

# High throughput conservation and landscape genomics

Erik D. Enbody, Cade Mirchandani, Russell Corbett-Detig  
Biomolecular Engineering, University of California, Santa Cruz

## Overview



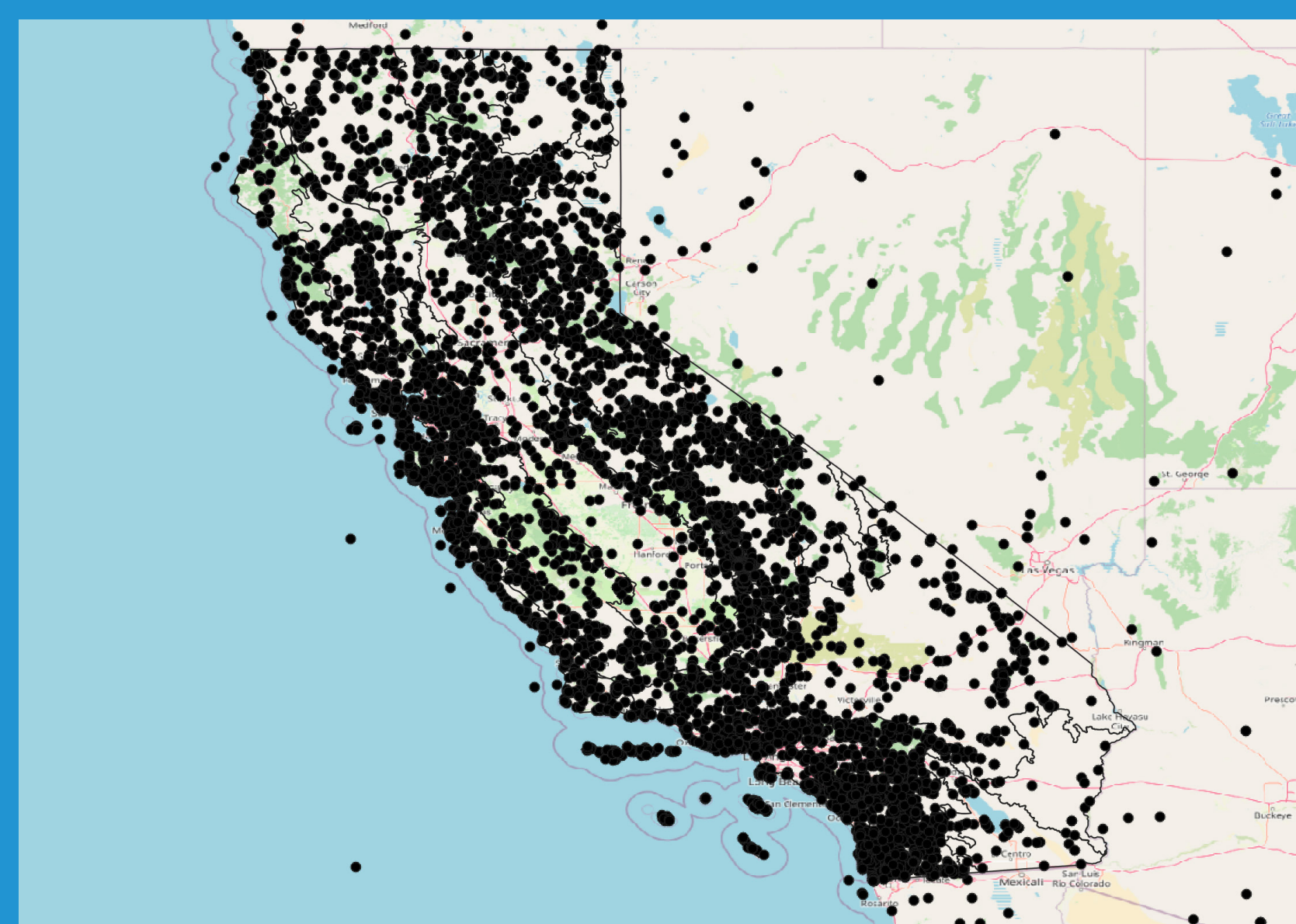
**CALIFORNIA  
CONSERVATION  
GENOMICS  
PROJECT**

### Goal:

To produce the most comprehensive multispecies genomic dataset ever assembled to help manage and protect regional biodiversity in the face of climate change.

### Dataset:

230 species of 150 samples each, ~20,000 whole-genome samples

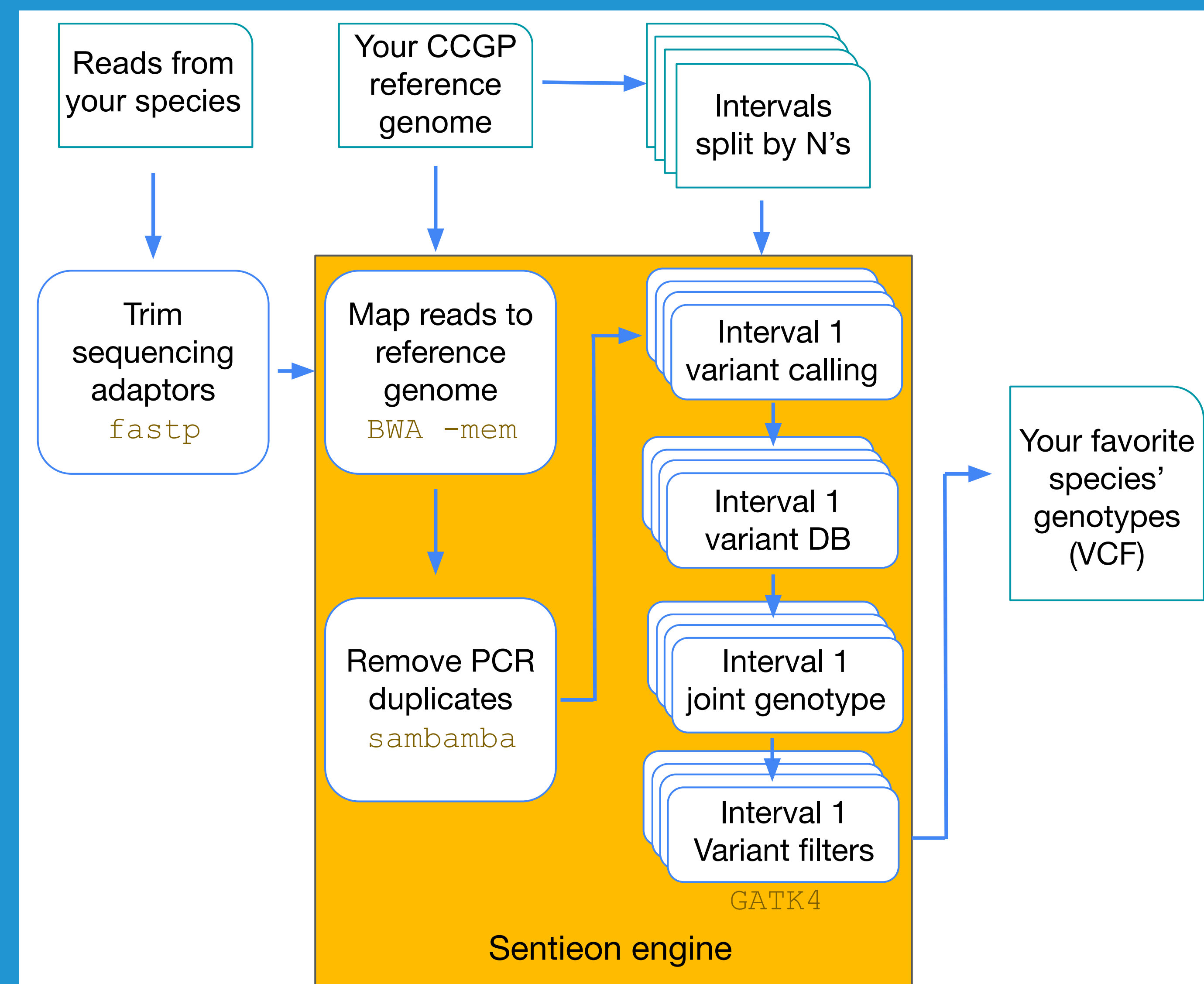


### Problem:

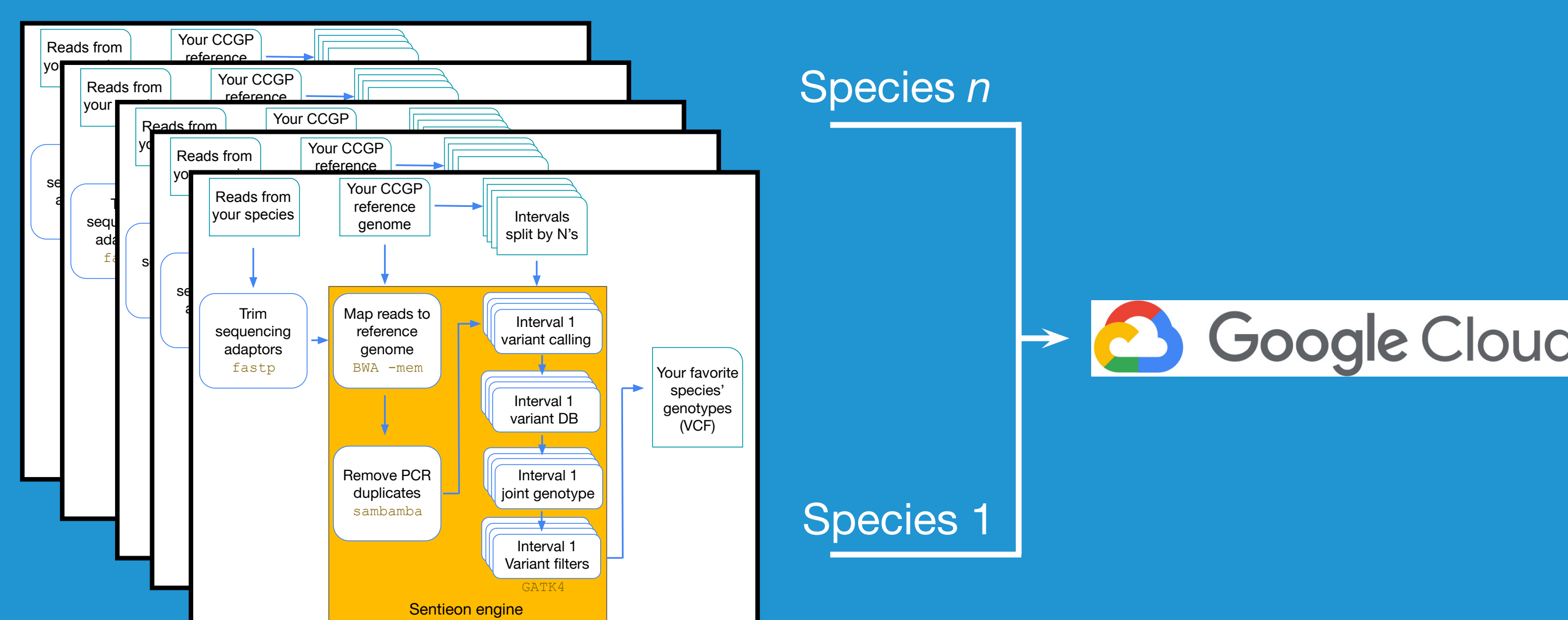
Generate reliable variants for 20,000 samples across non-model organisms:

- Reproducible and portable
- Fast and scalable
- Quality control

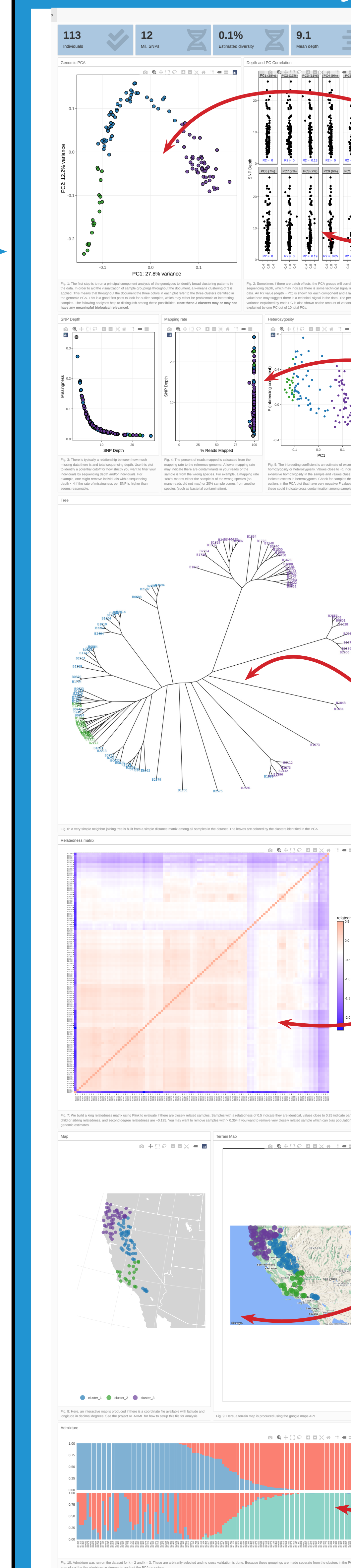
## Snakemake Variant Calling Pipeline Reproducible and Portable



## Cloudified Fast and Scalable (1-2 days)



## Variant Dashboard Quality Control



Black Bear QC on 100k SNPs:

Genomic PCA

PCs loading with depth

Missingness, depth, and heterozygosity

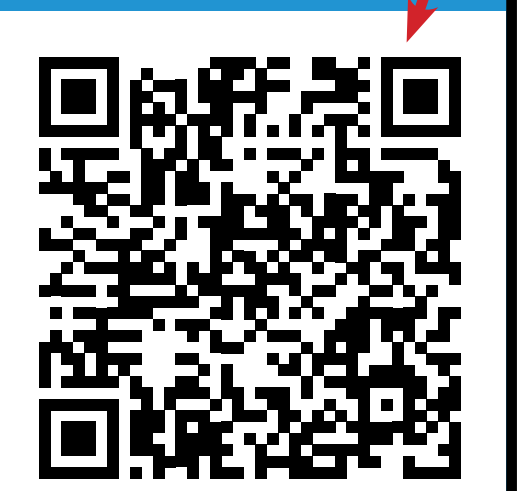
Simple neighbor-joining tree

Relatedness matrix

Maps

Interactive, try it!

Structure



A portable, scalable, and user-friendly pipeline for high-throughput genomic variant calling with interactive output for quality control



GitHub Repo:  
[https://github.com/cademirch/ccgp\\_workflow](https://github.com/cademirch/ccgp_workflow)

See also:  
<https://github.com/harvardinformatics/snpArcher>