

# Species richness and endemism in the native flora of California<sup>1</sup>

Bruce G. Baldwin<sup>2</sup>, Andrew H. Thornhill, William A. Freyman, David D. Ackerly, Matthew M. Kling, Naia Morueta-Holme, and Brent D. Mishler

**PREMISE OF THE STUDY:** California's vascular flora is the most diverse and threatened in temperate North America. Previous studies of spatial patterns of Californian plant diversity have been limited by traditional metrics, non-uniform geographic units, and distributional data derived from floristic descriptions for only a subset of species.

**METHODS:** We revisited patterns of sampling intensity, species richness, and relative endemism in California based on equal-area spatial units, the full vascular flora, and specimen-based distributional data. We estimated richness, weighted endemism (inverse range-weighting of species), and corrected weighted endemism (weighted endemism corrected for richness), and performed a randomization test for significantly high endemism.

**KEY RESULTS:** Possible biases in herbarium data do not obscure patterns of high richness and endemism at the spatial resolution studied. High species richness was sometimes associated with significantly high endemism (e.g., Klamath Ranges) but often not. In Stebbins and Major's (1965) main endemism hotspot, Southwestern California, species richness is high across much of the Peninsular and Transverse ranges but significantly high endemism is mostly localized to the Santa Rosa and San Bernardino mountains. In contrast, species richness is low in the Channel Islands, where endemism is significantly high, as also found for much of the Death Valley region.

**CONCLUSIONS:** Measures of taxonomic richness, even with greater weighting of range-restricted taxa, are insufficient for identifying areas of significantly high endemism that warrant conservation attention. Differences between our findings and those in previous studies appear to mostly reflect the source and scale of distributional data, and recent analytical refinements.

**KEY WORDS** bioregions; California; endemism; flora; floristic diversity; floristic hotspots; herbarium data; sampling theory; species richness

Spatial patterns of floristic richness and endemism are critical for understanding the evolution and assembly of native plant diversity on a regional scale (e.g., Linder and Verboom, 2015; Nagalingum et al., 2015; Schmidt-Lebuhn et al., 2015; Thornhill et al., 2016), and for resolving areas of special value for conservation planning (González-Orozco et al., 2015), especially in light of ongoing anthropogenic climate change (Loarie et al., 2008). The recent, rapid growth of herbarium databases with geo-referenced collection localities (e.g., Markos et al., 2016) and accompanying development of refined spatial-diversity metrics (Crisp et al., 2001; Laffan and Crisp, 2003, 2016; Rosauer et al., 2009; Mishler et al., 2014) are

increasingly allowing for a more nuanced assessment of floristic hotspots than was possible earlier, to the benefit of conservation resource prioritization. In Australia, for example, application of such approaches has been recognized as an important component of decision-making about designation of national parks and other reserves, to ensure that land preservation encompasses as much otherwise unprotected biological diversity as possible (e.g., Laity et al., 2015; Pollock et al., 2015).

Advances in quantitative approaches to spatial biodiversity analysis have enhanced the value of such studies for examining regional patterns of richness and endemism (e.g., Crisp et al., 2001; Laffan and Crisp, 2003; Rosauer et al., 2009; González-Orozco et al., 2014; Mishler et al., 2014). For example, range restriction, which has been traditionally captured floristically by examining *absolute endemism* (i.e., complete restriction to a particular setting), can be explored with greater resolution by examining *relative endemism* (i.e., inverse weighting of taxa or lineages by their geographic range

<sup>1</sup> Manuscript received 6 September 2016; revision accepted 23 February 2017.

University and Jepson Herbaria and Department of Integrative Biology, 1001 Valley Life Sciences Building #2465, University of California, Berkeley, California 94720-2465 USA

<sup>2</sup> Author for correspondence (e-mail: bbaldwin@berkeley.edu)

doi:10.3732/ajb.1600326

size, with or without correction for overall area richness; see Laffan et al., 2016). Similar metrics and hypothesis tests are applicable for both species-based and clade-based spatial biodiversity assessments. Species-based approaches are particularly useful when phylogenetic data are too limited to allow for fine-scale pattern resolution using clades.

As one of 36 global-scale biodiversity hotspots (Mittermeier et al., 2011), the California Floristic Province (CA-FP) (Howell, 1957) as well as the State of California have been the foci of multiple studies on regional patterns of native vascular plant diversity and endemism (reviewed by Baldwin, 2014; also see Burge et al., 2016). The young mediterranean-like climate of much of California coupled with a dynamic climatic and geological history, topographic complexity, and spatial environmental heterogeneity in general have been implicated in high rates of speciation, low rates of extinction, or both (Raven and Axelrod, 1978; Lancaster and Kay, 2013). High species turnover (beta-diversity) within California has been long recognized, as reflected by early and recent efforts to subdivide the state into floristically distinct bioregions or endemism areas (e.g., Jepson, 1925; Stebbins and Major, 1965; Jepson Flora Project, 2016; see Fig. 1).

Previous attempts to identify areas of native floristic richness and endemism within California have assigned taxa to geographic units of varying size, circumscription, and number based on range descriptions of taxa in floristic treatments (Stebbins and Major, 1965; Thorne et al., 2009; Kraft et al., 2010). Increased spatial resolution and refined sampling of diversity in newer studies have led to revised conclusions about the precise locations and relative importance of areas of high taxonomic richness and endemism. For example, Stebbins and Major (1965), using a system of 10 subdivisions of California and 70 diverse genera, identified Southern California as the primary hotspot for vascular plant endemism in the state; Thorne et al. (2009), using a system of 228 Californian subdivisions and ~94% of recognized species, concluded that the Southern California hotspot was mostly accounted for by the Transverse Ranges. No studies to date, however, have examined California native vascular plant diversity using a uniform spatial scale, a comprehensive taxon data set, and a specimen-based approach to determining plant distributions.

Growth of the Consortium of California Herbaria (CCH) data set of georeferenced collections beyond 2 million Californian specimen records of native, naturalized, and waif vascular plant taxa (Markos et al., 2016) and recent revision of the California flora (Baldwin et al., 2012; Jepson Flora Project, 2016) now enable the first quantitative analysis of patterns of richness and endemism across California based on the full native vascular flora and equal-area spatial units (i.e., grid cells vs. a priori bioregions). Here, we undertook that analysis to ask the following questions: (1) Does a spatially objective, taxonomically comprehensive, and specimen-based approach to examining patterns of richness and endemism in the California flora corroborate and refine patterns reported in previous studies? (2) Do the patterns obtained using our approach differ across major clades of vascular plants or for all native species in comparison with absolute endemics to California? (3) What are the patterns of turnover (beta-diversity) among identified areas of significantly high endemism within California? To ask those questions, we first addressed the optimal choice of spatial resolution (i.e., grid cell size) based on the available specimen data, with attention to sampling issues resulting from different numbers of collections per grid cell. Those issues were considered using a rarefaction

approach based on random sampling theory (Heck et al., 1975) and in light of the nonrandom, diversity-biased sampling represented by herbarium specimens.

## MATERIALS AND METHODS

**Species taxonomy**—Sampling of spatial diversity across the State of California included all native vascular plant species recognized by the Jepson Flora Project (2016) and, for one set of analyses, native vascular plant species examined by Stebbins and Major (1965) for regional patterns of endemism in California. The Index to California Plant Names (Rosatti, 2003; [http://ucjeps.berkeley.edu/about\\_ICPN.html](http://ucjeps.berkeley.edu/about_ICPN.html)) as reflected by synonymy in the Jepson eFlora (<http://ucjeps.berkeley.edu/eflora>) was used to aid in taxonomic concept-matching between currently recognized species and the scientific names used by Stebbins and Major (1965) or appearing on specimen records. Unpublished lists of species from the large genera studied by Stebbins and Major (1965) were obtained from the George Ledyard Stebbins Papers (Special Collections, University of California, Davis Library; the species list is deposited in the UC Berkeley Dash repository [<https://dx.doi.org/10.6078/D1B885>]).

**Occurrence data**—The locality information from herbarium specimens was obtained from five different online sources. The majority of records were from a complete download of the Consortium of California Herbaria (CCH, <http://ucjeps.berkeley.edu/consortium>) on 5 August 2015. Additional records were downloaded from the Consortium of Pacific Northwest Herbaria (<http://www.pnwherbaria.org>), Australia's Virtual Herbarium ([avh.chah.org.au](http://avh.chah.org.au)), Canadensys (<http://www.canadensys.net>), and the Global Biodiversity Information Facility (<http://www.gbif.org>) on 6 August 2015. The combined data set contained over 1.49 million records of native Californian vascular plant species before the culling of unwanted records.

**Cleaning data and adding georeferences**—A number of data-manipulation steps were taken to improve the quality and integrity of the initial spatial data set. Python scripts updated binomial names so that they reflected the taxonomy of the Jepson eFlora and sorted records into their recorded counties. The scatter plot function in Google Refine version 2.5 (<https://github.com/OpenRefine/OpenRefine/releases/tag/2.5>) was used to map records with a geocode and where possible to manually correct the geocode of any record that lay in the ocean or was an outlier to the county of recorded occurrence. The correct latitude and longitude were determined by entering the locality information of each specimen into GeoLocate (<http://www.museum.tulane.edu/geolocate>) or Google Maps (<https://www.google.com/maps>) and checking the georeference of the resulting pinpoint. Approximately 10,000 records were corrected by this process in this stage of the cleaning.

Second, when possible, we added new geocodes to records lacking such data by cloning geocodes from existing records. Python scripts converted all abbreviations to full words, e.g., S into South, Rd into Road. The locality information was then clustered using the fingerprint and n-gram fingerprint algorithms in Google Refine so that records were regrouped that originally had the same locality information but for which the exact details had been entered slightly differently by different institutions. As an example, a collector made a field trip on 14 November 1976 and collected three



**FIGURE 1** Major geographic subdivisions of California (Jepson Flora Project, 2016) and select geographic features discussed in the text. CaR = Cascade Ranges. CW = Central western California. DMoj = Mojave Desert. DSon = Sonoran (or Colorado) Desert. GV = Great Valley. MP = Modoc Plateau. NW = Northwestern California. SN = Sierra Nevada. SNE = East of the Sierra Nevada. SW = Southwestern California. CaR, CW, GV, NW, SN, and SW are within the Californian portion of the California Floristic Province; MP and SNE are within the Great Basin Province; DMoj and DSon are within the Desert Province.

different plant taxa at the North Peak of Mount Diablo. The collections were split among three herbaria. When the locality information was databased at the three institutions, it was entered in three ways, as N Pk of Mt Diablo, Nth P Mt Diablo, and Mt Diablo, N Peak. By converting the abbreviations back into real words, the

clustering algorithm was able to identify all of the alternative ways of entering the same information and regroup them into one cluster. Once clustered, a Python script searched locality information for exact matches across two fields: locality and date collected. If a record without a geocode had an exact match with a record that

included one, then the geocode was copied to the record(s) lacking such information. To reduce the possibility of a geocode being cloned incorrectly when the locality information had a common name, e.g., Bear Creek, the script only cloned records within the same county. The cloning process produced geocodes for 124,460 records (26.5% of the CCH records that were not already georeferenced), equating to roughly 7% of the total number of specimens in the CCH. The 345,095 records that still lacked a geocode were excluded from the analysis.

Third, after correction and addition of georeference information, records were excluded if georeferenced occurrences were well outside the native geographic distribution of a species according to Jepson eFlora author-verified ranges and other expert sources. This procedure was done by manually checking the range maps of every species one-by-one in Google Refine. In this way, naturalized occurrences of native Californian species were excluded from all analyses. An extreme example of naturalized native plant occurrences is Monterey cypress (*Hesperocyparis macrocarpa*), which was originally restricted to the vicinities of Monterey Peninsula and Point Lobos before European settlement, but now occurs throughout much of California and is reflected as such by herbarium records in the CCH. The updated records were screened using two different processes (online Appendix S1, see the Supplemental Data with this article). The first process identified records that occurred outside of the Jepson geographic subdivisions indicated as constituting the distributional ranges of species in the Jepson eFlora. The second stage used climatic niche modeling to identify georeferences with climates highly dissimilar to the climates inhabited by conspecifics. Every record was scored for climatic suitability according to the prediction of a Maxent model (Phillips and Dudík, 2008), fit using four 1-km gridded macroclimatic variables representing major energy and water-related variables important for plant distributions: climatic water deficit, annual precipitation, and summer (June–August) and winter (December–February) mean temperatures. All climate data were obtained from the California 2014 Basin Characterization Model (1951–1980 averages; Flint et al., 2013) at 270-m resolution, and upscaled to 810-m resolution to better match the spatial uncertainty of occurrence records. Gaps in climatic water deficit due to inland water bodies were interpolated from their nearest neighbors. We used default Maxent parameters with no threshold or hinge features to reduce overfitting of the models. As background data, we used 10,000 points extracted from cells with species records, thus reducing effects of spatial sampling biases. An “outlier index” was calculated for each record by dividing its suitability score by the species maximum suitability score and taking the negative log-10 of that ratio; an index score of 1 or 2 thus respectively indicates 10% or 1% of maximum suitability.

The information from both of these processes was then manually assessed using Google Refine to determine which records should be excluded from the final analysis (Appendix S1). We individually checked every specimen with focus on the climate outliers, i.e., those with a value >1.5 in the Maxent outlier index, and either removed or corrected the georeference information for those specimens. In some cases, records were excluded because the locality information was vague, e.g., California, San Francisco, or Los Angeles, and the recorded error radius was large, an issue that was highlighted by Maldonado et al. (2015). In some cases, additional information was used, such as the distribution maps in *The Distribution of Forest Trees in California* (Griffin and Critchfield, 1972). If no clear error was detected, the records were left in.

After these cleaning and cloning processes, the final data set contained 1,383,762 occurrences of Californian vascular plant species and is deposited in the UC Berkeley Dash repository (<https://dx.doi.org/10.6078/D16K5W>).

**Spatial resolution and sampling intensity analyses**—The coordinate information of the data set was converted into Albers Equal units using the EPSG-3310 projection. Uniform grids with cells measuring 5, 10, 15, 20, 25, or 50 km on a side were compared as spatial areas for analysis of floristic richness and endemism across California using the software package Biodiverse (Laffan et al., 2010). Species richness (total number of species) and species redundancy [1 – (richness/number of specimens); Garcillán et al., 2003] per cell were each mapped across California for the six different scales to aid in assessing the optimal grid-cell size for spatial pattern resolution at sufficient sampling density. We also explored the impact of varying grid-cell size on two endemism indices proposed by Crisp et al. (2001): weighted endemism (WE) and corrected weighted endemism (CWE); WE weights species by the inverse of their ranges (1/number of cells occupied by a species, summed across all species in a cell) and CWE corrects weighted endemism for overall species richness in a cell (WE/cell richness).

Rarefaction curves (Heck et al., 1975) were generated for each grid cell at five grid cell sizes (5, 10, 15, 25, and 50 km on a side) using R to estimate the specimen sampling intensity (number of specimens) per grid cell at which species richness begins to show evidence of saturation under an assumption of random sampling. An association, if any, between the richness and the endemism indices introduced above and specimen sampling intensity was also plotted using R to explore the sensitivity of those measures to sample size. We also examined whether the spatial randomization test results were sensitive to sample size and whether range-restricted species might be oversampled as compared to widespread species.

**Spatial diversity analyses**—Spatial patterns of diversity and endemism across California were examined for seven sets of native species, including all vascular plants, only vascular plants that are completely restricted (absolutely endemic) to California, only vascular plants native to the California Floristic Province, only angiosperms, only gymnosperms, and only “pteridophytes” (i.e., ferns and lycophytes). Lycophytes were included with ferns to examine diversity patterns for free-sporing vascular plants, with the understanding that they constitute a grade rather than a clade; separate spatial diversity analysis of the two clades was infeasible based on the small number of collections available for lycophytes. We also analyzed spatial patterns of diversity for the set of California species belonging to the diverse genera examined by Stebbins and Major (1965). All subset species lists and the R script used to extract the subsets from the master spatial file are available from the UC Berkeley Dash repository (<https://dx.doi.org/10.6078/D1G010>).

For each of the above analyses, Biodiverse was used to estimate species richness, WE, CWE, and results of a spatial randomization test designed to detect significantly high endemism in each grid cell (Rand END). The significance test involved generating a null expectation for each grid cell based on 999 replicates of randomly reassigning species to grid cells, without replacement, thus keeping constant the total number of cells occupied per species and the total richness per cell, and then determining whether the actual endemism value for a cell fell within the uppermost 5% of the null distribution for that cell. For the main analysis of all vascular plants, the

Biodiverse results for each grid cell, including a summary of the randomization results, are available from the DASH repository (<https://dx.doi.org/10.6078/D12S3Z>).

**Spatial turnover (beta-diversity) analysis**—For the analysis including all native vascular plant species in California, similarity in floristic composition among areas discovered to have significantly high endemism was examined by clustering those grid cells using Sørensen's index (SI) and range weighted turnover (RWT; Laffan et al., 2016) and mapping each cluster of interest in Biodiverse. This approach was taken primarily to assess whether adjacent grid cells with significantly high endemism represent floristically distinct regions and more generally to explore patterns of similarity across areas of significantly high endemism in California.

## RESULTS

**Spatial resolution and sampling intensity analyses**—Based on the final cleaned data set, species richness and species redundancy using grid cell sizes of 5 × 5 km and 10 × 10 km were so low in some areas that diversity patterns could not be assessed across multiple adjacent cells, especially in the San Joaquin Valley and Mojave Desert (Fig. 2, online Appendix S2). At larger grid cell sizes (15, 20, 25, or 50 km on a side), both species richness and species redundancy were sufficiently high that only two cells at most were unsampled and most other cells received modest to high scores for both measures. The same patterns were observed for WE and CWE (online Appendices S3, S4). The patterns of richness and endemism were similar at all the larger grid cell sizes (15, 20, 25, or 50 km on a side). Based on those findings, the 15 × 15 km grid size was chosen for all further analyses to maximize resolution of spatial patterns of diversity.

For the 15 × 15 km grid cells ( $N = 1959$ ), the number of specimens per grid cell across California ranged from 0 to 12,727 (median = 342); diversity per cell ranged from 1 to 787 (median = 151). The rarefaction analyses showed diversity increased rapidly with sampling in each grid cell, especially at low collection numbers (Fig. 3); however, it appeared that most accumulation curves began to flatten out near their ends for the larger grid cell sizes starting at 15 × 15 km (online Appendix S5). Scatterplots of richness and WE vs. specimen sampling intensity across all grid cells (Fig. 4A and 4B) showed a tight, nearly linear relationship, especially below about 200 specimens. In contrast, a scatterplot of CWE vs. specimen sampling intensity (Fig. 4C) showed that CWE is relatively insensitive to sampling intensity, as are the results of the test for significantly high endemism (Rand END). A slight bias in Rand END was evident at the lowest sample sizes (i.e., below 100 collections per grid cell) because there appeared to be an excess number of cells judged significantly high in endemism in that part of the distribution (Fig. 4D). A slight bias in Rand END was also evident at large sample sizes, because there appeared to be an excess number of cells judged insignificant in that part of the distribution (Fig. 4D). A plot of species range size vs. the average sampling density per cell (online Appendix S6) gives an indication in these data that range-restricted taxa are more likely to be collected than common taxa.

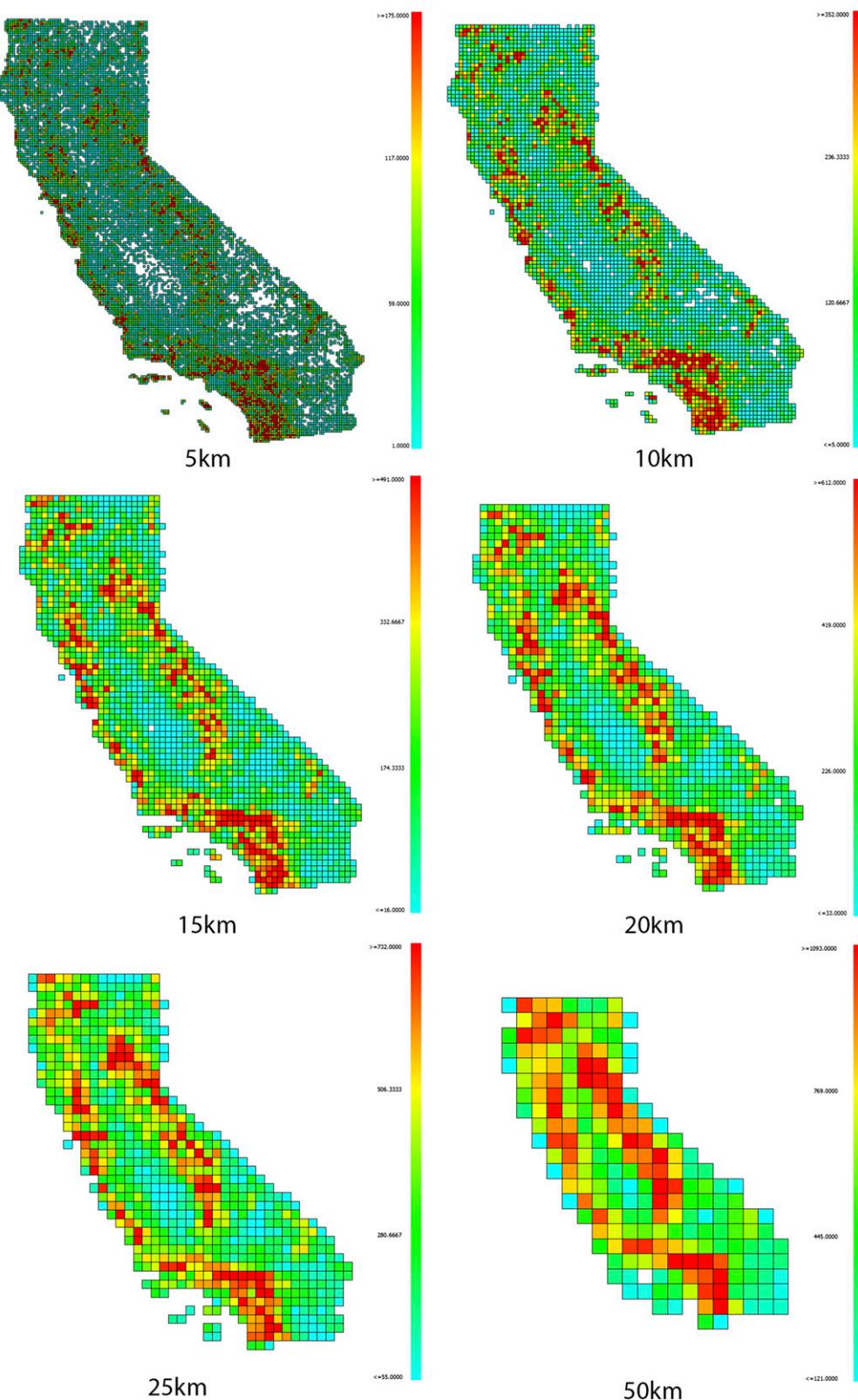
**Spatial diversity analyses**—For each of the sets of Californian species examined, spatial patterns of species richness and WE were more similar to one another than to the patterns of CWE and re-

sults of the randomization test for significant endemism (Rand END), which in turn were quite similar to each other. Those findings were as expected based on incorporation of species richness in the calculation of WE, and holding of species per cell constant in Rand END, which effectively corrects for differences in richness between cells, as does CWE in a different way.

In general, species richness was found to be strongly concentrated within the California Floristic Province (CA-FP) for all sets of species examined, including (1) all vascular plants (Fig. 5), (2) only vascular plants that are completely restricted to California (online Appendix S7), (3) only vascular plants native to the CA-FP (online Appendix S8), (4) only angiosperms (online Appendix S9), (5) only gymnosperms (online Appendix S10), and (6) only “pteridophytes” (online Appendix S11). In comparison with areas of high species richness, areas of high WE in general were more limited within the CA-FP, often in higher montane or more coastal settings, and were more extensive outside the CA-FP, mostly in the higher ranges of the Great Basin and Mojave Desert. For both species richness and WE, patterns resolved for the entire vascular flora were not substantially different than those resolved for the 70 genera data set of Stebbins and Major (1965; online Appendix S12) or for Californian angiosperms alone (Appendix S9).

Patterns of species richness and WE for gymnosperms (Appendix S10) and pteridophytes (Appendix S11), although preliminary given the relatively low representation of specimens per grid cell for these groups, followed the trends noted above but differed from the full vascular flora (Fig. 5) and angiosperm (Appendix S9) patterns sufficiently to warrant brief elaboration. Gymnosperms (Appendix S10) had high species richness mostly concentrated in the Klamath and Transverse ranges and Sierra Nevada, and in limited parts of the Cascade, North Coast, Panamint (Mojave Desert), Peninsular, and South Coast (Santa Lucia) ranges and Warner (Modoc Plateau) mountains. Areas of high WE include parts of the Klamath, Cascade (Mt. Shasta vicinity), northern North Coast, and Peninsular ranges, southern Sierra Nevada, North and southern South coast, northern Channel Islands, Santa Lucia Range (South Coast Ranges), Monterey Peninsula/Point Lobos area (Central Coast), and Santa Cruz (San Francisco Bay Area), White (east of Sierra Nevada), and eastern Mojave Desert mountains. For pteridophytes (Appendix S11), the most extensive areas of high species richness were the Klamath, Peninsular, Transverse, and South Coast ranges, Sierra Nevada, the San Francisco Bay Area (e.g., Santa Cruz Mountains), parts of the Cascade Ranges, coast (e.g., Monterey Peninsula), and northern Channel Islands. Areas of high WE for pteridophytes include the Cascade and Klamath ranges and to a lesser extent other areas of high species richness, and also included parts of the White Mountains (east of the Sierra Nevada) and northern (Panamint) and eastern high ranges of the Mojave Desert.

For the full data set (all vascular plants; Fig. 5), high CWE was particularly concentrated in the far northern, eastern, and southwestern (Channel Islands) periphery of California for most of the analyses, or at least to a greater extent than for species richness or WE, e.g., for plants occurring at least in part in the CA-FP (Appendix S8). Results of the Rand END closely resembled those for CWE. Areas of high species richness or WE overlapped with areas of high CWE or significance in Rand END primarily in the Klamath Ranges (e.g., Marble, Scott, Siskiyou, and Trinity mountains), the Mt. Shasta region of the Cascade Ranges, High Sierra Nevada, Channel Islands, the high Peninsular (e.g., Santa Rosa Mountains) and Transverse ranges (San Bernardino Mountains), and the high



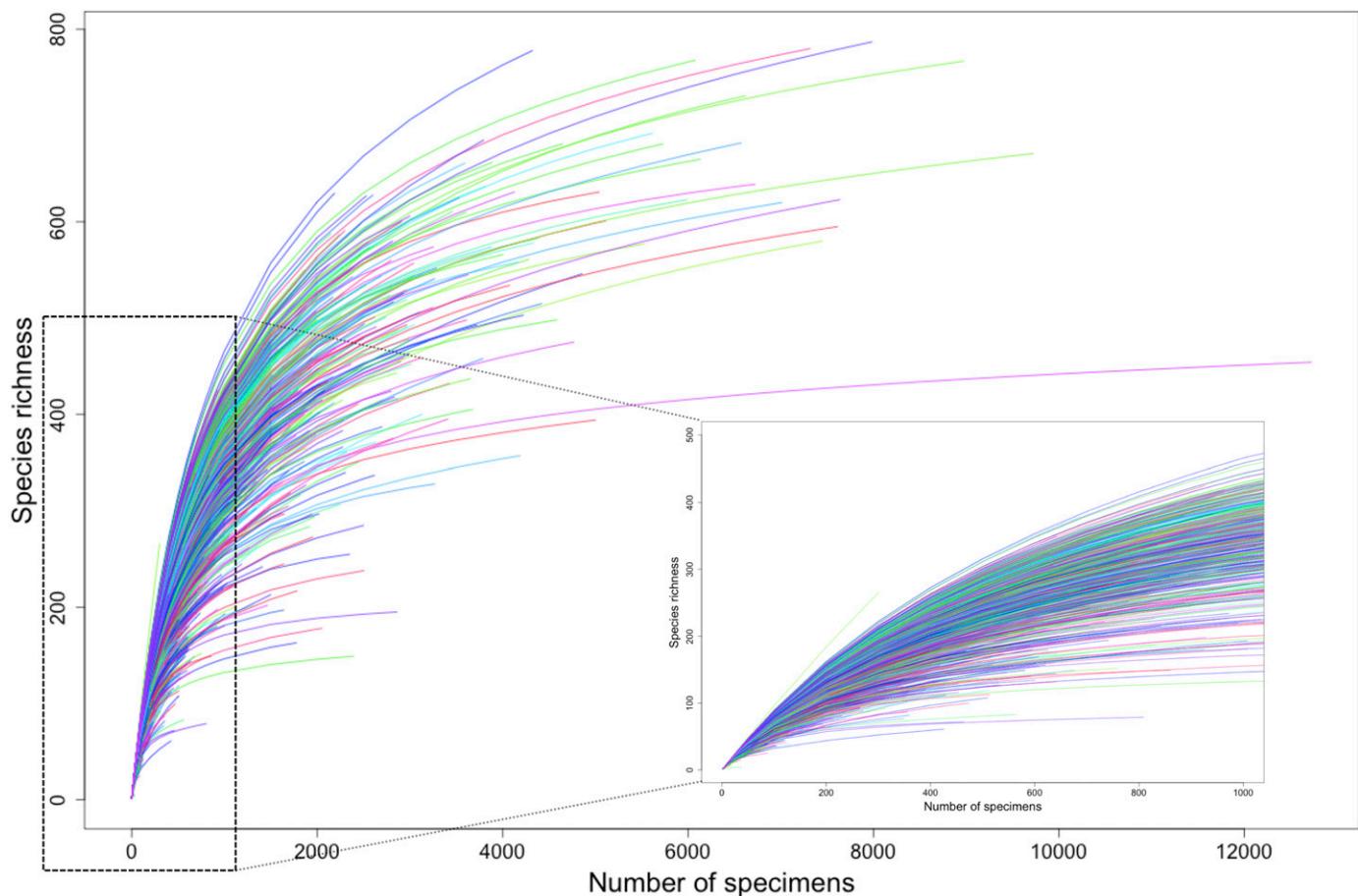
**FIGURE 2** Species richness for all Californian vascular plant species plotted at  $5 \times 5$ ,  $10 \times 10$ ,  $15 \times 15$ ,  $20 \times 20$ ,  $25 \times 25$ , and  $50 \times 50$  km scales.

ranges of the Great Basin (e.g., Sweetwater Mountains and White/Inyo Range) and Mojave Desert (e.g., Providence and New York Mountains, and Clark Mountain, Kingston, and Panamint ranges).

In gymnosperms (Appendix S10), spatial patterns of high CWE resembled those of high WE, although were less extensive in the Klamath and North Coast ranges and southern Sierra Nevada, and included some areas in the northernmost Sierra Nevada and in the Inyo Mountains (east of the Sierra Nevada) but not the Panamint Range (northern Mojave Desert). Areas of significantly high endemism were similar to areas of high CWE except for absence of significant cells in the southern Sierra Nevada and White Mountains. For pteridophytes (Appendix S11), areas of high CWE and significantly high endemism were almost identical to one another and comparable to patterns of high WE at the scale of major geographic subdivisions, although less extensive in the Klamath Ranges and Sierra Nevada and nearly absent in the Peninsular and Transverse ranges. Outside the CA-FP, areas of high CWE and significantly high endemism for pteridophytes were more extensive than areas of high WE in the eastern Mojave Desert and were of limited extent in the Sonoran Desert and Modoc Plateau, where areas of high WE were lacking.

**Spatial turnover (beta-diversity) among concentrations of endemism**—Clustering grid cells with significantly high endemism (from Rand END), and mapping the clusters, yielded similar overall patterns of floristic affinities under both SI and RWT. In general, however, RWT resolved clusters of greater geographic integrity than did SI, especially at finer spatial scales, as might be expected given that RWT emphasizes range-restricted species when measuring similarity, appropriate for evaluating concentrations of endemism. Therefore, we emphasize RWT (Fig. 6) in the Discussion.

When RWT was used, major clusters of grid cells were resolved

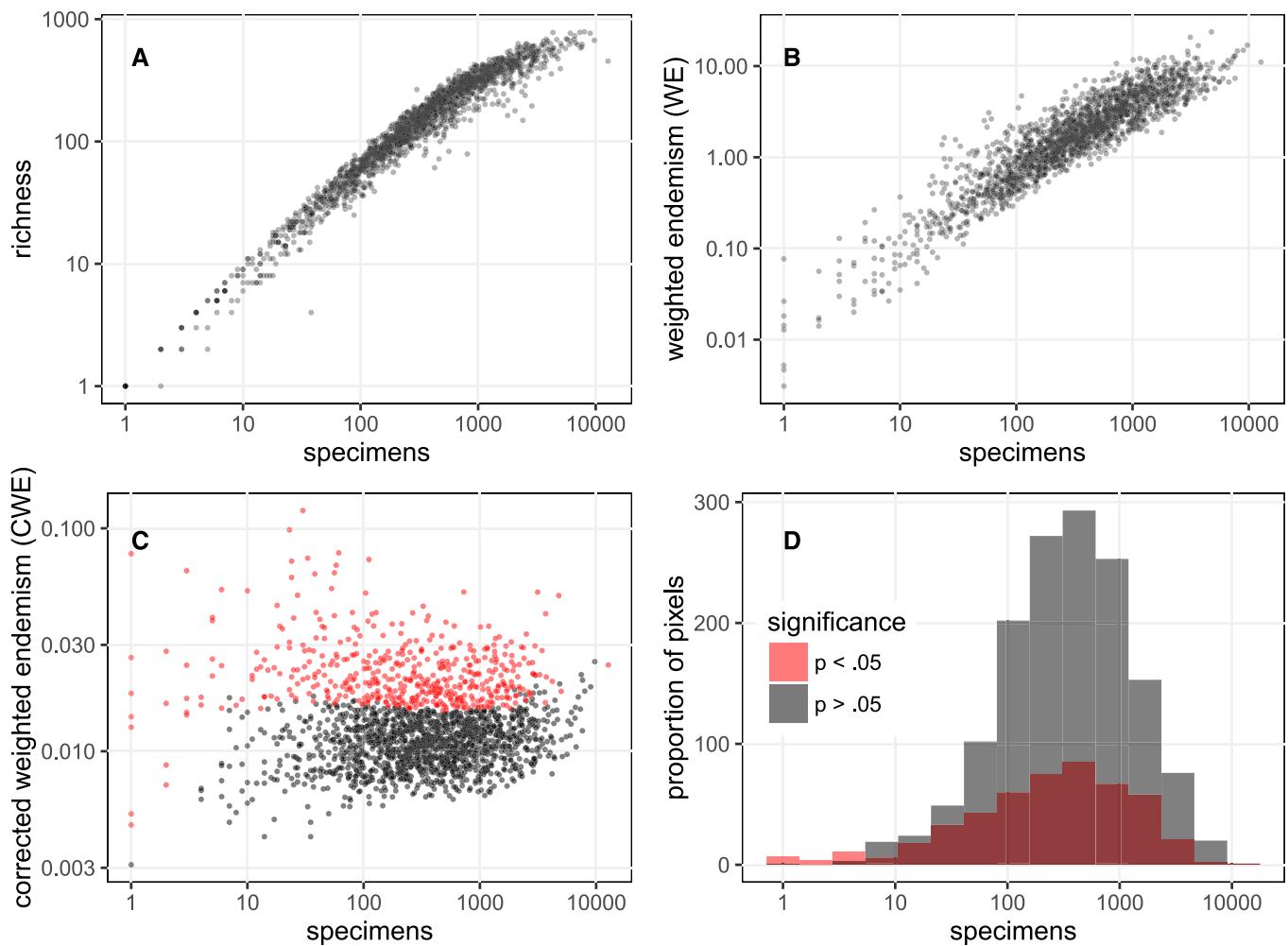


**FIGURE 3** Rarefaction curves illustrating species richness vs. number of samples, based on random sampling from the observed specimen records, for each of the 1959 spatial grid cells in California. Curves that saturate (i.e., flatten out near the end) suggest that the majority of the taxa occupying a grid cell have been collected. Inset shows the curves up to 1000 specimens. Comparisons of curves at constant sample number allow direct comparisons of diversity levels, independent of sampling intensity; see text for discussion of sampling biases that might impact the sampling-richness relationship.

that span the following geographic areas (see Fig. 6): (1) Great Valley, Sierra Nevada foothills, and South Coast Ranges; (2) North, Central, and South Coast plus southernmost North Coast Ranges (inland to the Mayacamas Range), Channel Islands, outermost northern North Coast Ranges, and coastal edge of the Klamath Ranges; (3) San Bernardino Mountains and Desert Province; (4) Great Basin Province, central High Sierra Nevada crest, and southern High Sierra Nevada; and (5) interior North Coast Ranges and Klamath Ranges, Cascade Ranges, northern High Sierra Nevada, and central High Sierra Nevada west of the Sierra crest. Another deep cluster was resolved that included some of the transitional areas among the Klamath Ranges, Cascade Ranges, and North Coast Ranges resolved using Sørensen's index but without the area northeast of Lake Shasta. At a finer scale, separate clusters were resolved for the following areas: (a) South Coast Ranges; (b) Sierra Nevada foothills, Sacramento Valley, and San Joaquin Valley south to northern Merced County; (c) Channel Islands and South Coast; (d) Central and North Coast plus outermost North Coast Ranges and coastal edge of Klamath Ranges; (e) San Joaquin Valley north to southern Merced County; (f) San Bernardino Mountains and Mojave Desert; (g) Sonoran Desert; (h) Modoc Plateau; and (i) central High Sierra Nevada crest, southern High Sierra Nevada, and Great

Basin area east of the Sierra Nevada. Additional clusters (e.g., Channel Islands, Central Coast from Monterey Bay south, Warner Mountains) are discussed below.

When SI was used, major clusters of grid cells with significantly high endemism corresponded closely to the set of cells occurring in the following geographic areas of California (see Fig. 1): (1) Great Valley; (2) San Bernardino Mountains and Desert Province (e.g., Mojave and Sonoran deserts) plus the Inyo and southern White mountains and surrounding foothills; (3) Southwestern California (including the Channel Islands), the Central Coast from Monterey Bay south, South Coast Ranges, and Sierra Nevada Foothills; (4) Central Coast north of Monterey Bay, northwestern California, Cascade Ranges, northern High Sierra Nevada, and the central High Sierra Nevada west of the Sierra crest; and (5) Great Basin Province (e.g., Modoc Plateau and east of the Sierra Nevada, minus the Inyo and southern White mountains and surrounding foothills), central High Sierra Nevada crest, and southern Sierra Nevada. Another deep cluster was resolved that included the easternmost Klamath Ranges (northeast of Lake Shasta) and transitional areas in the nearby Cascade Ranges and North Coast Ranges. Within the northern and southern clusters that include coastal regions,



**FIGURE 4** Relationship between specimens per grid cell (on a log scale) and (A) species richness (on a log scale), (B) weighted endemism (on a log scale), (C) corrected weighted endemism (on a log scale; red indicates significant cells), and (D) frequency of significant endemism.

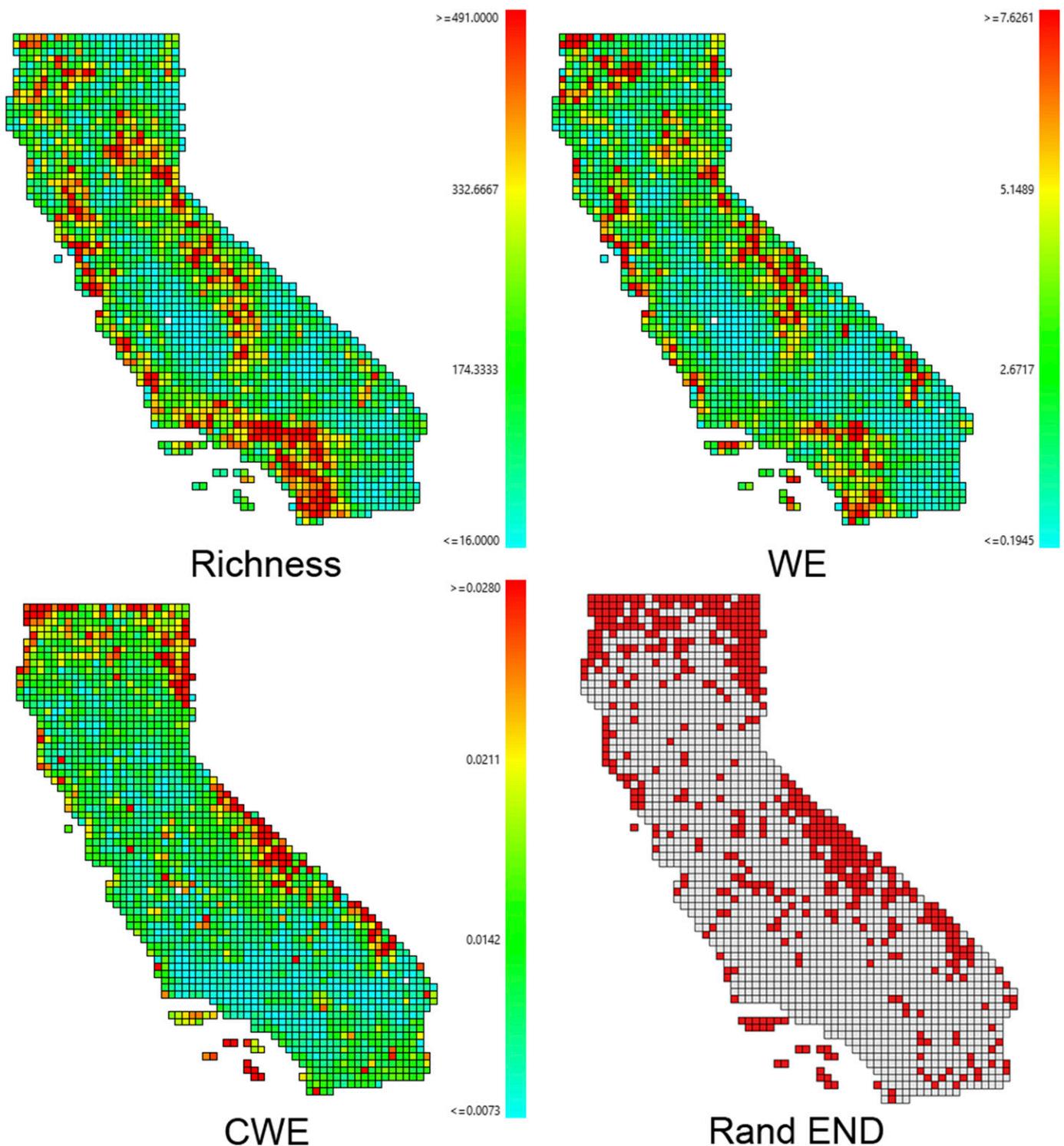
finer-scale clusters were resolved that united the set of cells occurring in (a) the Central Coast from Monterey Bay southward, the South Coast, and the Channel Islands and (b) the North Coast, southernmost North Coast Ranges (inland to the Mayacamas Range), outermost northern North Coast Ranges, and coastal edge of the Klamath Ranges.

## DISCUSSION

**Spatial resolution and sampling issues involving herbarium specimens**—Herbarium data, because they have a physical specimen attached, provide the best-documented source of information available for distributions of plant taxa, especially at broad geographic scales. However, it is important to consider carefully what biases might be present in such data. We observed that grid-cell diversity generally increases with sampling intensity (Figs. 3, 4A). Given the very high level of plant diversity in California, an exceptionally high density of sampling would be required to document all taxa occupying each grid cell, if one assumes that random sampling

theory can be applied to herbarium data. How reliable, therefore, are the differences we saw in richness across California?

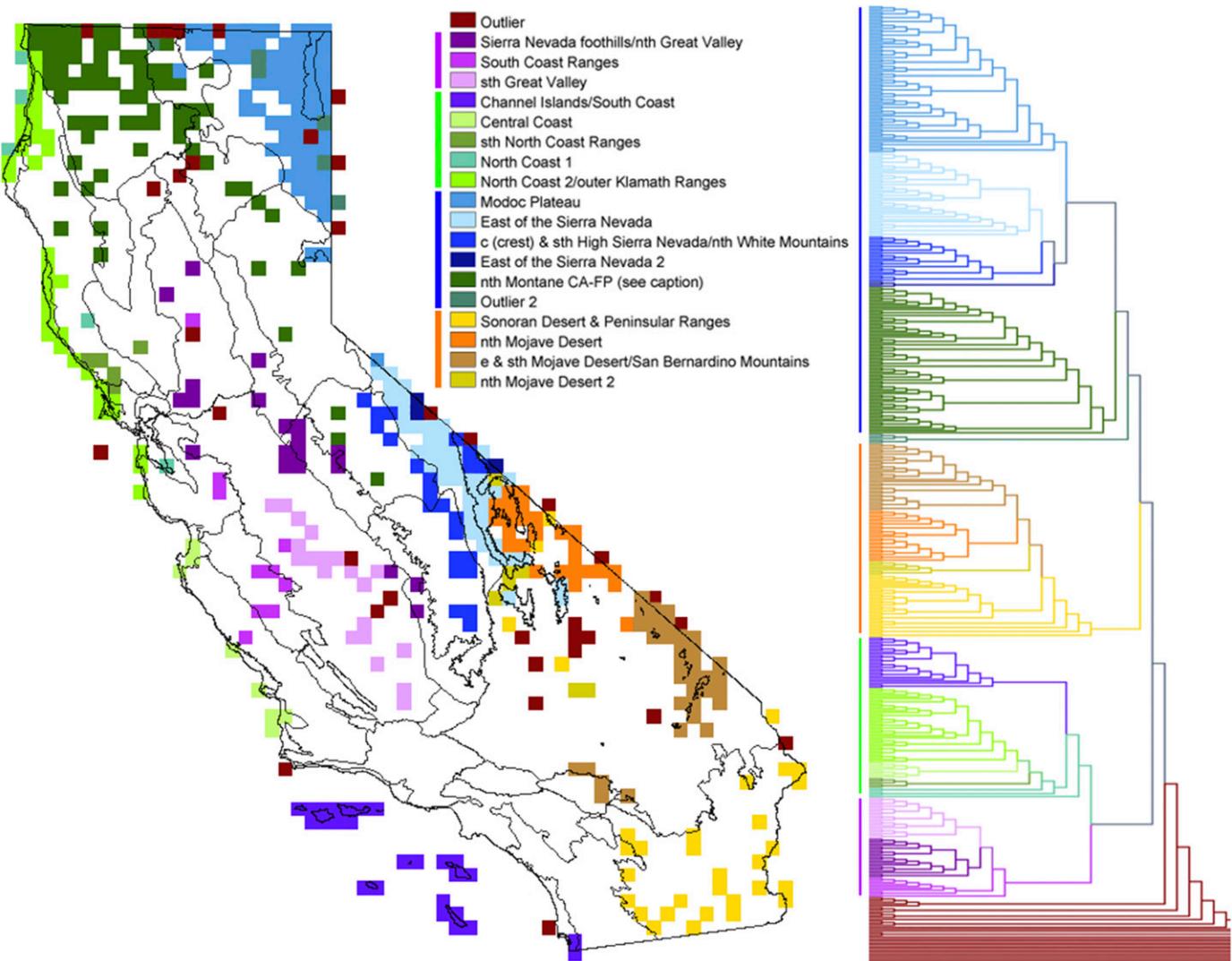
Rarefaction methods provide one approach under random sampling theory to obtaining comparable measures of diversity, correcting for sampling intensity. The results shown in Fig. 3 might be taken to mean that variation in observed richness, when sample sizes were less than about 1000, is mostly just a function of sampling intensity and that one actually has a poor idea what the true richness is for a cell if it were to be heavily sampled. However, such an interpretation is only applicable based on the assumption that herbarium specimens represent a systematic or random sample of diversity, akin to random sampling within an ecological plot. We believe this assumption is not appropriate for herbarium data because of three types of systematic bias in museum collections that collectively enhance the utility of collections for mapping of richness and endemism: (1) geographic focus in collecting intensity, with collectors concentrating on regions of high diversity and endemism; (2) a nonrandom collection strategy that emphasizes floristic documentation, with acquisition of only one or a few specimens per species in a location, and not random sampling in proportion to abundance (see Guralnick and Van Cleve, 2005); and (3) a systematic focus in favor of collecting unusual as opposed to common species. The last



**FIGURE 5** Results for species richness (Richness), weighted endemism (WE), corrected weighted endemism (CWE), and results of the spatial randomization (Rand END) for all native vascular plants in California.

two points were documented in a study by ter Steege et al. (2011) that used simulation studies and comparisons with plot data to show that collectors tend to follow a “never the same species twice” rule, resulting in a rapid species accumulation curve as compared to random sampling.

Herbarium collections in the United States are more representative of the activities of systematists (i.e., scientists trying to document the entire diversity of their study clade) and floriculturists (i.e., scientists trying to document the entire flora of their study area) than vegetation scientists (i.e., scientists recording data from their



**FIGURE 6** A cluster analysis using range-weighted turnover, comparing all grid cells that showed significantly high endemism, with the location of the major clusters shown on the map of California. Outlines on the map indicate the major geographic subdivisions of California described in Fig. 1.

plots or transects using random sampling), so the assumption of a diversity bias in collecting appears warranted. Collectors usually are motivated to find new plants in an area and to skip taxa they have already collected in the vicinity (ter Steege et al., 2011). An indication of the bias of collectors toward range-restricted taxa is shown in Appendix S6. These biases tend to lead to there being few collections of each taxon in an area and an expectation of a nearly linear curve of taxon accumulation for a location. In addition, sampling intensity may be the *effect*, rather than the *cause*, of diversity: if collectors are more likely to visit and intensively collect in high-diversity regions, then undersampled areas are unlikely to represent unrecognized concentrations of diversity.

Furthermore, any bias due to low specimen numbers in some grid cells is asymmetrical: poor sampling cannot lead to overestimation of richness, only underestimation, and calculation of WE is buffered against poor sampling of any particular grid cell by considering species occurrences across all grid cells. Similar patterns of richness and endemism were found in all the larger grid cell sizes examined (15, 20, 25, or 50 km on a side), indicating that even with much higher sampling in the large grid cells, the broad patterns

across the landscape are the same. For all these reasons, results for species richness were presented for all grid cells, with the caveat that cells of low richness might be relatively undersampled and accordingly are not emphasized here; instead we emphasize areas of high richness, some of which may be of even higher richness than indicated by our data.

Inspection of the scatterplot of specimen sampling intensity vs. CWE and the results of the randomization test (Fig. 4C, 4D) show that these measures of endemism are relatively robust to differences in amount of sampling per grid cell, by comparison with the measures of species richness or WE (Fig. 4A, 4B). This robustness to sampling differences may be partly because of the “endemism bias” of herbarium data, as discussed above, and perhaps more importantly to the fact that relative endemism measures are mostly dependent on the range of the species *outside* of the grid cell in question. From this perspective, patterns of CWE or significant endemism are much less likely to be misleading based on poor sampling of grid cells than results for species richness and WE.

However, it should be kept in mind that at extremely low levels of sampling, i.e., less than 10 collections per grid cell, there appeared

to be an excess of significant cells detected by the randomization test (Fig. 4D). Likewise, at high collection numbers (which are probably focused in areas of high richness) there appeared to be an excess of insignificant cells detected by the randomization test (Fig. 4D). Given that the randomization process randomly reassigned any terminal to a grid cell up to the original richness value (with each terminal keeping its original range size), if there is a large fraction of the total terminals in a grid cell originally, then there will tend to be a large number of range-restricted taxa added in many cycles of the randomization, which could make it harder to judge any concentration of range-restricted taxa in the original as significantly high.

In summary, we are interested in using the randomization to detect places that truly have an unusually high co-occurrence of range-restricted taxa—this is easy (sometimes too easy) to detect with very low richness, and harder (sometime too hard) to detect at very high richness. Thus, one needs to be cautious that there may be some false negative results at high richness and some false positive results at low richness.

**General spatial patterns of richness**—The concentration of species richness in the CA-FP (Figs. 1, 5) for the entire vascular flora corroborates earlier findings by Stebbins and Major (1965) and Thorne et al. (2009), who also found much lower overall species richness in the Great Basin and Desert provinces than in the CA-FP. The areas of highest species richness outside the CA-FP resolved here are in high montane areas, including the higher ranges of the northern, eastern, and southern Mojave Desert and the Warner and White mountains of the Great Basin. Within the CA-FP, the high species richness in the Klamath Ranges, northernmost Outer North Coast Ranges, southern Inner North Coast Ranges, Sierra Nevada, San Francisco Bay Area, Outer South Coast Ranges, Peninsular Ranges, and Transverse Ranges and low richness in the Great Valley mirrors in part the findings of Thorne et al. (2009), who used coarser map units and distributional data based primarily on floristic descriptions to estimate species ranges, and by Kraft et al. (2010) and Burge et al. (2016), who found highest species richness in the Sierra Nevada and in Northwest, Central West, and Southwest California.

The main exception to the general pattern found in the spatial analyses that areas of high WE are more restricted in spatial extent than areas of high species richness is the Channel Islands (Figs. 1, 5), where the flora has been long noted for a high diversity of range-restricted species that are also confined to California. Higher WE relative to species richness in montane areas at the northern, eastern, and southern periphery of California, including parts of the Klamath Ranges, northern Cascade Ranges, highest reaches of the central and southern Sierra Nevada, and high ranges of the Great Basin and Mojave Desert, apparently reflects a rich diversity of species that are range-restricted within California but more widespread outside the state. None of those areas is high in WE for species that occur only in California, unlike the Channel Islands, which are high in WE for all native species and for species that do not occur outside California. The Channel Islands does conform to the general pattern of high WE in coastal (and high montane) areas of California.

**General spatial patterns of relative endemism**—Range-restricted species co-occur much more often than would be expected by chance: 25% of cells had significantly high endemism, which is 5-fold more than what would be expected under the null hypothesis

of the randomization test. Previous studies of Californian plant diversity have considered spatial patterns of absolute endemism (species exclusively restricted to California) and even WE (Thorne et al., 2009) but not CWE or significant endemism (e.g., Stebbins and Major, 1965; Thorne et al., 2009; Burge et al., 2016), as explored here. The concentration of areas of high CWE and significantly high endemism at the northern and eastern periphery of California (Fig. 5) does not appear to be explained simply by a correspondingly high proportion of species that are rare in California but have distributions that extend into the adjacent states of Oregon and Nevada (that is, an artifact of political rather than floristic boundaries), except for the Modoc Plateau. Areas of high CWE and significantly high endemism remained concentrated in high montane parts of northwestern California, the southern Great Basin (east of the Sierra Nevada), and the northern and eastern Mojave Desert in analyses focused exclusively on species that are restricted to California (absolute endemics), although areas of significantly high endemism were less extensive (Appendix S7). Significantly high endemism for species restricted to California in the Siskiyou Mountains along the Oregon border was unexpected based on the high floristic similarity between northwestern California and southwestern Oregon (Burge et al., 2016), as was significantly high endemism in the Funeral, Grapevine, New York, and Sweetwater mountains and the Clark Mountain, Last Chance, Kingston, and White-Inyo ranges, along the Nevada border, in light of the vast extent of Great Basin and Mojave Desert montane habitat outside California (Appendix S7).

**Spatial turnover among areas of significantly high endemism**—Clusters of grid cells with significantly high endemism using RWT sometimes corresponded to geographic subdivisions of California previously treated as floristic areas (e.g., Jepson Flora Project, 2016; <http://ucjeps.berkeley.edu/eflora/geography.html>). Other clusters represent novel groupings of geographic subdivisions or parts thereof, sometimes spanning floristic provinces (Fig. 6). For example, areas of significantly high endemism along the central Sierra Nevada crest and southern High Sierra Nevada clustered with areas throughout the Great Basin Province, rather than with the rest of the High Sierra Nevada, which instead clustered with cells in the Cascade Ranges and the interior Klamath Ranges. Burge et al. (2016), who treated the Sierra Nevada as a single geographic unit, found highest floristic similarity (using Jaccard-similarity clustering of all CA-FP vascular species) between the Sierra Nevada and the Cascade Ranges, with both of those ranges in turn most similar to northwestern California (and southwestern Oregon), comparable to results obtained here for floristic similarities of areas of significantly high endemism in the northern and central High Sierra Nevada, minus the central Sierra Nevada crest. In our study, the (west-slope) Sierra Nevada Foothills clustered with the South Coast Ranges, on the opposite side of the Great Valley, somewhat reminiscent of Jepson's (1925) low foothill endemic area, which surrounded the Great Valley. In turn, the Sierra Nevada Foothills and South Coast Ranges clustered with the Great Valley.

The detected area of significantly high endemism in the Transverse Ranges, in the San Gorgonio Mountain region of the eastern San Bernardino Mountains, was more similar floristically to the Desert Province than to other parts of Southwestern California or the California Floristic Province in general. The mostly high montane areas of significantly high endemism in the eastern and southern Mojave Desert were more similar floristically to the high San

Bernardino Mountains than to the northern Mojave Desert or (lower elevation) Sonoran or Colorado Desert.

Within the CA-FP, floristic similarity between areas of significantly high endemism was greater among coastal regions (North, Central, and South Coast) and between those regions and the Channel Islands than reflected by geographic subdivisions of California that treat the coast plus adjacent mountains as units that break along a north-south gradient into southwestern, central western, and northwestern areas (e.g., Jepson Flora Project, 2016; <http://ucjeps.berkeley.edu/eflora/geography.html>). From that perspective, the steep climatic gradient from coast to interior that prevails across California appears to be better reflected by breaks in floristic similarity compared to patterns along the latitudinal climatic gradient. Latitudinally, Monterey Bay is evidently a more important break in floristic similarity among coastal areas of significantly high endemism than is San Francisco Bay based on our results. Whether that floristic break is coincidentally or causally associated with a phylogeographic break at Monterey Bay seen in some animal lineages remains to be determined (see Calsbeek et al., 2003).

**Regional patterns**—The following discussion giving details of finer-scale spatial patterns of diversity is organized by the 10 major geographic subdivisions of California as delimited by the Jepson Flora Project (2016; Fig. 1). As the foregoing discussion indicates, however, floristic similarities among the areas of concentrated endemism are not always reflected by this organizational framework.

**Southwestern California (South Coast, Peninsular Ranges, Transverse Ranges, and Channel Islands)**—Stebbins and Major (1965) concluded that Southwestern California was an area especially rich in native and California-restricted species. Subsequent studies found that the Transverse Ranges accounted for much of the endemism reported by Stebbins and Major (Thorne et al., 2009) and that the San Bernardino Mountains in particular were the primary endemic area there (Kraft et al., 2010). Spatial patterns detected in our study corroborate high species richness throughout the montane areas of Southwestern California, especially in the Peninsular Ranges (e.g., San Jacinto Mountains, Palomar Mountains) and the higher Transverse Ranges (e.g., San Gabriel, San Bernardino, eastern Santa Ynez, southeastern San Rafael, and San Emigdio mountains), as well as some areas along the South Coast, such as the San Diego region. Areas of high WE were less extensive, in the Laguna, Santa Rosa, and San Jacinto mountains of the Peninsular Ranges and primarily in the San Bernardino Mountains of the Transverse Ranges, and in the San Diego region of the South Coast and southern Peninsular Ranges and some Channel Islands (San Clemente, Santa Catalina, Santa Cruz, and Santa Rosa islands). In contrast, areas of high CWE and significantly high endemism were strongly concentrated in the Channel Islands, with only isolated areas of significantly high endemism on the mainland, e.g., in the San Bernardino Mountains, northern Santa Rosa Mountains, and La Jolla vicinity on the South Coast. Lack of areas of significantly high endemism on the mainland of Southwestern California likely reflects high richness in species that are not range-restricted rather than low richness in range-restricted species; that is, many range-restricted species are indeed present, but grid cells are not judged by the randomization test to contain significantly high endemism because many widespread native species are also present. As discussed above, some false negative results from the randomization at high richness are to be expected, but this is not likely to account for

the widespread lack of significant concentrations of endemism in this region.

**Central Western California (Central Coast, South Coast Ranges, and San Francisco Bay Area)**—Burge et al. (2016) noted that minimum-rank taxa range-restricted to the CA-FP are most diverse in Central Western California, which is highly isolated from the northern, eastern, and southern boundaries of the province. Stebbins and Major (1965) focused on fine-scale areas judged to be particularly rich in species range-restricted to California in the area from Monterey County north, their “endemic areas of the Central Coast Ranges” (p. 24), which extended somewhat into Northwestern California, as treated by the Jepson Flora Project (2016; see below). In our analyses, areas of high species richness in Central Western California for all native species and also for species range-restricted to California were in general more extensive than those proposed by Stebbins and Major (1965). In the area south of Stebbins and Major’s endemic areas, Hoover’s (1970) “Obispoan pocket of endemism”, including the San Luis Range (Irish Hills), southernmost Santa Lucia Range, Seven Sisters, and other areas in the vicinity of San Luis Obispo, where serpentine exposures are frequent, was resolved here to be part of an area of high species richness for all natives and California-restricted species that extends further west and south to include extensive coastal dunes. To the north, areas of resolved species richness for all natives and for California-restricted species in the Santa Lucia Range and Monterey Peninsula correspond in part to Jepson’s (1925) Lucian Area of endemism, and other areas of high richness to the interior and farther north fall within his Franciscan endemic area. Consideration of only California-restricted species, to be comparable to Stebbins and Major’s (1965) study, corroborates their local endemic areas throughout Central Western California, including their Diablo, Hamilton, Monterey, San Carlos, Santa Cruz, Santa Lucia, and Tamalpais areas but also indicates other local areas of species richness, such as the San Francisco Peninsula. Areas of high WE for all species include Hoover’s (1970) “Cruzian pocket of endemism”, in the vicinity of Arroyo de la Cruz and his Obispoan area, the Nacimiento River area of southwest Monterey County, the San Francisco Peninsula, and areas overlapping with most of Stebbins and Major’s local endemic areas (e.g., Diablo, Hamilton, Monterey, Santa Cruz, and Tamalpais). Some of Stebbins and Major’s local endemic areas were also resolved as areas of significantly high endemism (i.e., Hamilton, Monterey, and San Carlos), as were other areas, such as the Guadalupe and Morro dunes, Arroyo de la Cruz/Piedra Blanca Ranch area, and San Francisco Peninsula north to San Bruno Mountain.

**Northwest California (North Coast, Klamath Ranges, North Coast Ranges)**—Within the CA-FP, Northwestern California is exceeded in species richness only by the Sierra Nevada (Burge et al., 2016) and, as noted above, our results for areas of significantly high endemism indicate that at least much of the interior Klamath Ranges and North Coast Ranges share high floristic similarity with the Sierra Nevada and the Cascade Ranges. Areas of high species richness and WE are mostly away from the immediate coast, with a major center corresponding to both Jepson’s (1925) and Stebbins and Major’s (1965) Napa Lake endemic areas, including the Mayacamas Range. Further north, the Snow Mountain vicinity of the southern High North Coast Ranges was resolved as another species-rich area previously noted as phytogeographically important (Heckard and Hickman, 1984). Stebbins and Major’s other endemic area in this

region, Pitkin Bodega, overlaps with an area of resolved species richness and WE that extends inland between Bodega Bay and the mouth of the Russian River. Species richness and WE for California-restricted species is also high throughout both of the above regions. The limited extent of areas of high CWE and significantly high endemism within these same regions, with some exceptions (e.g., the Mayacamas Range, including Mt. St. Helena and the Palisades), contrasts with high CWE and significantly high endemism in some of the northern North Coast Ranges and Klamath Ranges that also are high in species richness and WE, such as the Marble, Scott, Siskiyou, and Trinity mountains and The Eddys, where the floras are relatively rich in geographically restricted species, including California-restricted species. The Klamath Ranges in general have been long regarded as both a refugium, with a relatively equable, moist climate since the Pleistocene, and a cradle of evolutionary diversification, with extensive climatic, topographic, and edaphic complexity and geological dynamism (Raven and Axelrod, 1978). In contrast to the general pattern of low species richness and WE along the North Coast, the coastline from the Russian River to Cape Vizcaino and the coastline from Cape Mendocino to the Oregon border are areas of significantly high endemism.

**Cascade Ranges**—Although the Cascade Ranges in California have been noted as having higher species richness per unit area than the other five major geographic subdivisions of the CA-FP, i.e., Northwest, Central West, and Southwest California, and the Sierra Nevada and Great Valley (Burge et al., 2016), the areas of high richness and endemism resolved here were relatively limited in extent. Of the areas with high species richness and WE, the highest peak in California's Cascade Ranges, Mt. Shasta, also stood out for high CWE and significantly high endemism. Although only the southwest slopes of Mt. Shasta were resolved as high in species richness and WE, the larger area including all of Mt. Shasta, The Whaleback, and Black Butte was resolved as having high CWE and significantly high endemism, even for California-restricted species. Some areas west and south of Mt. Lassen also showed either high species richness or significantly high endemism, such as the area northwest of Lake Almanor. The southernmost Cascades, in Butte County, were also resolved as high in species richness.

**Sierra Nevada**—Evidence presented here indicating that areas of significantly high endemism in the foothills, montane, and crest regions of the Sierra Nevada are less similar floristically to each other than to areas outside the Sierra Nevada or even the CA-FP may help to explain why the Sierra Nevada has the highest species richness of any of the major geographic subdivisions of California or the CA-FP (Thorne et al., 2009; Burge et al., 2016). Areas of high species richness or WE were resolved throughout the High Sierra, including drainages of the Feather, Yuba, American, Carson, Stanislaus, Tuolumne, Merced, San Joaquin, Kings, Kaweah, and Kern rivers. Examples of such areas include much of the Feather River country; the region encompassing the North Fork American and Bear rivers, Donner Pass, and northwest Lake Tahoe; upper South and Silver forks of the American River and Carson Range; Sonora Pass region; much of Yosemite National Park (except the southeastern part) and upper Merced River drainage west of the park; Big and Kaiser creek drainages (eastern tributaries of San Joaquin River); much of Sequoia National Park, including Giant Forest, Mineral King, and the southern Great Western Divide; and northern Greenhorn Mountains and Kern River country north of Lake

Isabella. Overlap with areas of high CWE and significantly high endemism was mostly along the central and southern Sierra Nevada crest, where floristic similarity was greater with the Great Basin than elsewhere in the Sierra, as noted above, from south of Mammoth Lakes to south of Mt. Langley. Other sizable blocks of significantly high endemism in the Sierra Nevada included two areas west of Yosemite National Park in the Sierra Nevada foothills, one in the vicinity of Mariposa and the other, farther west, within the Stanislaus and Tuolumne drainages (in the vicinity of Jamestown and Chinese Camp) and extending into the Great Valley.

**Great Valley**—Lack of areas of high species richness or WE in the Great Valley corroborates results of previous studies that found the area to be the least species-rich in California or the CA-FP (Stebbins and Major, 1965; Thorne et al., 2009; Burge et al., 2016). In the Sacramento Valley, the Sutter Buttes, which rise to ~650 m a.s.l. and are topographically and vegetationally diverse, and the area bordering the Sacramento-San Joaquin delta in the vicinity of Fairfield and Suisun City were resolved as moderately rich for all vascular plants and for species range-restricted to California. Using RWT, areas of significantly high endemism in the Sacramento Valley, northern San Joaquin Valley, and the Sierra Nevada foothills were resolved as a distinct cluster from those in the southern San Joaquin Valley. Hoover (1937) recognized a major break in the flora of the upper and lower San Joaquin Valley, as well, with the western boundary between his Kern (upper valley) and San Joaquin (lower valley) endemic areas at the northern boundary of Fresno County, in keeping with our findings. Hoover's basis for the boundary was mostly floristic, although he noted that annual precipitation differed between the two regions. The largest area of significantly high endemism in the Sacramento Valley was resolved in the lower Sacramento-San Joaquin delta region (southeastern Solano County), including the Montezuma Hills and Jepson Prairie. In the northern San Joaquin Valley, the largest area of significantly high endemism was along the lower Merced and Tuolumne rivers in the vicinities of La Grange, Merced Falls, and Snelling, contiguous with the area discussed at the end of the preceding section in the Sierra Nevada foothills (including Chinese Camp and Jamestown). In the southern San Joaquin Valley, areas of significantly high endemism were widely scattered, including a large contiguous region north of the Tulare Lake bed in southeastern Merced, western Madera, western Fresno, northern Kings, and northwestern Tulare counties and another to the east and south of the Tulare Lake bed in northwestern Kern, southern Kings, and western Tulare counties. Another sizable area of significantly high endemism was resolved southwest of Bakersfield and east of the Buena Vista Lake bed.

**Modoc Plateau**—Species richness and WE on the Modoc Plateau were limited to the high, southern Warner Mountains, with WE extending farther north, to beyond Cedar Pass and somewhat south of the high, southern end of the range. Raven (1977) noted that the diverse flora of the Warner Mountains combines "Californian" and 'extra-Californian' elements". Areas of high CWE and, especially, significantly high endemism for all species were extensive, but not for species range-restricted to California, which were limited mostly to scattered individual grid cells (e.g., Clear Lake/Timber Mountain; Mt. Vida/Goose Lake; west of Madeline Plains/Whitinger Mountain/northern Grasshopper Valley; southeast of Eagle Lake). Evidently, most of the species of the Modoc Plateau that are range-restricted within California also occur outside of California, as noted above.

*East of the Sierra Nevada*—Species richness in this part of the Great Basin was generally low, with resolution of only modestly high richness in parts of the White Mountains. Areas of high WE and CWE and significantly high endemism were more extensive, including the Sweetwater Mountains, southern Bodie Hills, and much of the White-Inyo Range, for all species and for species range-restricted to California. Despite proximity to the Nevada border, much of the diversity in the higher ranges of this region is evidently limited in range to California, unlike the flora of the Modoc Plateau. Previous studies have established that the Sweetwater Mountains and White-Inyo Range have distinctive floras and strong floristic similarities to the High Sierra Nevada (Lloyd and Mitchell, 1973; Lavin, 1983; Morefield, 1992; Baldwin and Moe, 2002).

*Mojave Desert*—High species richness in the Mojave Desert was mostly limited to the high ranges, including the Granite, New York, and Providence Mountains, of the eastern Mojave and transitional areas such as the desert immediately east of Owens and Sawtooth peaks in the southern Sierra Nevada and the desert at the north edge of the San Bernardino Mountains (including the town of Lucerne Valley). Areas of high WE included all of the above areas plus additional ranges, such as the Clark Mountain, Kingston, and northern and central Panamint ranges. High CWE and significantly high endemism were more extensive than either species richness or WE, even for species restricted to California, including most of the Mojave Desert north of the latitude of the central Panamint Range and encompassing much of the Death Valley region. Another large area of significantly high endemism farther south in the eastern Mojave included the Tecopa vicinity; Lanfair, Mesquite, and Pahrump valleys; Clark Mountain, Kingston, and Nopah ranges; Bristol, Granite, Ivanpah, New York, and Providence Mountains; and the Mid Hills. Other such areas include the transitional zone at the north edge of the San Bernardino Mountains noted above for high species richness, and scattered sites in the western and central Mojave. The Desert Mountains sensu Jepson Flora Project (2016; Fig. 1), California's Death Valley region, and some of the areas in the western and central Mojave Desert identified here as areas of significantly high endemism have been noted previously for concentrations of locally range-restricted taxa (Jepson, 1925; Thorne et al., 1981, 2009; Baldwin and Moe, 2002).

*Sonoran Desert*—California's Sonoran or Colorado Desert has been indicated as being particularly low in species diversity (Stebbins and Major, 1965; Thorne et al., 2009), and no areas of high species richness or WE were resolved there in our study. Some areas of significantly high endemism and, to a lesser extent, high CWE were discovered across the region, however. Near the Colorado River, the southern Chocolate Mountains were found to be an area of significantly high endemism for all species and even for species restricted entirely to California. The Orocopia Mountains and northwest Chocolate Mountains (northeast of Salton Sea) were also resolved as having significantly high endemism for California-restricted species. Areas with significantly high endemism for all species (but not for CA-restricted species) include the Big Maria, central Chocolate, Palo Verde, southern Turtle, Vallecito, and Whipple mountains, Borrego and Palo Verde valleys, Algodones Dunes, parts of the Imperial Valley and flanking mesas, and scattered sites near the western desert edge from Palm Desert to the Mexican border. Species of California's Sonoran Desert that are restricted in range to this region or to California are relatively few

compared with the number of locally or California-restricted species in the Mojave Desert flora (Baldwin and Moe, 2002).

## CONCLUSIONS

Spatial patterns of species richness and endemism in the California flora presented here corroborate and extend results of previous studies and highlight the importance to conservation biology of not focusing attention exclusively on areas of high floristic diversity. As expected, areas high in WE were not always found to have significantly high endemism, i.e., the range restriction of species present, while high, was about as expected given the richness present. The concentration of significantly high endemism in areas near the periphery of California, even in analyses focused only on California-restricted species, demonstrates the high conservation value of the floras of northwestern California and California's deserts. The next steps in understanding these patterns will require adding an evolutionary perspective that allows for consideration of phylogenetic relationships among species in estimates of alpha- and beta-diversity and for resolution of areas of concentrated neo- and paleo-endemism in the California flora (A. Thornhill et al., unpublished manuscript). In addition, patterns of spatial turnover (here examined only among areas of significantly high endemism) warrant further scrutiny of floristic similarities across California among all grid cells to examine and possibly refine bioregional boundaries (B. Mishler et al., unpublished manuscript).

The present results evaluating internal patterns of richness and endemism within California eventually need to be set into more global contexts. For example, in a study of North America as a whole, most grid cells in California (except probably in the Modoc Plateau) would be judged significantly high in range-restricted taxa in that broader context. The comparison group of taxa changes as the study scale grows larger, as do the questions addressed and the likely processes operating. Thus, a full understanding of biodiversity will require examination of pattern at different spatial scales and different evolutionary scales.

## ACKNOWLEDGEMENTS

The authors thank David Baxter for help in organizing data from the Consortium of California Herbaria (CCH), member institutions of CCH for sharing their data, Nunzio Knerr and Shawn Laffan for help with Biodiverse, Dick Olmstead and Ben Legler for access to the Californian herbarium records of the Consortium of Pacific Northwest Herbaria, Naomi Brydon for assisting in data cleaning, and two anonymous reviewers for helpful comments on the manuscript. This research was supported by NSF grant DEB-1354552 to B.D.M., B.G.B., and D.D. A. and a Villum Postdoctoral Fellowship to N.M.H.

## LITERATURE CITED

- Baldwin, B. G. 2014. Origins of plant diversity in the California Floristic Province. *Annual Review of Ecology, Evolution and Systematics* 45: 347–369.
- Baldwin, B. G., D. H. Goldman, D. J. Keil, R. Patterson, T. J. Rosatti, and D. H. Wilken [eds.], 2012. The Jepson manual: Vascular plants of California, second edition. University of California Press, Berkeley, California, USA.
- Baldwin, B. G., and R. L. Moe. 2002. Floristic diversity in the California deserts. In B. G. Baldwin, S. Boyd, B. J. Erter, R. W. Patterson, T. J. Rosatti, and D. H. Wilken [eds.], The Jepson desert manual: Vascular plants of southeastern California. University of California Press, Berkeley, California, USA.

- Burge, D. O., J. H. Thorne, S. P. Harrison, J. R. Shevock, E. R. Alverson, L. K. Hardison, J. D. Rodríguez, et al. 2016. Plant diversity and endemism in the California Floristic Province. *Madroño* 63: 3–206.
- Calsbeek, R., J. N. Thompson, and J. E. Richardson. 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: The California Floristic Province. *Molecular Ecology* 12: 1021–1029.
- Crisp, M. D., S. Laffan, H. P. Linder, and A. Monro. 2001. Endemism in the Australian flora. *Journal of Biogeography* 28: 183–198.
- Flint, L. E., A. L. Flint, J. H. Thorne, and R. Boynton. 2013. Fine-scale hydrologic modeling for regional landscape applications: The California Basin Characterization Model development and performance. *Ecological Processes* 2: 25.
- Garcillán, P. P., E. Ezcurra, and H. Riemann. 2003. Distribution and species richness of woody dryland legumes in Baja California, Mexico. *Journal of Vegetation Science* 14: 475–486.
- González-Orozco, C. E., M. C. Ebach, S. Laffan, A. H. Thornhill, N. J. Knerr, A. N. Schmidt-Lebuhn, C. C. Cargill, et al. 2014. Quantifying phytogeographical regions of Australia using geospatial turnover in species composition. *PLoS One* 9: e92558.
- González-Orozco, C. E., B. D. Mishler, J. T. Miller, S. W. Laffan, N. J. Knerr, P. Unmack, A. Georges, et al. 2015. Assessing biodiversity and endemism using phylogenetic methods across multiple taxonomic groups. *Ecology and Evolution* 5: 5177–5192.
- Griffin, J. R., and W. B. Critchfield. 1972. The distribution of forest trees in California. Pacific Southwest Forest and Range Experimental Station (USDA Forest Service Research Paper PSW-82) (reprinted with supplement, 1976), Berkeley, California, USA.
- Guralnick, R., and J. Van Cleve. 2005. Strengths and weaknesses of museum and national survey data sets for predicting regional species richness: Comparative and combined approaches. *Diversity & Distributions* 11: 349–359.
- Heck, K. L. Jr., G. van Belle, and D. Simberloff. 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56: 1459–1461.
- Heckard, L. R., and J. C. Hickman. 1984. The phytogeographical significance of Snow Mountain, North Coast Ranges, California. *Madroño* 31: 30–47.
- Hoover, R. F. 1937. Endemism in the flora of the Great Valley of California. Ph.D. dissertation, University of California, Berkeley, California, USA.
- Hoover, R. F. 1970. The vascular plants of San Luis Obispo County, California. University of California Press, Berkeley, California, USA.
- Howell, J. T. 1957. The California flora and its province. *Leaflets of Western Botany* 8: 133–138.
- Jepson, W. L. 1925. Manual of the flowering plants of California. Associated Students Store, Berkeley, California, USA.
- Jepson Flora Project [eds.]. 2016. Jepson eFlora. Website <http://ucjeps.berkeley.edu/eFlora/>.
- Kraft, N. J. B., B. G. Baldwin, and D. D. Ackerly. 2010. Range size, taxon age and hotspots of neoendemism in the California flora. *Diversity & Distributions* 16: 403–413.
- Laffan, S. W., and M. D. Crisp. 2003. Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *Journal of Biogeography* 30: 511–520.
- Laffan, S. W., E. Lubarsky, and D. F. Rosauer. 2010. Biodiverse, a tool for spatial analysis of biological and related diversity. *Ecography* 33: 643–647.
- Laffan, S. W., D. F. Rosauer, G. DiVirgilio, J. T. Miller, C. E. Gonzalez-Orozco, N. J. Knerr, A. H. Thornhill, and B. D. Mishler. 2016. Range-weighted metrics of species and phylogenetic turnover can better resolve biogeographic transition zones. *Methods in Ecology and Evolution* 7: 580–588.
- Laity, T., S. W. Laffan, C. E. González-Orozco, D. P. Faith, D. F. Rosauer, M. Byrne, J. T. Miller, et al. 2015. Phylogenetic diversity to inform conservation policy: An Australian example. *The Science of the Total Environment* 534: 131–143.
- Lancaster, L. T., and K. M. Kay. 2013. Origin and diversification of the California flora: Re-examining classic hypotheses with molecular phylogenies. *Evolution* 67: 1041–1054.
- Lavin, M. 1983. Floristics of the upper Walker River, California and Nevada. *Great Basin Naturalist* 43: 93–130.
- Linder, H. P., and G. A. Verboom. 2015. The evolution of regional species richness: The history of the Southern African flora. *Annual Review of Ecology, Evolution, and Systematics* 46: 393–412.
- Loarie, S. R., B. E. Carter, K. Hayhoe, S. McMahon, R. Moe, C. A. Knight, and D. D. Ackerly. 2008. Climate change and the future of California's endemic flora. *PLoS One* 3: e2502.
- Lloyd, R. M., and R. S. Mitchell. 1973. A flora of the White Mountains, California and Nevada. University of California Press, Berkeley, California, USA.
- Markos, S., R. L. Moe, and D. Baxter. 2016. A powerful resource for plant conservation efforts: The Consortium of California Herbaria reaches two million specimens. *Fremontia* 44: 16–19.
- Mishler, B. D., N. J. Knerr, C. E. González-Orozco, A. H. Thornhill, S. W. Laffan, and J. T. Miller. 2014. Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian flora. *Nature Communications* 5: 4473.
- Mittermeier, R. A., W. R. Turner, F. W. Larsen, T. M. Brooks, and C. Gascon. 2011. Global biodiversity conservation: The critical role of hotspots. In F. E. Zachos and J. C. Habel [eds.], *Biodiversity hotspots: Distribution and protection of conservation priority areas*, 3–22. Springer-Verlag, Berlin, Germany. Available at [http://www.cepf.net/news/top\\_stories/Pages/Announcing-the-Worlds-36th-Biodiversity-Hotspot.aspx](http://www.cepf.net/news/top_stories/Pages/Announcing-the-Worlds-36th-Biodiversity-Hotspot.aspx).
- Maldonado, C., C. I. Molina, A. Zizka, C. Persson, C. M. Taylor, J. Albán, E. Chilquillo, N. Rønsted, and A. Antonelli. 2015. Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? *Global Ecology and Biogeography* 24: 973–984.
- Morefield, J. D. 1992. Spatial and ecologic segregation of phytogeographic elements in the White Mountains of California and Nevada. *Journal of Biogeography* 19: 33–50.
- Nagalingum, N. S., N. J. Knerr, S. W. Laffan, C. E. González-Orozco, A. H. Thornhill, J. T. Miller, and B. D. Mishler. 2015. Continental scale patterns and predictors of fern richness and phylogenetic diversity. *Frontiers in Genetics, Evolutionary and Population Genetics* 6: 1–14.
- Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* 31: 161–175.
- Pollock, L. J., D. F. Rosauer, A. H. Thornhill, H. Kujala, M. D. Crisp, J. T. Miller, and M. A. McCarthy. 2015. Phylogenetic diversity meets conservation policy: Small areas are key to preserving eucalypt lineages. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 370: 20140007.
- Raven, P. H. 1977. The California flora. In M. G. Barbour and J. Major [eds.], *Terrestrial vegetation of California*, 109–137. John Wiley, New York, New York, USA.
- Raven, P. H., and D. I. Axelrod. 1978. Origin and relationships of the California flora. *University of California Publications in Botany* 72: 1–134.
- Rosatti, T. J. 2003. Jepson Online Interchange for California Floristics. *Fremontia* 33: 23–29.
- Rosauer, D. F., S. W. Laffan, M. D. Crisp, S. C. Donnellan, and L. G. Cook. 2009. Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology* 18: 4061–4072.
- Schmidt-Lebuhn, A. N., N. J. Knerr, J. T. Miller, and B. D. Mishler. 2015. Phylogenetic diversity and endemism of Australian daisies (Asteraceae). *Journal of Biogeography* 42: 1114–1122.
- Stebbins, G. L., and J. Major. 1965. Endemism and speciation in the California flora. *Ecological Monographs* 35: 1–35.
- ter Steege, H., P. P. Haripersaud, O. S. Banki, and F. Schieving. 2011. A model of botanical collectors' behavior in the field: Never the same species twice. *American Journal of Botany* 98: 31–37.
- Thorne, J. H., J. H. Viers, J. Price, and D. M. Stoms. 2009. Spatial patterns of endemic plants in California. *Natural Areas Journal* 29: 344–366.
- Thorne, R. F., B. A. Prigge, and J. Henrickson. 1981. A flora of the higher ranges and the Kelso Dunes of the eastern Mojave Desert in California. *Aliso* 10: 71–186.
- Thornhill, A. H., B. D. Mishler, N. J. Knerr, C. E. González-Orozco, C. M. Costion, D. M. Crayn, S. W. Laffan, and J. T. Miller. 2016. Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *Journal of Biogeography* 43: 2085–2098.