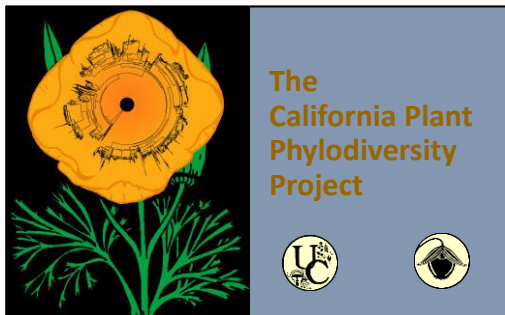


Spatial phylogenetics of the native California flora: integrating ecology, evolution, and conservation

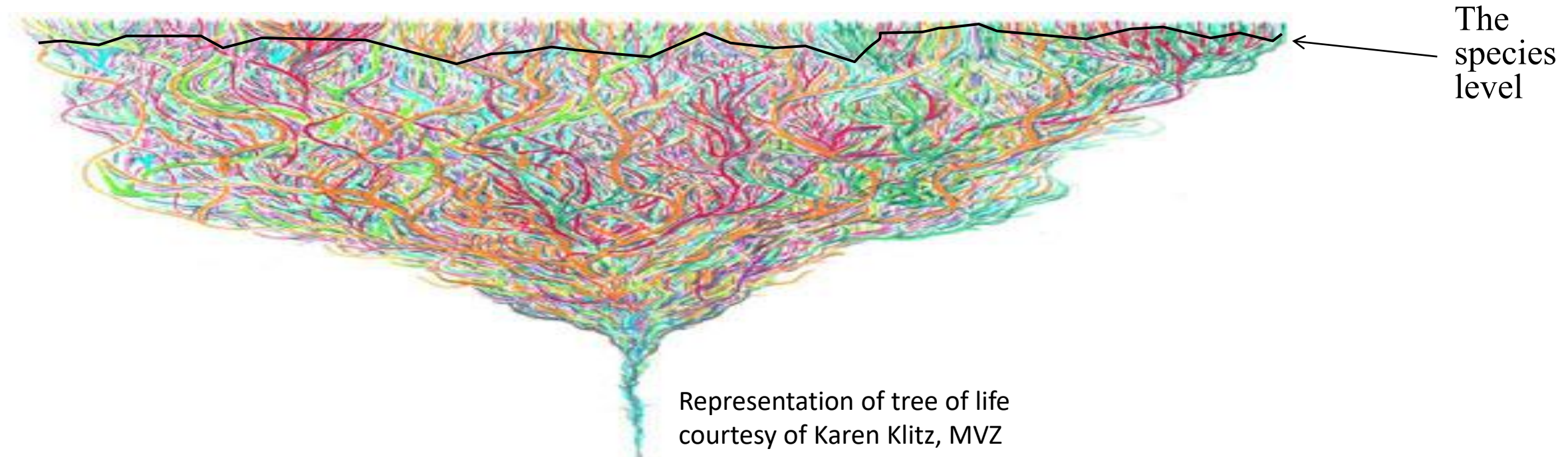
Brent D. Mishler, Matthew Kling, Bruce Baldwin,
David Ackerly, and Andrew Thornhill

University and Jepson Herbaria
University of California, Berkeley



Biodiversity is not just species

- Biodiversity is the *whole tree of life*, not just named species.
- There are lineages smaller and larger than the traditional species level.
- Species are not comparable among lineages, just an arbitrary cut-off somewhere along a branch in the tree of life.



Endemism isn't just about species, either

- Likewise, endemism is not just about species, even though virtually all studies on endemism to the present focus solely on species.
- Clades at all levels can be endemic to a greater or lesser extent, and all levels are relevant to discovery and evaluation of centers of endemism (note: we're using a concept of *relative endemism* throughout, i.e., the inverse of range size).
- Endemism, rather than being species-centric, should be more broadly defined to mean *the geographic rarity of that portion of a phylogenetic tree found in a given area*.
- Thus for a full understanding of diversity and endemism, *we need to look at the whole tree*.

Spatial phylogenetics

Investigations on species alone lacks the depth of a phylogenetic approach.

Spatial Phylogenetics combines two main elements, a phylogeny and a spatial dataset representing phylogeny terminals.

Can be applied at any taxonomic and geographic level.

Measures diversity and endemism based on branch length and phylogenetic relatedness.

Because the approach is rank free it doesn't matter what taxonomic levels the terminals represent, as long as they are monophyletic.

Relatively robust to lumping and splitting decisions.

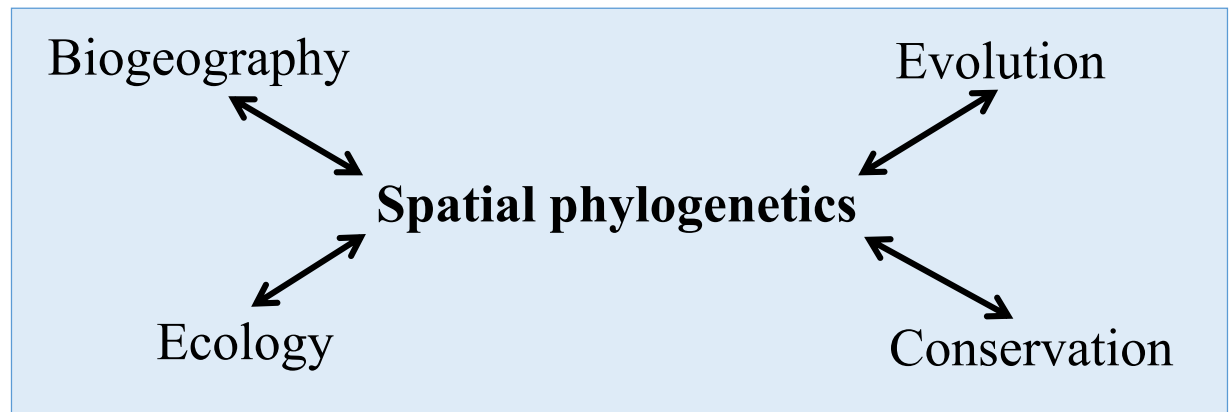
→ *Placing the tree of life on maps*

A “big data” approach enabled by technological advances:

Advances in digitization and availability of natural history museum specimen data

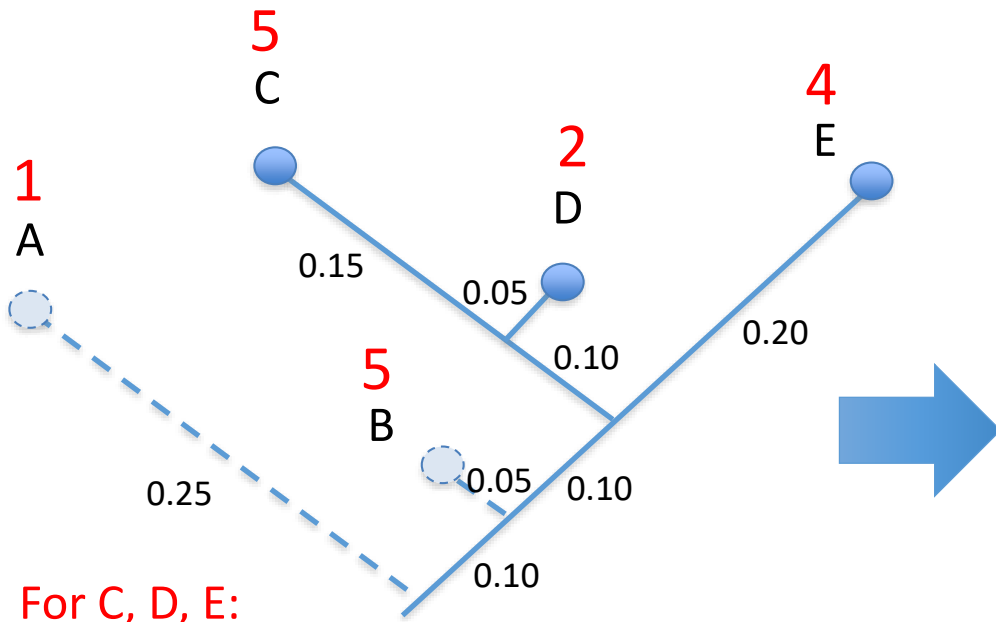
Plethora of DNA in GenBank; advances in mining software.

Major advances in computational methods for both tree-building and tree-using e.g. RAxML



Phylogenetic Diversity

PD (Faith, 1992) is the sum of the branches that connect all terminals in an area to the root of the tree.



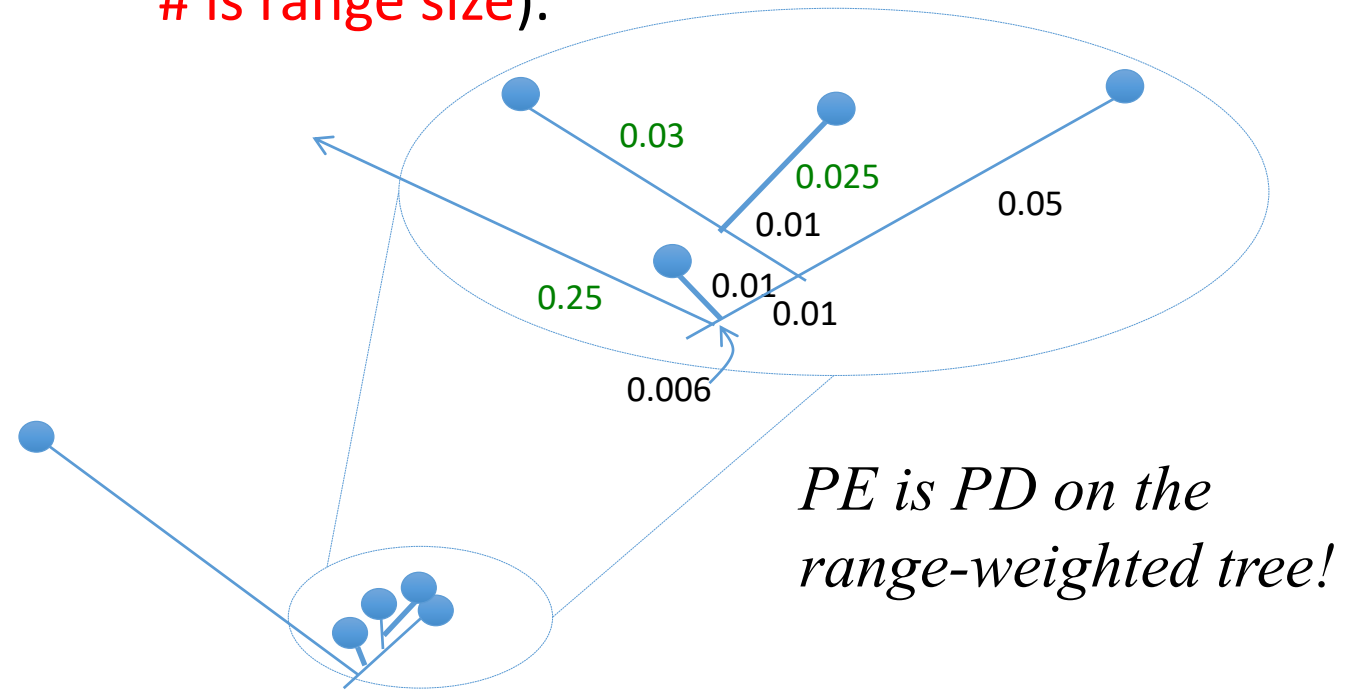
For C, D, E:

Species Richness = 3

PD = 0.70

Phylogenetic Endemism

PE Rosauer et al., (2009) uses a range-weighted tree, i.e., each branch length on the original tree is divided by its range (# is range size).



PE is PD on the range-weighted tree!

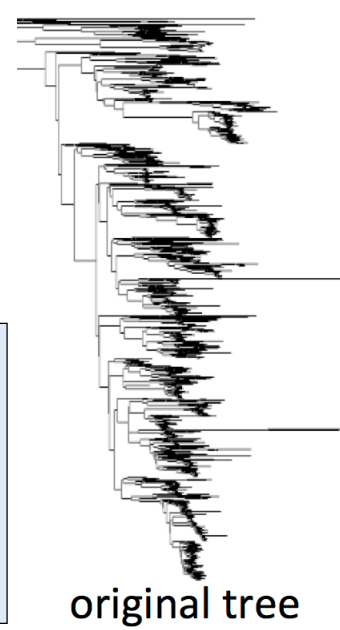
This is a range-weighted tree!

Relative Phylogenetic Diversity and Relative Phylogenetic Endemism

RPD is ratio of A/C:

$$\text{RPD} - \text{relative phylogenetic diversity} = \frac{\text{PD on the original tree}}{\text{PD on a comparison tree with all branch lengths equal}}$$

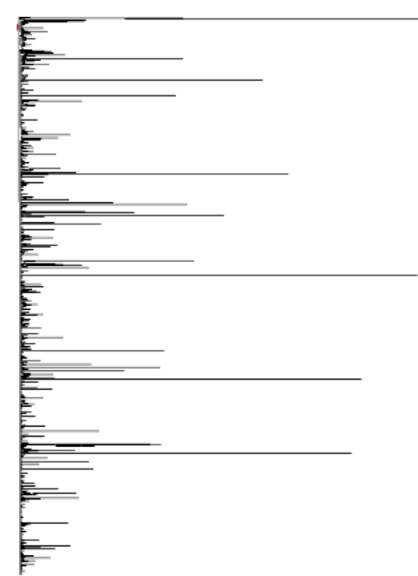
A



original tree



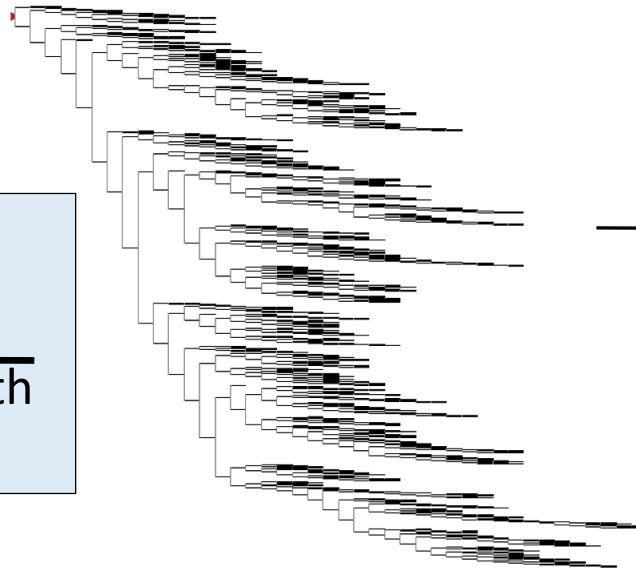
B



range-weighted original tree

RPE is ratio of B/D:

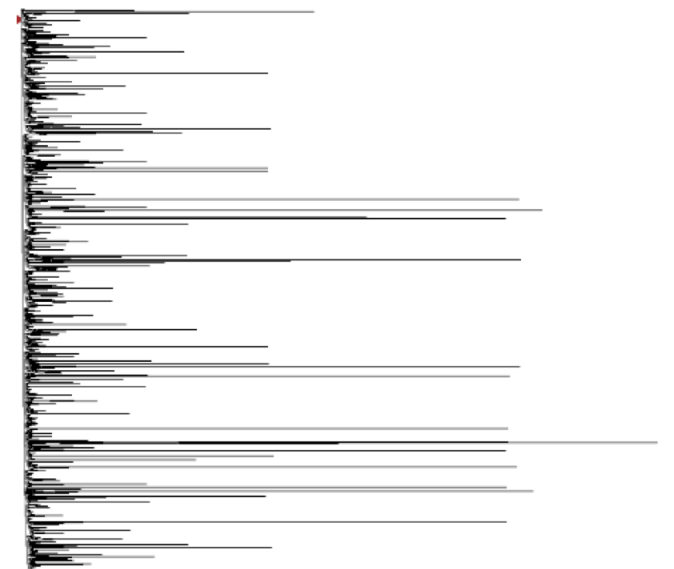
$$\text{RPE} - \text{relative phylogenetic endemism} = \frac{\text{PD on the range-weighted original tree}}{\text{PD on a range-weighted comparison tree with all branch lengths equal}}$$



comparison tree

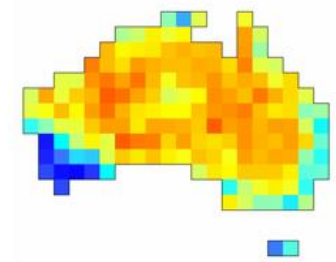


D

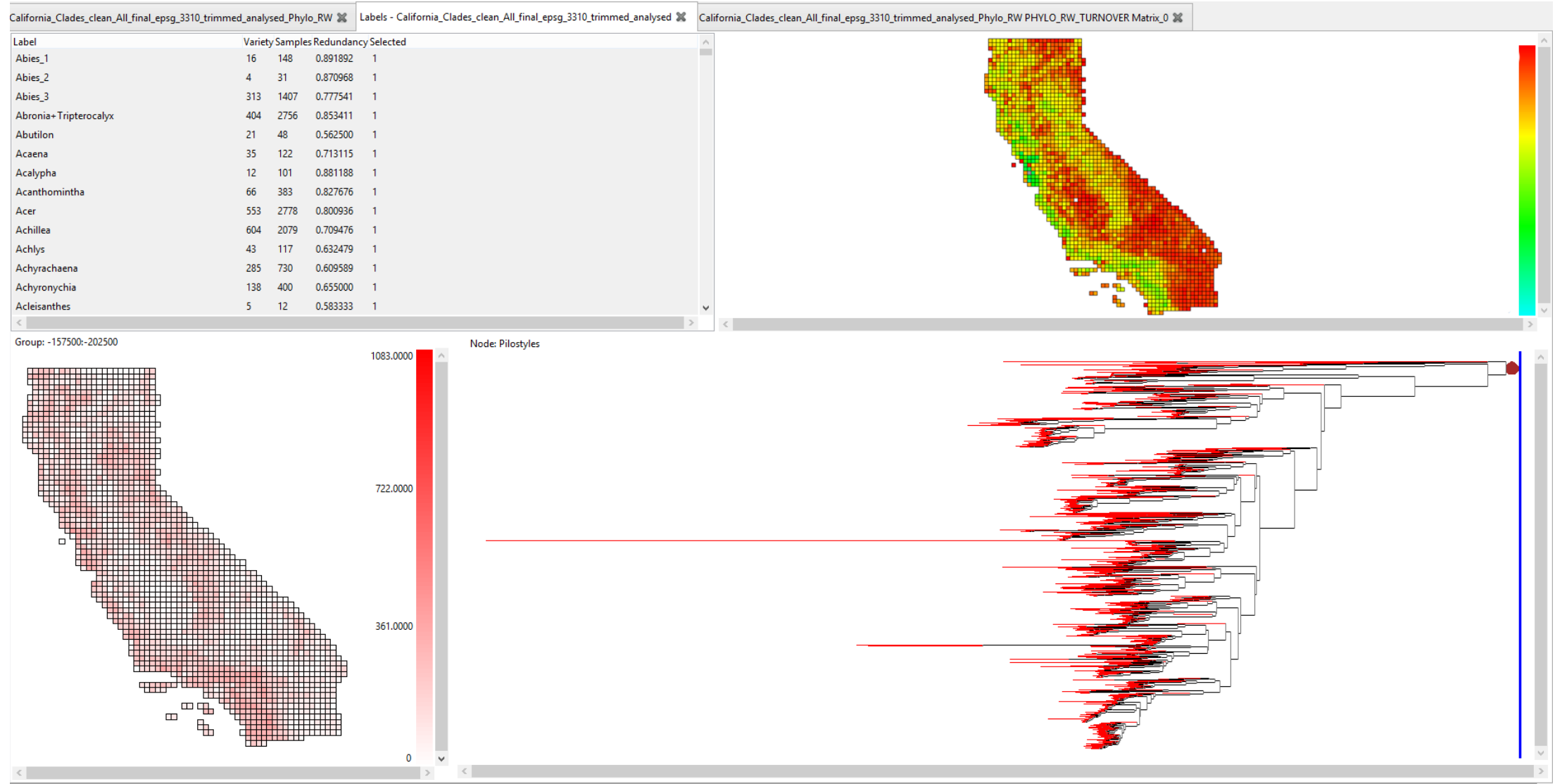


range-weighted comparison tree

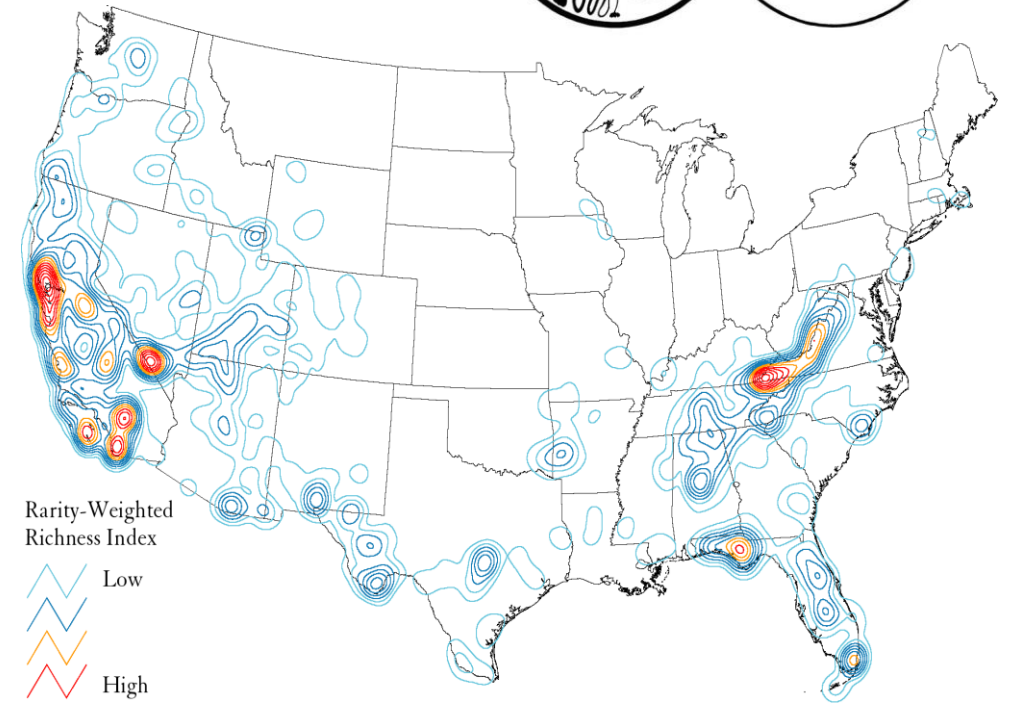
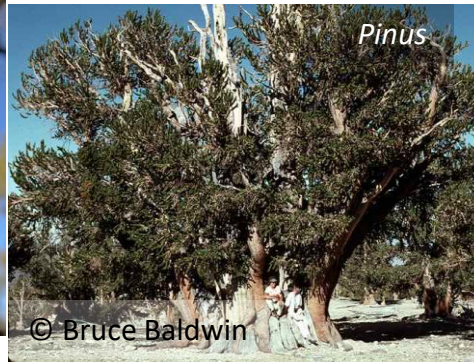
Diversity and endemism metrics calculated using Biodiverse



<https://github.com/shawnlaffan/biodiver>



California: Global-scale floristic hot spot



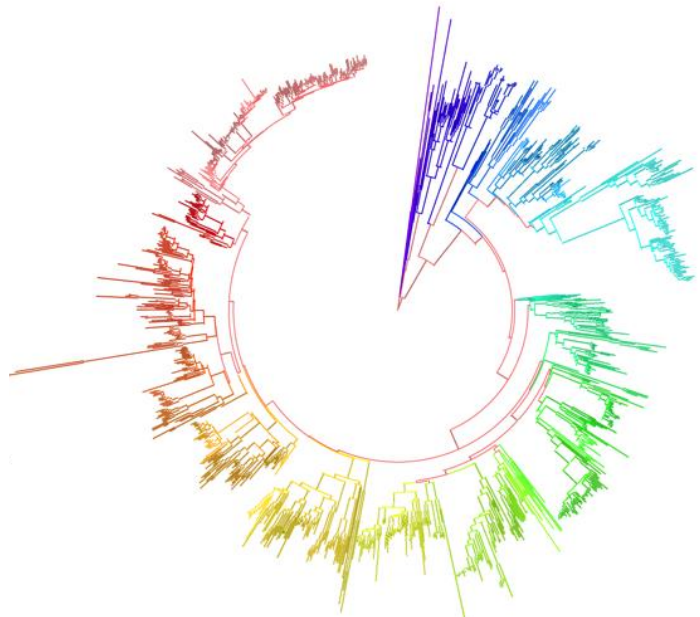
Source: *Precious Heritage* (2000) © TNC, NatureServe

- Isolated, young Mediterranean climate
- Pockets of stable, equable climate (refugia)
- Substrate diversity (e.g., serpentine)
- Topographic and climatic diversity



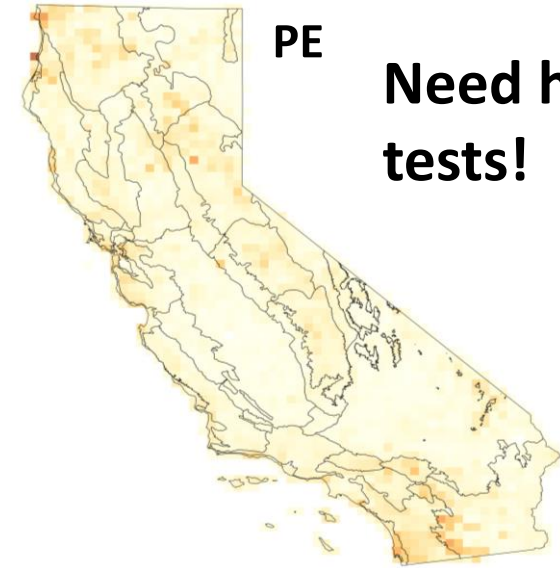
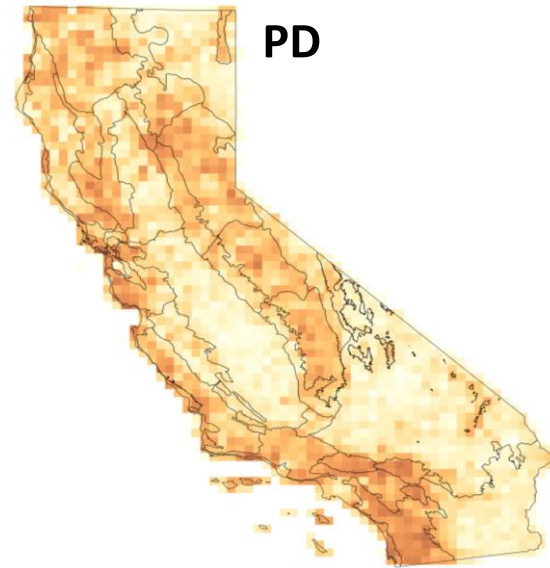
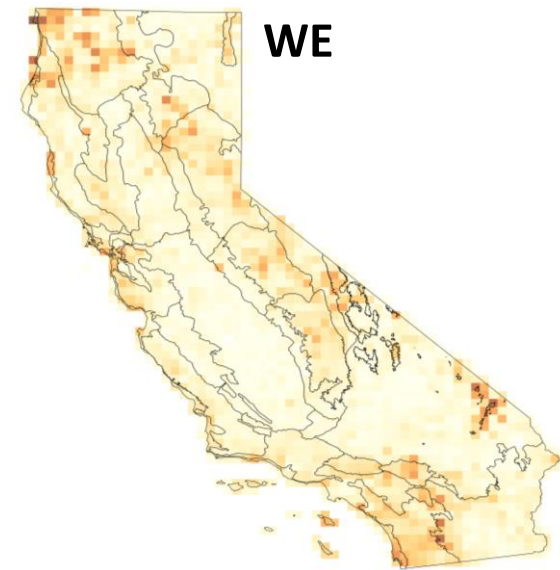
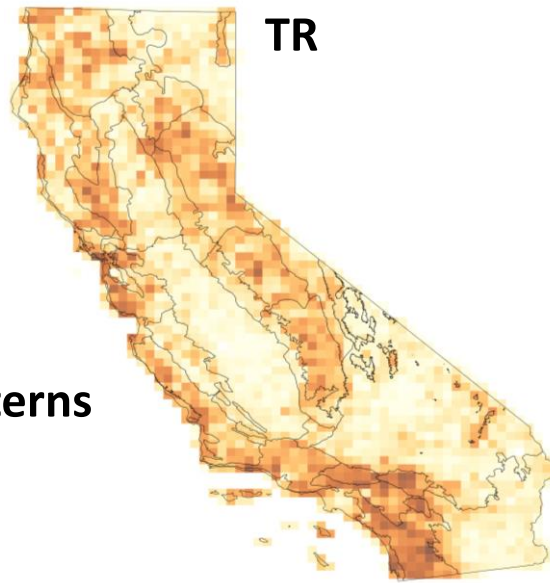
Consortium of California Herbaria

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

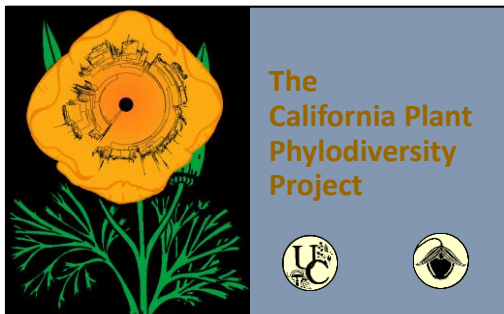


- Bruce G. Baldwin, David D. Ackerly
- Andrew H. Thornhill, Naia Morueta-Holme
- William A. Freyman, Matthew M. Kling, Thomas Madsen
- Sonia Nosratinia

Raw patterns



Need hypothesis tests!



The California Plant
Phylodiversity
Project



Supported by DEB-1354552

Thornhill et al. 2017. *Spatial phylogenetics of the native California flora*. BMC Biology 15:96

We test statistical significance using spatial randomizations of the terminal taxa on the map, subject to two constraints: richness of each grid cell and range size of each taxon remains constant.

PD randomized

Compares the observed PD with the PD measured on the same tree for 999 randomly selected sets of the same number of taxa.

Are the co-occurring taxa in a grid cell more or less closely related to each other than would be expected by random?

Significant locations

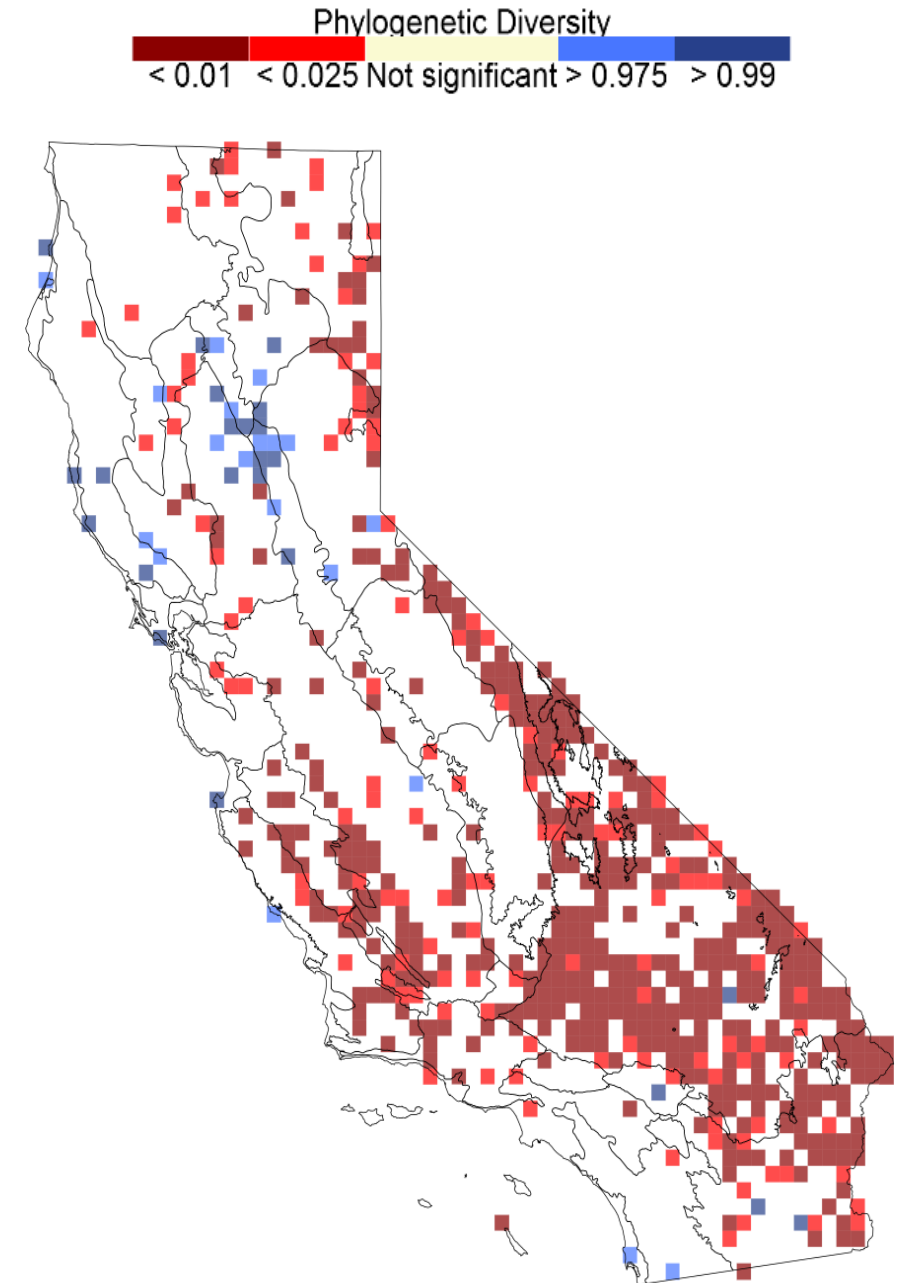
may have an ecological explanation.

Phylogenetic over-dispersion (blue)

close relatives may exclude each other.

Phylogenetic clustering (red)

clades may have evolutionarily conservative habitat preferences and thus close relatives co-occur.



Relative Phylogenetic Diversity (RPD)

PD calculated using the original tree

PD calculated using a comparison tree with all branch lengths equal

Is there an overrepresentation of long branches or short branches in a grid cell as compared to what would be expected if the same number of taxa had been selected at random?

Significant locations

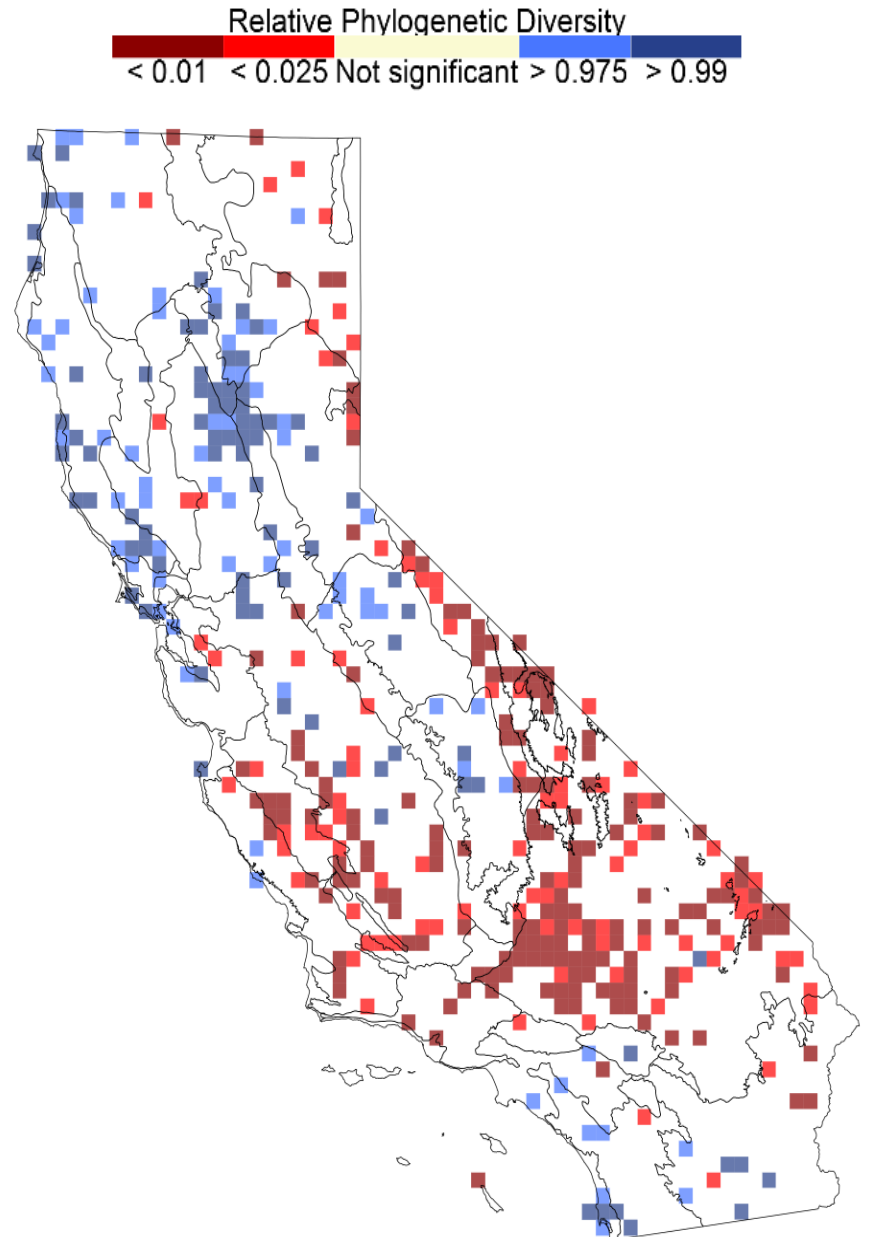
more likely to relate to biogeographic and evolutionary processes.

Concentration of long branches (blue)

a refugium, or the dispersal of a few members of large clades that mainly occur outside of the study region.

Concentration of short branches (red)

recent evolutionary divergence.



CANAPE: Categorical Analysis of Neo- And Paleo-Endemism

- Since RPE is a ratio, we need to be sure that there is indeed a high amount of endemism.
- Statistical significance tested with spatial randomizations of the terminal taxa on the map, subject to two constraints: richness of each grid cell and range size of each taxon remains constant.
- **Two-step** process for CANAPE analysis :
- First, to determine if a place is a center of significantly high PE, a grid cell needs to be significantly high (one-tailed test) in either the numerator, the denominator, or both.
- Second, if (and only if) a grid cell passes one of those tests, we divide the centers of endemism into three meaningful, non-overlapping categories: neo-endemism, paleo-endemism, and mixed endemism.

Phylogenetic measures of geographic similarity

Beta-diversity is important also!

Typical turnover measures look at matching in species composition, measured via a dissimilarity index such as:

Jaccard

$$= 1 - \frac{A}{A + B + C}$$

where A is the count of species found in both locations, B is the count unique to location 1, and C is the count unique to locations 2.

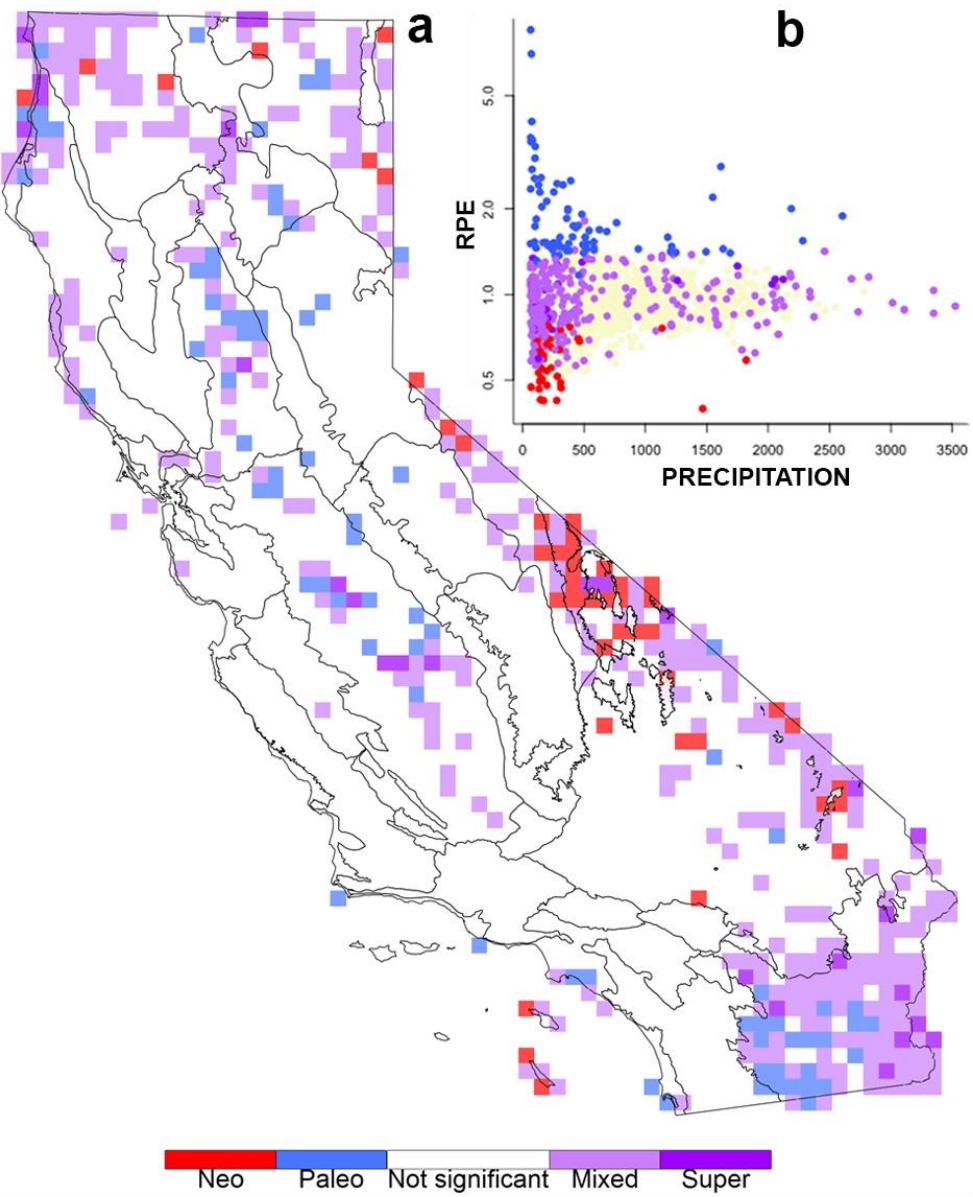
Sorensen

$$= 1 - \frac{2A}{2A + B + C}$$

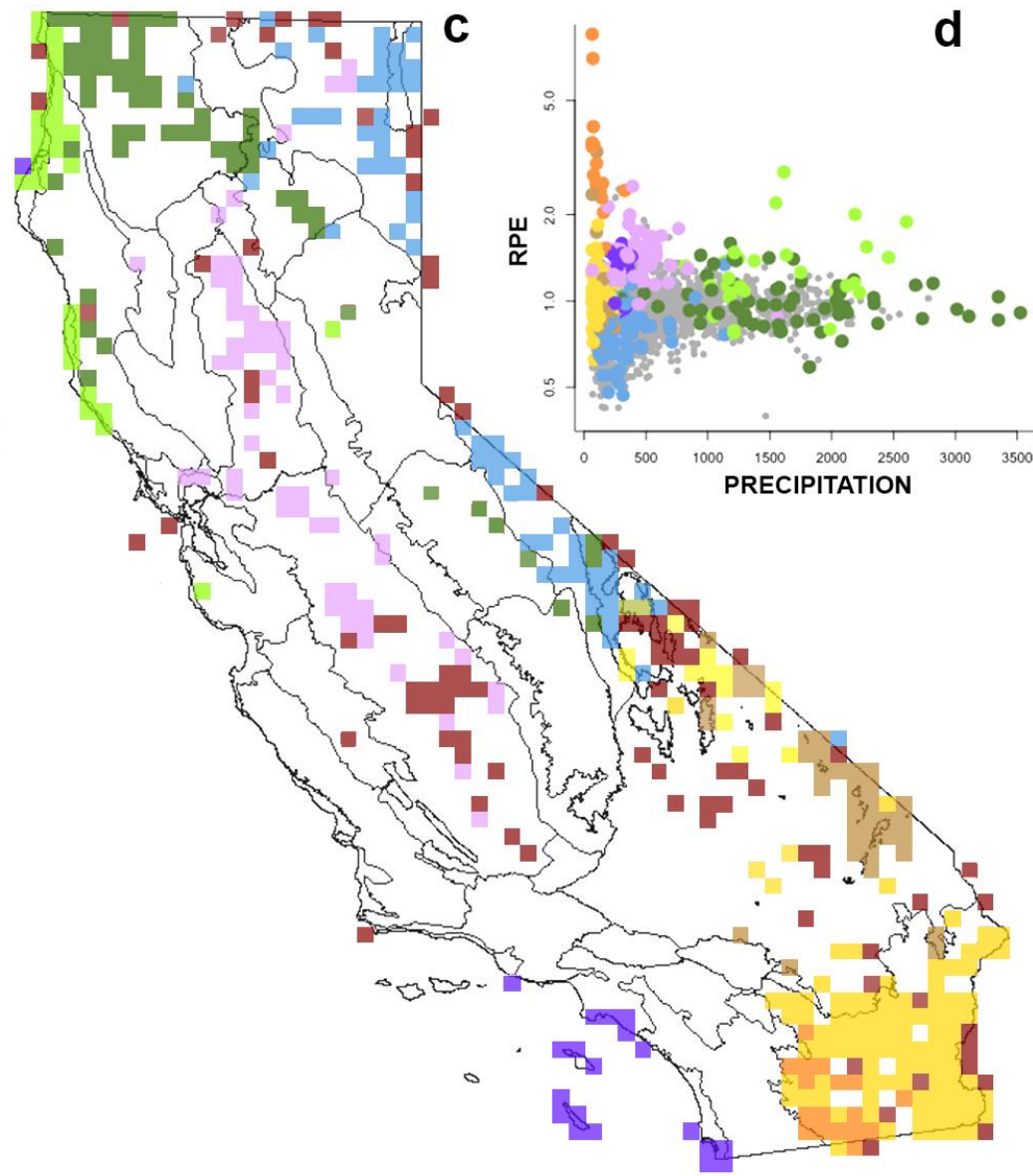
There is an exact phylogenetic analog of these indices (e.g., Phylo-Jaccard and Phylo-Sorensen) where A is the length of shared branches, and B and C are the length of branches found only in neighbor sets 1 and 2. **This is phylogenetic turnover.

There is a new measure that uses the range-weighted tree to measure turnover. This represent the amount of phylogenetic endemism, shared between two locations. **This is range-weighted phylogenetic turnover.

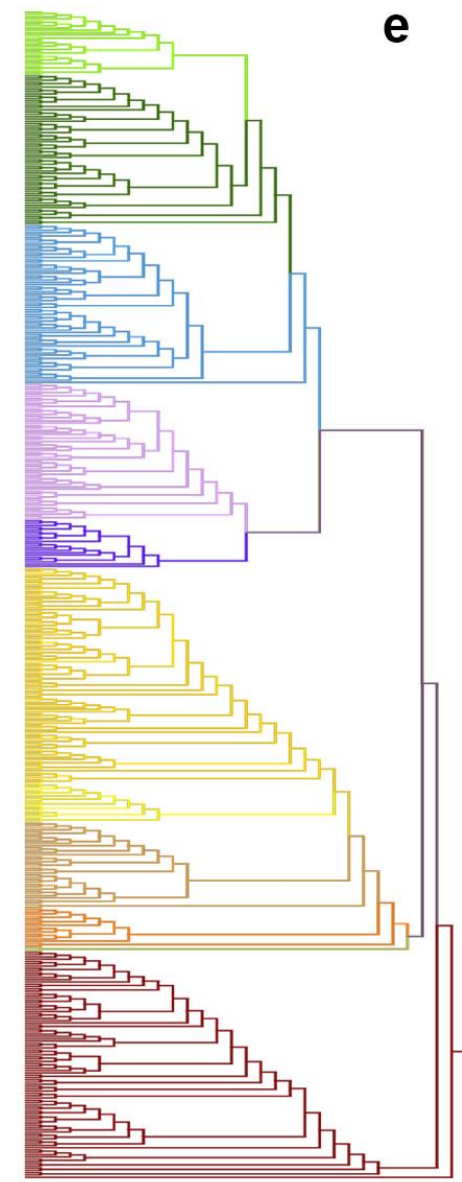
(See: S.W. Laffan, D.F. Rosauer, G. Di Virgilio, J.T. Miller, C.E. González-Orozco, N. Knerr, A.H. Thornhill, and B.D. Mishler. 2016. Range-weighted metrics of species and phylogenetic turnover can better resolve biogeographic transition zones. *Methods in Ecology and Evolution* 7: 580-588.)



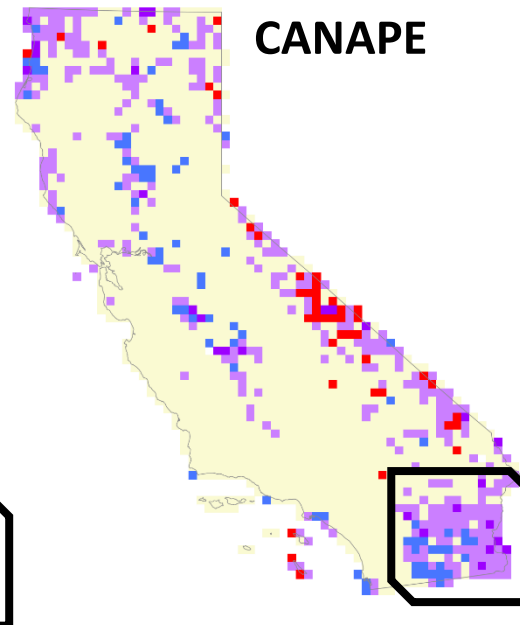
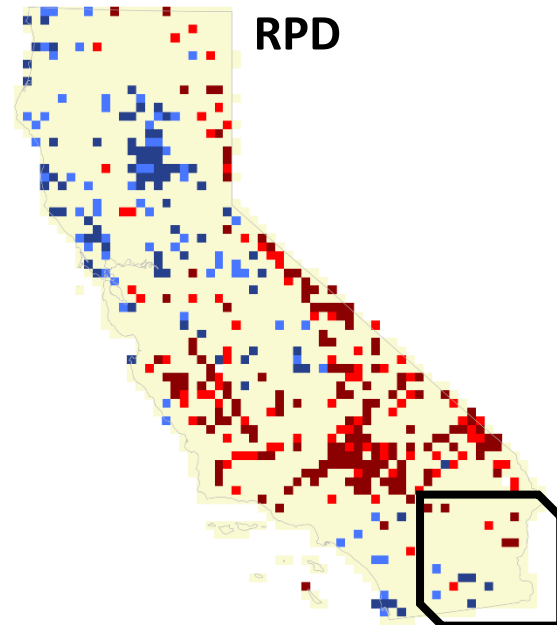
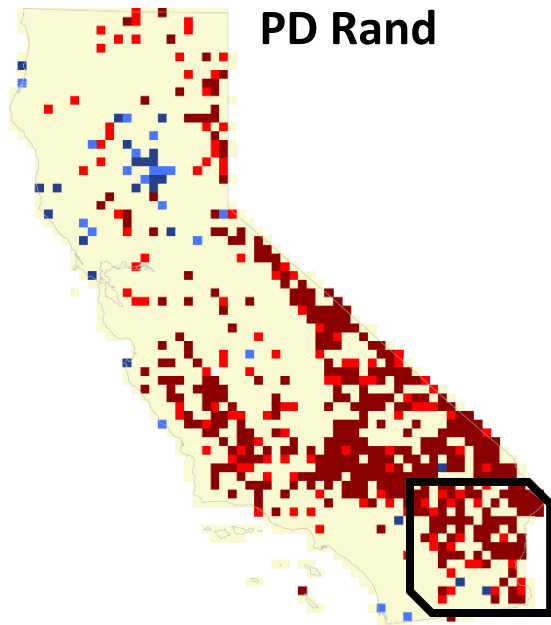
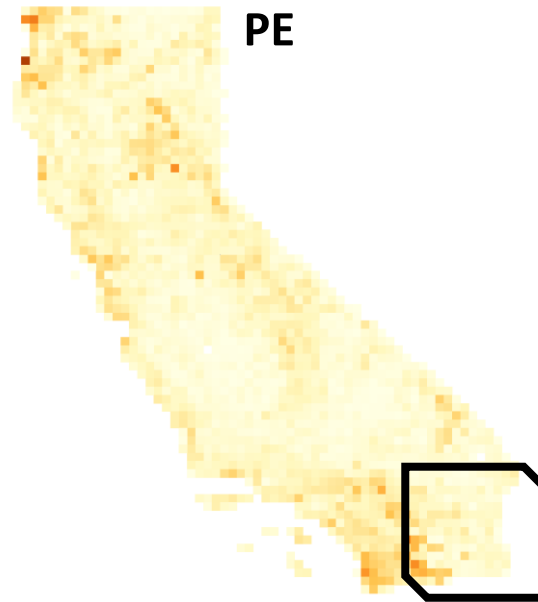
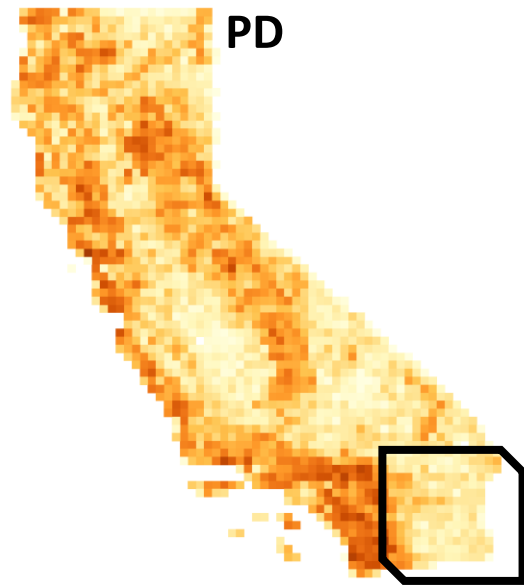
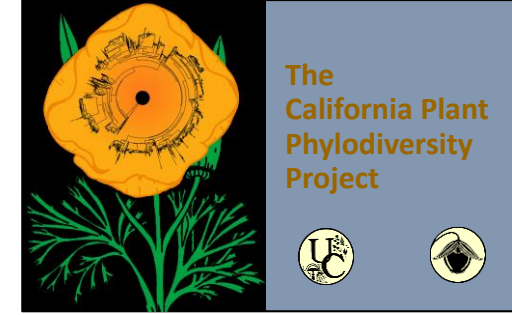
CANAPE



RW Turnover among significant centers of endemism found by CANAPE



Sonoran Desert



Patterns:

Low observed PD

Low observed PE

Signif. low PD

No significant RPD

Paleo- or Mixed-endemism

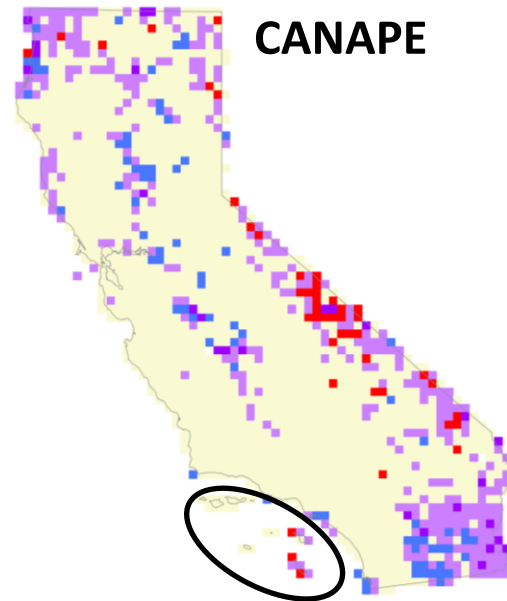
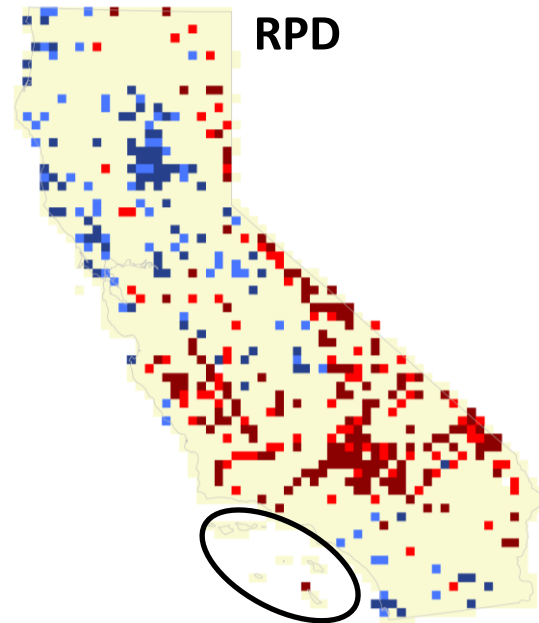
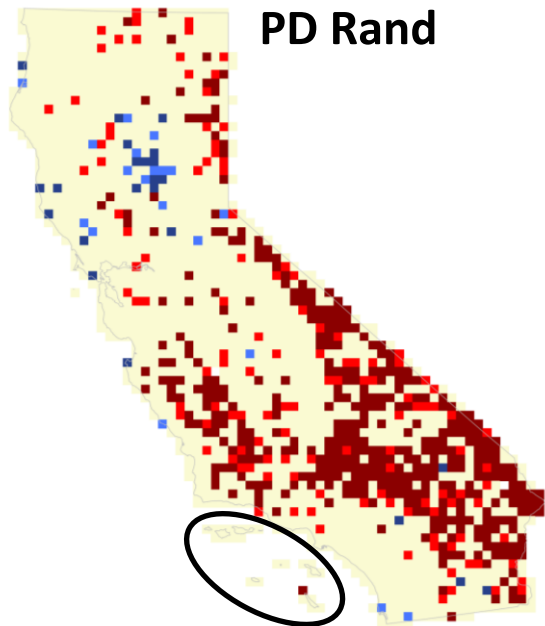
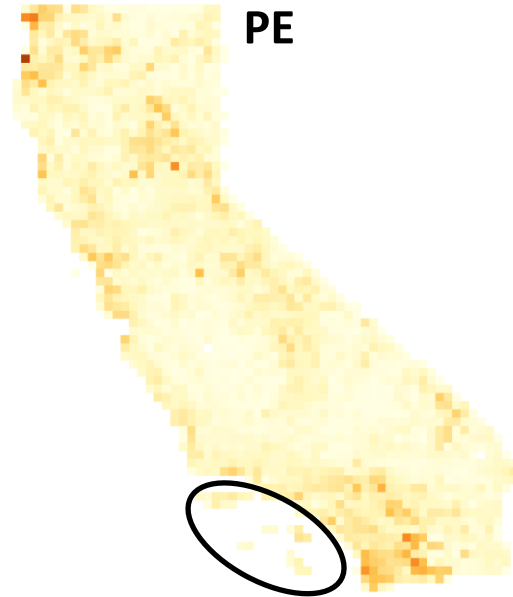
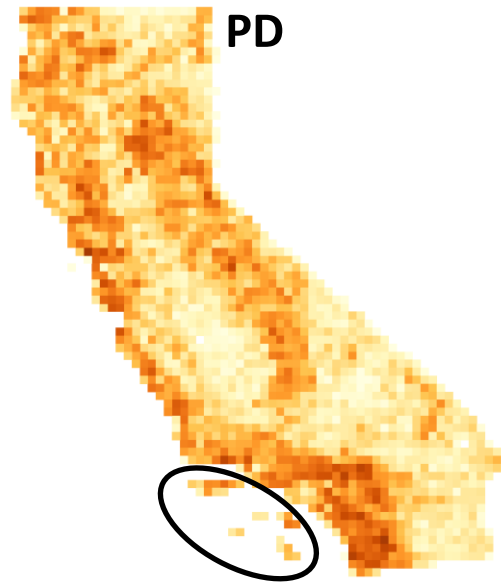
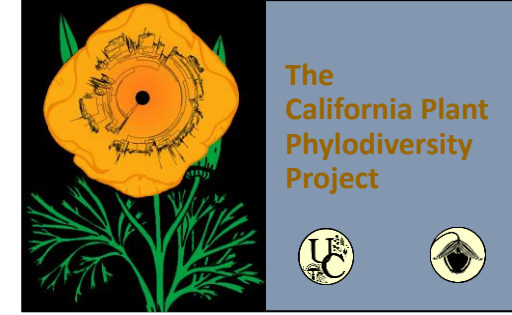
Possible Processes:

Desert specialists

Desert endemism = neo

Tropical remnants = paleo

Channel Islands



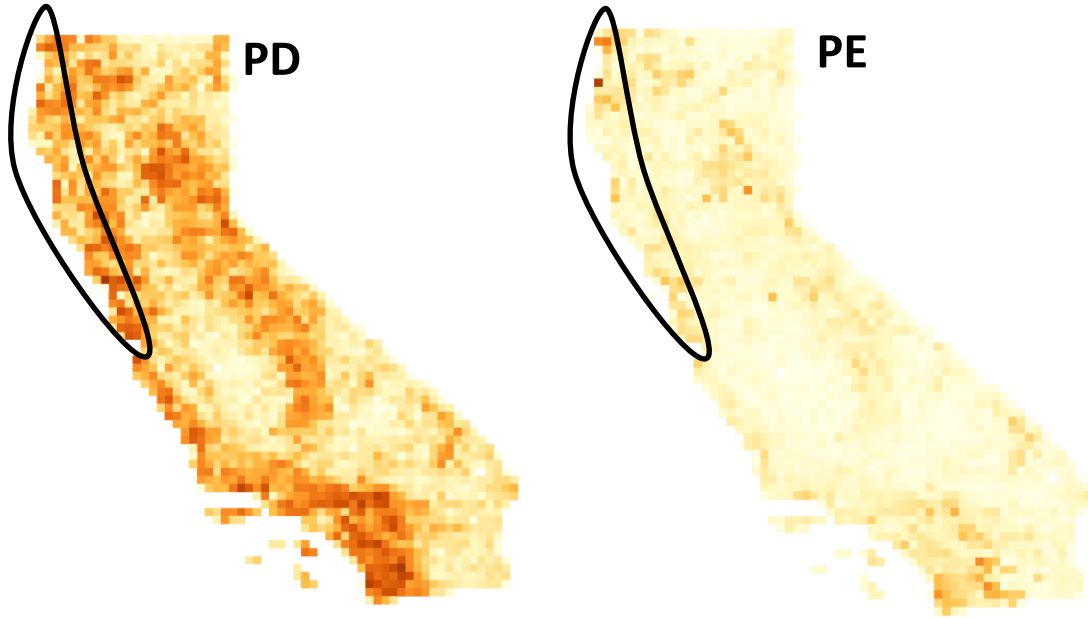
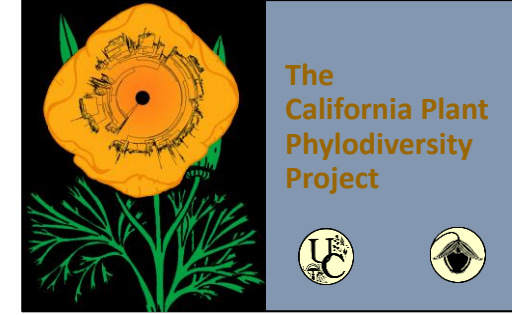
Patterns

Moderate observed PD
Low observed PE
No significant PD rand
No significant RPD
Neo-endemism

Possible Process

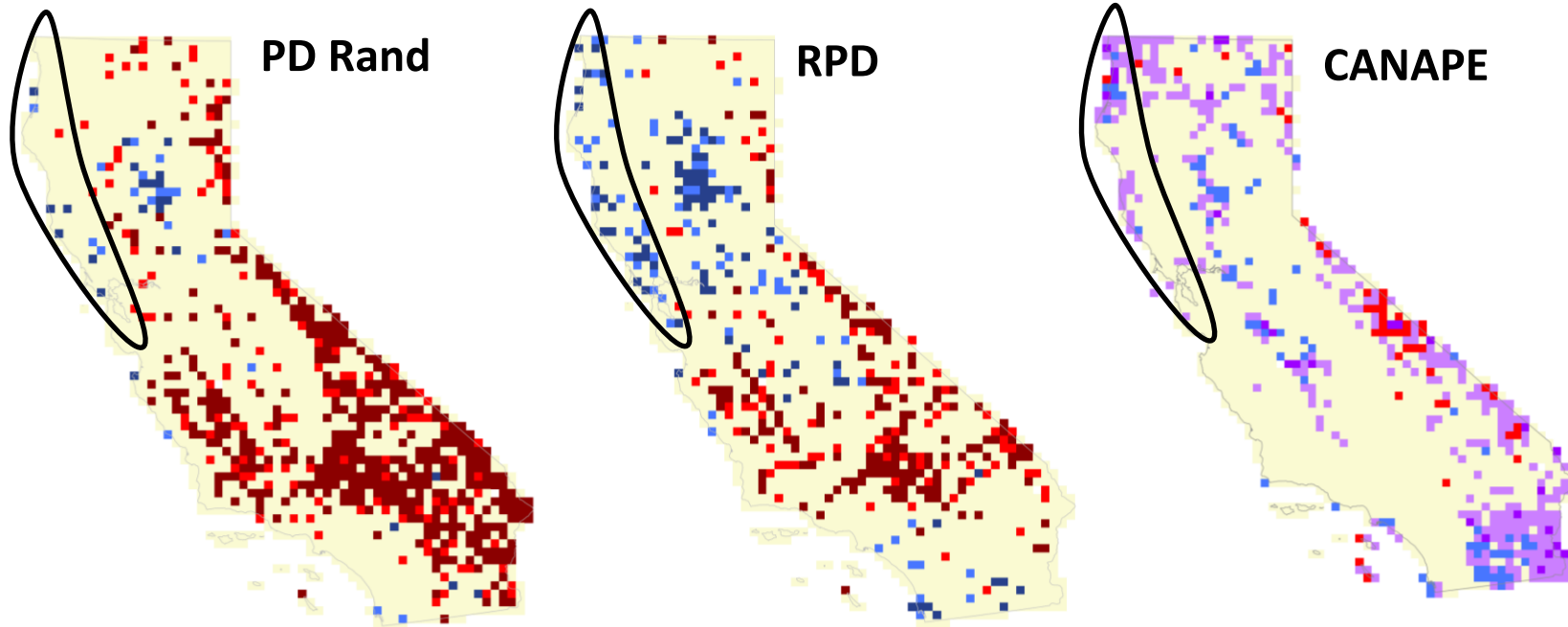
Recent divergence from
mainland sources
Southern islands more
isolated through time

North Coast



Patterns

- High observed PD
- Some observed PE
- No significant PD
- High RPD
- Super-endemism + Paleo

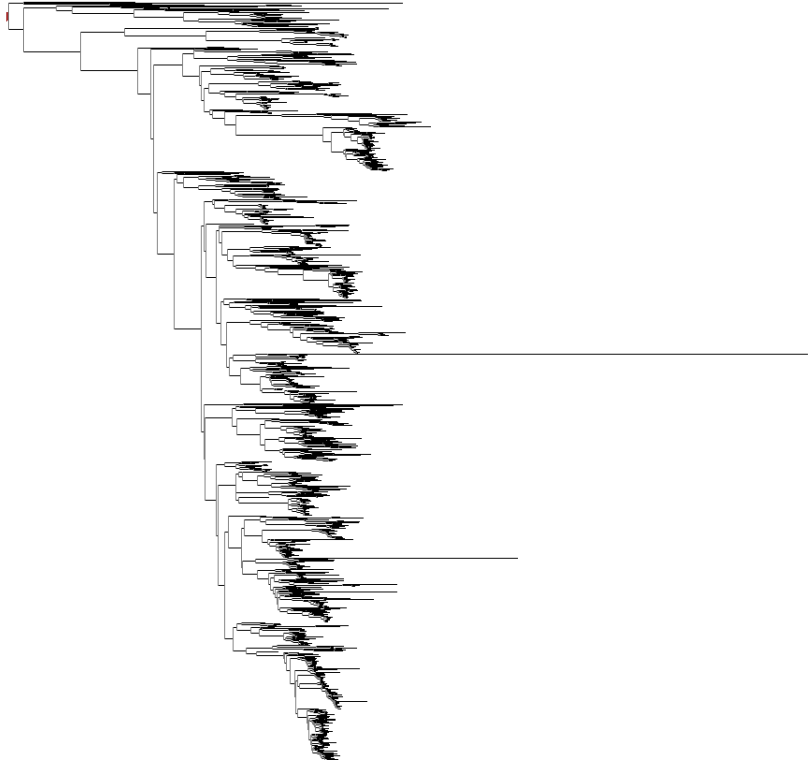


Possible Process

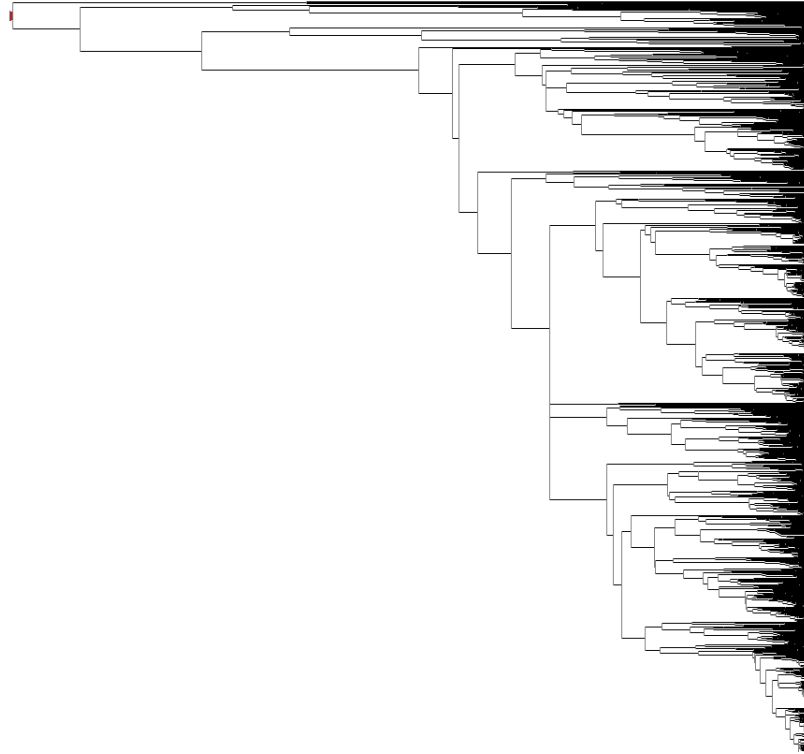
- More equable climate through time
- Refugial taxa, some range-restricted

The affect of time calibrations

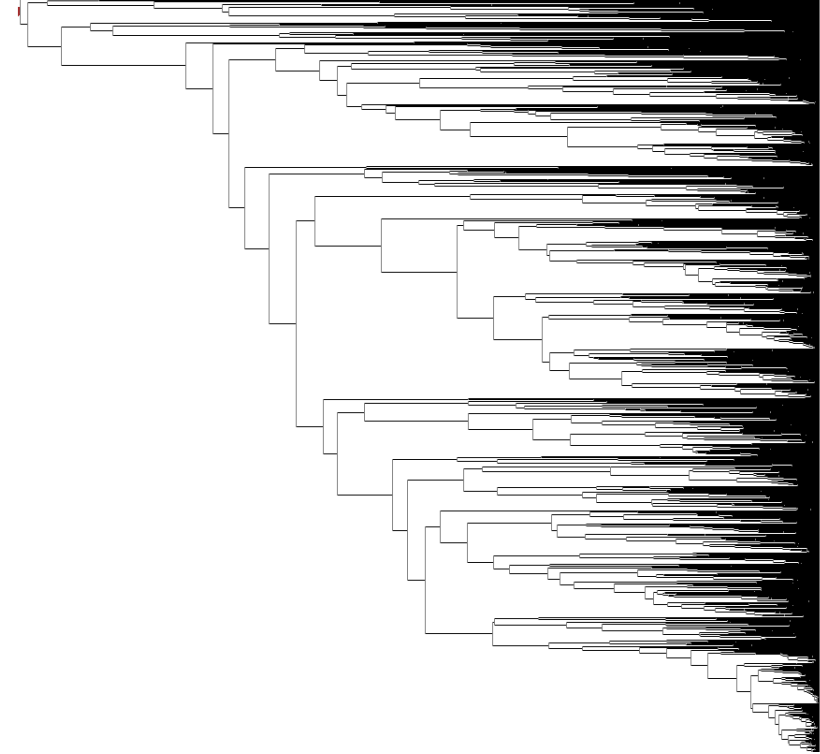
Uncalibrated tree



Root calibrated tree



Fully calibrated tree



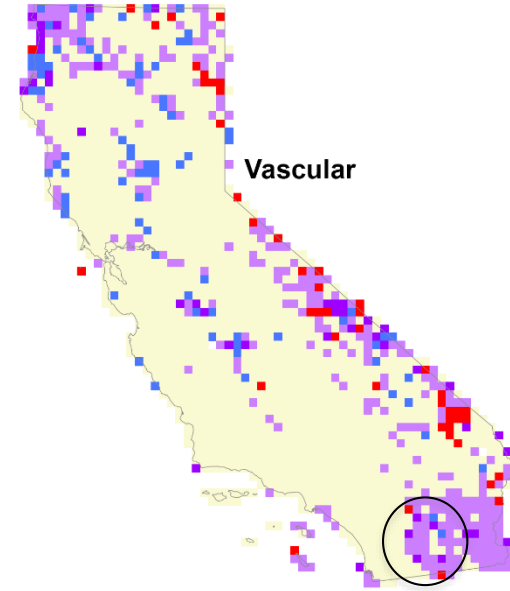
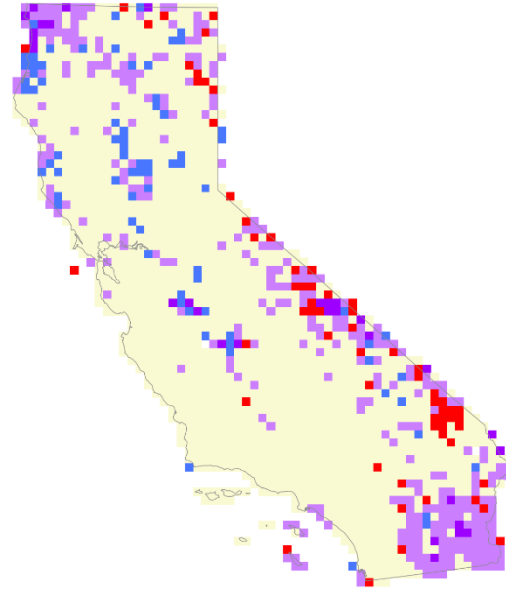
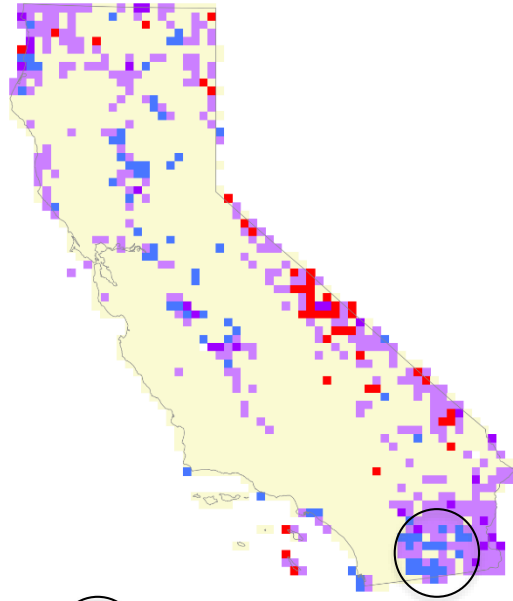
Information for 55 calibrations
from Magallón et al., 2013; 2015

The affect of time calibrations

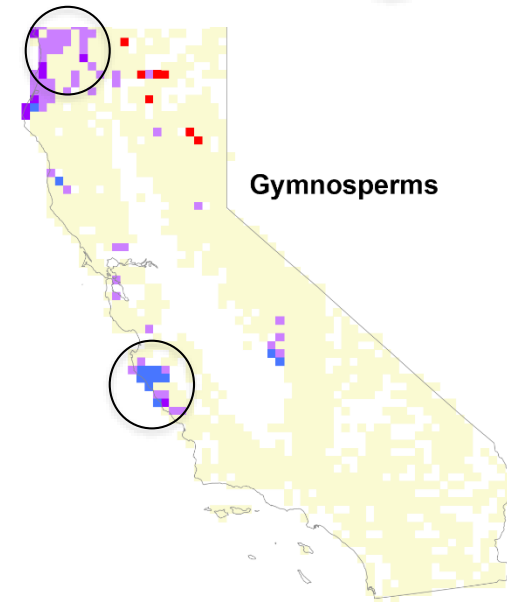
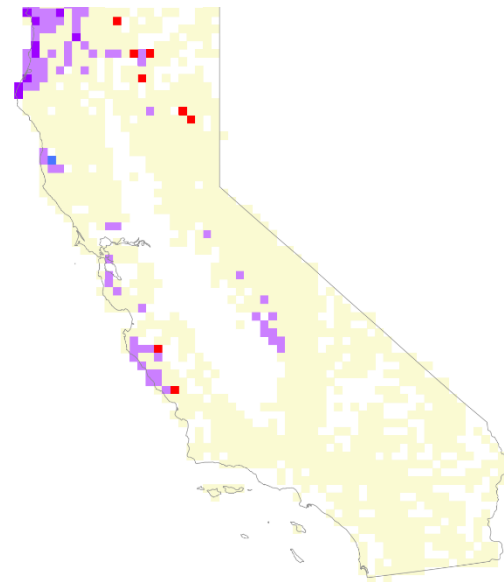
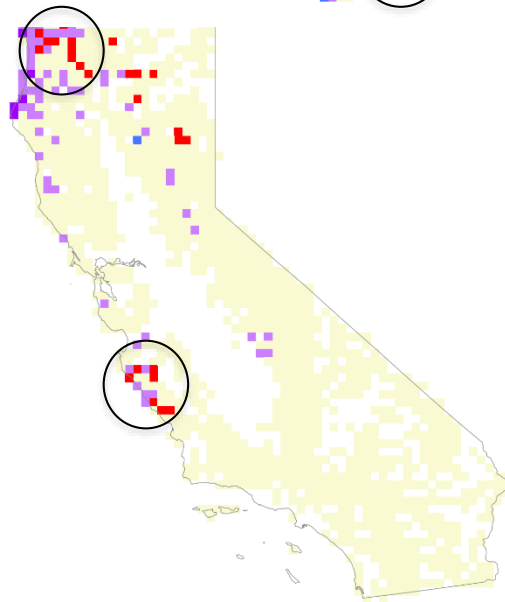
No calibrations

Root only

All calibrations



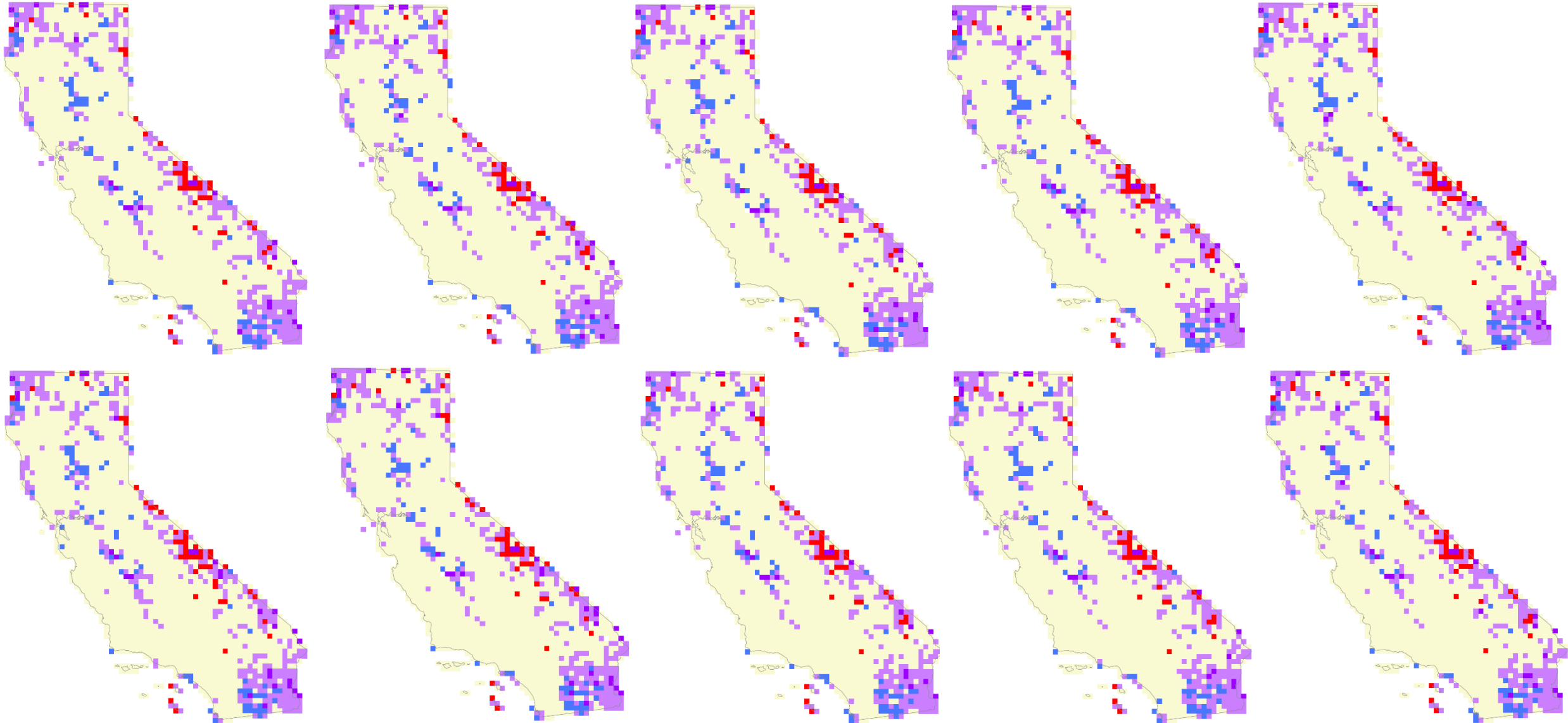
Vascular

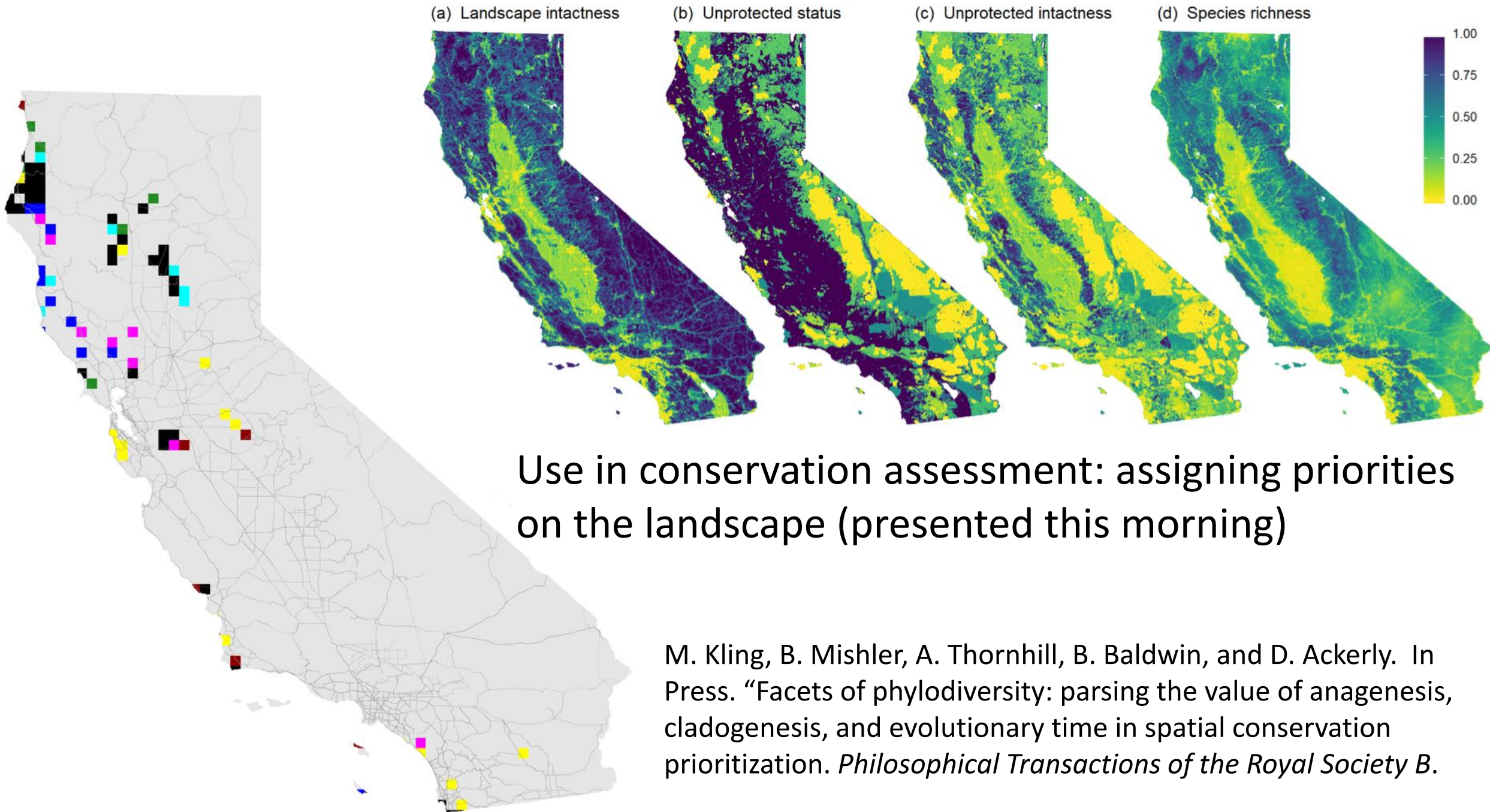


Gymnosperms

Tree uncertainty

Ten different RAxML topologies assessed in CANAPE





Spatial phylogenetics

- Era of large-scale, evolutionarily-based biodiversity studies.
- Spatial phylogenetics will continually improve with:
 - Denser sampled phylogenies
 - Better spatial data / finer scale grid cells
- Observed and randomised results should be interpreted together.
- Spatial phylogenetics is applicable to
 - Evolution
 - Ecology
 - Biogeography
 - Conservation
- Tree uncertainty has basically no affect.
- Chronogram vs. phylogram: has little affect on locations of significant PE in CANAPE, but can affect interpretations.
- **Acknowledgments:** Will Freyman, David Baxter, Tom Madsen, Sonia Nosratinia, Dominique Turcotte, Eric Chuang, Naia Morueta-Holme

