Contrasting whole-genome and reduced representation sequencing for population demographic inference: an alpine mammal example

Daria Martchenko¹ and Aaron Shafer¹

¹Trent University

January 11, 2022

Abstract

Genomic approaches to the study of population demography rely on accurate SNP calling and by-proxy the site frequency spectrum (SFS). Two main questions for the design of such studies remain poorly investigated: do reduced genomic sequencing summary statistics reflect that of whole genome, and how do sequencing strategies and derived summary statistics impact demographic inferences? To address those questions, we applied the ddRAD sequencing approach to 254 individuals and whole genome resequencing approach to 35 mountain goat (*Oreamnos americanus*) individuals across the species range with a known demographic history. We identified SNPs with 5 different variant callers and used ANGSD to estimate the genotype likelihoods (GLs). We tested combinations of SNP filtering by linkage disequilibrium (LD), minor allele frequency (MAF) and the genomic region. We compared the resulting suite of summary statistics reflective of the SFS and quantified the relationship to demographic inferences by estimating the contemporary effective population size (N_e), isolation-by-distance and population structure, F_{ST} , and explicit modelling of the demographic history with $\delta a \delta i$. Filtering had a larger effect than sequencing strategy, with the former strongly influencing summary statistics. Estimates of contemporary N_e and isolation-by-distance patterns were largely robust to the choice of sequencing, pipeline, and filtering. Despite the high variance in summary statistics, whole genome and reduced representation approaches were overall similar in supporting a glacial induced vicariance and low N_e in mountain goats. We discuss why whole genome resequencing data is preferable, and reiterate support the use of GLs, in part because it limits user-determined filters.

Hosted file

Contrasting_whole-genome_and_reduced_representation_sequencing_pipelines.docx available at <https://authorea.com/users/454966/articles/552422-contrasting-whole-genome-and-reduced-representation-sequencing-for-population-demographic-inference-an-alpine-mammal-example>