa, could be termed centres of evolutionary dynamism and stability, respectively. We identified priority areas of dynamism in the mediterranean and desert regions of southern and eastern California (yellow–red cells, figure 3), and priority areas of stability along the central and northern coast and the northern interior of the state (cyan–blue cells). These phylogenetic patterns correlate with a history of paleoclimatic dynamism versus stability of these regions [57,58], hinting at a connection between environmental change and rates of character divergence [59], independent of previously published connections with rates of speciation and extinction [57].

Phylogenetic diversity is a potentially useful conservation prioritization tool for two distinct but not mutually exclusive goals: preserving the present diversity and function of the biota, and preserving its future evolutionary potential and ecological resilience. Applications should evaluate the three branch-length metrics discussed here against each of these two goals, both for theoretical relevance as discussed in the introduction and for empirical confidence based on uncertainty in the phylogeny used.

In the case of our empirical analysis, there is phylogenetic uncertainty associated with each of the three branch-length metrics. For the cladogram, a primary concern is missing taxa—while our phylogeny represents almost every described native plant species in California, it includes only California species, and many terminals represent multiple species. Both aspects will cause underestimation of cladogram PD, which could be biased if missing taxa are non-randomly distributed in space or on the phylogeny. Assuming phylogenetic clustering at broad spatial scales due to niche conservatism or dispersal limitation, the largest bias from excluding non-California species would be expected near the state borders; the fact that most areas with high diversification relative to survival time occurred near the state border runs counter to this expectation and suggests this issue may not be unduly influencing the results. Uncertainty in cladogram PD will decline as the extent and resolution of phylogenies continue to rapidly improve in future studies.

For the phylogram, a relatively small number of lineages have notably long branches indicating rapid molecular change. While there is inevitable uncertainty in inferred molecular branch lengths and their connection to functional traits, the qualitative observation that the lineages with the longest branches include many parasitic or carnivorous plants (e.g. members of the genera *Arceuthobium*, *Cuscuta*, *Drosera*, *Pilostyles* and *Utricularia*) representing major changes in ecological function strengthens the interpretation that these branches correspond to large phenotypic changes.

Chronograms with time-based branch lengths are perhaps the most commonly used phylogenies in PD analyses [16]. Because the chronogram has dating constraints on tips and on numerous internal nodes and has branch lengths minimally affected by missing taxa, our PD uncertainty for persistence time is arguably lower than for the other branch length metrics. This does not mean, however, that chronogram PD is the best predictor of functional or genomic diversity. Indeed, though survival time does represent the potential for character divergence [11], an important recent study found that in fully a third of cases, chronogram PD was a worse surrogate for functional diversity than was species richness [60], though it did not evaluate alternative branch metrics. Further research should focus on whether chronogram or phylogram PD best predicts standing diversity in genes and functional traits [16,17].

Further research should also address empirical uncertainties surrounding the use of PD as a predictor for future macroevolutionary and ecological resilience, beyond its use as a surrogate for standing diversity. Studies have found that species extinction risk is phylogenetically clustered [61,62] and that range size is negatively correlated with evolutionary distinctiveness [37], suggesting that macroevolutionary history can be a key predictor of vulnerability. But while the three facets of phylogenetic diversity have theoretical relevance to future conservation, it remains largely unknown how the diversification, divergence and independent survival histories of individual lineages or species assemblages will predict their responses to Anthropocene environmental change.

Data accessibility. The herbarium specimen geolocations and phylogenetic trees used as inputs to our analysis are available online [41,46]. R scripts implementing our reserve selection algorithm, a raster file containing the current land conservation scores developed for this study, and a CSV file of our final prioritization rankings are available online at https://github.com/matthewkling/facets-of-phylodiversity. Spatial data including the modelled geographical ranges of every species and larger clade are accessible through the web application associated with this paper at http://ucjeps.berkeley.edu/phylodiversity/.

Authors' contributions. M.M.K, B.D.M. and D.D.A. conceived the study. A.H.T. and B.G.B. prepared the herbarium data and constructed the phylogeny. M.M.K. conducted the spatial analyses and drafted the manuscript. All authors contributed to manuscript revision and gave final approval for publication.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by National Science Foundation-grant no. DEB-1354552 to D.D.A, B.G.B. and B.D.M. M.M.K. was supported by a National Science Foundation Graduate Research Fellowship.

Acknowledgements. We are grateful for substantive comments on the manuscript from two anonymous reviewers.

References

1. Barnosky AD *et al.* 2011 Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57. (doi:10.1038/nature09678)
2. Bottrill MC *et al.* 2008 Is conservation triage just smart decision making? *Trends Ecol. Evol.* **23**, 649–654. (doi:10.1016/j.tree.2008.07.007)
3. Margules CR, Pressey RL. 2000 Systematic conservation planning. *Nature* **405**, 243–253. (doi:10.1038/35012251)

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024

4. Brooks TM, Mittermeier RA, da Fonseca GAB, Gerlach J, Hoffmann M, Lamoreux JF, Mittermeier CG, Pilgrim JD, Rodrigues ASL. 2006 Global biodiversity conservation priorities. *Science* **313**, 58–61. (doi:10.1126/science.1127609)
5. Purvis A, Hector A. 2000 Getting the measure of biodiversity. *Nature* **405**, 212–219. (doi:10.1038/35012221)
6. Jarzyna MA, Jetz W. 2016 Detecting the multiple facets of biodiversity. *Trends Ecol. Evol.* **31**, 527–538. (doi:10.1016/j.tree.2016.04.002)
7. Pollock LJ, Thuiller W, Jetz W. 2017 Large conservation gains possible for global biodiversity facets. *Nature* **546**, 141–144. (doi:10.1038/nature22368)
8. Faith DP. 1992 Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10. (doi:10.1016/0006-3207(92)91201-3)
9. Isaac NJ, Mallet J, Mace GM. 2004 Taxonomic inflation: its influence on macroecology and conservation. *Trends Ecol. Evol.* **19**, 464–469. (doi:10.1016/j.tree.2004.06.004)
10. Srivastava DS, Cadotte MW, MacDonald AAM, Marushia RG, Mirotchnick N. 2012 Phylogenetic diversity and the functioning of ecosystems. *Ecol. Lett.* **15**, 637–648. (doi:10.1111/j.1461-0248.2012.01795.x)
11. Mooers A, Heard SB, Chrostowski E. 2005 Evolutionary heritage as a metric for conservation. In *Phylogeny and conservation* (eds A Purvis, JL Gittleman, T Brooks), pp. 120–138. Cambridge, UK: Cambridge University Press.
12. Cadotte MW, Davies TJ. 2010 Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Divers. Distrib.* **16**, 376–385. (doi:10.1111/j.1472-4642.2010.00650.x)
13. Diniz-Filho JAF. 2004 Phylogenetic diversity and conservation priorities under distinct models of phenotypic evolution. *Conserv. Biol.* **18**, 698–704. (doi:10.1111/j.1523-1739.2004.00260.x)
14. Isaac NJ, Turvey ST, Collen B, Waterman C, Baillie JE. 2007 Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* **2**, e296. (doi:10.1371/journal.pone.0000296)
15. Winter M, Devictor V, Schweiger O. 2013 Phylogenetic diversity and nature conservation: where are we? *Trends Ecol. Evol.* **28**, 199–204. (doi:10.1016/j.tree.2012.10.015)
16. Elliott MJ, Knerr NJ, Schmidt-Lebuhn AN. 2018 Choice between phylogram and chronogram can have a dramatic impact on the location of phylogenetic diversity hotspots. *J. Biog.* **45**, 2190–2201. (doi:10.1111/jbi.13399)
17. Mooers AO, Redding DW. 2009 Where the rare species are. *Mol. Ecol.* **18**, 3955–3957. (doi:10.1111/j.1365-294X.2009.04312.x)
18. Bromham L, Penny D. 2003 The modern molecular clock. *Nat. Rev. Genet.* **4**, 216–224. (doi:10.1038/nrg1020)
19. Welch JJ, Bininda-Emonds OR, Bromham L. 2008 Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol. Biol.* **8**, 53. (doi:10.1186/1471-2148-8-53)
20. Magallón S, Hilu KW, Quandt D. 2013 Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am. J. Bot.* **100**, 556–573. (doi:10.3732/ajb.1200416)
21. Wysocki WP, Clark LG, Attigala L, Ruiz-Sanchez E, Duvall MR. 2015 Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. *BMC Evol. Biol.* **15**, 50. (doi:10.1186/s12862-015-0321-5)
22. FitzJohn RG. 2010 Quantitative traits and diversification. *Syst. Biol.* **59**, 619–633. (doi:10.1093/sysbio/syq053)
23. Sandel B, Arge L, Dalsgaard B, Davies RG, Gaston KJ, Sutherland WJ, Svenning J-C. 2011 The influence of Late Quaternary climate-change velocity on species endemism. *Science* **334**, 660–664. (doi:10.1126/science.1210173)
24. Harrison S, Noss R. 2017 Endemism hotspots are linked to stable climatic refugia. *Ann. Bot.* **119**, 207–214. (doi:10.1093/aob/mcw248)
25. Procheş Ş, Ramdhani S, Perera SJ, Ali JR, Gairola S. 2015 Global hotspots in the present-day distribution of ancient animal and plant lineages. *Sci. Rep.* **5**, 15457. (doi:10.1038/srep15457)
26. Vane-Wright RI, Humphries CJ, Williams PH. 1991 What to protect? Systematics and the agony of choice. *Biol. Conserv.* **55**, 235–254. (doi:10.1016/0006-3207(91)90030-D)
27. Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igić B. 2010 Species selection maintains self-incompatibility. *Science* **330**, 493–495. (doi:10.1126/science.1194513)
28. Davis EB, Koo MS, Conroy C, Patton JL, Moritz C. 2008 The California Hotspots Project: identifying regions of rapid diversification of mammals. *Mol. Ecol.* **17**, 120–138. (doi:10.1111/j.1365-294x.2007.03469.x)
29. Bowen BW. 1999 Preserving genes, species, or ecosystems? Healing the fractured foundations of conservation policy. *Mol. Ecol.* **8**, S5–S10. (doi:10.1046/j.1365-294X.1999.00798.x)
30. Kraft NJ, Baldwin BG, Ackerly DD. 2010 Range size, taxon age and hotspots of neoendemism in the California flora. *Divers. Distrib.* **16**, 403–413. (doi:10.1111/j.1472-4642.2010.00640.x)
31. Meyer C, Weigelt P, Kreft H. 2016 Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* **19**, 992–1006. (doi:10.1111/ele.12624)
32. Jepson eFlora. 2018 See http://ucjeps.berkeley.edu/eflora/.
33. Consortium of California Herbaria. 2018 See ucjeps.berkeley.edu/consortium/.
34. Baldwin BG, Thornhill AH, Freyman WA, Ackerly DD, Kling MM, Morueta-Holme N, Mishler BD. 2017 Species richness and endemism in the native flora of California. *Am. J. Bot.* **104**, 487–501. (doi:10.3732/ajb.1600326)
35. Thornhill AH, Baldwin BG, Freyman WA, Nosratinia S, Kling MM, Morueta-Holme N, Madsen TP, Ackerly DD, Mishler BD. 2017 Spatial phylogenetics of the native California flora. *BMC Biol.* **15**, 96. (doi:10.1186/s12915-017-0435-x)
36. Purvis A, Gittleman JL, Cowlishaw G, Mace GM. 2000 Predicting extinction risk in declining species. *Proc. R. Soc. Lond. B* **267**, 1947–1952. (doi:10.1098/rspb.2000.1234)
37. Tucker CM, Cadotte MW, Davies TJ, Rebelo TG. 2012 Incorporating geographical and evolutionary rarity into conservation prioritization. *Conserv. Biol.* **26**, 593–601. (doi:10.1111/j.1523-1739.2012.01845.x)
38. Rosauer DF, Laffan SW, Crisp MD, Donnellan SC, Cook LG. 2009 Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* **18**, 4061–4072. (doi:10.1111/j.1365-294x.2009.04311.x)
39. Moilanen A, Franco AM, Early RI, Fox R, Wintle B, Thomas CD. 2005 Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proc. R. Soc. B* **272**, 1885–1891. (doi:10.1098/rspb.2005.3164)
40. Ball IR, Possingham HP, Watts M. 2009 Marxan and relatives: software for spatial conservation prioritisation. In *Spatial conservation prioritisation: quantitative methods and computational tools* (eds A Moilanen, KA Wilson, H Possingham), pp. 185–195. Oxford, UK: Oxford University Press.
41. Thornhill AH, Baldwin BG, Freyman WA, Nosratinia S, Kling MM, Morueta-Holme N, Madsen T, Ackerly D, Mishler B. 2017 *Master spatial file for Spatial phylogenetics of the native California flora*. Berkeley, CA: University of California Berkeley. See https://doi.org/10.6078/D1KX0V.
42. Phillips SJ, Dudík M, Schapire RE. 2004 A maximum entropy approach to species distribution modeling. In *Proc. 21st Int. Conf. Machine Learning, 4–8 July 2004*, p. 83. New York, NY: ACM.
43. Weber MM, Stevens RD, Diniz-Filho JAF, Grelle CEV. 2017 Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. *Ecography* **40**, 817–828. (doi:10.1111/ecog.02125)
44. Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. 2000 Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858. (doi:10.1038/35002501)
45. Degagne R, Brice H, Gough M, Sheehan T, Strittholt J. 2016 *Terrestrial landscape intactness (1 km), California*. Conservation Biology Institute. From DataBasin.org: https://databasin.org/datasets/e3ee00e8d94a4de58082fdbc91248a65.
46. Thornhill AH, Baldwin BG, Freyman WA, Nosratinia S, Kling MM, Morueta-Holme N, Madsen T, Ackerly D, Mishler B. 2017 *Sequence matrix and tree files for Spatial phylogenetics of the native California flora*. Berkeley, CA: University of California Berkeley. See https://doi.org/10.6078/D1VD4P.
47. Pollock LJ, Rosauer DF, Thornhill AH, Kujala H, Crisp MD, Miller JT, McCarthy MA. 2015 Phylogenetic diversity meets conservation policy: small areas are

key to preserving eucalypt lineages. *Proc. R. Soc. B* **370**, 20140007. (doi:10.1098/rstb.2014.0007)
48. California Protected Areas Database. 2017 Version 2017a (accessed 29 November 2017). See http://www.calands.org/data.
49. California Conservation Easement Database. 2016 December 2016 release (accessed 29 November 2017). See http://www.calands.org/cced.
50. National Wilderness Preservation System. 2018 (accessed 23 February 2018). See https://www.wilderness.net/nwps/geography.
51. R Core Team. 2013 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
52. Daru BH *et al.* 2018 Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytol.* **217**, 939–955. (doi:10.1111/nph.14855)
53. Newbold T. 2010 Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Prog. Phys. Geogr.* **34**, 3–22. (doi:10.1177/0309133309355630)
54. Steege Ht, Haripersaud PP, Bánki OS, Schieving F. 2011 A model of botanical collectors' behavior in the field: never the same species twice. *Am. J. Bot.* **98**, 31–37. (doi:10.3732/ajb.1000215)
55. Mishler BD, Knerr N, González-Orozco CE, Thornhill AH, Laffan SW, Miller JT. 2014 Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian *Acacia*. *Nat. Commun.* **5**, 4473. (doi:10.1038/ncomms5473)
56. Tucker CM, Cadotte MW. 2013 Unifying measures of biodiversity: understanding when richness and phylogenetic diversity should be congruent. *Divers. Distrib.* **19**, 845–854. (doi:10.1111/ddi.12087)
57. Lancaster LT, Kay KM. 2013 Origin and diversification of the California flora: re-examining classic hypotheses with molecular phylogenies. *Evolution* **67**, 1041–1054. (doi:10.1111/evo.12016)
58. Raven PH, Axelrod DI. 1978 Origin and relationships of the California flora. *Univ. Calif. Publ. Bot.* **72**, 1–134.
59. Mazel F, Wüest RO, Gueguen M, Renaud J, Ficetola GF, Lavergne S, Thuiller W. 2017 The geography of ecological niche evolution in mammals. *Curr. Biol.* **27**, 1369–1374. (doi:10.1016/j.cub.2017.03.046)
60. Mazel F *et al.* 2018 Prioritizing phylogenetic diversity captures functional diversity unreliably. *Nat. Commun.* **9**, 2888. (doi:10.1038/s41467-018-05126-3)
61. Purvis A, Agapow PM, Gittleman JL, Mace GM. 2000 Nonrandom extinction and the loss of evolutionary history. *Science* **288**, 328–330. (doi:10.1126/science.288.5464.328)
62. Thuiller W, Lavergne S, Roquet C, Boulangeat I, Lafourcade B, Araujo MB. 2011 Consequences of climate change on the tree of life in Europe. *Nature* **470**, 531–534. (doi:10.1038/nature09705)

Downloaded from https://royalsocietypublishing.org/ on 09 May 2024