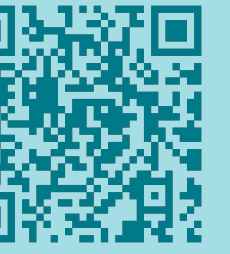


# Comparative mutation load in the California Conservation Genomics Project



Anne Nakamoto<sup>1,2</sup>, Erik Enbody<sup>3</sup>, Anne Chambers<sup>4</sup>, CCGP Consortium, Brad Shaffer<sup>5</sup>, Russ Corbett-Detig<sup>1,2</sup>



<sup>1</sup>Department of Biomolecular Engineering, UC Santa Cruz, CA, USA; <sup>2</sup>Genomics Institute, UC Santa Cruz, CA, USA; <sup>3</sup>Department of Computational Biology, Cornell University, Ithaca, NY, USA; <sup>4</sup>Department of Environmental Science, Policy, & Management, UC Berkeley, CA, USA; <sup>5</sup>Department of Ecology & Evolutionary Biology, UC Los Angeles, CA, USA

✉ [aanakamo@ucsc.edu](mailto:aanakamo@ucsc.edu)

Corbett-Detig Lab

## Background

### A landscape genomics approach to biodiversity conservation

The **California Conservation Genomics Project (CCGP)**<sup>1</sup> 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<sup>2...</sup>

- ↑ 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.<sup>3</sup>

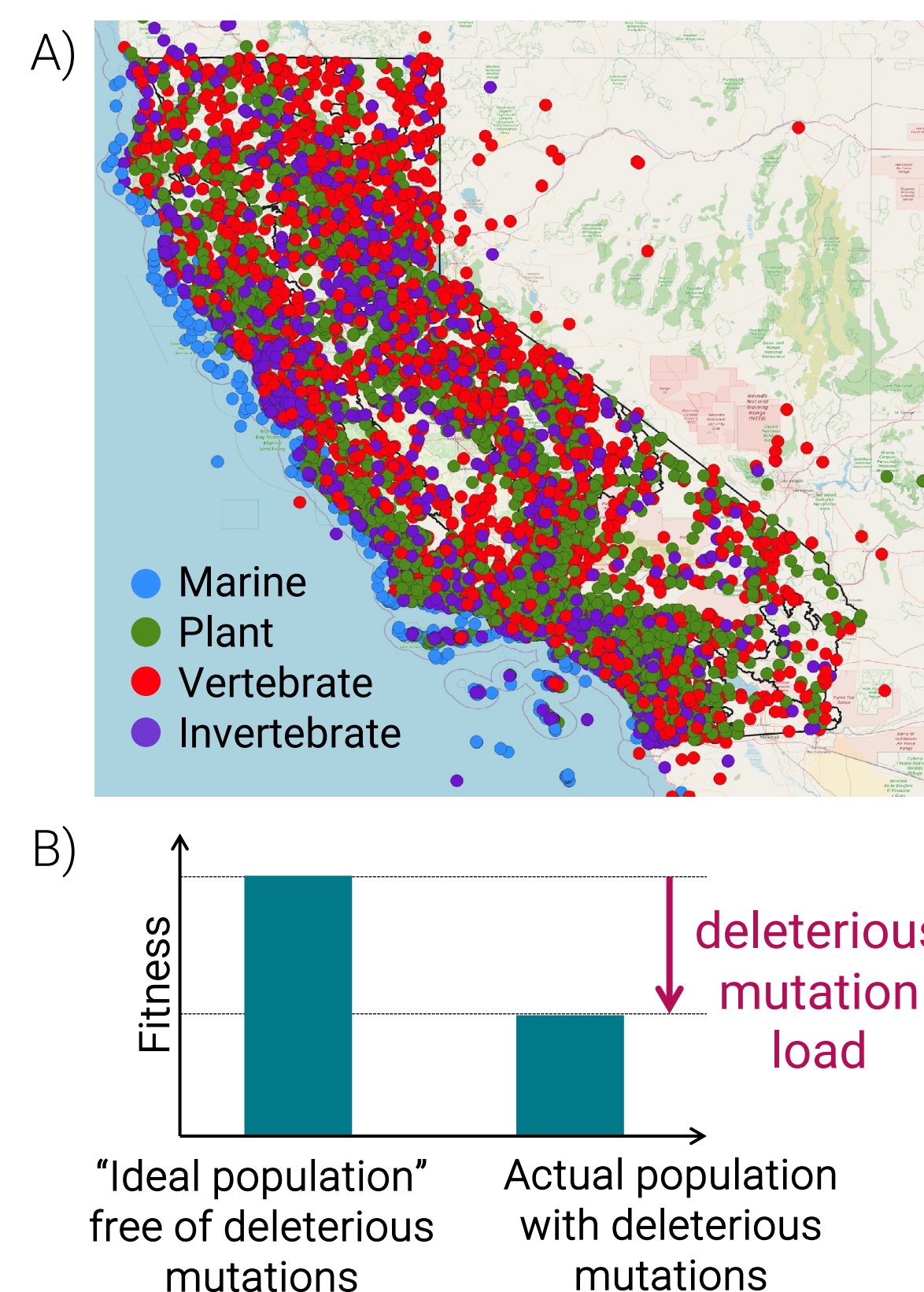


Figure 1: A) CCGP sampling. B) Simplified definition of deleterious mutation load.

## Results

### The extent to which load is spatially structured varies across 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**.<sup>6</sup> We can count these mutations in individuals throughout the landscape, and infer load at unsampled points using **kriging, a spatial interpolation method**.<sup>6</sup> This reveals species with strong (Fig. 3A-C) vs weak (Fig. 3D-F) spatial structure, and varies across species. (Fig. 3I) We also find potential hotspots of realized load across many species within taxonomic classes. (Fig. 3G,H)

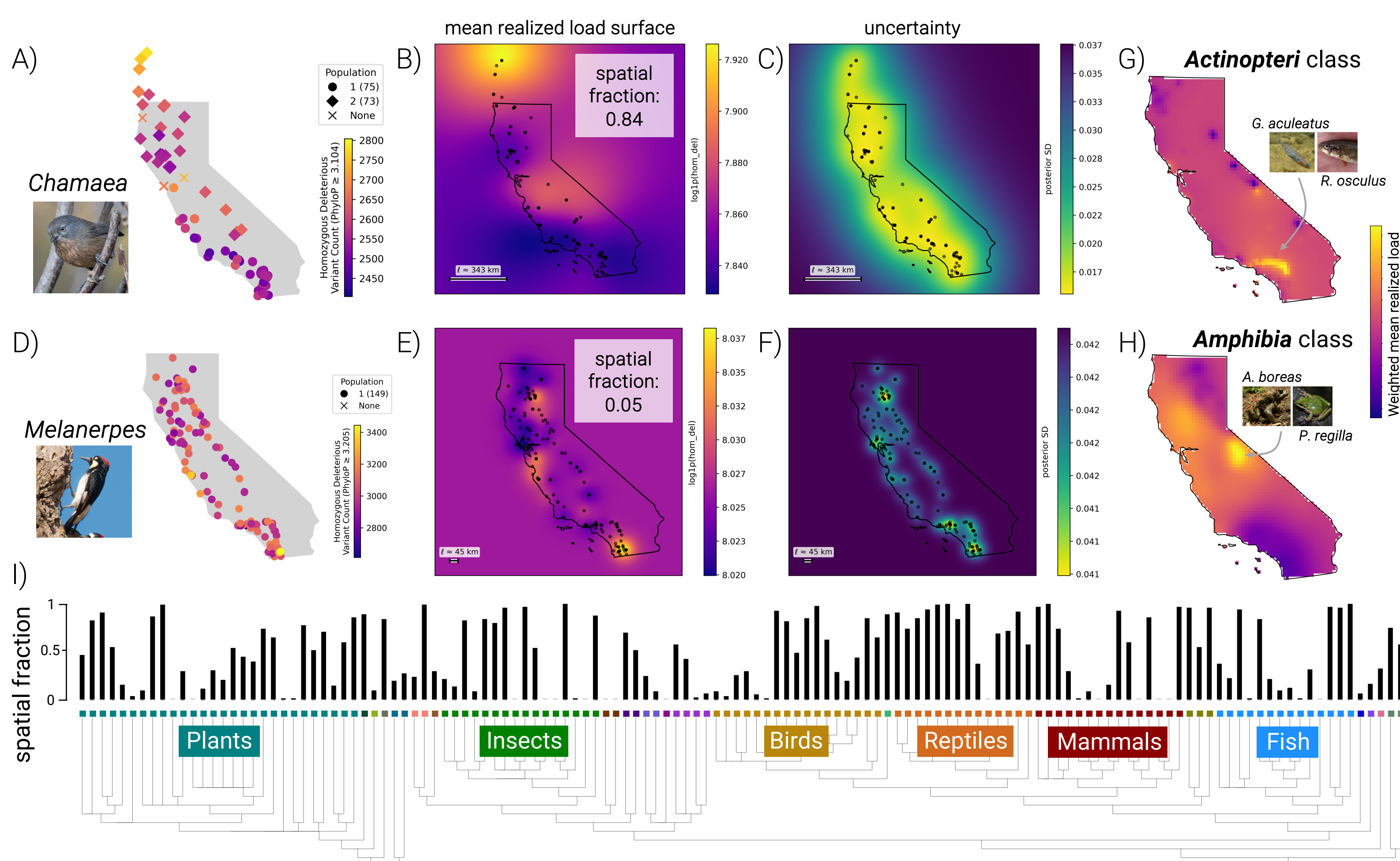


Figure 3: Spatial structure of realized load. A) Homozygous deleterious variant counts per individual and B) mean interpolated load surface with C) uncertainty for *Chamaea fasciata* (strong spatial structure) and D,E,F) *Melanerpes formicivorus* (weak spatial structure). Average of many surfaces weighted by uncertainty reveals load hotspots for G) Actinopteri and H) Amphibia classes. I) Spatial fraction (spatial variance / total) across all CCGP species.

### Inbreeding and segregating load are strong predictors of realized load

Across all CCGP species populations, we showed that **inbreeding** (Froh) acts on **segregating load** and redistributes it as **realized** (homozygous) or **masked** (heterozygous) load. Recent landscape processes, such as **human impact**, **range edge**, and **migration** influence realized load through their more direct effect on **inbreeding**. (Fig. 4)

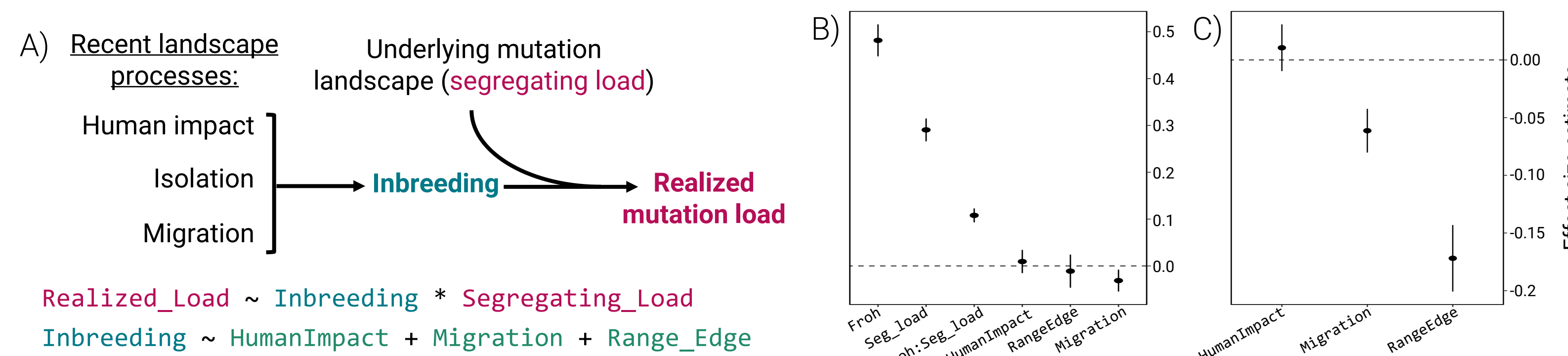


Figure 4: A) Models of realized load and inbreeding. Predictor effect sizes for B) realized load and C) inbreeding.

## 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.

## Acknowledgements

We thank the 70+ PIs who have contributed to the CCGP, which is funded by the State of California. 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](http://ccgp.org). Funding: UC Santa Cruz, NSF-GRFP, NIH (T32HG012344)

## Methods

### Calling putatively deleterious variation based on sequence conservation

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

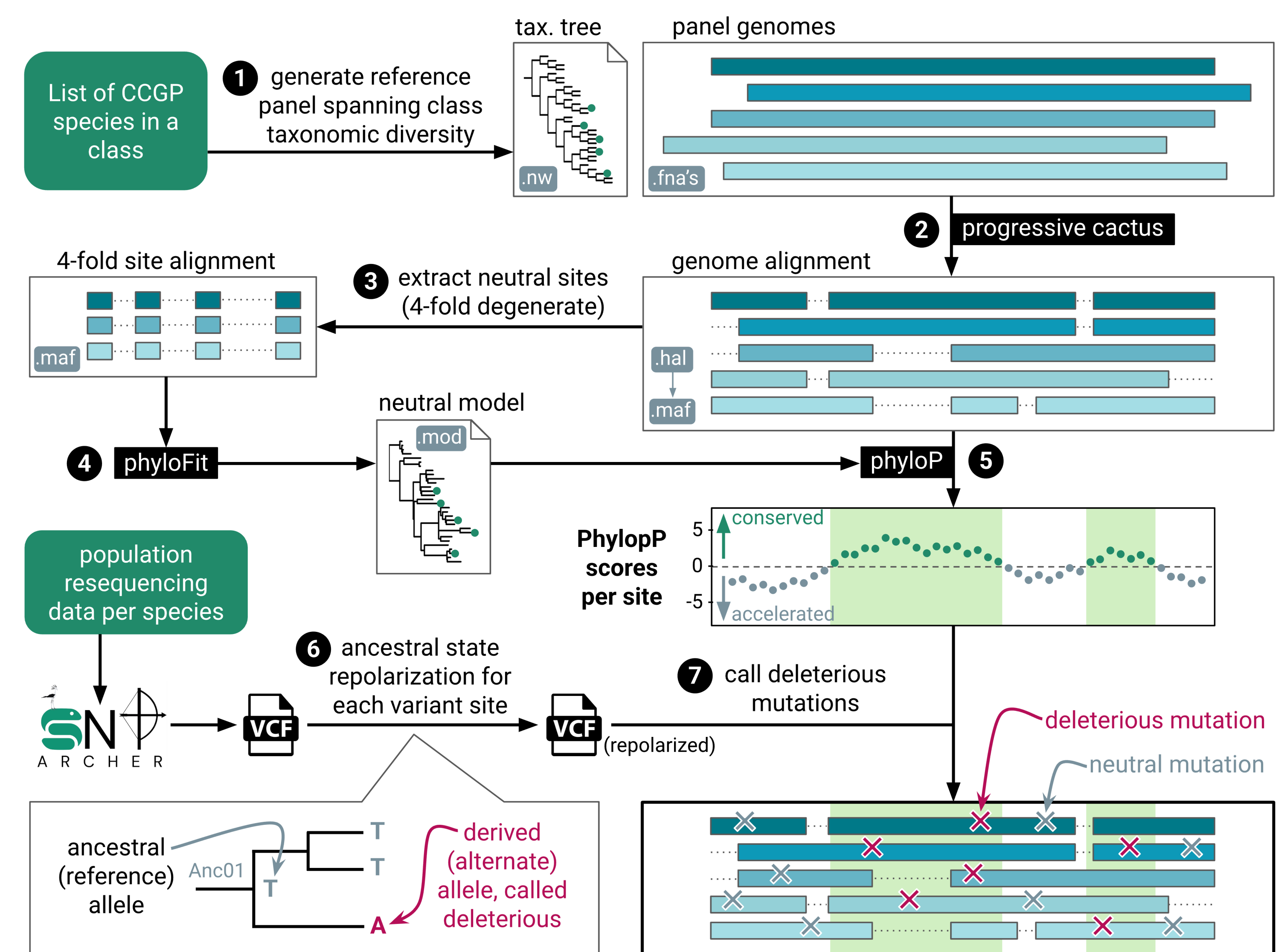


Figure 2: Computational workflow used to (1) generate a reference panel of genomes for each CCGP class, (2) perform whole-genome alignment with Progressive Cactus, (3-5) compute PhyloP sequence conservation scores based on a neutral model of evolution from 4-fold degenerate sites, (6) repolarize variants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

## Discussion & Ongoing Work

### Function of mutation load across taxa

How do the functions mutation load is associated with vary across species? We began by looking at **coding vs noncoding sites**, and find a large fraction of realized load in non-coding regions for many species. (Fig. 6) The next step is to characterize functions using gene ontology.

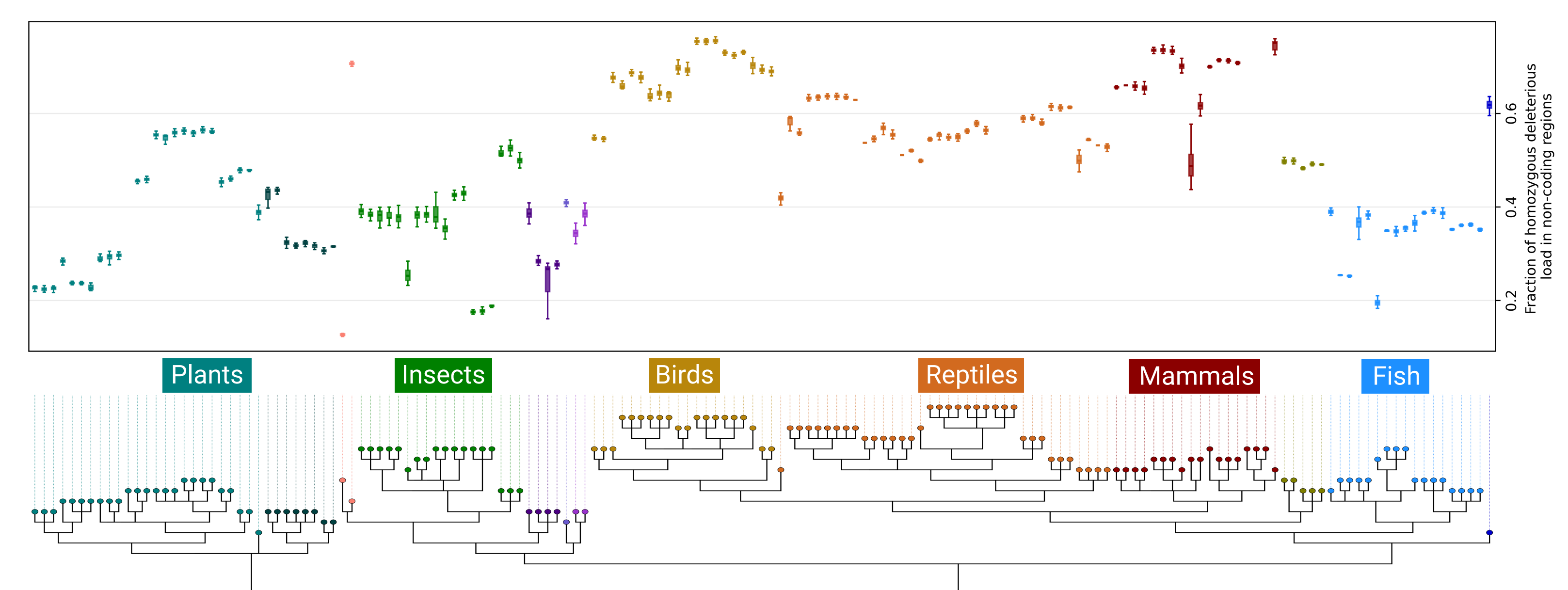


Figure 6: Proportion of realized load at non-coding genomic sites (as opposed to coding sites) across CCGP taxa. Each leaf in the taxonomic tree represents a species population, colored by taxonomic class.

### Informing on hotspots of recent genomic concern across California

Species range edges often occur along California's ecoregion boundaries, (Fig. 5A) which represent **transitional environments of potential conservation interest**. Our preliminary analyses show potential hotspots of realized load in many species at certain range edges. (Fig. 5B)

### Effects of population history

While recent landscape processes affect realized load, **deeper-time historical demography** ( $N_e$ ) likely impacts the underlying landscape of segregating load in modern populations. We will investigate this relationship further to understand how mutation load can be used to make predictions about the future health of these populations.

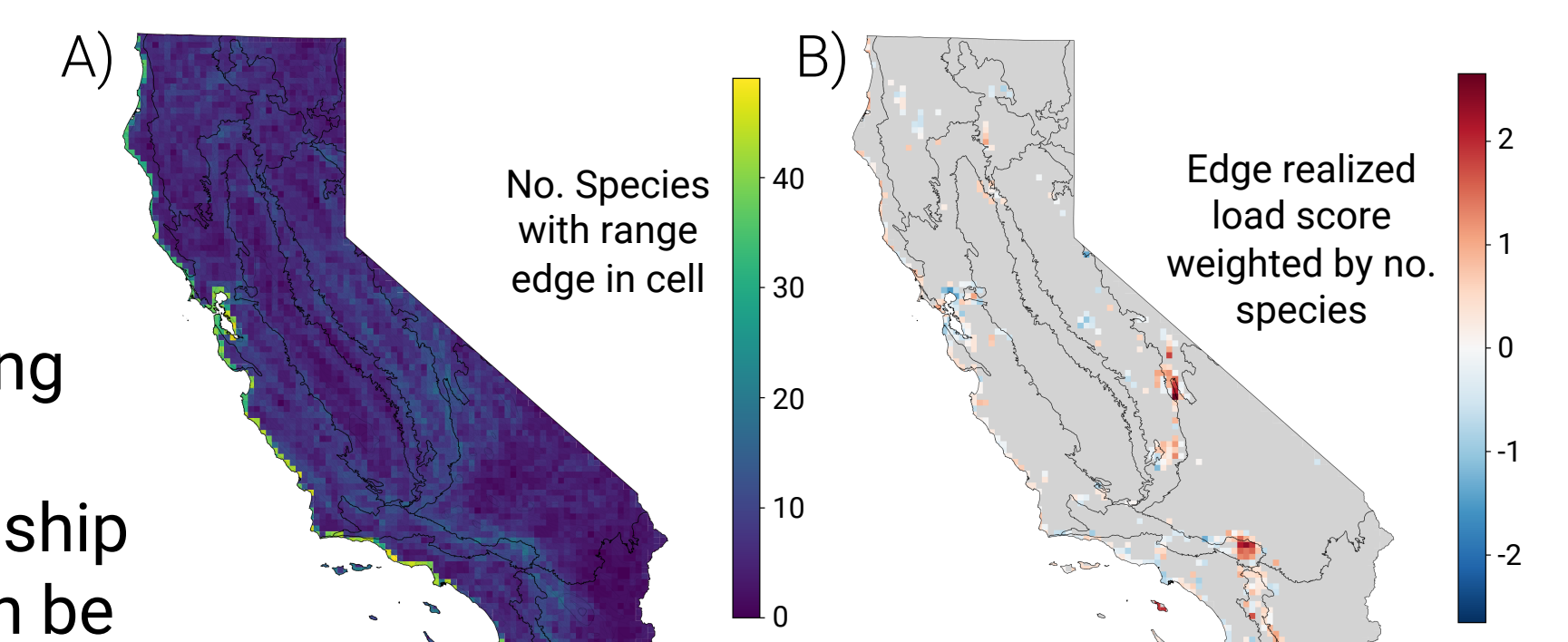


Figure 5: A) Species range edge richness. B) Effect size and direction of range edge load hot/cold spots.

See Erik Enbody's poster 160W for more on inbreeding!