

CalJep: A Spatial Distribution Database of CalFlora and Jepson Plant Species

JOSHUA H. VIERS, JAMES H. THORNE, AND JAMES F. QUINN¹

ABSTRACT

CalJep is a spatially enabled database that reconciles or cross-walks the two prominent electronic plant distribution lists for California: CalFlora and Jepson. We intersected the distribution information from the two data sources to create a refined spatial distribution repository that can be used to examine patterns of plant diversity, distribution ranges of individual plant species or infrataxa, or vegetation associations. These data will allow scientists and resource managers to examine potential range maps for non-native plants, create range maps for plant species of restoration interest, and corroborate lines of evidence for determining appropriate management and conservation activities. We present here a detailed description of the methods used to create the CalJep geodatabase, data rendered from its creation, and a discussion of its applicability to a wide range of biogeographical and ecological questions, including restoration planning and adaptive management for the Bay-Delta ecosystem. CalJep records 7,887 plant species, subspecies, and varieties mapped onto 228 ecological subunits with corresponding distributional information for vascular plant species at varying levels of confidence. Information derived from this geodatabase is inherently as accurate as the digital floras used to create it; hence, its utility is best realized when implemented at the regional or statewide scale. CalJep provides a previously unavailable service to vegetation science in California and to resource managers operating within the Bay-Delta ecosystem.

KEY WORDS

California; Bay-Delta; taxonomic databases; GIS; geodatabases; flora; endemism; natives; invasives; range maps

INTRODUCTION

Accurate descriptions of the geographical distributions of vegetation types have been of interest to ecologists since von Humboldt and Wallace in the 19th century (in Lomolino and others 2004). Floristic checklists are one of the most important tools in developing these descriptions (McLaughlin 1989). In many cases these lists form the basis for subsequent

taxonomic floras of different regions. California has a tradition of formal vegetation classification and plant biogeography dating from at least Wieslander (1935). Munz (1968), Whittaker (1961), Stebbins and Major (1965), Griffin and Critchfield (1972), Raven and Axelrod (1978), and Major and Barbour (1988) are among those who have since documented

1. Department of Environmental Science and Policy, University of California, Davis

patterns of plant distribution and the high species richness and levels of endemism found in the state of California and the California Floristic Province.

Efforts to map the botanical biogeography of the world, and thus California, have fallen into two fields: potential and actual vegetation. Potential vegetation maps predict the distribution of vegetation as driven by a variety of environmental factors, and are exemplified by the works of Kùchler (1964, 1977), Bailey and Cushwa (1982), Omernik (1986), Bailey and others (1998), and Ricketts (1999). Actual vegetation maps directly record existing vegetation and are based on imagery or field observations. Examples of actual vegetation maps for California include Wieslander (1935), Davis and others (1998), and Thorne and others (2004), wherein each effort relies upon both field observation and synoptic mapping techniques to qualify and demarcate dominant vegetation at a given point in time at a given place.

Maps of individual species ranges can also be considered potential or actual in this same vein. Systematic development of synoptic species-level range maps for vascular plants in California has not been undertaken to our knowledge, although written descriptions of species ranges and habitats are found in floras written for the state, including Munz and Keck (1959), Munz (1968), and Hickman (1993). The Jepson Manual breaks California into 10 ecoregions with 35 subcoregions intended to reflect major vegetation boundaries and underlying biogeographic patterns in climate, soils, elevation, and landforms. The distribution of a specific plant is described as present or absent in each of these subcoregions. The earlier Munz flora (1968) describes species ranges by county. A digital version of that flora was developed by Lum (1975) and Lum and Richerson (1980), who subdivided some counties that cover large

gradients in environmental conditions to create described ranges on 96 units for the state. This later database was subsequently further developed by Dennis (2000) and is now known as CalFlora.

The Jepson Manual (Hickman 1993) records represent the combined efforts of many researchers who generally published by genus within the book. The nomenclature and range estimates are continually being updated. The Jepson Online Interchange for California Floristics² (hereafter Jepson Online Interchange) serves as the interim mechanism for published updates to the Jepson Manual. Initially put on line in 2001, these data represent the latest published updates and corrections to nomenclature, evolutionary origins, and present distribution. In parallel, CalFlora was constructed to serve as an electronic repository of floristic data within California (Dennis 2000). The CalFlora database includes occurrence information, references to photographs, and numerous ancillary data to support the mission of providing public access to readily available information in support of scientific and conservation efforts. Similar to the Jepson Online Interchange, the CalFlora database represents dynamic data generally cataloged from 1995 through 2002. Most of the original range information is derived from an earlier authoritative flora for California (Munz and Keck 1959; Munz 1968). Our efforts focused only on taxonomic, distributional, and life form data for floristic taxa.

We constructed a relational geospatial database that reconciles the taxonomic names and geographic descriptors between two prominent reference compendia for California plants—the Jepson Manual and its electronic counterpart, the Jepson Online Interchange, and CalFlora. Hereafter we will refer the

2. <http://ucjeps.berkeley.edu/interchange.html>

electronic versions of these resources as CalFlora and Jepson, representing the databases of Dennis (2000) and Moe and others (2000), respectively.

The new geodatabase, CalJep, is a combination of a traditional database with a geospatial component, and permits a variety of spatial renditions of the California flora within a geographical information system (GIS). CalJep permits summary statistics of a variety of species classifiers across geospatial units. This paper presents CalJep, explains the methods used in its development, and illustrates some of its utility.

Indeed, CalJep has already been used to examine patterns of regional plant alpha and beta diversity (Harrison and others 2000; Harrison and Inouye 2002), endemism (Harrison and others 2004), extinction risk from non-native plants (Seabloom and others, in press), patterns of species homogenization (Schwartz and others 2006) and anthropogenic impacts on species richness (Williams and others 2005). Additionally, the original subcounty level allocation (Lum 1975; Richerson and Lum 1980) has been used as the basis for assessing remotely sensed indices as correlates to regional species composition (Walker 1992; Qi and Yang 1999; Fairbanks and McGwire 2004).

MATERIALS AND METHODS

CalJep Database Compilation

We obtained electronic lists of plant names and distribution records from Jepson and CalFlora. We used the full nomenclature identified in each database as the basis for matching taxa across the databases. From Jepson we utilized the attributes: family, full taxonomic name, genus, species, infrataxon, authority, origin (native or non-native), elevation range, and presence/absence in each of the Jepson subcoregions. From

CalFlora we utilized: full taxonomic name, genus, species, infrataxon, authority, origin, life form, elevation, and qualified distribution for each subcounty. We used Jepson as the foremost authority on the botany and biogeography of California's higher plants, over-riding CalFlora where there were discrepancies.

Species recorded in each of these databases were cross-referenced (i.e., cross-walked), producing a list of 7,887 plant names in common for which we could estimate distribution across the intersected 228 map-units. These map-units, called ecological subunits or ESUs, were created by intersecting the subcoregions used in Jepson with the subcounties used in CalFlora. Using this data, we could make estimates of species ranges from the two datasets in a variety of ways, permitting the viewing of consensus range maps which may offer a better estimate of the range of the species. The new range maps, being digital, are comparable to museum and other records for validation purposes.

Note that Jepson nativity and CalFlora origin, life form, and distribution are taxa descriptors not used to establish cross-referenced lists, but for subsequent analysis.

More recent changes to the CalFlora database structure removed reference to a critical piece of our undertaking: distributional reference to California's subcounties (Lum 1975). Specifically, REGIONCODE information was removed in 2002 (Dennis 2000); thus, our version predates this action. Our use of Jepson data is predicated on the publication date of The Jepson Manual, as edited by Hickman (1993). Both CalFlora and Jepson databases continue to be updated. Nevertheless, use of the circa 2000 geographic range descriptors gives a good overview of distribution for most California plant taxa, and certainly reflects patterns of

biogeographic richness and heterogeneity in the California flora.

Taxonomic Cross-walking

The original databases each contain information that we wanted to summarize and analyze within a spatial context. The original attributes found in CalFlora and Jepson included range estimates, information on plant nativity (endemic or introduced), and plant life form (e.g., whether the plant is woody or herbaceous, etc.). To take advantage of data types available in the two datasets, we needed to develop a common index. We used the taxonomic names as the cross-walk mechanism, and conducted an iterative process for name matching.

Taxonomic names from of each species list, Jepson ($n = 8,412$) and CalFlora ($n = 8,363$), were matched using family, genus, species, and infrataxon (i.e., variety and subspecies) information. We preserved these pairings with the creation of a master cross-walk—CalJep—housing a unique record identifier (caljepID) and the respective unique identifiers for the two electronic lists. This primary key is central to permitting cross database queries such as determining the spatial extent of a species as recorded in each database.

We were provided with a digital version of Jepson in ASCII format (as published in 1993), but we have since attempted to update the record entries to taxonomic names represented on the online version of Jepson, which accounts for changes in nomenclature. The tabular data from CalFlora³ were retrieved electronically in ASCII format. Each file was uploaded into Access (Microsoft 2003) and treated as a separate table.

3. <http://www.calflora.org>

Methods for cross-walking consisted initially of matching queries based solely on full taxonomic names, which resulted in 3,562 original matches. These data were appended with subsequent matching routines that used combinations of family, genus, species and infrataxon information and verified through examination. About 90% of the names in the two databases matched using these methods. We examined the remainder and were able to match an additional 413 names using ancillary information to account for name changes, misspellings for a total name cross-walk of 7,887, or about 94% of the records in both Jepson and CalFlora.

Taxa that Jepson lists as “unresolved” are either left out or are included but identified as unresolved. Additionally, 511 records in Jepson do not contain distributional information by subecoregion; however, we included them in the cross-walk to provide a placeholder for future improvements and as a point of reference for other researchers. We also flagged taxonomic homonyms for record entries where a subspecies or variety exists. For example, *Lupinus albifrons* was flagged as a taxonomic homonym as its subspecies surrogate *Lupinus albifrons albifrons* exists in CalJep. All taxonomic homonyms were removed from summaries of infrataxonomic records by ESU ($n = 641$). For record entries that had a single instance in one source and related, but multiple, records in the other source, we coded them as taxonomic equivalents and included only one entry in the relational cross-walk. For example, due to name changes, *Chenopodium hians* is the standard within Jepson, but has two corresponding records in CalFlora: *Chenopodium hians* and *Chenopodium incognitum*. We use only *C. hians* to *C. hians* as the record level association for CalJep and do not include *C. incognitum*'s attributes in the cross-walked table. However, information from

C. incognitum is kept in an associated table in CalJep for future reconciliation efforts ($n = 76$).

Native California species were recorded as such where both data sets agreed (i.e., CalFlora native = True and Jepson nativity = Native); similarly, where both data sets agreed that the taxon was non-native, it was tallied as such. We reviewed all records where combinatorial values from the two sources diverged ($n = 72$), using Jepson to confirm nativity. In some cases, taxa were denoted as naturalized in Jepson, such as the popular cucurbit *Citrullus colocynthis* var. *lanatus* (watermelon). We categorized naturalized entries as non-native. Jepson does not have a category for California endemic plants, these were solely identified by CalFlora, where Range = "CA Endemic." We defined aquatic taxa as ones with the term "aquatic" embedded within their life form notation from CalFlora. We defined herbaceous taxa as ones whose CalFlora life form description included "herb," otherwise we considered the taxon woody. All taxa designations were conducted at the taxonomic level of the combination of these two datasets, therefore these designations apply to the entire range of each species portrayed in the spatial component of CalJep.

CalJep Spatial Cross-walking and Geodatabase Construction

The boundaries of subcoregions defined in The Jepson Manual (Hickman 1993) were originally created with Küchler's (1977) lines as the primary reference. These subcoregions were digitized. Digital boundaries were determined using 1990 Landsat imagery and ancillary digital data at a 1:100,000 map scale (Davis 1995; Davis and others 1998). Digital maps representing CalFlora's county and subcounty boundaries defined by Lum (1975), and formalized by Richerson and Lum (1980), were created and used for a variety of purposes (e.g. Walker 1992; Qi and Yang 1999). There are 58 county-level records, with

18 units subdivided to provide a total of 94 CalFlora map units, where San Francisco and San Mateo counties are combined and an additional unit is introduced to represent the region surrounding Lake Tahoe. Subcounty level divisions were largely defined by topographic boundaries or features within counties, and were generally applied to the larger counties. For example, Tehama County is subdivided by the Sacramento River extending from north to south to create western and eastern subcounty units.

CalJep has a geodatabase construction with two primary components, spatial data, based on vector topology, and tabular data. The spatial data are housed in ArcGIS9.0 personal geodatabase format (ESRI 2005), used to portray and analyze the remainder of the data, and housed in an Access database (Microsoft 2003) with spatial references. We created three spatial feature classes within the geodatabase. These feature classes consist of subcoregions and subcounties, as above, and the intersection of these data to create ecological subunits. There are 35 subcoregions (Hickman 1993) and 94 subcounties (Richerson and Lum 1980). The intersection of the two spatial data resulted in 284 unique combinations of ecological subunits (ESUs) which we further reduced in number by removing small units at the margins of boundary overlap.

Each ESU was assigned a unique, non-sequential numeric identifier; this primary key is central to data retrieval due to the interpretive rules of the geodatabase format (ESRI 2005). It is also worth noting that each ESU may contain several polygons; this is particularly the case for the ESUs comprising the Channel Islands and desert mountain ranges. We removed the "Bay" polygon from the CalFlora subcounty spatial data, as CalFlora does not maintain this as a valid attribute value in the tabular source data.

| Name | Description |
|--------------------------|--|
| CalfloraSubcounties | ESRI GDB Feature Class of California Subcounties |
| CalJepEcoSubUnits | ESRI GDB Feature Class of CalJep Ecological Sub Units |
| JepsonSubcoregions | ESRI GDB Feature Class of Jepson Subcoregions |
| tblCalflora | operational Calflora table |
| tblCalJep | California -> Jepson Cross Walk |
| tblCalJepCalfloraDist | CalJepID by California Subcounties used for California Allocation |
| tblCalJepESUGenDefs | Genera Allocation Table to ESUs [Jepson Ecoregion x California Subcounty] concatenated definition values |
| tblCalJepESUSppDefs | Species Allocation Table CalJepID to ESUs [Jepson Ecoregion X California Subcounty] concatenated definition values |
| tblCalJepESUTaxDefs | Taxa Allocation Table Taxa to ESUs [Jepson Ecoregion X California Subcounty] concatenated definition values |
| tblCalJepExtern | Table of possible taxonomic matches for future CalJep expansion. |
| tblESUAquaTaxaCount | CalJep Aquatic Taxa Count by Possible, Probable, Present Definitions |
| tblESUEndemicTaxaCount | CalJep Endemic Taxa Count by Possible, Probable, Present Definitions |
| tblESUGenCount | CalJep Genera Count by Possible, Probable, Present Definitions |
| tblESUHerbTaxaCount | CalJep Herbaceous Taxa Count by Possible, Probable, Present Definitions |
| tblESULandscape | ESU Landscape Variables |
| tblESUNativeTaxaCount | CalJep Native Taxa Count by Possible, Probable, and Present Definitions |
| tblESUNonNativeTaxaCount | CalJep NonNative Taxa Count by Possible, Probable, and Present Definitions |
| tblESUSppCount | CalJep Species Count by Possible, Probable, Present Definitions |
| tblESUTaxaCount | CalJep Taxa Count by Possible, Probable, Present Definitions |
| tblESUTerraTaxaCount | CalJep Terrestrial Taxa Count by Possible, Probable, Present Definitions |
| tblESUWoodyTaxaCount | CalJep Woody Taxa Count by Possible, Probable, Present Definitions |
| tblGenusQuercus | Example of Species Counts within Genus Distribution by Probable, Possible, Present [Quercus] |
| tblGenusQuercusSpp | Example of Species Level Distribution within Genus by ESU [Quercus] |
| tblGenusRubusSpp | Example of Species Level Distribution within Genus by ESU [Rubus] |
| tblGenusVerbenaSpp | Example of Species Level Distribution within Genus by ESU [Verbena] |
| tblJepson | operational Jepson Table |

Figure 2. CalJep geodatabase with tabular descriptions. This graphic shows the contents of the MS Access database tables, formatted as a Personal Geodatabase, with descriptions of each table. A number of representative queries are also included as an example of database uses and contents, including those queries used to create Tables 3–10

that contains spatial distributional definitions (e.g. *tblCalJepESUTaxDefs*). In *Jepson* the distributional definition for a taxonomic record to its subcoregions is stored as a binary response of 0 for absence and 1 for presence (Moe and others 2000). In *CalFlora* the value is stored as one of four possibilities where each value conveys both a sense of confidence and basis of determination (Dennis 2000). A value of 1 is defined as present; a value of 2 is defined as distribution uncertain; a value of 6 is defined as presence inferred, generally from Hickman (1993); and a value of F is defined as not present (Table 2, Part B). As such, the combined values for each species record for each ESU confer a level of confidence to the prediction of presence or absence. We constructed four definitions of distribution,

presented in Table 2, based on the combinatorial possibilities.

CalJep uses four distribution definitions: “present,” “probable,” “possible,” and “not recorded” (Table 2). These definitions are necessary to understand limits of the respective databases. A “possible” distribution includes any combination of presence information, including prediction of presence by one but not the other, and any level of agreement. A “probable” distribution contains *Jepson* value of 1 and *CalFlora* values of 1, 2 or 6, and “present” indicates that *Jepson* indicates presence (value of 1) and *CalFlora* indicates the highest level of certainty as well (value of 1). The “not recorded” category indicates that there is insufficient information

available to determine a recorded presence. In CalJep, this definition represents the absence of the taxon from a given subregion, remembering that none of the subregions has been completely inventoried. Each of the three positive levels of distribution can be spatially mapped for any of the biogeographic themes such as: species range maps, the summation

of species or subspecies by ESU and various views of endemic taxa counts by ESU. In addition, we tallied the number of infrataxa and species per ESU and have constructed queries to tabulate the number of California state endemic, non-native, woody, herbaceous, aquatic, and terrestrial taxa.

Table 1. Relevant tabular data tables in CalJep. The following entries describe the primary attribute tables in CalJep, which are related to the spatial feature classes (Figure 1) through a relational database structure using primary fields as key indexes

| <i>Table Name</i> | <i>Description</i> | <i>Relationship</i> |
|---------------------|---|---|
| tblCalJep | Cross-walk of tblJepson and tblCalflora | uses caljepID as primary key; relates to tblJepson via jepsonID and tblCalflora via CALRECNUM ; this is the primary cross-walk |
| tblJepson | Tabular data from <u>Jepson</u> | uses an imposed jepsonID as primary key, which is also contained in tblCalJep |
| tblCalflora | Tabular data from <u>CalFlora</u> | uses CALRECNUM as primary key, which is also contained tblCalJep |
| tblESULandscape | Landscape variables by ESU | Uses ESUKey as primary key; includes positional information, areal estimates, etc. |
| tblCalJepESUTaxDefs | ESU definitions from <u>Jepson</u> and <u>CalFlora</u> for taxonomic records | uses both caljepID and ESUKey as relationship keys to tblCalJep and CalJepEcoSubUnits , respectively |
| tblCalJepESUSppDefs | ESU definitions from <u>Jepson</u> and <u>CalFlora</u> for unique genus-species | uses ESUKey as relationship key to CalJepEcoSubUnits ; uses Genus and Species to relate to tblJepson |
| tblCalJepESUGenDefs | ESU definitions from <u>Jepson</u> and <u>CalFlora</u> for unique genera | uses ESUKey as relationship key to CalJepEcoSubUnits ; uses Genus to relate to tblJepson |
| tblESUTaxaCount | Tally of taxonomic records, less taxonomic homonyms, for each ESU | uses ESUKey as relationship key to CalJepEcoSubUnits |
| tblESUSppCount | Tally of unique genus and species for each ESU | uses ESUKey as relationship key to CalJepEcoSubUnits |
| tblESUGenCount | Tally of unique genera for each ESU | uses ESUKey as relationship key to CalJepEcoSubUnits |
| tblCalJepExtern | Record entries from either <u>CalFlora</u> or <u>Jepson</u> with multiple, related taxonomic entities | Records both CALRECNUM and jepsonID for future expansion of the cross-walk database |

Table 2. CalJep Species Distribution Categories. CalJep identifies four categories for the distribution of a species. The categories refer to a taxon's status within each map unit, and are derived from crossing the Presence/Absence (1 or 0) records in Jepson with the four levels of species presence information recorded in CalFlora (Part B). CalJep's four categories of species presence indicate: 1) a species is "present" when Jepson indicates a presence for a map unit and there is a herbarium record (CalFlora); 2) a species is "probable" when there is a presence record in Jepson for the map unit and is noted in a published report, or thought to be there (CalFlora) (note that the probable category is inclusive of the present category but not vice versa); 3) a "possible" species is one indicated to be present in one or both of the original datasets (Jepson or CalFlora, or both); and 4) both datasets agree the species is absent from the map unit, which we note as "not recorded."

| Part A | CalFlora Level 1 | CalFlora Level 2 | CalFlora Level 6 | CalFlora Level F |
|-------------------|---|------------------|------------------|------------------|
| Jepson Present, 1 | Present | Probable | Probable | Possible |
| Jepson Absent, 0 | Possible | Possible | Possible | Not Recorded |
| Part B | CalFlora Species Distribution Categories (adapted from Dennis 2000) | | | |
| 1 | Probably present or definitely present. For data transcribed from CNPS inventory data, presence in a particular subcounty region may be less certain than indicated if QUAD data was not available for that county. CNPS inventory data are considered verified at the county level in all cases, but not necessarily at the subcounty level. | | | |
| 2 | Presence uncertain or distribution or identity from Dennis (2000) | | | |
| 6 | Presence implied by <u>Jepson</u> manual range description but not confirmed by other sources. | | | |
| F | Not present | | | |

Validation

To provide some sense as to the accuracy of using CalJep, inherently a combination of two prominent data sources in CalFlora and Jepson, we compared its potential distribution values to a fixed locale with a relatively well known species list of vascular plants. We chose the Cosumnes River Preserve as our test case as it represents a well studied location within the California Bay-Delta and it is managed for its abundant riparian and oak woodland vegetation in a semi-natural setting. We compared vascular plant infrataxa known to exist on the preserve ($n = 388$, via Tu 2000 and Keller 2003) against CalJep distribution definitions for the same geographic location, or ecological subunit. We examined rates of agreement across definitions and determined an overall accuracy for this particular comparison to give readers and geodatabase

users a measure of robustness as translated from the intersection and union of CalFlora and Jepson.

RESULTS

CalJep Taxonomic Summary

From the original species lists, Jepson ($n = 8,412$) and CalFlora ($n = 8,363$), our cross-walk resulted in 7,887 unique record matches (94% success). Of these matches, we identified taxonomic homonyms and removed them from the analysis (e.g., removed *Quercus garryana* and retained *Quercus garryana garryana* when analyzing subspecies distribution patterns), which resulted in 7,224 unique infrataxonomic matches at the subspecies and variety level where possible. Taxa that could not be matched at the subspecies level were matched at the species level. At the genus-species level, subsuming

infrataxa, we recorded 5,988 unique entries for CalJep. CalJep has records for 1,278 genera and 172 families. We also reconciled attribute values contained in both primary data sources treating nativity. Of the 7,887 records cross-walked in CalJep, 72 records had some level of disagreement, either containing null values or mismatching values. We reviewed each of the 72 records using the Jepson Online Interchange or The Jepson Manual (Hickman 1993) as records of decision. Results of this review are shown in Table 3.

CalJep Summary Statistics by ESU

Maps 1-28 represent the accumulated information represented by the versions of CalFlora and Jepson used. The information has an unknown level of accuracy, which may vary from species to species, and within and across sub-ecological units. The maps derived

Table 3. Designation of CalJep nativity by combinations of reported values from CalFlora and Jepson by number of taxonomic records in each category. The total number of native taxa identified is 6,828 and total number of non-native taxa identified is 1,059.

| <i>CalJep Native Designation</i> | <i>CalFlora Native Column</i> | <i>Jepson Nativity Column</i> | <i>No. of Records</i> |
|--|---------------------------------------|---------------------------------------|---------------------------|
| FALSE | f | intro | 1,033 |
| FALSE | f | naturalized | 11 |
| FALSE -- Reviewed | f | | 3 |
| FALSE -- Reviewed | t | | 7 |
| FALSE -- Reviewed | t | intro | 5 |
| TRUE | t | native | 6,771 |
| TRUE -- Reviewed | f | | 2 |
| TRUE -- Reviewed | f | native | 1 |
| TRUE -- Reviewed | t | | 25 |
| TRUE -- Reviewed | t | intro | 29 |

from CalJep are inherently as accurate as the digital floras used to create them; hence, the mapped data—and the utility of CalJep—are best realized when implemented and interpreted over broad areas such as in regional or statewide contexts.

We tallied a variety of CalJep records by ESU: the number of species; the number of infrataxa (subspecies, varieties or species); the number of state endemic infrataxa; the number of native infrataxa; the number of non-native infrataxa; the number of woody and herbaceous infrataxa; and the number of aquatic and terrestrial infrataxa. We placed this cross-referenced information into tables and generated corresponding maps.

Because the length of the data in the tables is too large to be legibly presented in the body of this paper, we present a brief description of each table and associated maps instead. The actual tables and maps are provided as associated Adobe PDF files, which are hyperlinked to this document at first reference.

CalJep Spatial Summary

The geographic combining of Jepson Subcoregions ([Map 1](#)) and CalFlora Subcounties ([Map 2](#)) resulted in 228 unique combinations of ecological subunits contained within the CalJep geodatabase ([Map 3](#)). ESUs vary in size (minimum 71.2 km²; maximum 12,857.4 km²; mean 1795.4 km²; standard deviation 1875.0 km²; *n* = 228) with physical attributes reflecting the state of California in regards to elevation range, proximity to coast, latitude and longitude.

CalJep Biogeographic Summary by ESUs

Table 4. Spatial Definition of CalJep ESUs

This table presents the unique codes from Jepson and CalFlora spatial units used to create the CalJep ESUs. The attributes of this table include the primary numeric key (i.e.,

ESUKEY) with corresponding symbol codes for Jepson (JEP CODE) and CalFlora's subcounties (CAL CODE). This table also includes ancillary information about ESU Area (km²), the longitude and latitude of each ESU centroid (in decimalized degrees), distance from coast (km), minimum and maximum elevation (m), and the number of polygons comprising the ESU. Distance to coast is calculated from ESU centroids, so anomalies do occur such as the measures for the two Channel Island units where the centroid is located over the Pacific Ocean. Distance to coast in this case is the distance from the centroid to the closest island ($n = 2$). There are also a number of ESUs with more than one polygon, as listed in this table; therefore, each of the centroids representing these multi-feature units will have a geographic center that may or may not fall within the ESU boundary.

[Table 5. Total Number of Species and Infrataxa Recorded for each ESU](#)

The total for species across the three classes of distribution agreement between Jepson and CalFlora is given first, followed by three columns that show the total number of species and infrataxa (subspecies, varieties) across the three levels. ESU key codes and codes for Jepson and CalFlora map units are provided for reference. Species counts from Table 5 are illustrated in [Maps 4–9](#).

[Table 6. Total Number of Native Infrataxa in each ESU](#)

The sum of native subspecies and varieties in each CalJep unit are recorded for each of the three classes of spatial distribution. ESU key codes and codes for Jepson and CalFlora map units are provided for reference.

[Table 7. Total Number of California Endemic Subspecies or Varieties](#)

The endemic subspecies or varieties in each ESU are presented. Their totals are calculated for the three classes of spatial

distribution. Data from Table 7 are illustrated in [Maps 10–12](#).

[Table 8. Number of Non-native Infrataxa per ESU](#)

The number of non-native infrataxa, as defined in Table 2, is recorded for each ESU according to the three classes of spatial distribution. This table corresponds to [Maps 13–15](#).

[Table 9. Number of Woody and Herbaceous Infrataxa Tallied for each ESU](#)

Subspecies and varieties of woody and herbaceous plants are tallied by ESU according to the three classes of spatial distribution.

[Table 10. Number of Terrestrial and Aquatic Species Present in each ESU](#)

Terrestrial and aquatic subspecies and varieties are summed per ESU according to the three classes of spatial distribution.

CalJep Range Maps

Range maps in CalJep can be produced at the genus, species, or subspecies level. At each taxonomic level, the range map can represent possible, probable, or present range estimates at the same time. (See Table 2 for spatial distribution definitions.) In [Maps 4–15](#) taxonomic richness was displayed by different distribution definitions. Similarly, in [Maps 16–18](#), we show richness for the genus *Quercus* at varying levels of confidence. However, at the taxonomic level, we portray these distinctive levels as a singular range map. For example, and as tabulated in Table 11, we present range maps for specific taxa within *Quercus*, four species and one subspecies. These distributional ranges are shown in [Maps 19–23](#).

Table 11. Area (km²) of distribution of selected taxa in California. Areal estimates for four species and a subspecies of oak and their genus within California according to the four levels of distribution certainty in CalJep and for two native and two non-native plants. These data also correspond with Maps 19–27.

| <i>CalJep Taxonomic Record</i> | <i>Present</i> | <i>Probable</i> | <i>Possible</i> | <i>Not Recorded</i> |
|--|----------------|-----------------|-----------------|---------------------|
| <i>Quercus</i> spp. | 321,820.9 | 321,820.9 | 391,539.6 | 17,803.9 |
| <i>Quercus douglasii</i> | 104,119.9 | 108,486.7 | 186,275.5 | 223,068.0 |
| <i>Quercus kelloggii</i> | 183,033.0 | 184,026.2 | 260,837.3 | 148,506.2 |
| <i>Quercus lobata</i> | 149,943.8 | 154,079.5 | 211,861.4 | 197,482.1 |
| <i>Quercus garryana</i> | 66,319.6 | 98,042.8 | 188,088.3 | 221,255.2 |
| <i>Quercus garryana</i> var. <i>garryana</i> | 0.0 | 70,876.0 | 102,587.1 | 306,756.4 |
| <i>Rubus discolor</i> | 0.0 | 0.0 | 267,896.4 | 141,447.1 |
| <i>Rubus ursinus</i> | 125,715.2 | 252,534.4 | 278,733.4 | 130,610.1 |
| <i>Verbena californica</i> | 1,550.1 | 1,550.1 | 6,807.5 | 402,536.0 |
| <i>Verbena litoralis</i> | 21,289.5 | 21,289.5 | 80,512.3 | 328,831.2 |

We also present two combinations of native and introduced species to illustrate the differences in map utility. *Rubus ursinus* and *R. discolor* represent a species that are, respectively, native and invasive to California (Table 11, [Maps 24 and 25](#)). Note that the invasive species is only represented by the possible class distribution as it has no better resolution records in the CalFlora data. *Verbena californica* represents a relatively narrow endemic, recorded in five ESUs for the possible distribution definition and only one ESU for the present definition. *Verbena litoralis*, on the other hand, is an introduced species with more extensive records in both databases (Table 11, [Maps 26 and 27](#)).

One application that may be of use to resource managers working within CALFED Bay-Delta Program jurisdictional management zones is presented in [Map 28](#), a graphic of the total numbers of non-native vascular plants recorded in each ESU, overlaid by CALFED management zones. As shown, the Central

and South San Francisco Bay management zone harbors an excess of 500 non-native infrataxa. The number of non-native infrataxa diminishes as one continues east and increases in elevation, showing the tangible front of invasion. Maps such as this one, when coupled with additional spatial information, can be used to plan eradication efforts by resource managers with the CALFED Bay-Delta program. Alternately, a similar map could be constructed indicating expected floral diversity of wetland obligates, which could then be juxtaposed against biodiversity restoration goals in wetland habitats.

Validation

We successfully identified 388 taxa listed as present on the Cosumnes Preserve represented in the CalJep geodatabase. We determined the location of the Cosumnes River Preserve to be in the intersection of CalFlora's Sacramento subcounty and Jepson's Sacramento Valley subecoregion (ESU 292). In a cross-comparison of known presences on

the Cosumnes River Preserve with the cataloged distribution of those taxa within the CalJep geodatabase we observed 90 infrataxa at the “present” designation, 97 at the “probable” level, and 152 at the “possible” definition. There were 49 taxonomic records listed as present at the Cosumnes River Preserve that were “not recorded” by either CalFlora or Jepson. In all, when compared to the 388 known taxa at the preserve, the omitted records represent 12.63% of the list (see Table 12). This represents an overall accuracy of 82.3% using the cross-walked taxonomy in CalJep to a known and near exhaustive species list at a given locale of restoration and management importance.

DISCUSSION

The conceptual framework of CalJep is simple: range estimates derived from two

sources are linked to GIS for spatially explicit digital representation. CalJep is, to our knowledge, the first database application designed to make the Jepson and CalFlora data cross-compatible and rendered to a spatial template for display. The ability to identify narrower species ranges through the intersection of occurrence agreement between the two databases provides a new tool for biogeographers interested range modeling questions. The utility of this approach is evident when examining diversity, endemism, and invasive plants by ESU.

CalJep range estimate categories are analogous to actual and potential ranges, although they are strictly speaking all potential. “Present,” the highest level of certainty, is the

Table 12. Contingency Table of Validation Effort. Contingent frequencies of Cosumnes River Preserve taxonomic records against potential CalJep infrataxa distribution at the “possible” designation for the selected ecological subunit (ESU 292). Numbers in bold represent the number of taxonomic records in each category; these data indicate an overall agreement of 82.3% for potential vegetation at the Cosumnes River Preserve.

| | <i>Taxa Absent at CRP According to CRP List</i> | <i>Taxa Present at CRP According to CRP List</i> | | |
|--|---|--|------------------|--------------------------|
| | 6139 | 49 | 6188 | No. of Taxonomic Records |
| Taxa Absent within CRP ESU According to <u>CalJep</u> | 77.84 | 0.62 | 78.46 | % of Taxonomic Records |
| | 81.86 | 12.63 | | CRP % |
| | 99.21 | 0.79 | | <u>CalJep</u> ESU % |
| | 1360 | 339 | 1699 | No. of Taxonomic Records |
| Taxa Present within CRP ESU According to <u>CalJep</u> | 17.24 | 4.3 | 21.54 | % of Taxonomic Records |
| | 18.14 | 87.37 | | CRP % |
| | 80.85 | 19.95 | | <u>CalJep</u> ESU % |
| No. of Taxonomic Records | 7499 | 388 | Overall Accuracy | |
| % of Taxonomic Records | 95.08 | 4.92 | | |

most well-documented range map measurement, but it may lack collection effort or specimens have not been entered into the CalFlora version used here. The “probable” category, where both datasets agree but CalFlora indicates lower level of certainty, is the closest to the concept of actual range and is the best category for identifying native species’ ranges. This is because actual records of species locations are relatively poorly documented, whereas the expert knowledge represented in the original datasets is actually quite good. “Probable” range estimates are closer to the concept of a potential range as they represent the broadest estimate of where a species might occur. The “Possible” category, where only one or the other dataset need contain information on a species’ distribution (or both), is best used for poorly documented species such as non-native or invasive species. Many non-natives are simply not collected for herbarium collections so their ranges are best known from anecdotal accounts. However, these species may already be widely distributed or have large potential for range expansion, features that are best portrayed in CalJep by the “possible” range estimate category.

The ESUs vary in size and thus are non-uniform samples and cannot be used as such (e.g., ordination) without additional manipulation to account for area differences. Some ESUs include relatively small areas of elevated lands, but are attributed with species that inhabit higher elevations; many boundaries, driven by the political origins of counties, are non-biophysical in form, thus lending a sense of arbitrariness to the resulting maps. However, as the ESUs permit an order of magnitude finer display of plant distribution patterns than was available prior, it is worth the introduction of some arbitrary lines.

The digital range maps in CalJep have many uses. They are analogous to the

California Wildlife Habitat Relationship Model (CWHHR, Mayer and Laudenslayer 1988; California Department of Fish and Game 2002) range maps for vertebrates thereby enabling resource managers throughout California to use CalJep. Regional management groups of the CALFED Bay-Delta Program will be able to glean information for their management areas, such as estimates of the spread of invasive species across the region (Map 24). CalJep can also be used to come up with lists of potential species in an ESU against which local species lists can be compared, an application which can start the process of comprehensive accuracy assessments for the state floristic repositories (but which is beyond the scope of this paper). ESU lists could further be refined to identify potential rare plants per ESU, to screen for regions in the state that have particularly high levels of endemism (Map 12) or species diversity (Map 6), and to identify areas of the state and/or species that have not been well documented or collected.

CalJep can be useful for restoration and remediation of impacted landscapes. Stohlgren and others (1999) have shown that regions rich in species and high in endemism are susceptible to exotic species invasion. The CalJep geodatabase can produce species counts for native, endemic and non-native species by ESU (similar to Dark 2000), which in turn can be coupled with spatial data depicting vulnerable riparian and wetland habitats adjacent to agricultural areas. This approach could be useful in the California Bay-Delta region where concentrated agricultural activities operate adjacent to riparian areas and vernal pools. The compilation of multiple species range maps, also possible, will permit an estimate of the distribution of range sizes found in California. Those species that extend beyond the borders of the state would have to be discarded in such an analysis, but we hope to shortly extend the boundaries of CalJep to include at least the California Floristic Province

(Stebbins and Major 1965; Raven and Axelrod 1978).

CalJep range maps do not convey where populations are more or less dense. In CalFlora, if a single observation is made in a county, that county then registers the presence of that taxon. Jepson catalogs its holdings in similarly coarse units, although the editors will generally exclude species that mostly stop at a border but for which there may be a few records on the “unpopulated” side. The utility of CalJep then is to refine both the areal unit of accounting and its level of specificity. Maps such as Griffin and Critchfield (1972), however, are better able to portray a mix of broad and sparse populations in the distribution range of a species. The CalJep range estimates can portray both the union and intersection of the range estimates from the original versions of CalFlora and Jepson; thus, each should be used for the analyses to which it is most suited. For example, common weedy species do not show up often in CalFlora at level 1 but do appear with a level 2 designation; the union of CalFlora and Jepson, as represented as a “possible” designation in CalJep is, therefore, the appropriate distribution definition to be used for analyzing those species distributions.

The range maps recorded in CalJep represent one of three types of portrayals of species distributions: range maps, point locality records, and modeled distributions based on a variety of predictor variables. Modeled plant ranges, such as derived from GARP (Stockwell 1999) can potentially be validated by CalJep. CalJep is currently being used to help inform modeling efforts of plant response to climate change in California (L. Hannah, Conservation International, pers. comm.). Furthermore, CalJep may have the potential to be used to represent the potential distribution of invasive species. Note that *Rubus discolor* (Map 25, a “possible” distribution) is shown to have been reported

across a wide portion of the state. Whether the plant has reached all corners of this distribution is not answered here, but the range map implies that it might be able to do so. The use of extensive herbaria records, when digital and georeferenced, can permit a refinement of the range maps presented here (see Riemann and Ezcurra 2005 for an example of the use of herbaria records in combination with polygon maps).

The CalJep database captures a snapshot of ongoing data development and has certain limitations, particularly the taxonomy and range maps of many species that are under revision in Jepson. These revisions include naming of taxa not yet described and the regrouping of taxa into finer (subspecies and varieties) or coarser (species) levels of organization. Researchers at the Jepson Herbarium have identified over 1,300 names of unresolved status which will be reviewed for possible inclusion in the next edition of The Jepson Manual (B. Baldwin, Jepson Herbarium, pers. comm.). We do not attempt to answer any questions that taxonomists are still working on nor reconcile differences created between databases due to taxonomic changes. We designed the database to be useable at either the species level or at the taxonomic record (subspecies or variety) level, in part to deal with the problem of changing taxonomic records. Examining CalJep at the species level permits the “rolling up” of the taxon list, avoiding the bulk of current taxonomic revisions that mostly involve subspecies and varieties. CalJep can potentially serve as a framework for future taxonomic and geographic information refinements since new data can be added to the system as data are made available.

Both CalFlora and Jepson are dynamic repositories of knowledge. The nomenclature is continually being updated and new records of species’ locations continue to modify range

estimates. This dynamic nature makes them the likely best sources of comprehensive plant data for the state and also makes them difficult to validate. Both represent accumulated knowledge of multiple individuals, which is continually under modification. The CalJep geodatabase is an effort to formalize some of the source records so they may be analyzed in various ways. We do not attempt to conduct a comprehensive validation of the described species ranges, relying instead on both original resources' own internal reviews. We recommend that, when the California herbaria have georeferenced their specimen collections, these be used to conduct a comprehensive accuracy assessment.

In conclusion, we hope that uses for the distribution data that we have not anticipated will be recognized by the release of the geodatabase.

ACKNOWLEDGEMENTS

We would like to thank a number of individuals who have made these data and this manuscript possible: Susan Harrison, Dick Moe, Bruce Baldwin, Ann Dennis, David Ackerly, Eric Seabloom, Jack Williams, Kendi Davies, Hugh Safford, Rachel Hutchinson, Sydney Vergis, Marat Gubaydullin, Ingrid Hogle, Mark Schwartz, Jon Price, Dean Fairbanks, David Stoms, Frank Davis, Fred Hrusha, and Diana Jacobs. We would also like to thank Tom Stohlgren and two anonymous reviewers for their insight and helpful comments. We would like to acknowledge the passing of David Hudson; a database programmer by trade, David was also a mentor, friend, and father.

The data in "tblJepson" come from The Jepson Manual (Hickman 1993), used by permission of the Jepson Herbarium, with names slightly revised with reference to the Jepson Online Interchange. The Jepson data

are copyrighted by the Regents of the University of California, and are not to be reproduced or stored in retrieval systems without prior written permission of the Jepson Herbarium. Note that the Jepson Herbarium is constantly revising the distribution of species as new data are acquired. Thus, the most authoritative versions of range estimates may have changed since the publication of the data used in the creation of the CalJep database. Please contact the Jepson Herbarium⁴ or visit the Jepson Online Interchange⁵ for updated information.

REFERENCES

- Bailey RG, Cushwa CT. 1982. Ecoregion, land-surface form, and hydrologic unit maps of the United States. Eastern Energy and Land Use Team. Kearneysville (WV): U.S. Fish and Wildlife Service.
- Bailey RG. 1995. Descriptions of the Ecoregions of the United States. Miscellaneous Publication Number 1391. Washington DC: U.S. Department of Agriculture. 108 p.
- Barbour MG, Major J, editors. 1988. Terrestrial vegetation of California. Berkeley (CA): California Native Plant Society.
- California Interagency Wildlife Task Group. 2002. CWHR version 8.0 personal computer program. Sacramento (CA): California Department of Fish and Game.
- Dark SJ. The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis. *Diversity and Distributions* 10:1-9.

4. <http://ucjeps.berkeley.edu>

5. <http://ucjeps.berkeley.edu/interchange.html>

- Davis FW, Stoms DM, Hollander AD, Thomas KA, Stine PA, Odion D, Borchert MI, Thorne JH, Grey MV, Walker RE, Warner K, Graae J. 1998. The California gap analysis project: final report. Santa Barbara (CA): University of California. Available at: http://www.biogeog.ucsb.edu/projects/gap/gap_rep.html.
- Davis FW, Stine PA, Stoms DM, Hollander AD. 1995. Gap analysis of the actual vegetation of California 1. The southwestern region. *Madroño* 42(1):40-78.
- Dennis A. 2000. CalFlora: California flora database. The CalFlora Database, a non-profit organization, Berkeley, California.
- [ESRI] Environmental Systems Research Institute, Inc. 2005. ArcGIS version 9.0. Redlands (CA): Environmental Systems Research Institute, Inc.
- Fairbanks DHK, McGwire KC. 2004. Patterns of floristic richness in vegetation communities of California: regional scale analysis with multi-temporal NDVI. *Global Ecology and Biogeography* 13:221-235.
- Harrison S, Inouye BD. 2002. High beta diversity in the flora of Californian serpentine "islands." *Biodiversity and Conservation* 11:1869-1876.
- Harrison S, Safford H, Wakabayashi J. 2004. Does the age of exposure of serpentine explain variation in endemic plant diversity in California? *International Geology Review* 46:235-242.
- Harrison S, Viers JH, Quinn JF. 2000. Climatic and spatial patterns of diversity in the serpentine plants of California. *Diversity and Distributions* 6:153-162.
- Hickman JC, editor. 1993. The Jepson manual: higher plants of California. University of California Press, Berkeley.
- Keller KE. 2003. Evaluation of riparian restoration success in the Central Valley, CA [Ph.D. thesis]. Available from: University of California, Davis.
- Küchler AW. 1964. Potential natural vegetation of the conterminous United States. Pages col. map; 95 x 149 cm. New York (NY): American Geographical Society.
- Küchler AW. 1977. Natural vegetation of California. Pages col. map; 109 x 194 cm.; Scale 101:101,000,000. Lawrence, Kansas.
- Griffin JR, Critchfield WB. 1972. The distribution of forest trees in California. Pacific Southwest Forest and Range Experiment Station, USDA Forest Service Research Paper PSW 83.
- Lum K-L. 1975. Gross patterns of vascular plant species diversity in California [MS thesis]. Available from: University of California, Davis.
- Mayer KE, Laudenslayer WF, Jr, editors. 1988. A guide to the wildlife habitats of California. Sacramento (CA): California Department of Fish and Game.
- McLaughlin SP. 1989. Natural floristic areas of the western United States. *Journal of Biogeography* 16:239-248.
- Moe RL, Wetherwax M, Jepson WL. 2000. Checklist and distribution maps of taxa covered by the Jepson manual. University and Jepson Herbaria, University of California, Berkeley.
- Munz PA. 1968. Supplement to a California flora. Berkeley (CA): University of California Press.
- Munz PA, Keck DD. 1959. A California flora. Berkeley (CA): University of California Press.

- Omernik JM. 1986. Ecoregions of the conterminous United States. Corvallis (OR): U.S. Environmental Protection Agency, Corvallis Environmental Research Laboratory.
- Qi Y, Yang Y. 1999. Topographic effect on spatial variation of plant diversity in California. *Geographic Information Sciences* 5(3):39-46.
- Raven PH, Axelrod DI. 1978. Origin and relationships of the California flora. Berkeley (CA): University of California Press. Republished in 1995 by California Native Plant Society.
- Richerson PJ, Lum K. 1980. Patterns of plant-species diversity in California—relation to weather and topography. *American Naturalist* 116:504-536.
- Ricketts TH. 1999. Terrestrial ecoregions of North America: a conservation assessment. Washington DC: Island Press.
- Riemann H, Ezcurra E. 2005. Plant endemism and natural protected areas in the peninsula of Baja, California, Mexico. *Biological Conservation* 122:141-150.
- Schwartz MW, Thorne JH, Viers JH. 2006. Biotic homogenization of the California flora in urban and urbanizing regions. *Biological Conservation* 127(3):282-291.
- Seabloom EW, Williams JW, Slayback D, Stoms DM, Viers JH, Dobson AP. In press. The invasion debt: exotic plants and extinction risk in the California flora. *Ecological Applications*.
- Stebbins GL, Major J. 1965. Endemism and speciation in the California flora. *Ecological Monographs* 35:1-35.
- Stockwell DRB, Peters DP. 1999. The GARP modeling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Systems* 13:143-158.
- Stohlgren TJ, Binkley D, Chong GW, Kalkhan MA, Schell LD, Bull KA, Otsuki Y, Newman G, Bashkin M, Son Y. 1999. Exotic plant species invade hot spots of native plant diversity. *Ecological Monographs* 69:25-46.
- Thorne JH, Kennedy JA, Keeler-Wolf T, Quinn JF, McCoy M, Menke J. 2005. A new vegetation map of Napa County using the manual of California vegetation classification and its comparison to other digital vegetation maps. *Madroño* 51(4):343-363.
- Tu IM. 2000. Vegetation patterns and processes of natural regeneration in periodically flooded riparian forests in the Central Valley of California [PhD dissertation]. University of California, Davis. 181 p.
- Walker RE. 1992. Community models of species richness: regional variation of plant community species composition on the west slope of the Sierra Nevada, California [MS thesis]. University of California, Santa Barbara.
- Whittaker RH. 1961. Vegetation history of the Pacific Coast states and the “central” significance of the Klamath region. *Madroño* 16:5-19.
- Wieslander AE. 1935. A vegetation type map of California. *Madroño*. 3(3):140-144.
- Williams JW, Seabloom EW, Slayback D, Stoms DM, Viers JH. 2005. Anthropogenic impacts upon plant species richness and NPP in California. *Ecology Letters* (8):127-137.