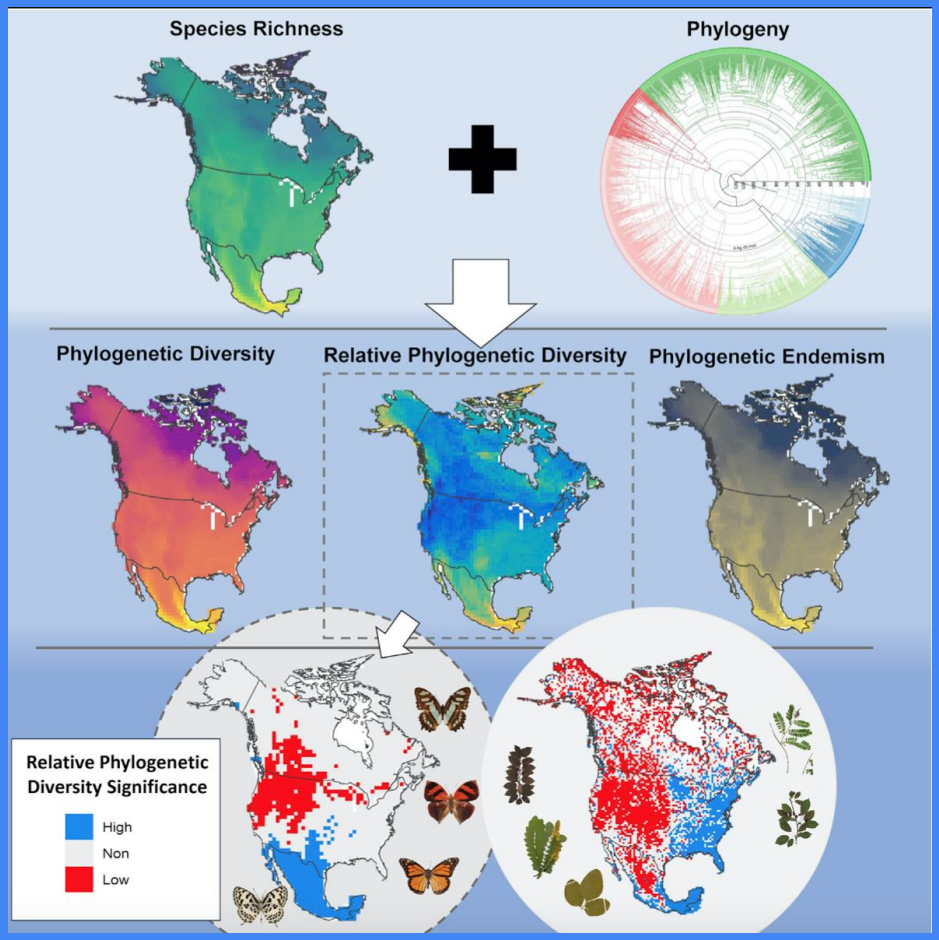
Welcome to the

Spatial Phylogenetics workshop



From: Earl et al. 2021. iScience 24 (4): 102239.

International Biogeography Society workshop series

October 12 – 13, 2022

Introduction of workshop instructors

1. **Israel Borokini**, Smith Postdoctoral Fellow, University and Jepson Herbaria, University of California, Berkeley
2. **Kyle Rosenblad**, PhD candidate, Department of Integrative Biology, University of California, Berkeley
3. **Matthew Kling**, Postdoctoral Fellow, Department of Biology, University of Vermont

Workshop lecture outline

1. What is spatial phylogenetics?
2. Metrics of alpha phylodiversity
3. Spatial randomizations
4. CANAPE
5. Phylogenetic turnover measures
6. Facets of phylodiversity
7. Conservation prioritization
8. Scale

Spatial phylogenetics → *Placing the tree of life on maps*

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

Can be applied at any taxonomic level or geographic scale.

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.

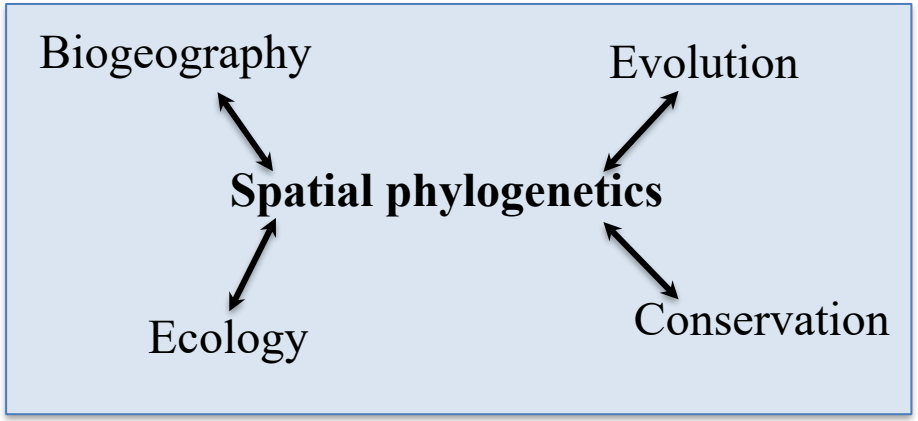
Relatively robust to lumping and splitting decisions.

A "big data" approach enabled by technological advances:

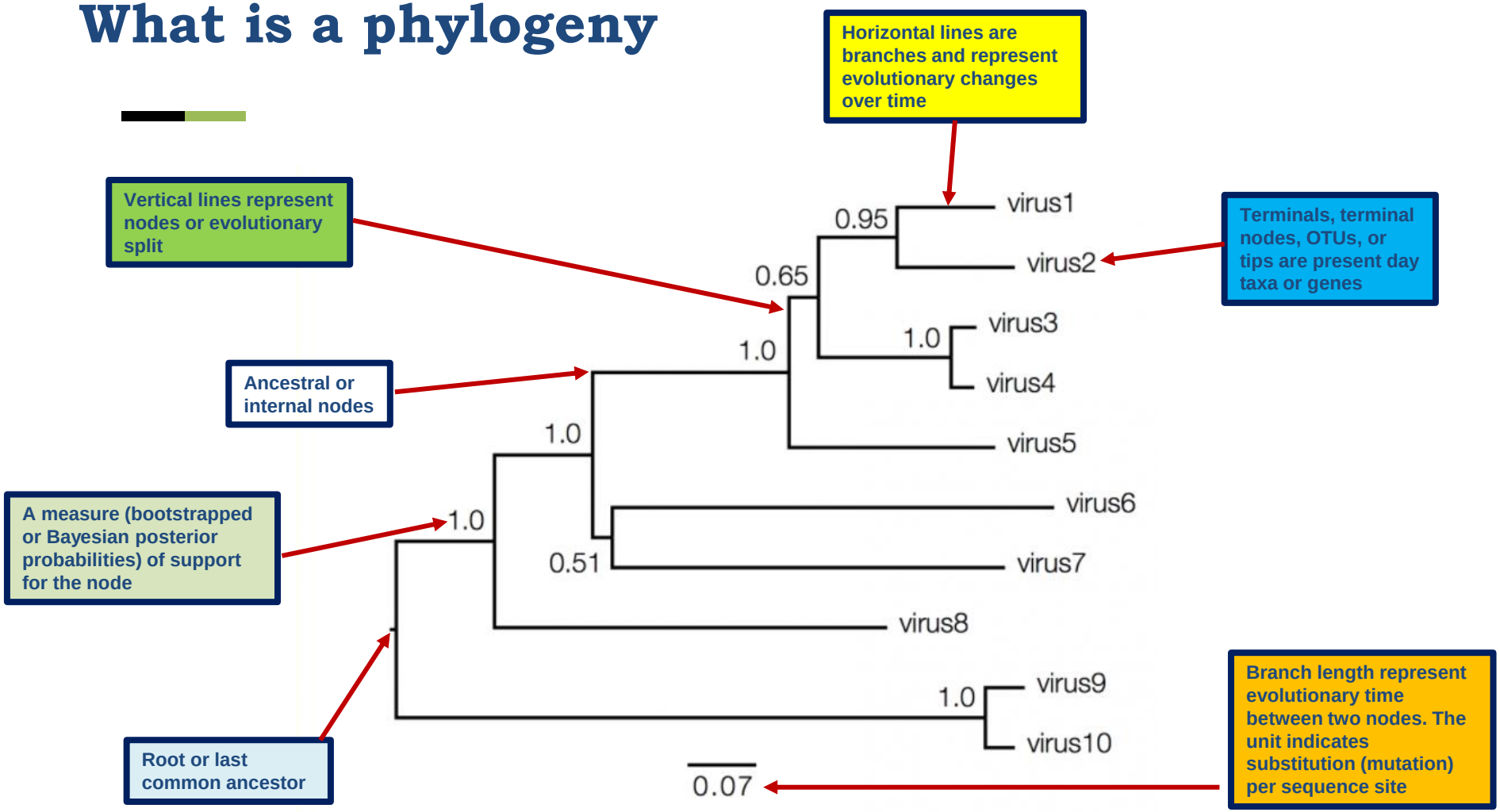
Advances in digitization and availability of natural history museums specimens

Plethora of DNA in GenBank; advances in mining software.

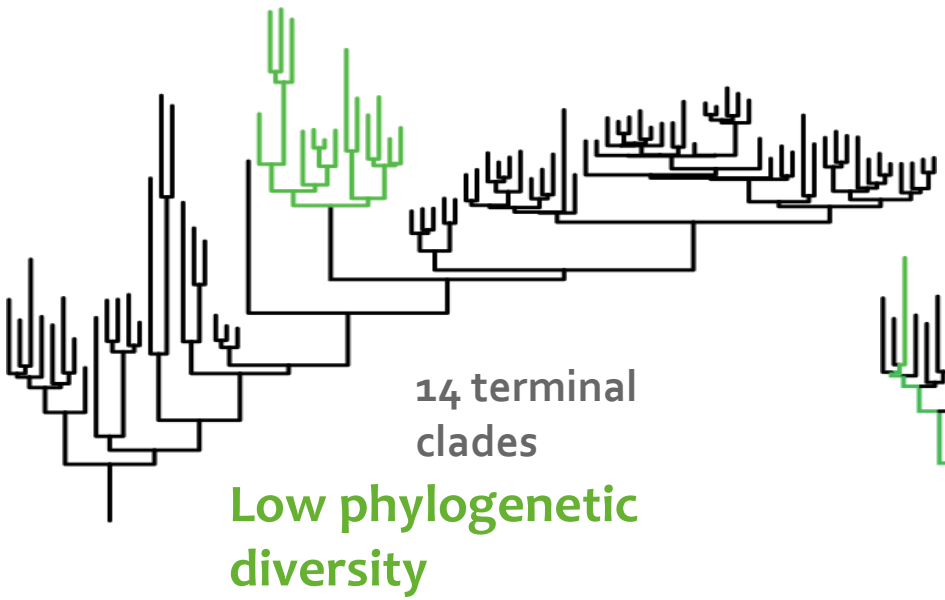
Major advances in computational methods for both tree-building and tree-using e.g. RAxML

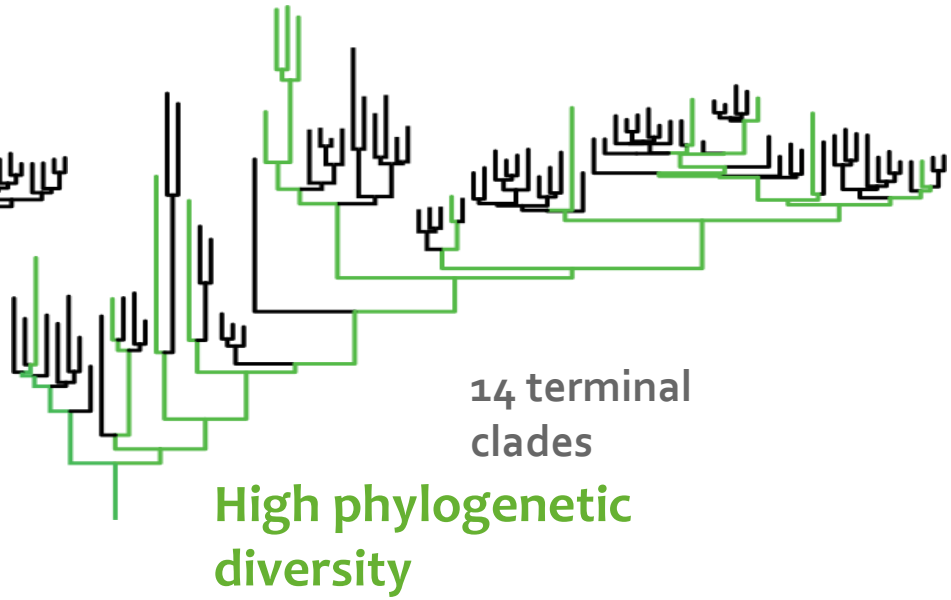


What is a phylogeny



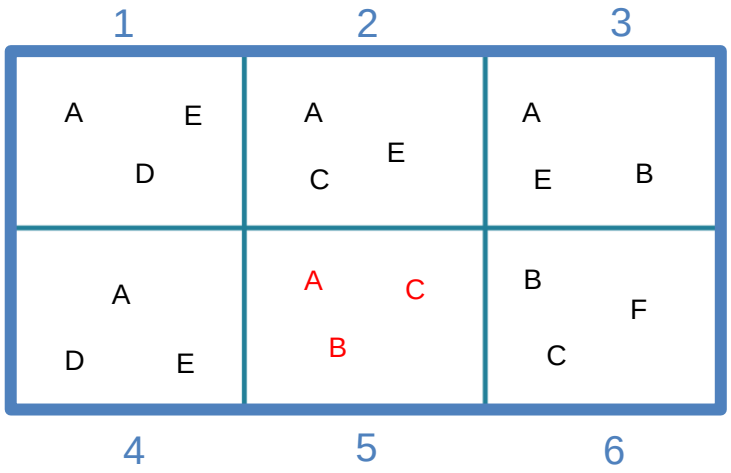
Measuring biodiversity & endemism





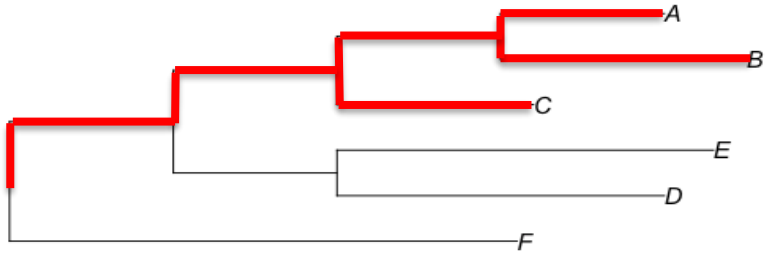
- Traditional metrics: **species richness** and **weighted endemism** (inverse of range size)
- Alternative metrics: **phylogenetic diversity** (PD), which is the sum of the branch lengths that connect all terminals in an area to the root of the tree, and **phylogenetic endemism** (PE), which is PD measured on a range-weighted tree (i.e., with each branch divided by its range size).

Phylogenetic diversity (PD)

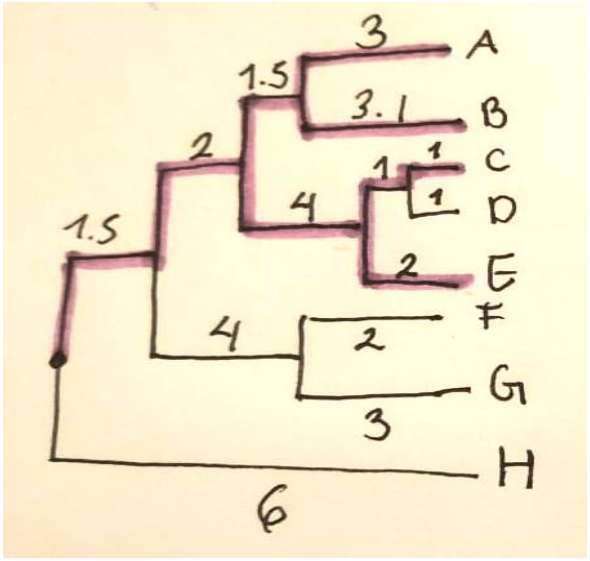


‘The **sum of the lengths of all those branches** that are members of the corresponding minimum spanning path’

Faith, 1992.

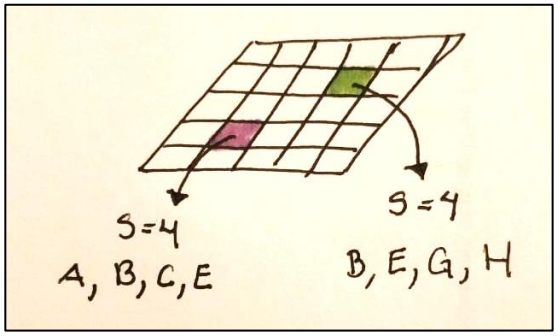


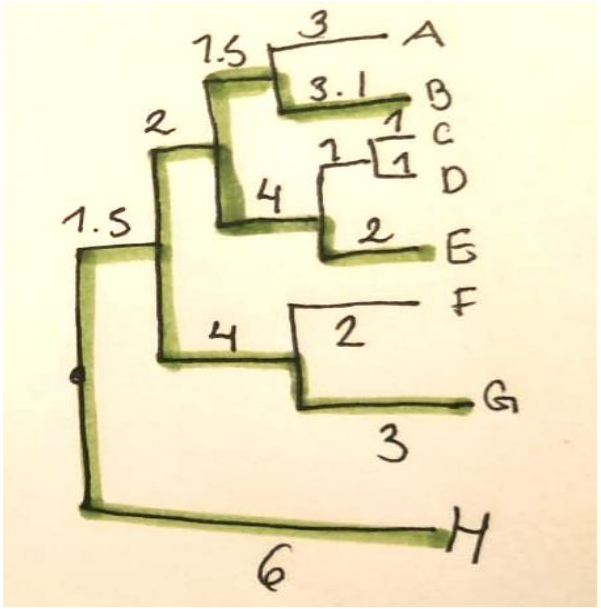
Calculates the shared evolutionary history of taxa in an area



PD = 19.1

Low phylogenetic diversity



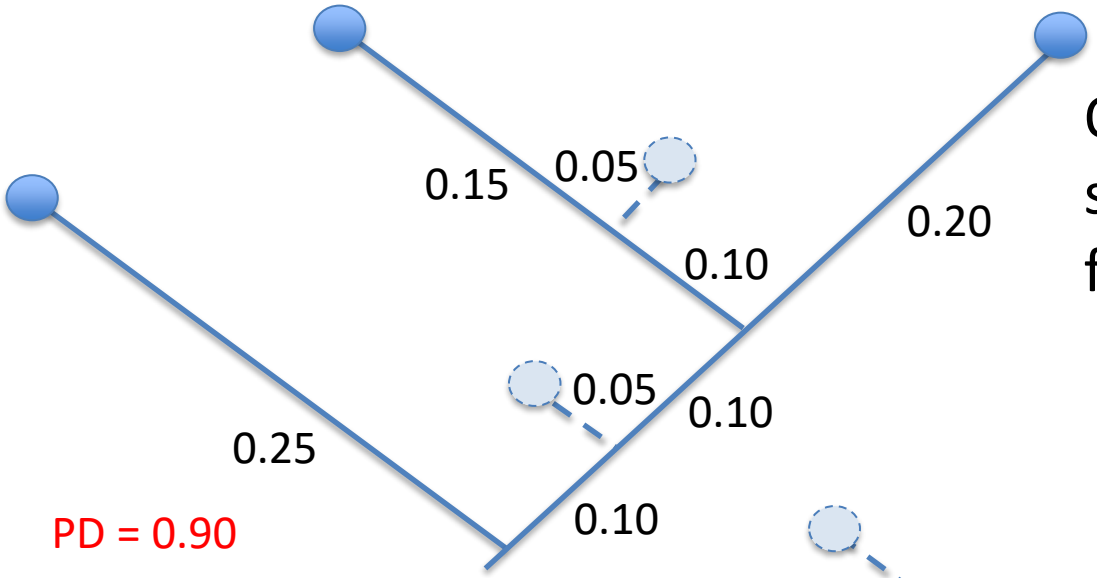


PD = 27.1

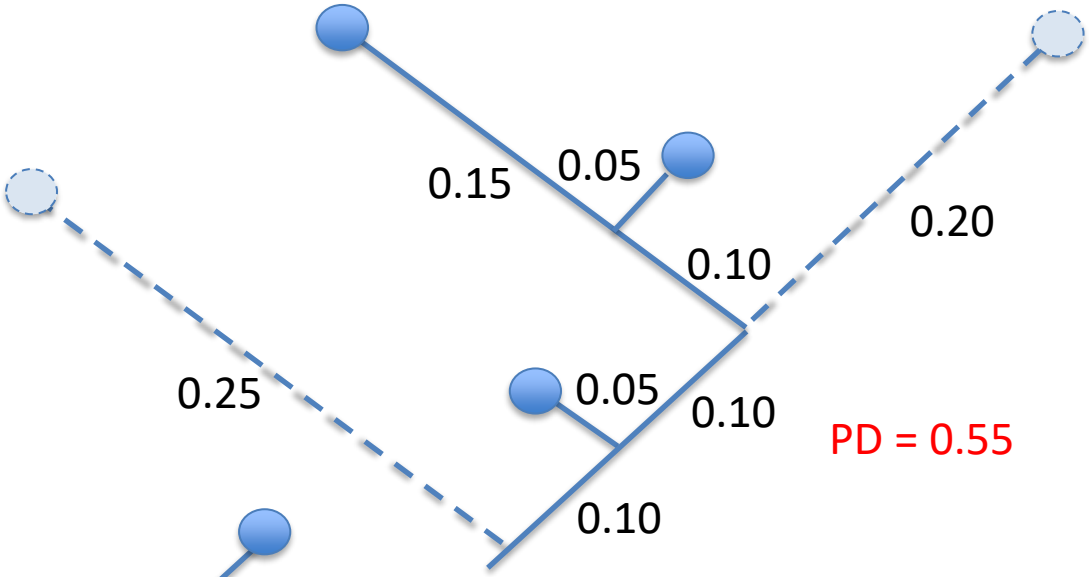
High phylogenetic diversity

How is PD calculated?

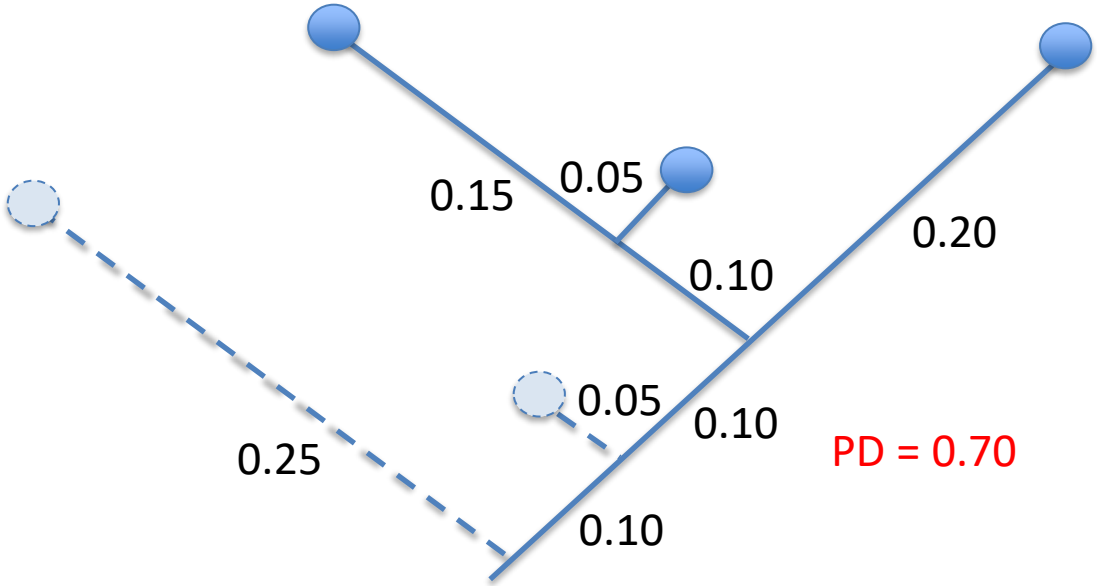
Original tree, with random selections of three out of five terminals



PD = 0.90



PD = 0.55



PD = 0.70

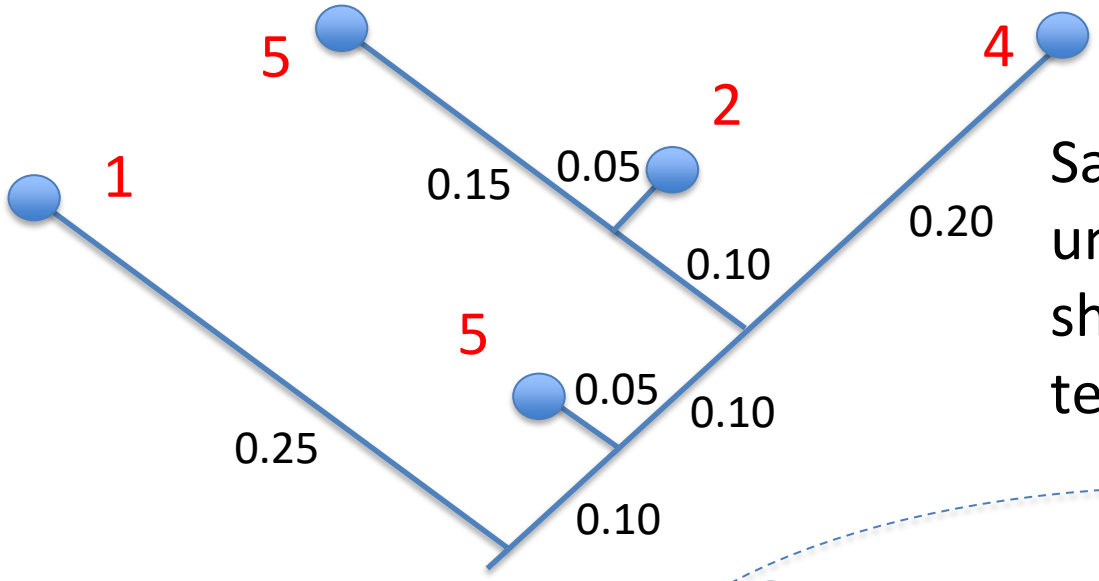
PD ranges between 0.55 and 0.90 on this particular topology

Phylogenetic endemism (PE)

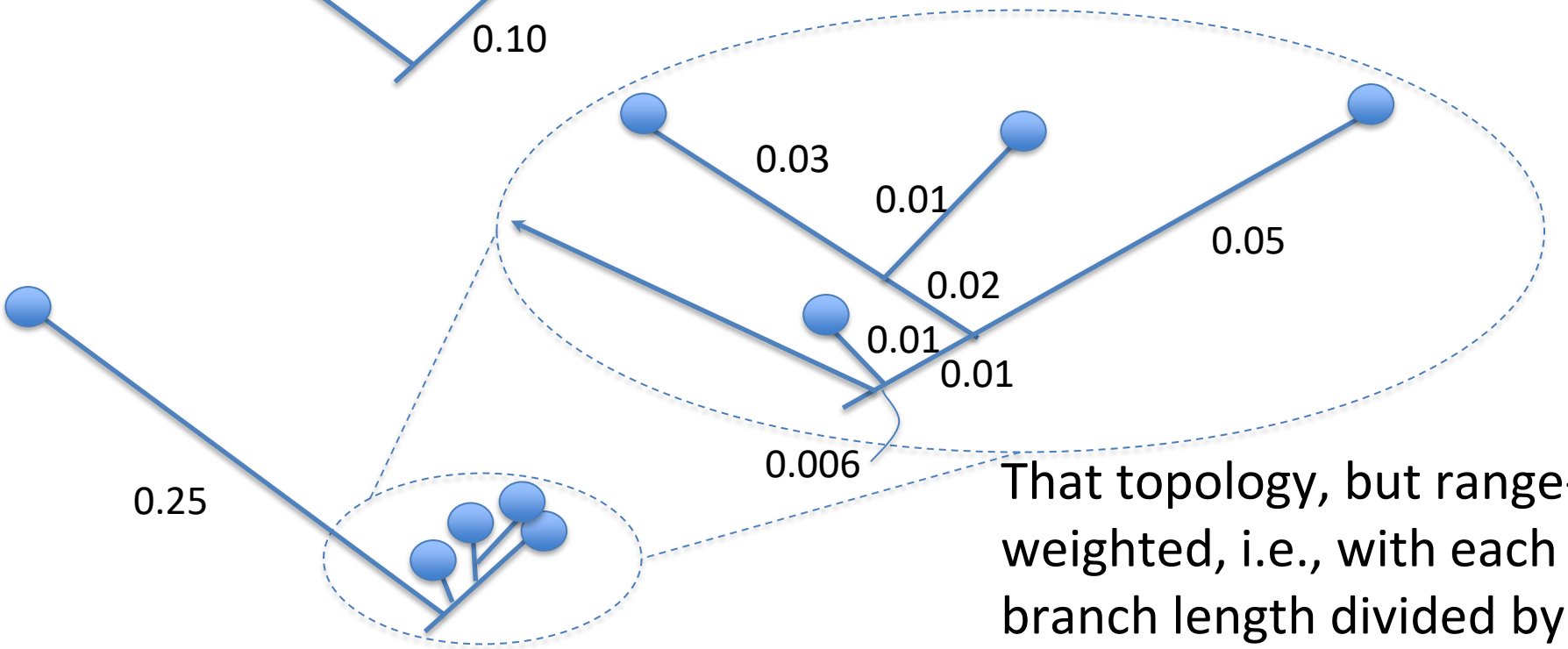
‘Summation of branch lengths weighted by range size, for each branch in a minimum spanning path linking species in a grid cell to the root of the tree’

Measure of endemism considering the evolutionary relationships between species as inferred from a phylogenetic tree, where larger values indicate the presence of older lineages

How is PE calculated?

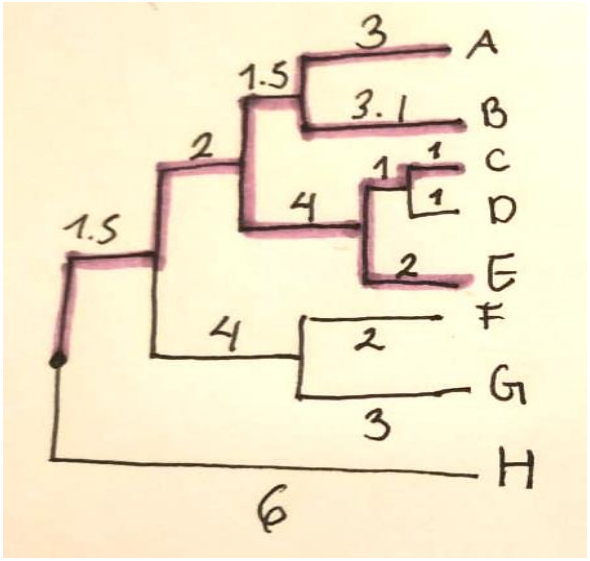


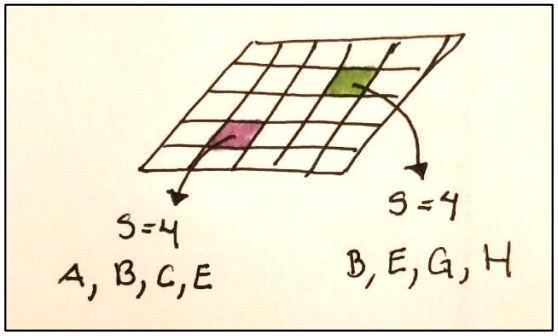
Same topology, with unmodified branch lengths, showing the range size of terminal taxa in red

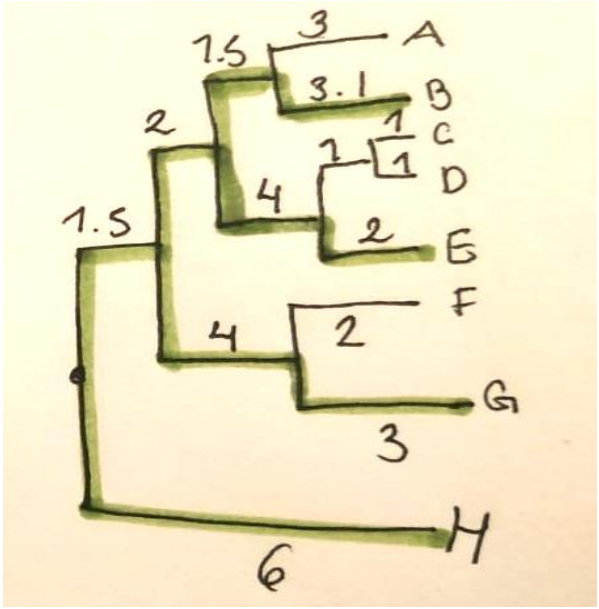


That topology, but range-weighted, i.e., with each branch length divided by its range = PE

So PE is PD on a range-weighted tree!







PD = 19.1

Low phylogenetic diversity

PD = 27.1

High phylogenetic diversity

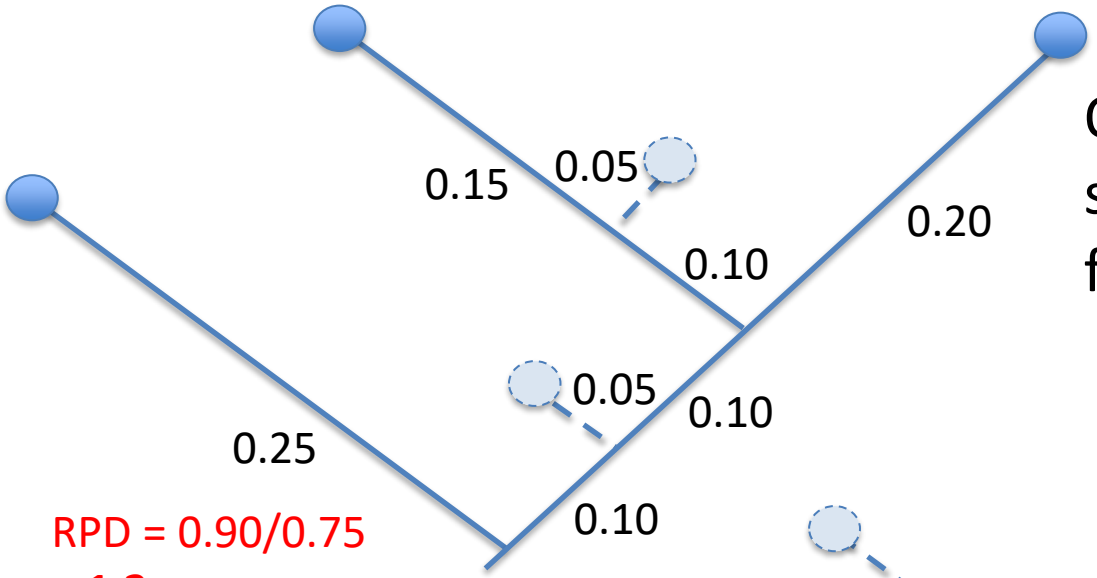
Low or high PD in these cases are context dependent

Relative Phylogenetic diversity (RPD)

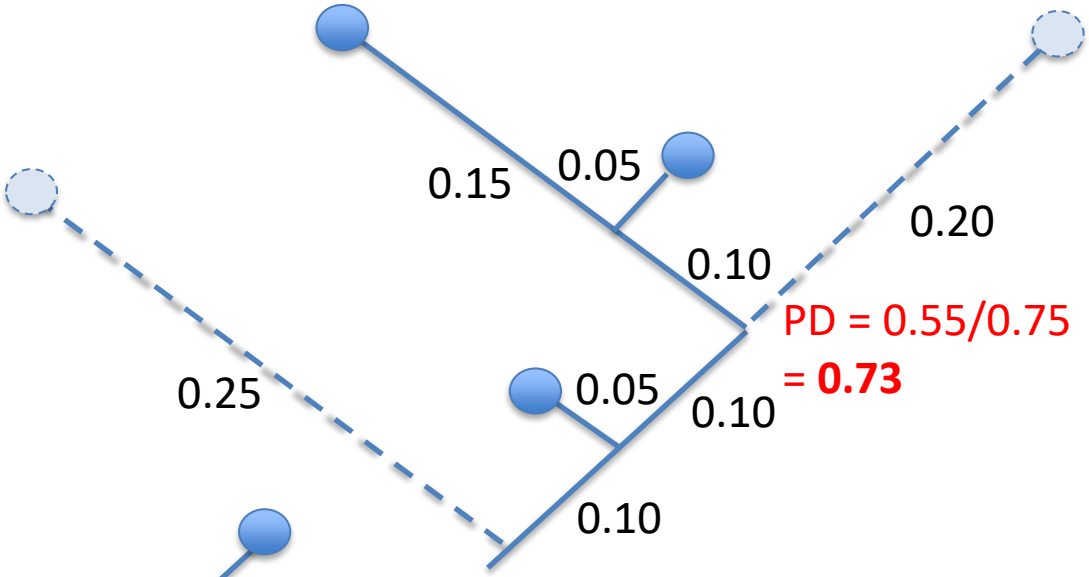
- Relative phylogenetic diversity (RPD) identifies locations with unusual concentrations of long or short branches (González-Orozco et al. 2021)
- Divide the PD by a phylogeny with exactly the same topology but equal branch lengths (cladogram)
- To identify areas with low or high PD than expected in null hypothesis
- RPD values greater than 1 represent areas with that have long branch lengths

How is RPD calculated?

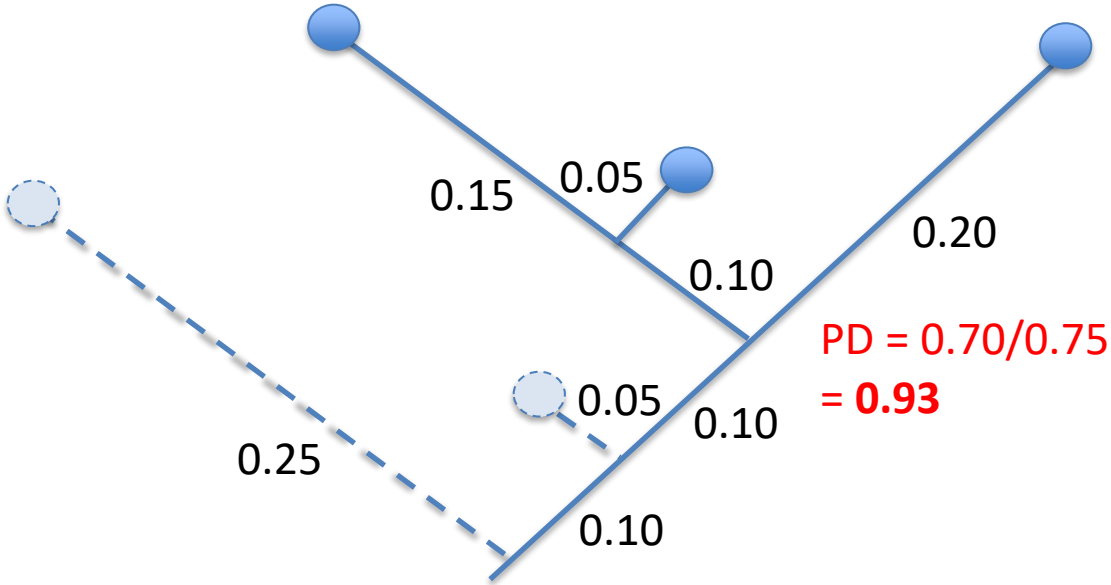
Original tree, with random selections of three out of five terminals



RPD = 0.90/0.75
= **1.2**

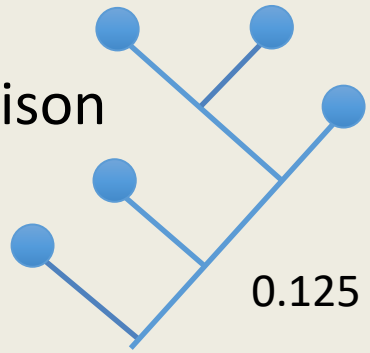


PD = 0.55/0.75
= **0.73**



PD = 0.70/0.75
= **0.93**

Comparison tree :



This is also used in the denominator of RPE

Relative Phylogenetic endemism (RPE)

- Measure of endemism considering the evolutionary relationships between species as inferred from a phylogenetic tree, where larger values indicate the presence of older lineages and/or restricted range
- Divide the PE by a phylogeny with exactly the same topology but equal branch lengths (cladogram)
- The metric favors lineages with restricted range sizes and long branch lengths
- What about lineages with short branch lengths and restricted range sizes?

PD randomization

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:

a) Phylogenetic over-dispersion - close relatives may exclude each other.

b) Phylogenetic clustering - clades may have evolutionarily conservative habitat preferences and thus close relatives co-occur.

RPD randomization

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

Concentration of short branches: recent evolutionary divergence.

CANAPE: Categorical Analysis of Neo- And Paleo-Endemism

- Use RPE to find and classify centers of endemism.
- Since RPE is a ratio, we need to be sure that there is indeed a high amount of endemism.

- **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; p= 0.05) in either the numerator, the denominator, or both (using the RPE)

- Second, if a grid cell passes step one in RPE, we divide the centers of endemism into three meaningful, non-overlapping categories

Mishler et al. (2014)

CANAPE: Categorical Analysis of Neo- And Paleo-Endemism

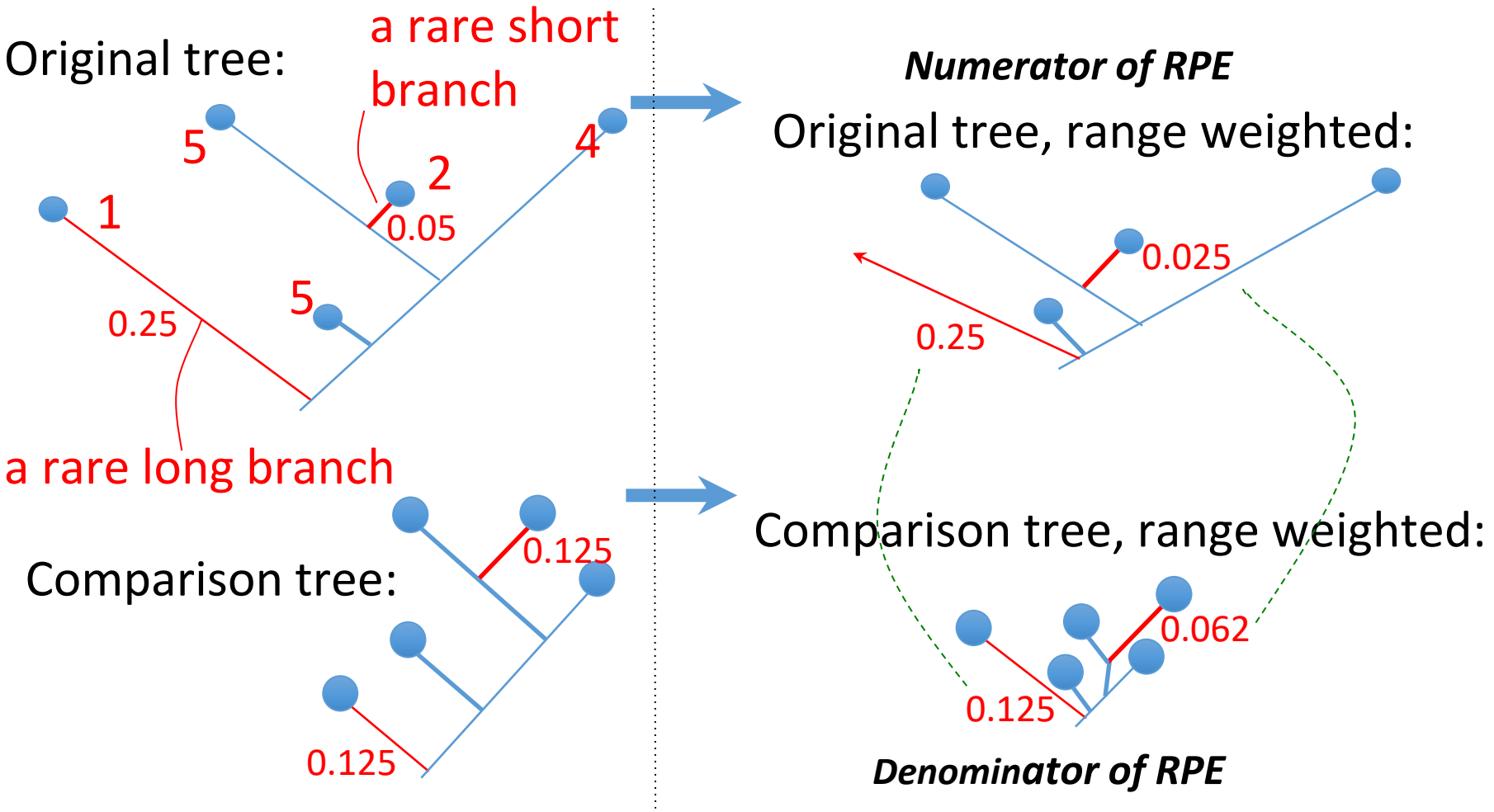
- Based on a two-tailed test ($p<0.05$), if a cell is:
  - significantly high RPE – center of paleo-endemism
  - Significantly low RPE – center of neo-endemism

- When the ratio is neither significantly high nor low, then it is a center of **mixed endemism** (mixture of both neo- and paleo-endemic lineages)

- If a mixed endemism cell is significantly high in both the numerator and the denominator of RPE (one-tailed test) at the $p= 0.01$ level, then we have termed it a center of **super-endemism**.

How RPE works in CANAPE

(Categorical Analysis of Neo- And Paleo-Endemism)



The rare long branch is **longer** in the RW original tree than expected under the null.

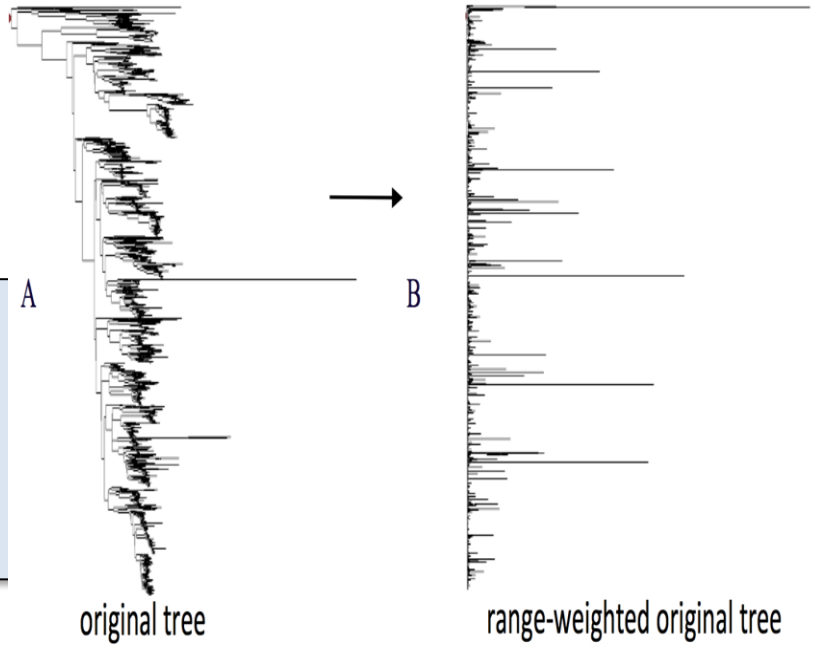
The rare short branch is **shorter** in the RW original tree than expected under the null.

Relative Phylogenetic Diversity and Relative Phylogenetic Endemism

RPD is ratio of A/C:

RPD - *relative phylogenetic diversity=*

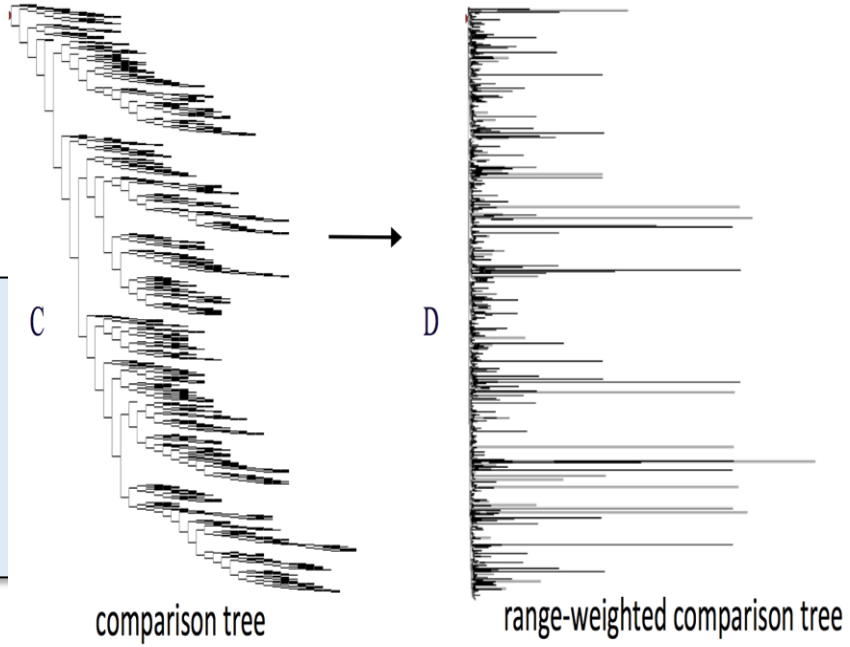
$$\frac{\text{PD on the original tree}}{\text{PD on a comparison tree with all branch lengths equal}}$$



RPE is ratio of B/D:

RPE - *relative phylogenetic endemism =*

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



Mishler et al., (2014)

Phylogenetic measures of geographic similarity

Beta-diversity

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

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

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

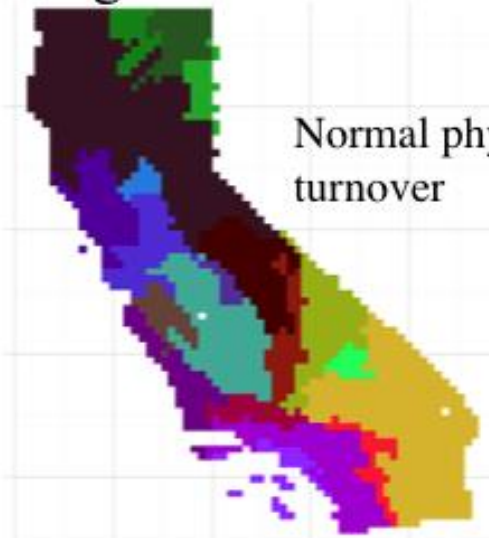
where A is the count of species found in both neighbor sets, B is the count unique to neighbor set 1, and C is the count unique to neighbor set 2.

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.

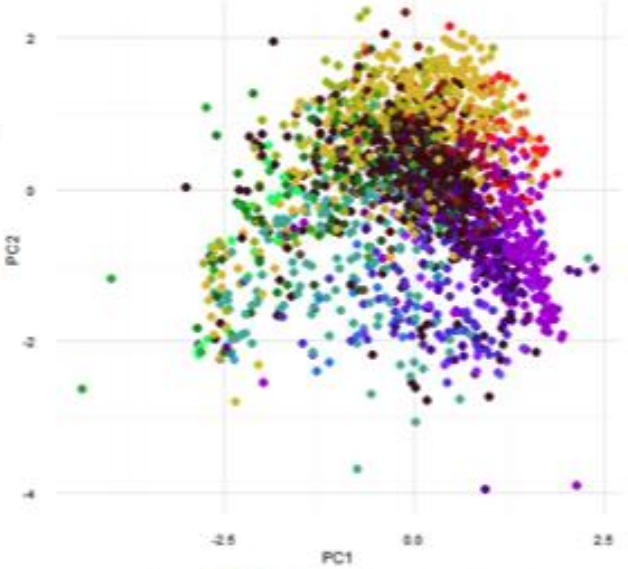
- It is PD turnover if we are using the original tree
- It is PE turnover if we are using the range-weighted tree

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.

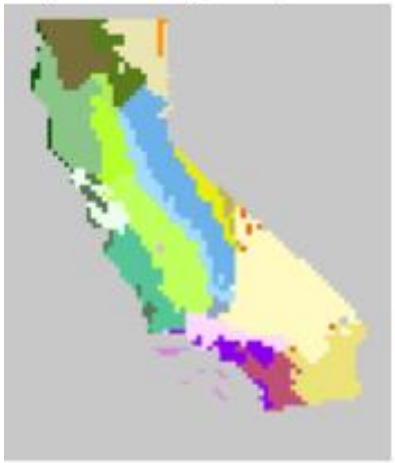
Biotic regionalization - SKATER

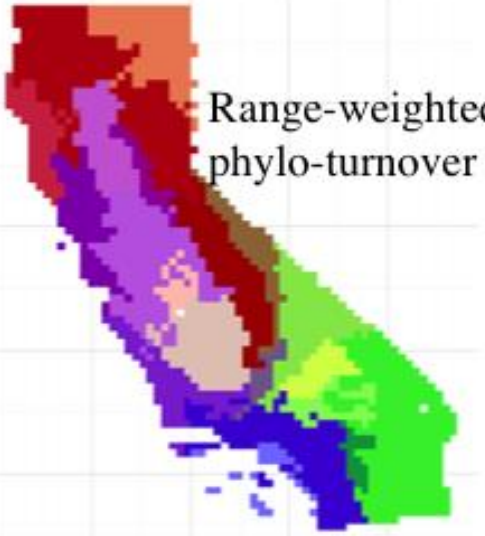


Normal phylo-turnover

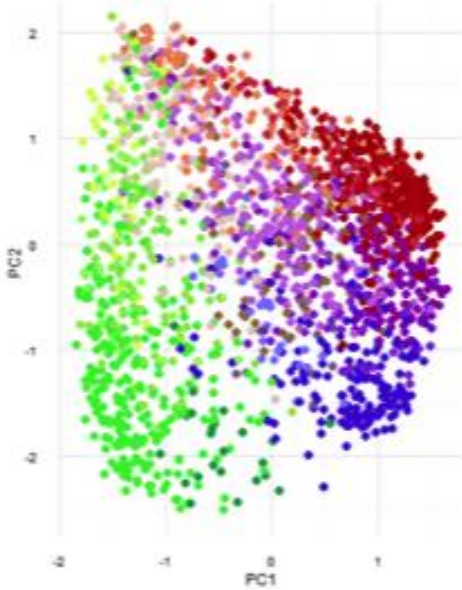


Jepson regions
(15 km grids)





Range-weighted phylo-turnover

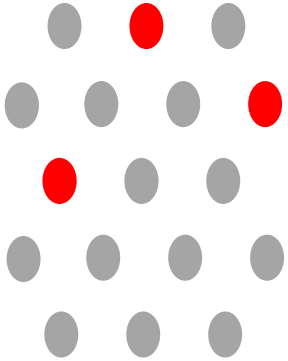


The California Plant Phylodiversity Project

SKATER analysis comparing normal phyloturnover (a) and range-weighted phyloturnover (b). Shown for each are a map and a biplot of PC 1 and PC2. The inset map at left shows current Jepson Biotic Regions for comparison.

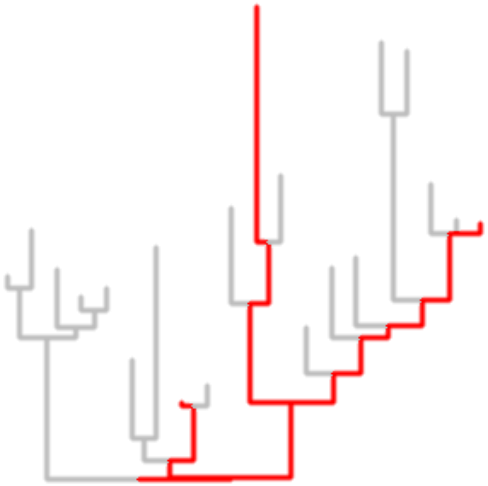
Facets of phylodiversity

Species richness



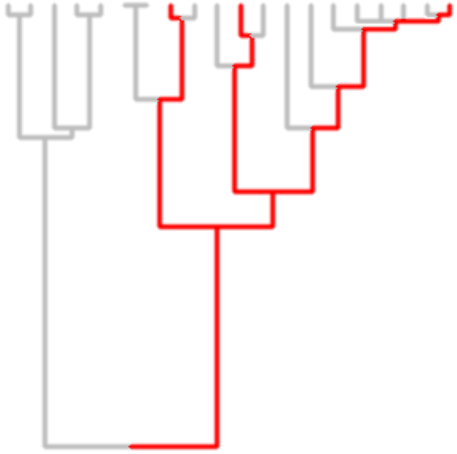
3 species

Character divergence



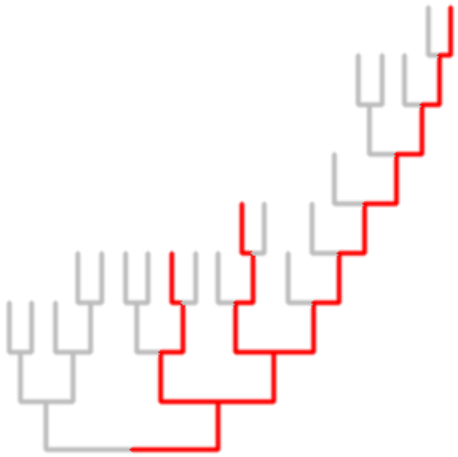
625 changes

Survival time



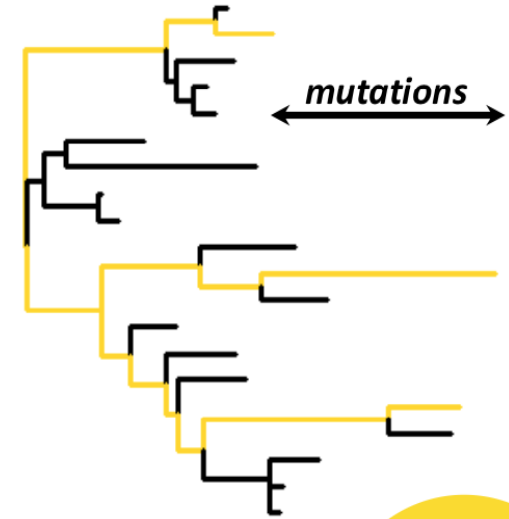
12M years

Cumulative diversification

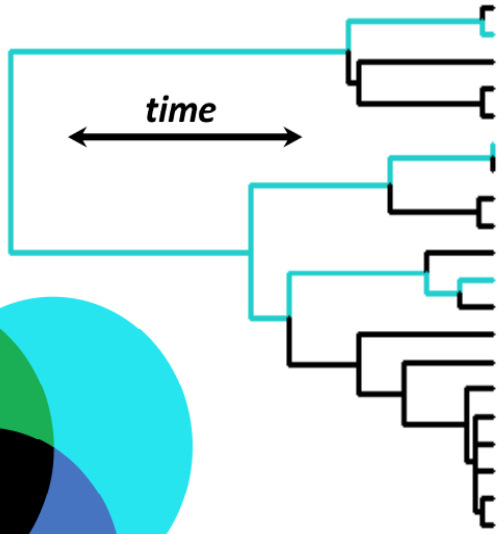


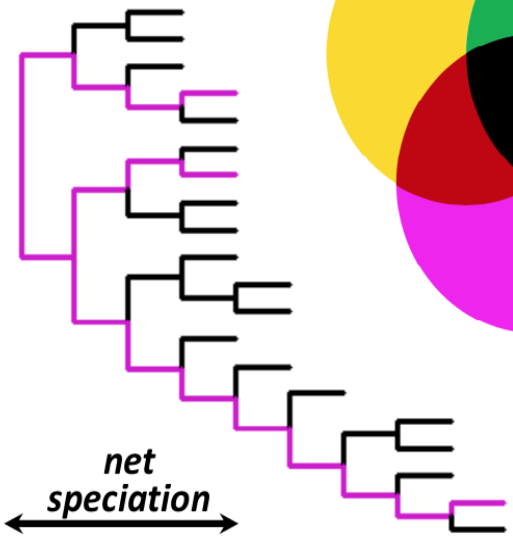
13 splits

DIVERGENCE:
high PD on phylogram =
disparate assemblages



SURVIVAL TIME:
high PD on chronogram =
experienced assemblages



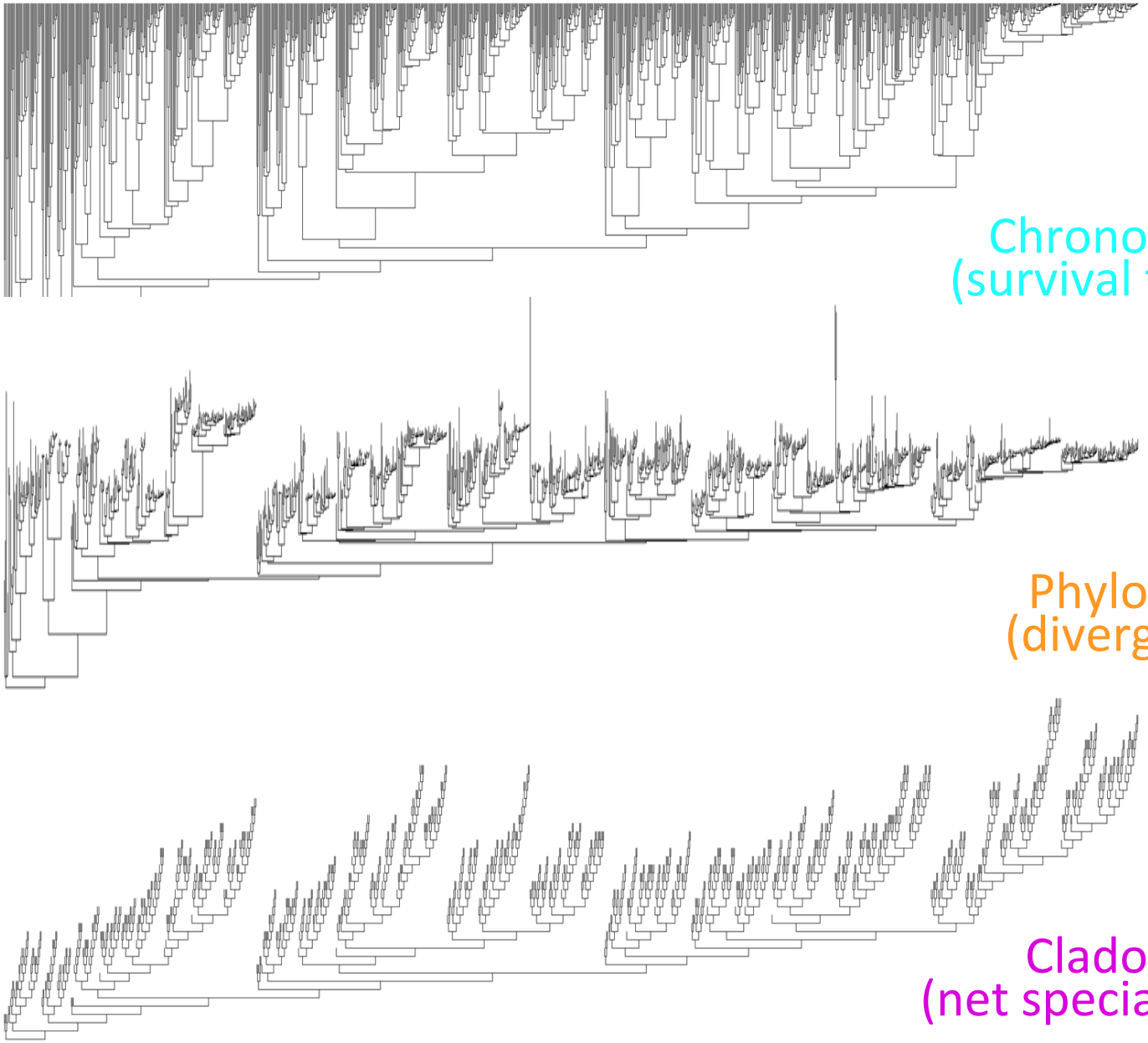


NET DIVERSIFICATION:
high PD on cladogram =
assemblages from
species-rich clades

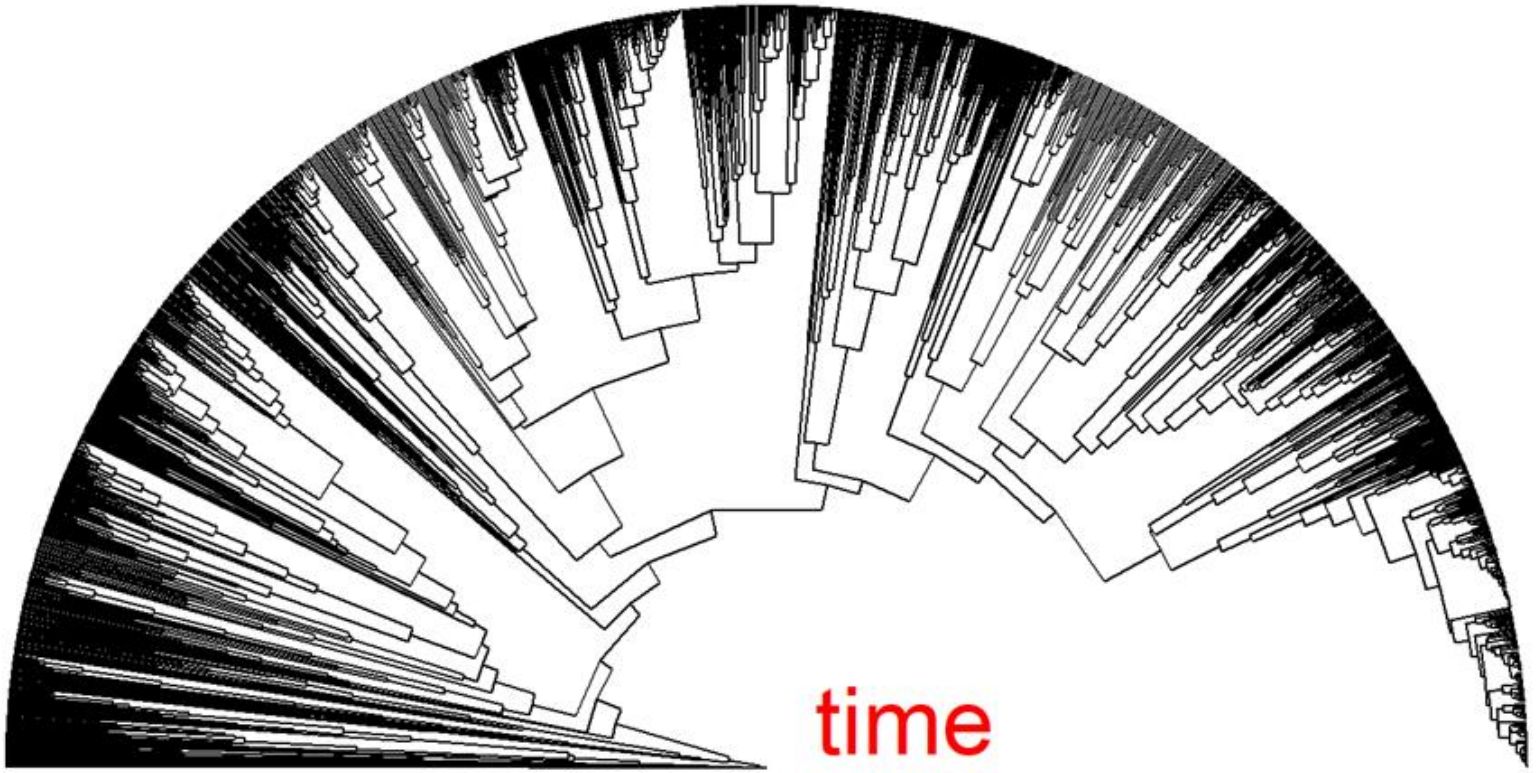
The three facets of phylodiversity

Kling et al. (2018),
Philosophical Transactions Royal Society B.

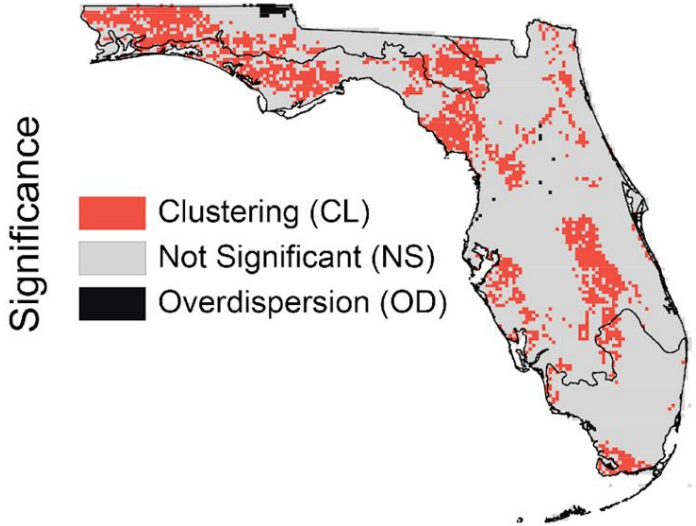
Example of these facets with the California flora



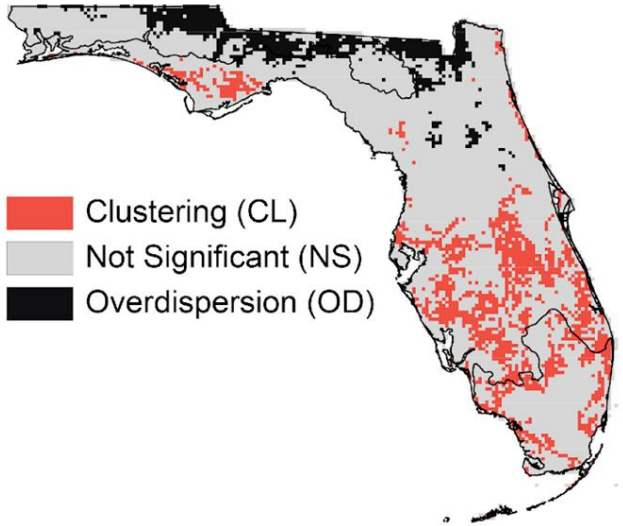
Three facets of the California plant phylogeny



Differences between phylograms and chronograms highlight unique processes that have shaped Florida's ecosystems

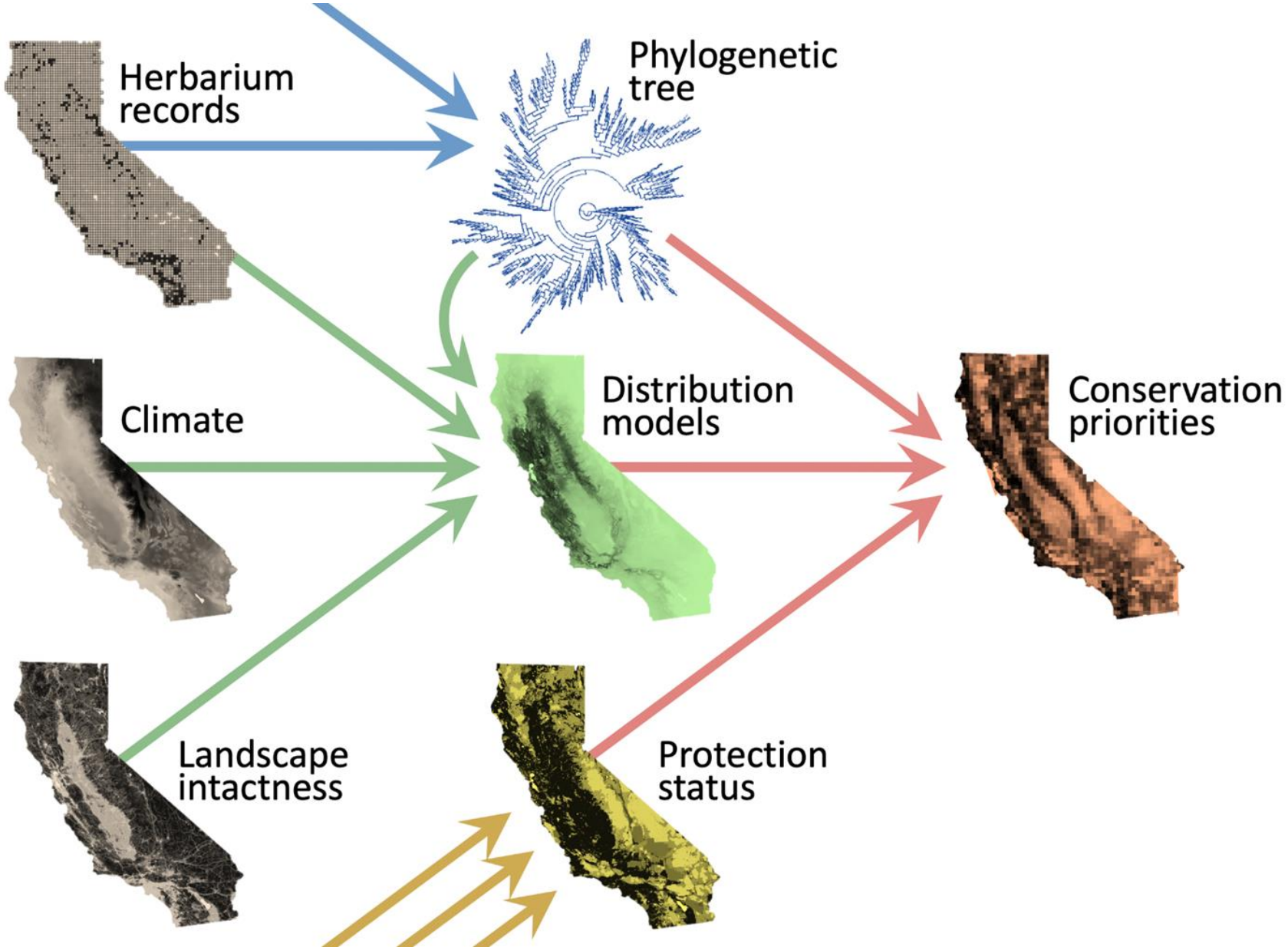


Character Divergence



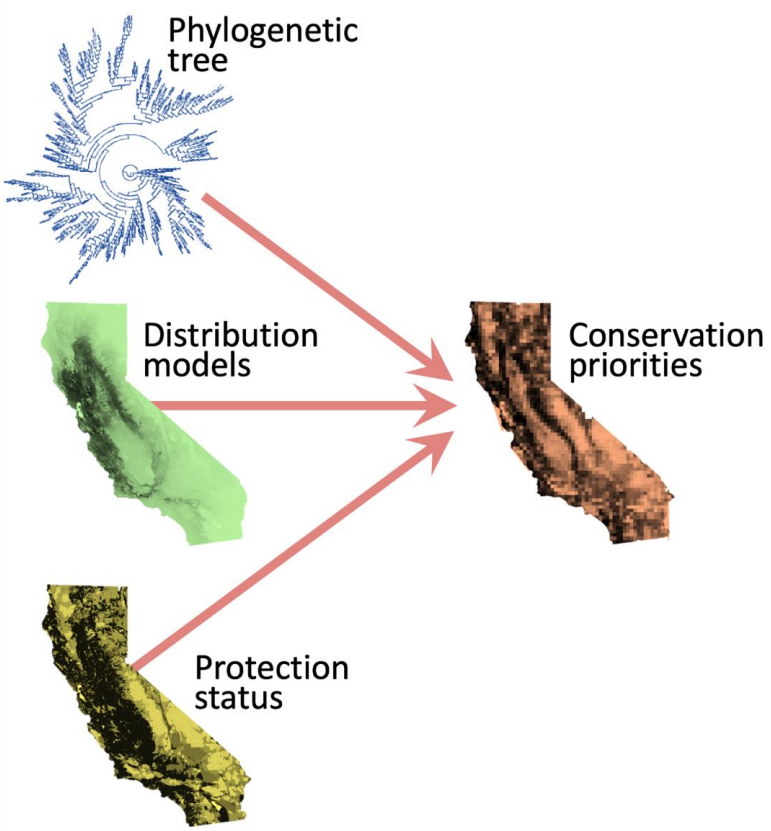
Survival Time

Conservation prioritization using spatial phylogenetics



Key considerations in conservation prioritization

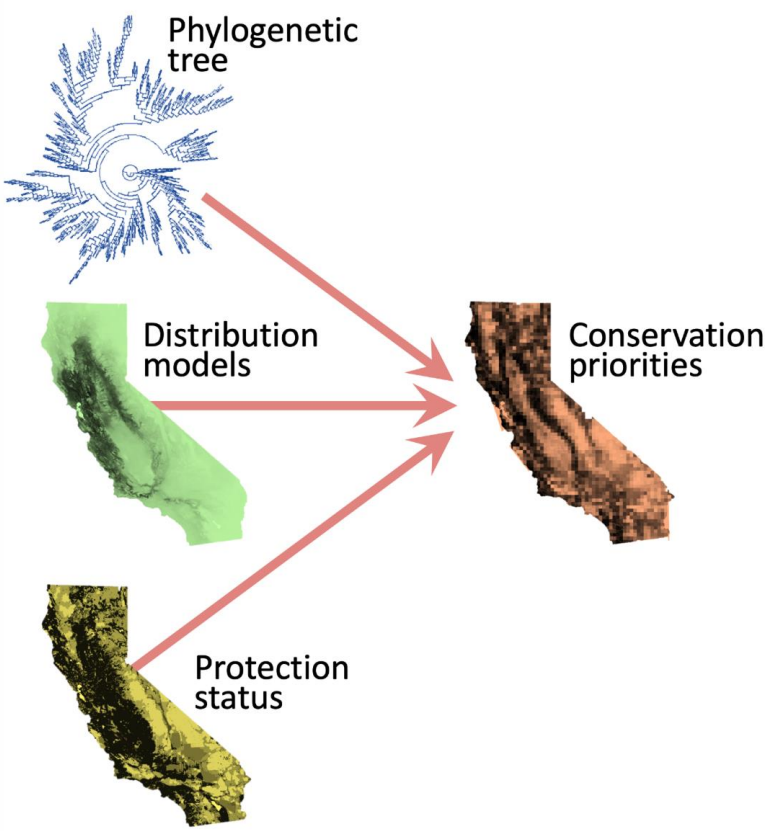
- Probability that taxa are present at site (range models)
- Unique evolutionary attributes of each taxon (branch length)
- Importance of site to each taxon (range size)
- Portion of taxon's range that's already protected
- Current site protection level
- Representation of taxa across proposed reserve network (complementarity)



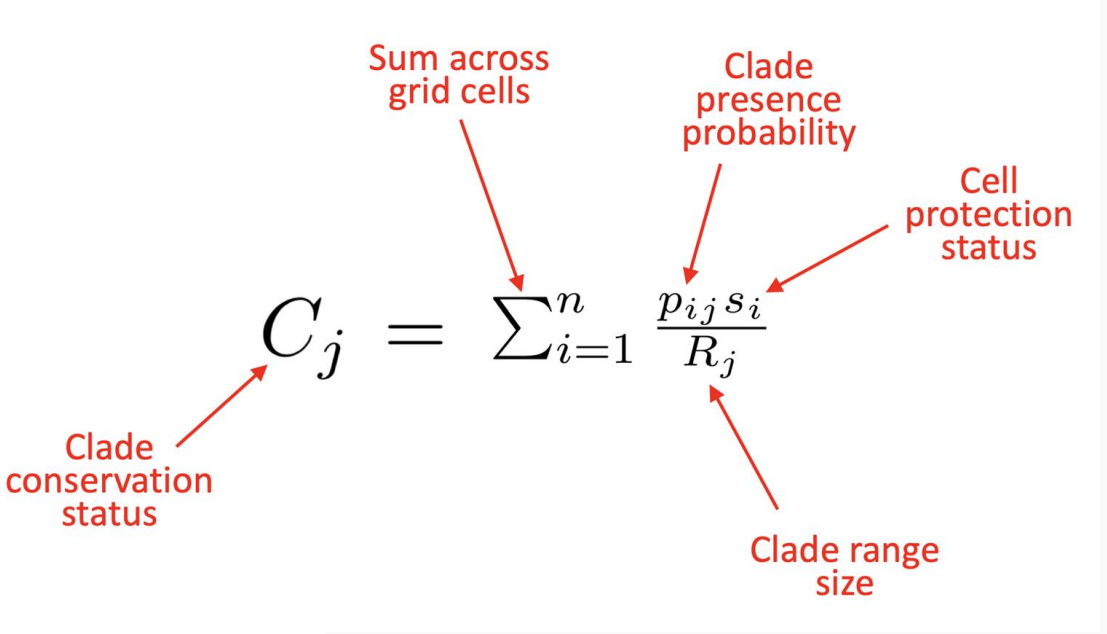
Stepwise optimization algorithm

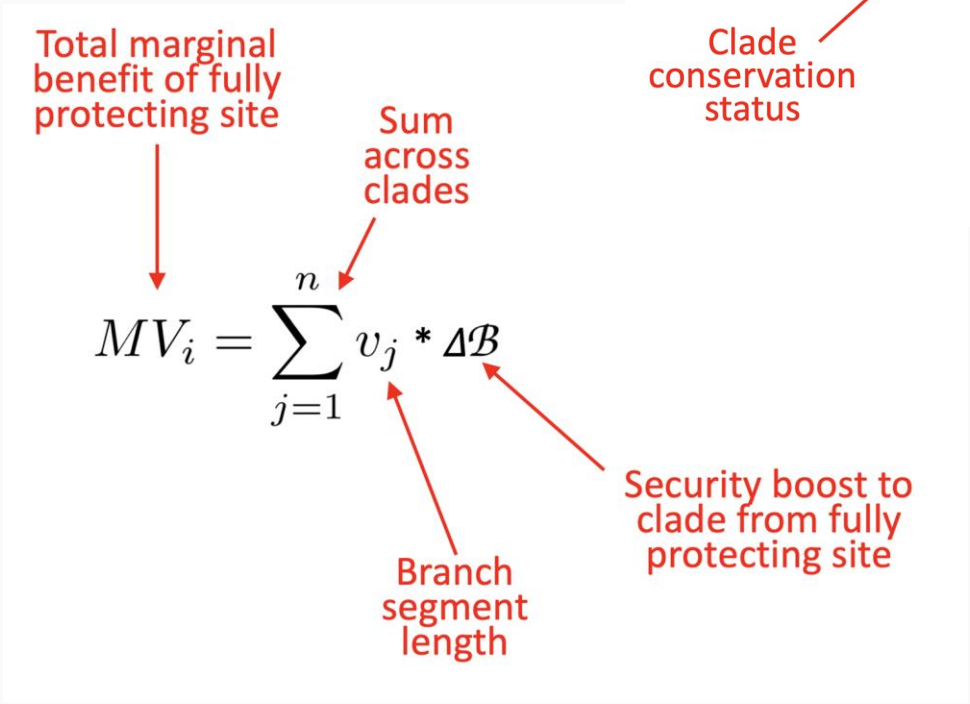
Computes an ordered ranking of new reserve sites that are:

- Poorly protected
- Contain many taxa with
  - Long branches
  - Small ranges
  - Poor range protection
- Complement other reserves



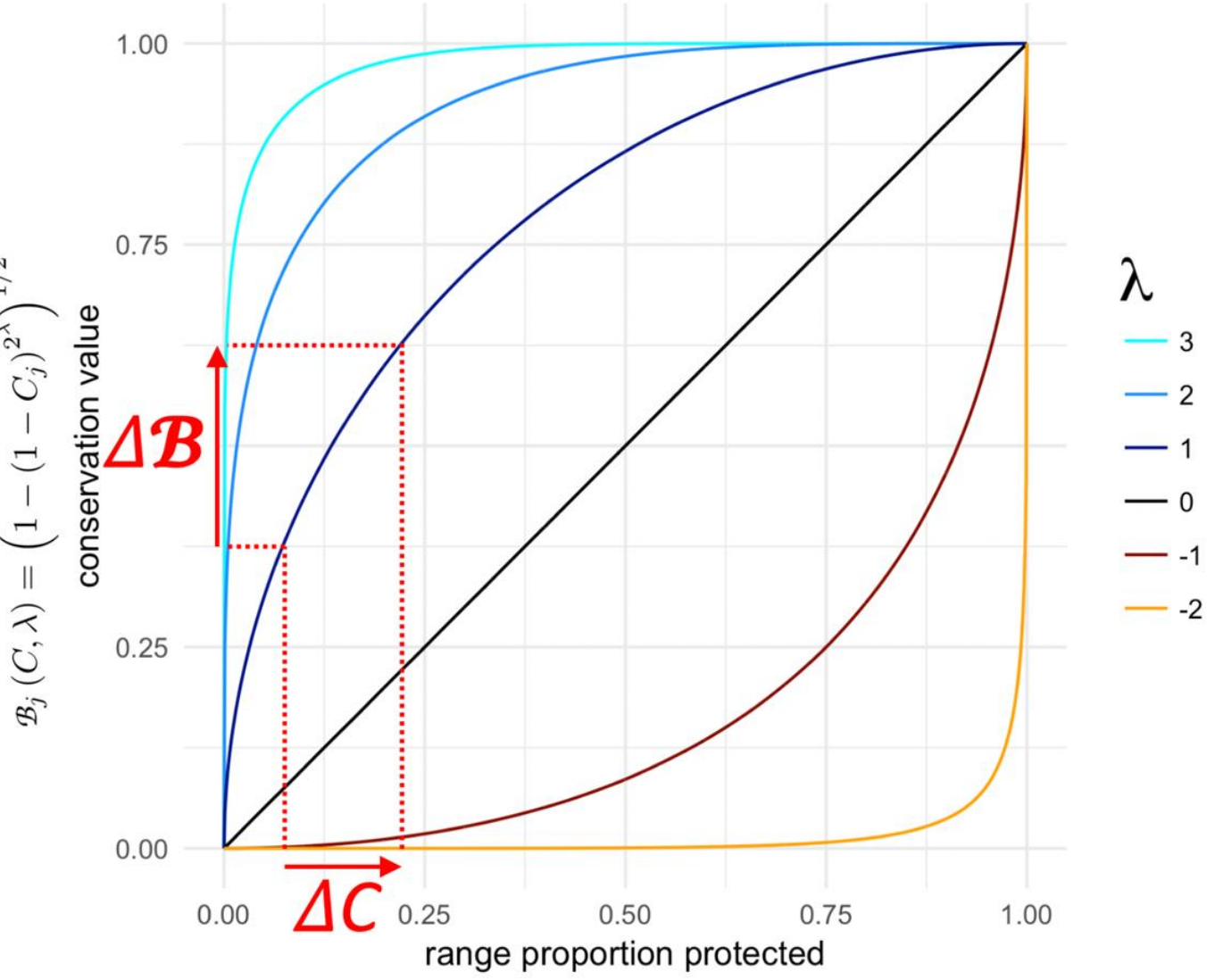
Stepwise optimization algorithm





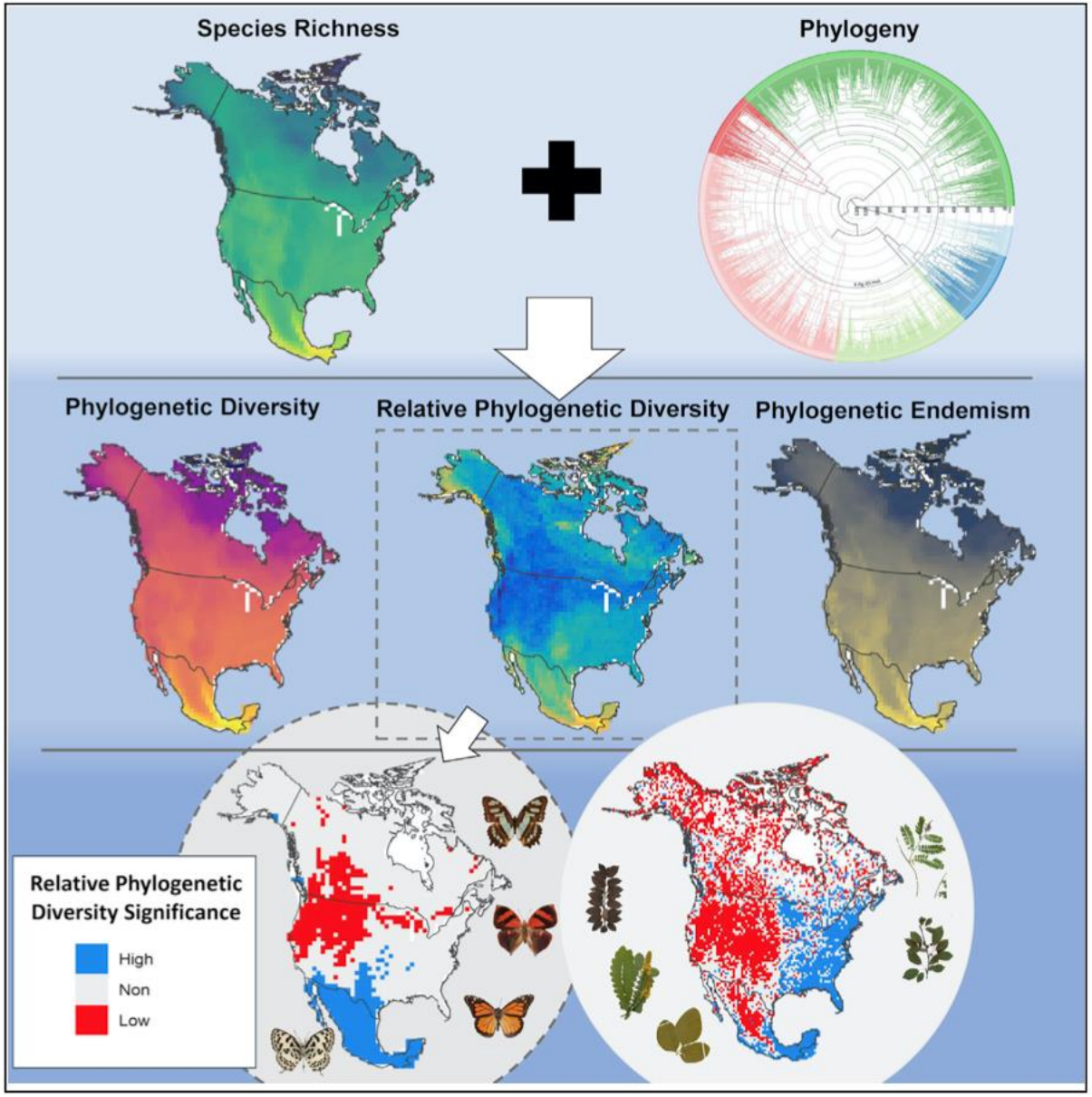
Stepwise optimization algorithm

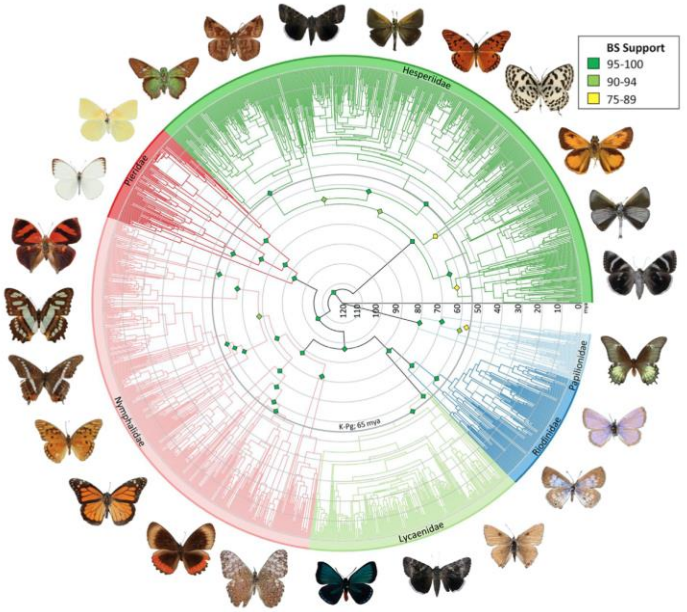
Taxon conservation benefit function



Testing hypotheses using spatial phylogenetics

- Compare phylodiversity patterns across taxonomic groups for envolutionary congruence, co-evolution, etc.
- Relate phylodiversity with paleo- and current climatic conditions to test biogeographic hypotheses
- Further investigations to identify eco-evolutionary drivers of neo- and paleo-endemism





Earl C., M.W. Belitz, S.W. Laffan, V. Barve, N. Barve, D.E. Soltis, J.M. Allen, P.S. Soltis, B.D. Mishler, A.Y. Kawahara, and R. Guralnick. 2021. Spatial phylogenetics of butterflies in relation to environmental drivers and angiosperm diversity across North America. iScience 24 (4): 102239.

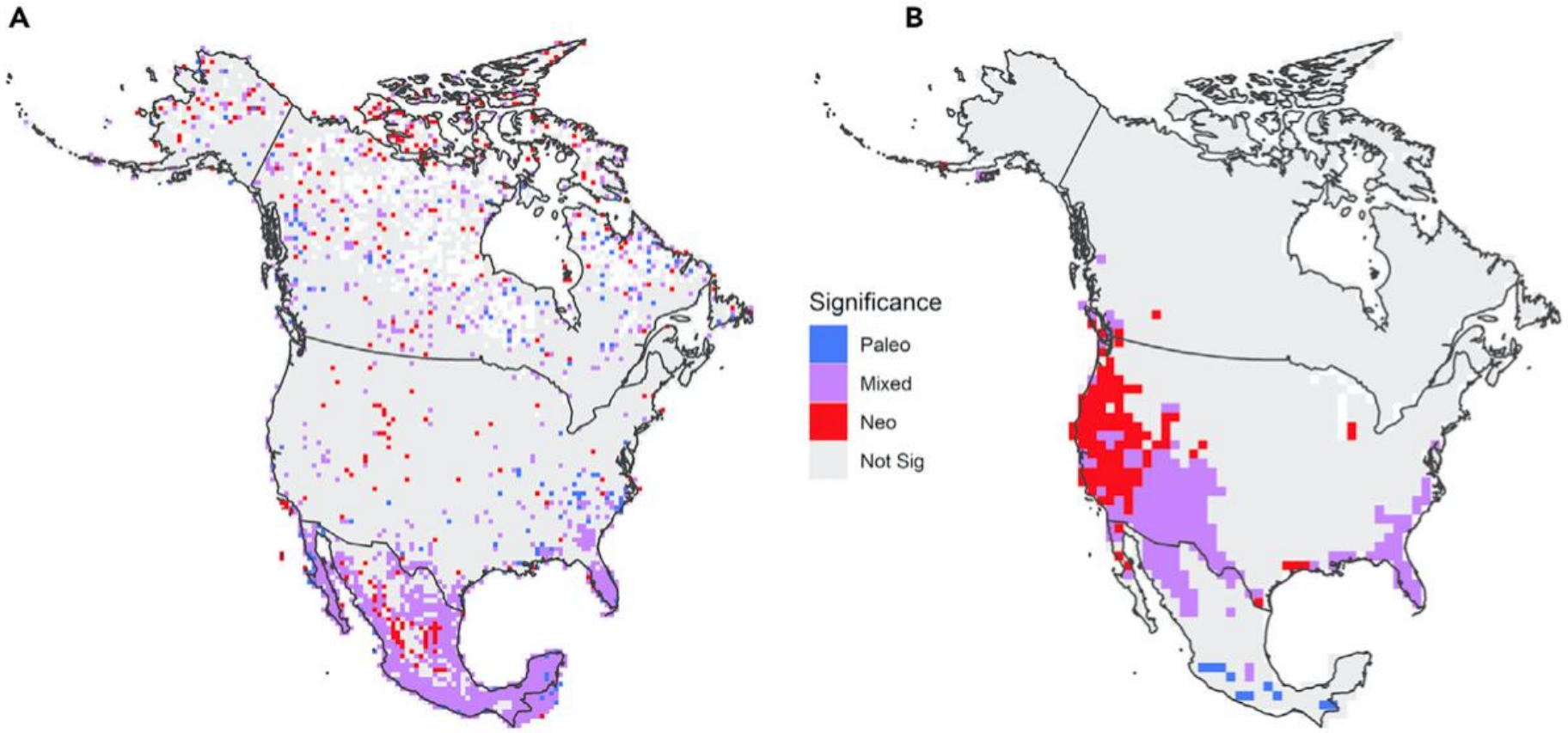
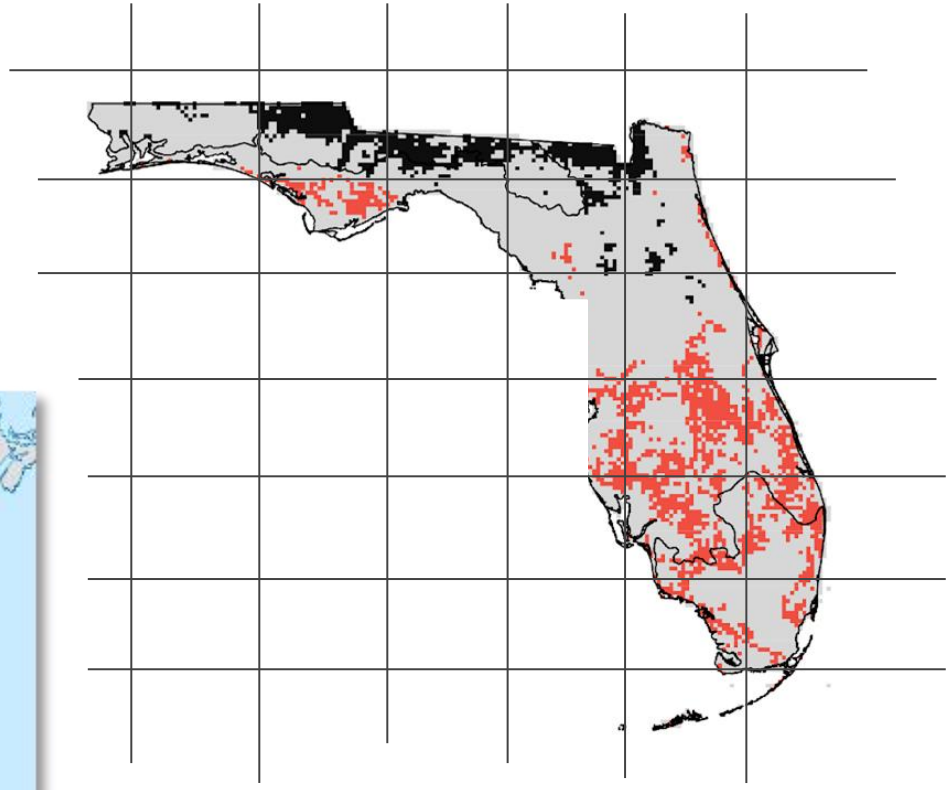


Figure 5. CANAPE results

CANAPE results showing statistically significant centers of phylogenetic endemism for (A) angiosperms and (B) butterflies. All cells that are colored have significantly high PE. Red cells have concentrations of rare short branches (neoendemism); blue cells have concentrations of rare long branches (paleoendemism), and purple cells have mixtures of neo- and paleoendemism.

Scale Matters: How do we choose the appropriate scale? **Gridsize**

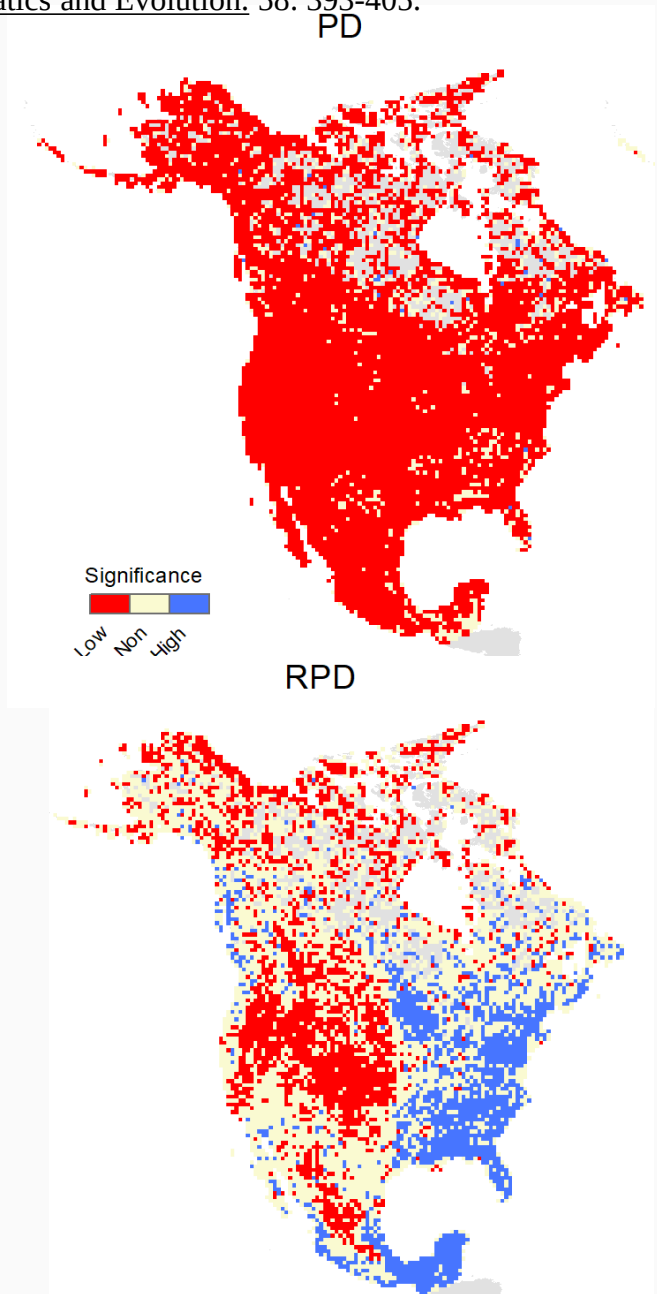
Scale in terms of grid size



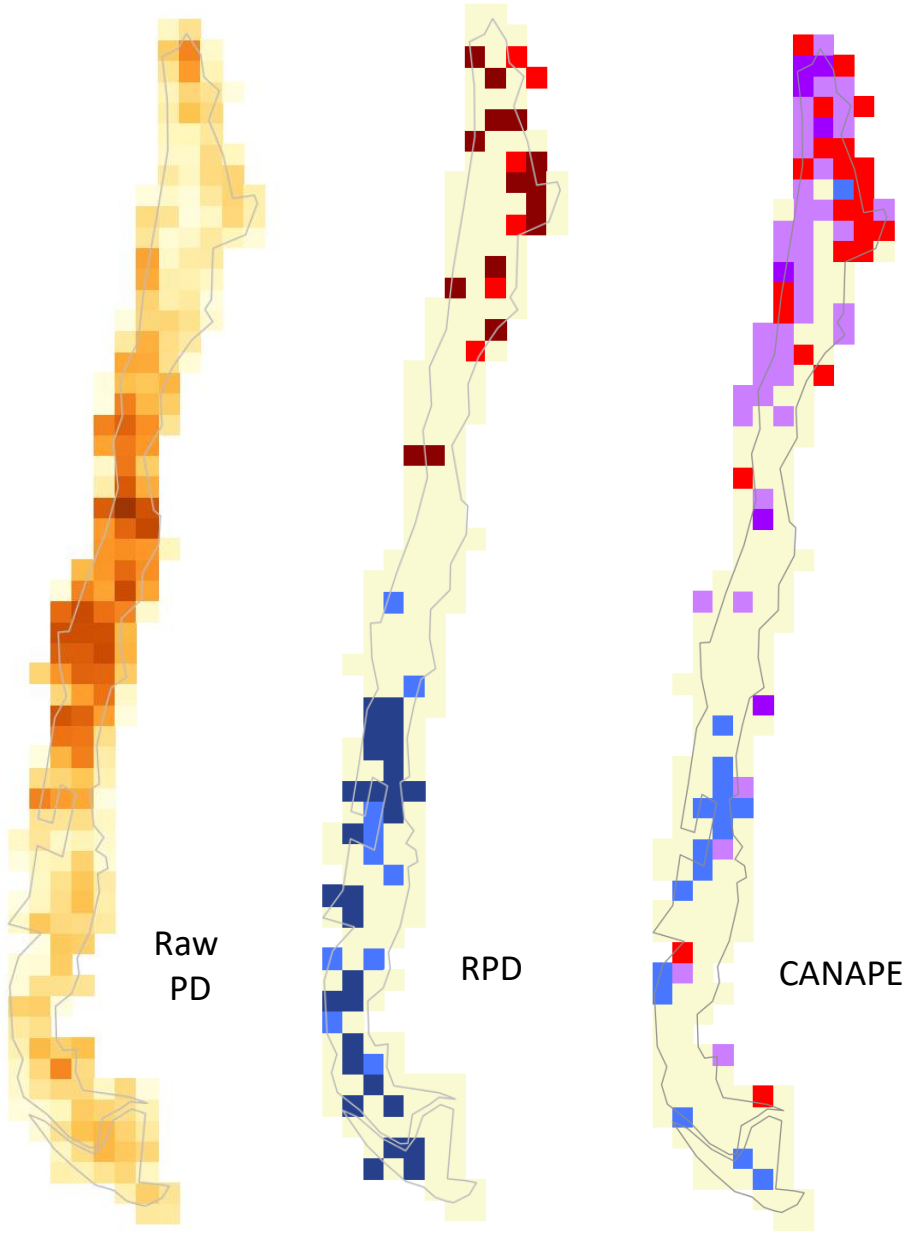
Be explicit with questions based on extent and measurements

- Results depend on the geographic extent and boundaries of analysis
- Results depend on the species pool and the phylogenetic tree that connects those species
- Questions and interpretations must consider the delimitations of your region and relationship among the species included within those delimitations

B.D. Mishler, R.Guralnick, P.S. Soltis, S.A. Smith, D.E. Soltis, N. Barve, J.M. Allen, and S.W. Laffan. 2020. Spatial phylogenetics of the North American flora. Journal of Systematics and Evolution. 58: 393-405.

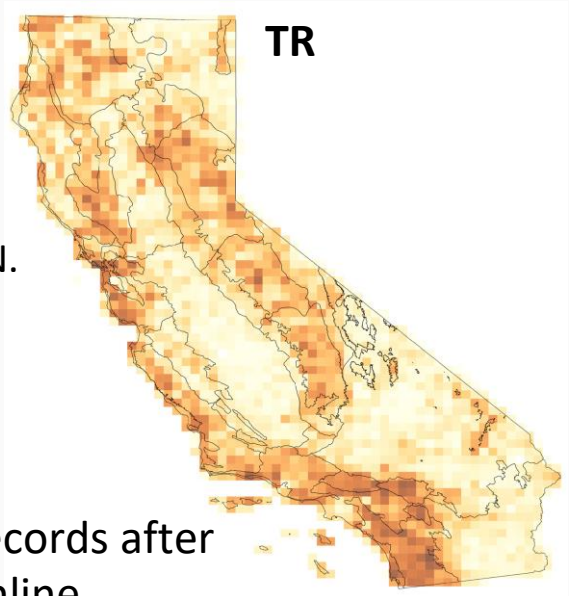


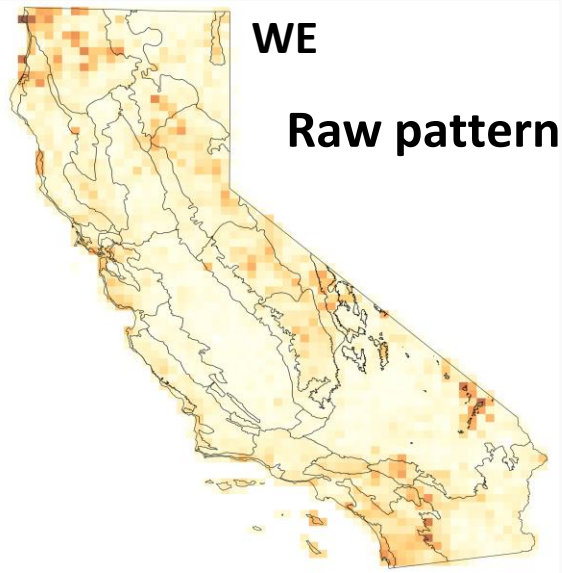
R.A. Scherson, A.H. Thornhill, R. Urbina-Casanova, WA. 2017. Freyman, P.A. Pliscoff, and B.D. Mishler. Spatial phylogenetics of the vascular flora of Chile. Molecular Phylogenetics and Evolution 112: 88-95.



Quick example from the California flora

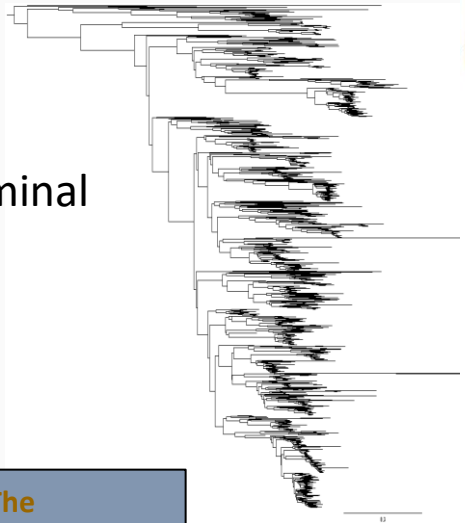
A.H. Thornhill, B.G. Baldwin, W.A. Freyman, S. Nosratinia, M.M. Kling, N. Morueta-Holme, T.P Madsen, D.D. Ackerly, and B.D. Mishler. 2017. **Spatial phylogenetics of the native California flora**. *BMC Biology* 15:96.



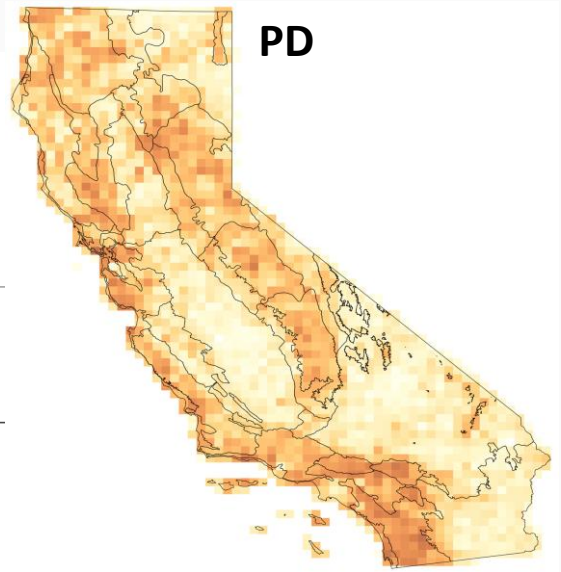


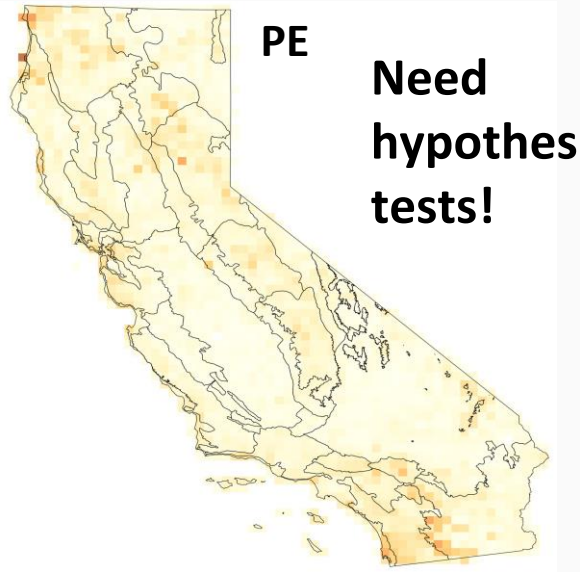
Raw patterns

1.4 million herbarium specimen records after cleaning – all datasets deposited online

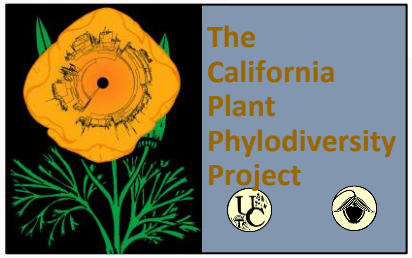


1083 terminal clades
9 genes

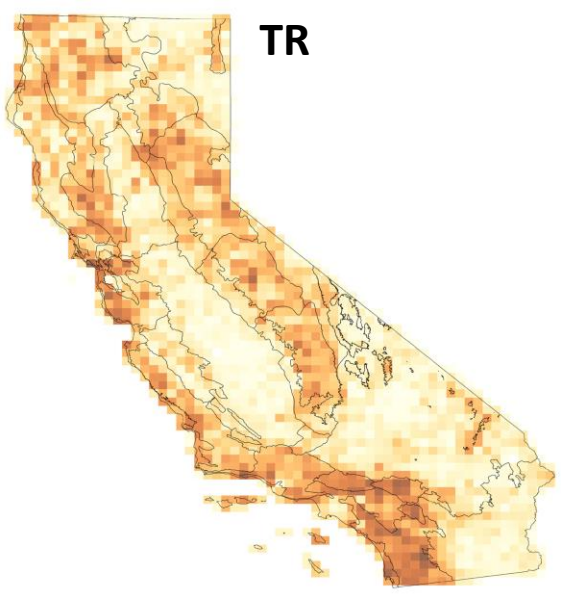


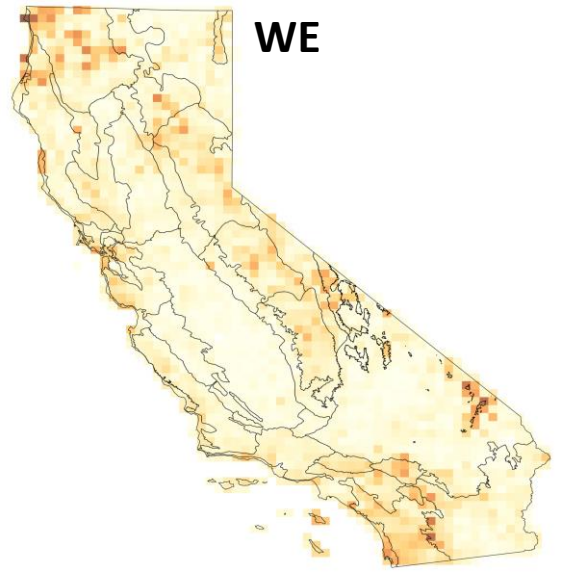


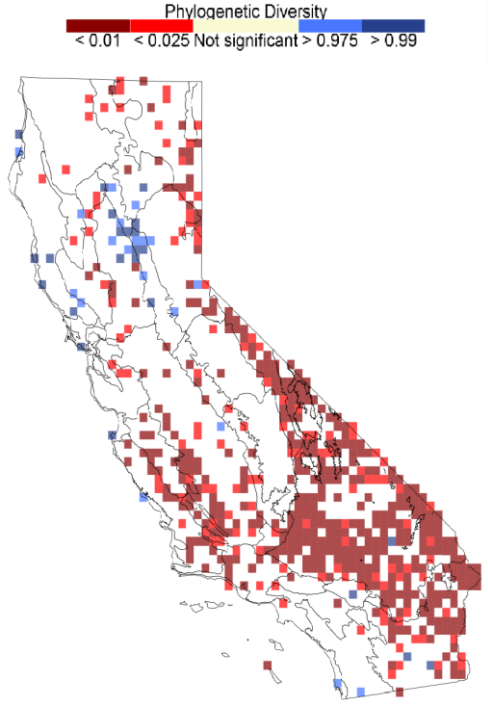
Need hypothesis tests!

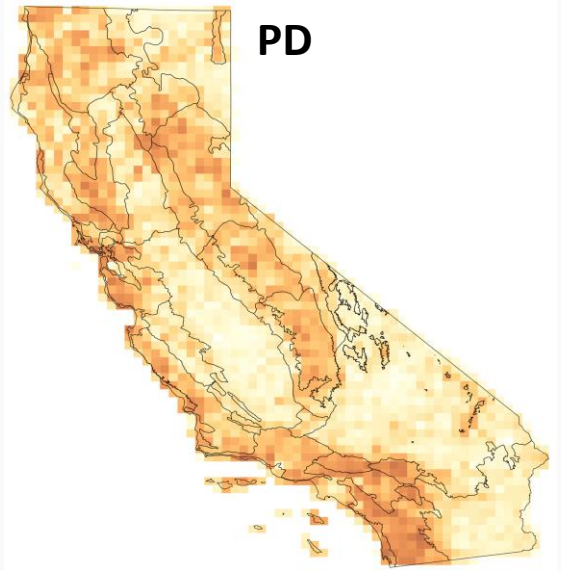


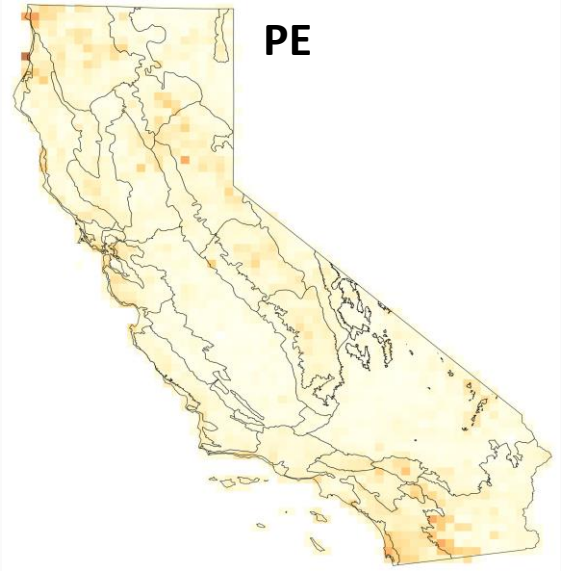
We test statistical significance using spatial randomizations of the terminal taxa on the map, subject to two constraints: richness of each grid cell and range size of each taxon remains constant.

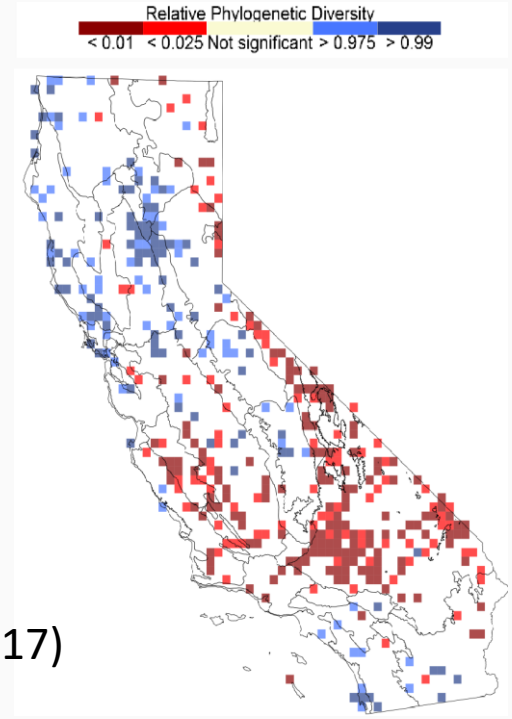




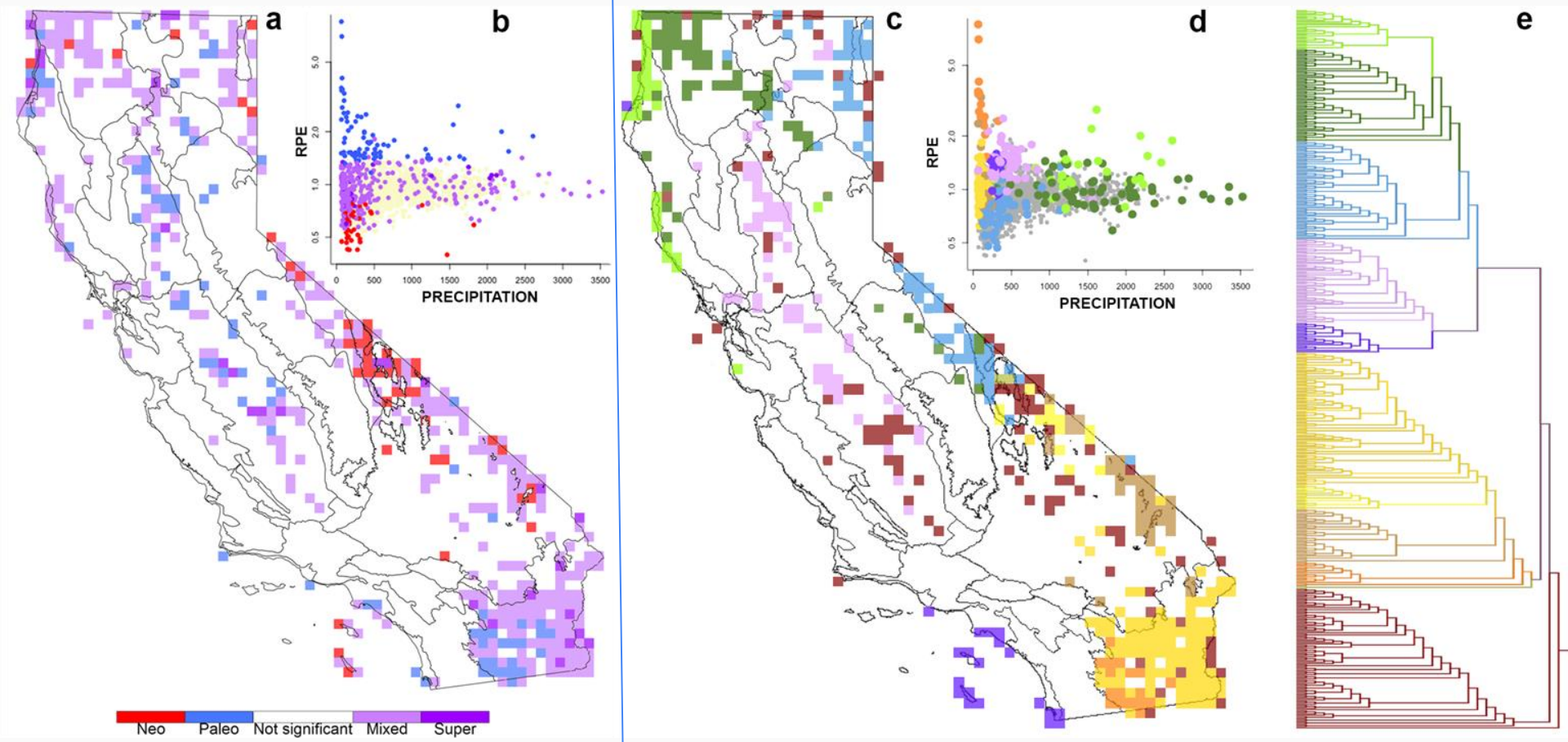








Thornhill et al. (2017)



CANAPE

Phylogenetic turnover among significant centers of endemism found by CANAPE

The California Plant Phylodiversity Project

Thornhill et al. (2017)