Comparative mutation load in the California Conservation Genomics Project

Anne Nakamoto[1,2], Erik Enbody[3], Russ Corbett-Detig[1,2]

[1]Department of Biomolecular Engineering, University of California Santa Cruz, CA, USA; [2]Genomics Institute, University of California Santa Cruz, CA, USA; [3]Department of Computational Biology, Cornell University, Ithaca, NY, USA

Corbett-Detig Lab
✉ aanakamo@ucsc.edu

Background

A landscape genomics approach to biodiversity conservation

The **California Conservation Genomics Project (CCGP)**[1] provides an extensive dataset of ~230 species sampled across California.(Fig. 1A) A major goal is to understand population health, which can be assessed using many metrics. Healthy populations often have[2]…

- ⇧ Effective population size & genetic diversity
- ⇩ Homozygosity & inbreeding depression
- ⇩ **Deleterious mutation load** (Fig. 1B)

Here, we investigate patterns in genetic load across populations in California's landscape

Focus has traditionally been placed on maximizing genetic diversity in efforts to rescue species of conservation concern. However, the genetic burden imposed by deleterious mutations may be a more relevant metric to recent time-scales, and to determining extinction risk.[3]

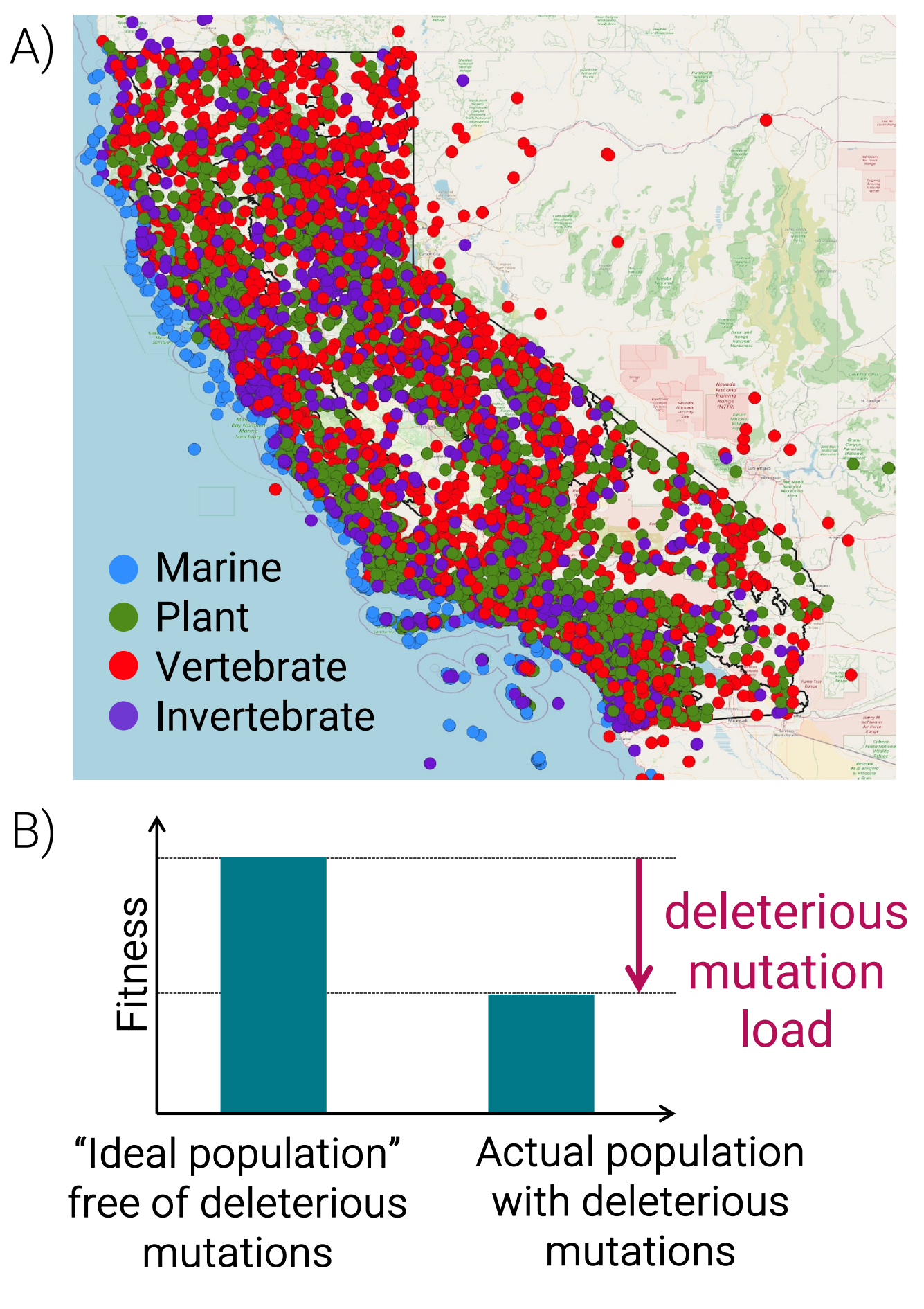

Figure 1: A) CCGP sampling. B) Simplified definition of deleterious mutation load.

Methods

Calling putatively deleterious variation based on sequence conservation

Mutations occurring at evolutionarily constrained genomic sites are likely to have deleterious effects.[2] We compute **PhyloP sequence conservation scores**,[4] which don't rely on the availability of gene annotations, and allow for calling deleterious variants in non-coding regions.(Fig. 2)

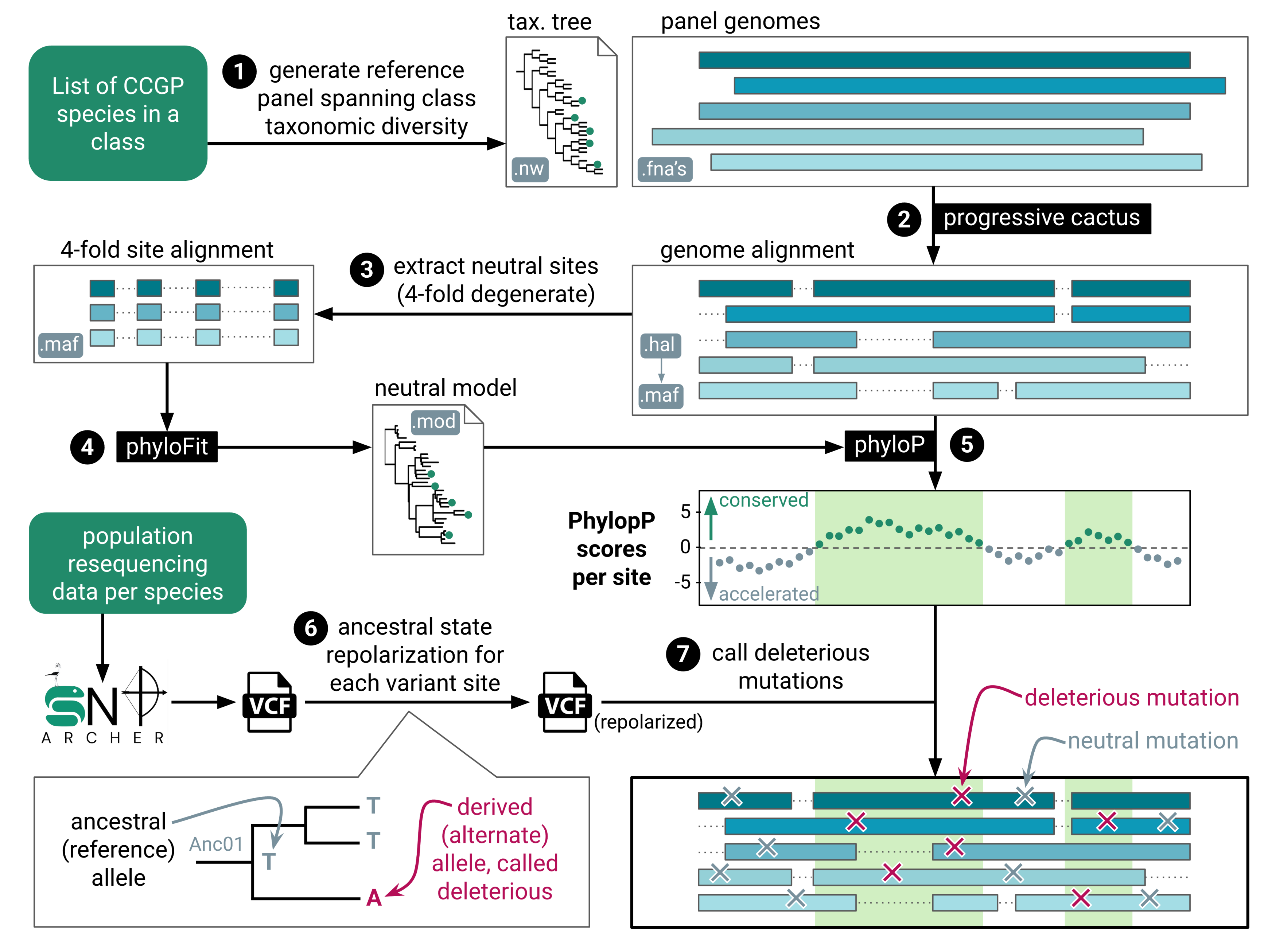

Figure 2: Computational workflow used to (1) generate a reference panel of genomes for each CCGP class, (2) perform whole-genome alignment with Progressive Cactus, (3-5) compute PhyloP sequence conservation scores based on a neutral model of evolution from 4-fold degenerate sites, (6) repolarize variants called by the snpArcher workflow[5] based on the inferred ancestral state, and (7) call putative deleterious variants at conserved sites.

Results

Realized mutation load varies within and across CCGP species

Strongly deleterious mutations are often recessive and have a large fitness effect. Thus, the number of homozygous variants at constrained sites estimates realized mutation load,[6] and can be normalized by the number of highly conserved sites in each genome. This ratio varies across species,(Fig. 3A) and reveals differences in load among populations of the same species.(Fig. 3B)

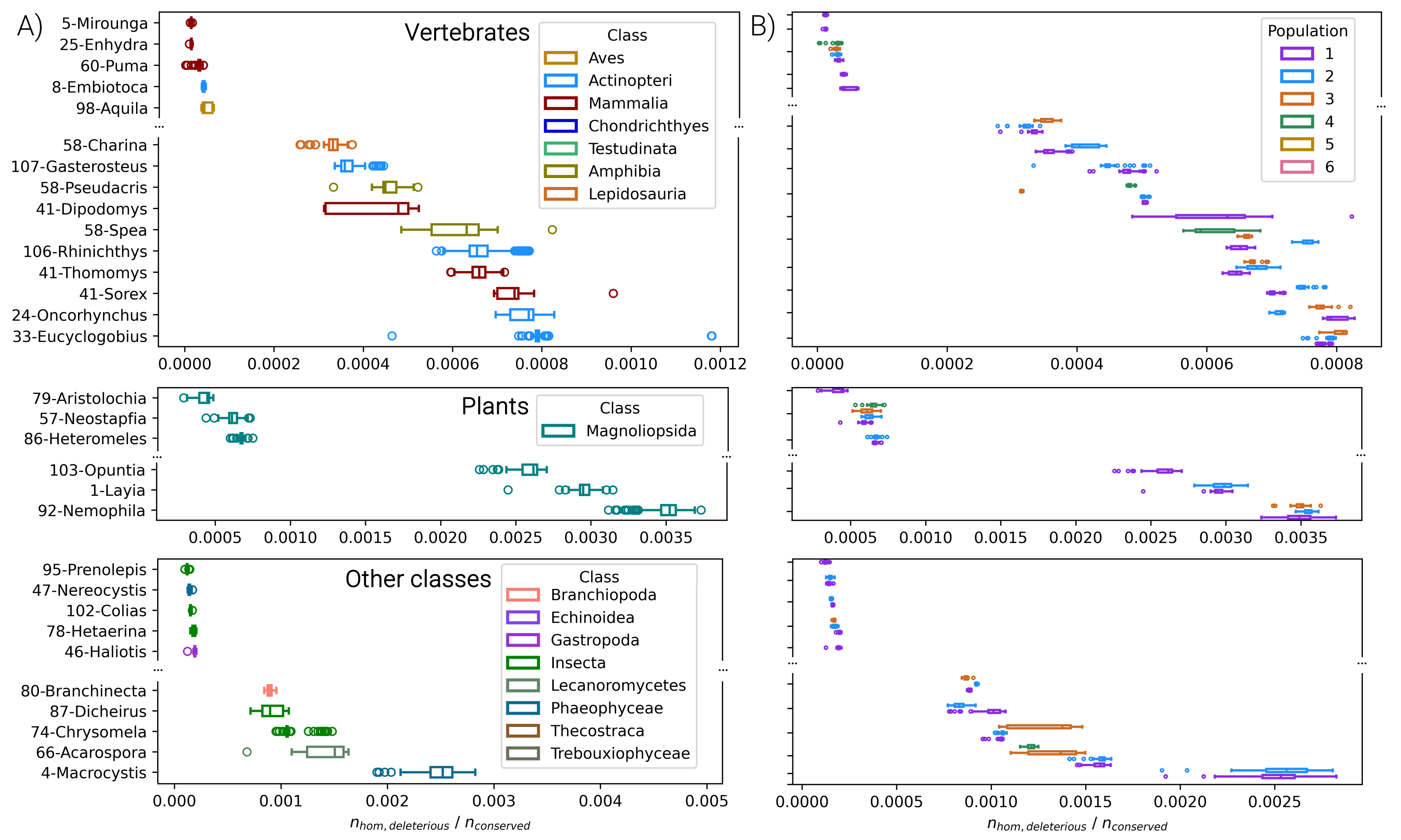

Figure 3: For a truncated set of CCGP species, the ratio of homozygous deleterious variant sites over the set of all conserved sites is shown, colored by A) taxonomic class and B) populations within species.

Interpreting mutation load metrics using the stickleback as an exemplar

The threespine stickleback has distinct marine and inland populations. The inland population is an endangered species that has experienced declines,[7] which is reflected by multiple mutation load metrics.(Fig. 4) Overall, the inland population has greater realized mutation load,(Fig. 4A) likely due to increased inbreeding(Fig. 4E) exacerbated by recent population decline.

Gasterosteus aculeatus
Threespine stickleback

Table 1: Population history metrics.

Population	Tajima's D	π	Fst
Marine (1)	-0.2341	0.0036	0.1007
Inland (2)	0.2308	0.0021	

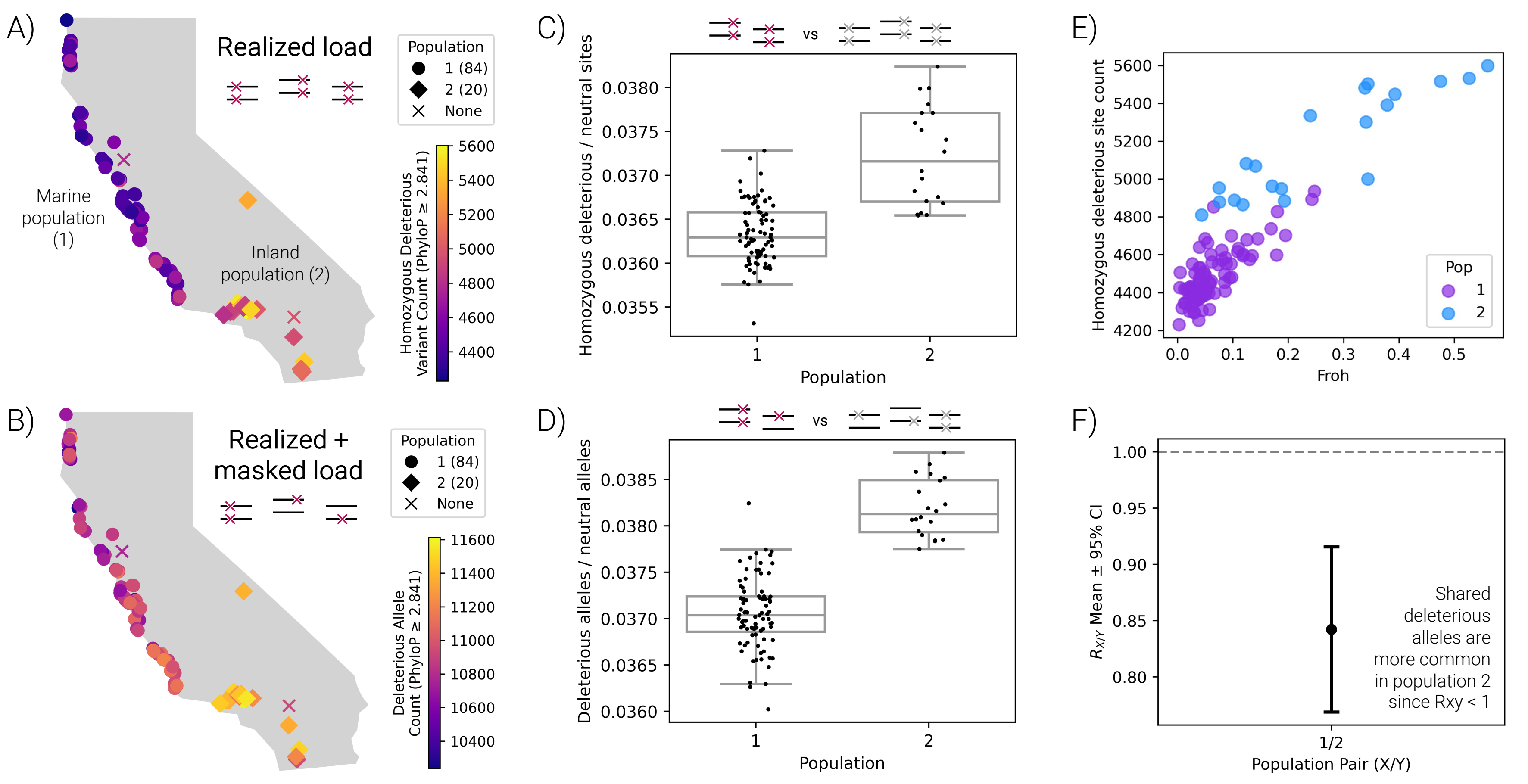

Figure 4: Metrics of mutation load for *G. aculeatus*. Counts of A) homozygous deleterious mutations and B) deleterious alleles per individual are shown, along with the ratio of deleterious (PhyloP ≥ 2.841, determined by FDR correction) over neutral (PhyloP ∈ [-0.1, 0.1]) variants for C) homozygous sites and D) all alleles. E) Inbreeding coefficient (Froh) versus homozygous deleterious site counts. F) Rxy statistic between population pairs.

Discussion and Ongoing Work

Investigating broad-scale geographic and phylogenetic patterns in mutation load

The next step of this project is to aggregate our mutation load results to address the following questions:

1. **Can the spatial pattern of mutation load in one species be predictive of load in another species?**
2. **How does the phylogenetic distance between species affect our ability to make such predictions?**

We have begun to look for hotspots of mutation load in the landscape by ranking relative load within species and their populations. For example, ranking the count of homozygous deleterious mutations within species of fish shows a hotspot for southern inland fish populations.(Fig. 5A) Similarly, there is a hotspot around Yosemite for amphibians.(Fig. 5B) Ranking within populations for reptiles shows a gradient of elevated load to the north, and elevated load in island lizard populations.(Fig. 5C)

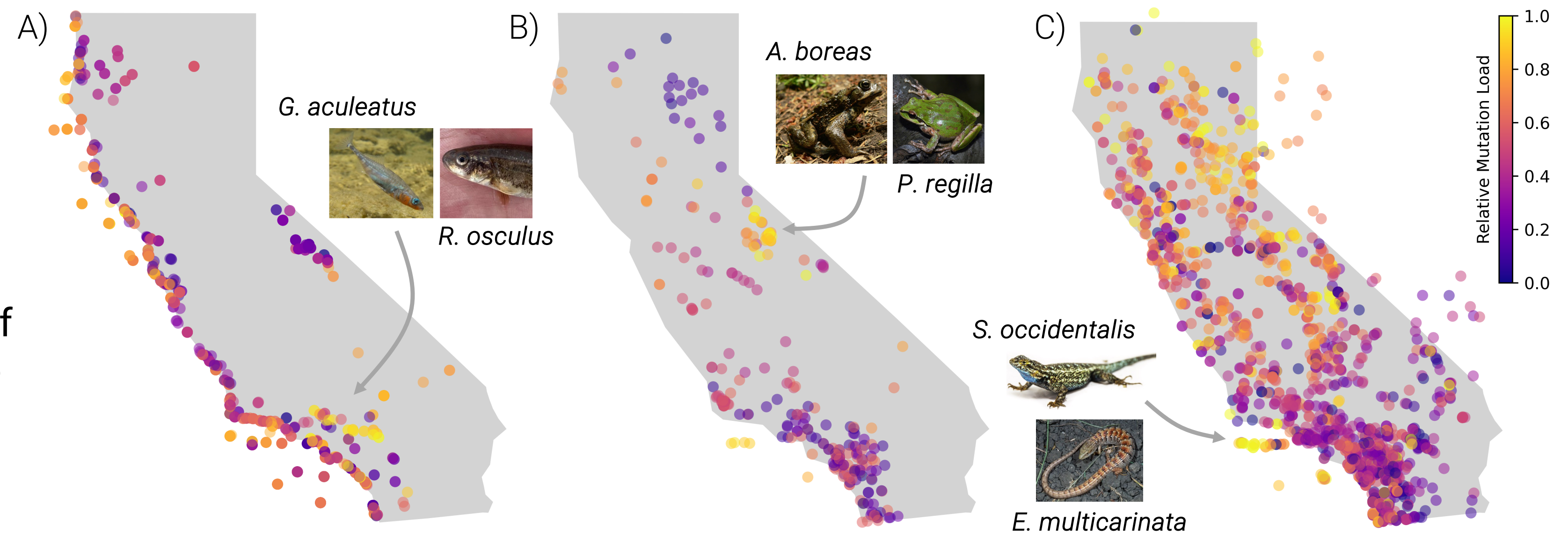

Figure 5: The relative homozygous deleterious mutation counts are ranked within species for A) *Actinopteri* (fish) and B) *Amphibia*, and ranked within species populations for C) *Lepidosauria* (reptiles). Species contributing to hotspots are indicated.

Reconciling the effect of population history on mutation load metrics

The ratio of deleterious to neutral sites(Fig. 4C,D) is often used to evaluate the efficacy of selection within populations. However, non-equilibria conditions can cause spurious signals of reduced selection.[8] For example, the inland stickleback population has a positive Tajima's D indicating a bottleneck,(Table 1) so it's not likely to be experiencing true reduced selection efficacy. Despite this, we can still conclude that the inland population's elevated realized mutation load is likely having a negative effect on its fitness. Keeping these limitations in mind will be important to understanding how mutation load data can be used to make predictions about the future health of these populations.

References

1) Shaffer HB, et al. *J Hered*, 2022. **2)** Robinson J, et al. *Annu Rev Anim Biosci*, 2023. **3)** Kyriazis CC, et al. *Evol Lett*, 2021. **4)** Pollard KS, et al. *Genome Res*, 2010. **5)** Mirchandani CD, et al. *Mol Biol Evol*, 2024. **6)** Femerling G, et al. *Mol Biol Evol*, 2023. **7)** Turba R, et al. Mol Ecol, 2022. **8)** Turba Brandvain Y, Wright IS. *Trends in Genetics*, 2016.

Acknowledgements

Thank you to Erik Enbody and Russ Corbett-Detig for their mentorship, and to the Corbett-Detig Lab for feedback and support. This project utilized the Phoenix computational cluster provided by the UCSC Genomics Institute, and the BioHPC computational resource from Cornell Research & Innovation. Species images are from ccgp.org.
Funding: UC Santa Cruz, NSF-GRFP, NIH (T32HG012344)