A beginner's guide to low-coverage whole genome sequencing for population genomics

Runyang Nicolas Lou[1], Arne Jacobs[1], Aryn Wilder[2], and Nina Overgaard Therkildsen[1]

[1]Cornell University
[2]San Diego Zoo Institute for Conservation Research

April 21, 2021

Abstract

Low-coverage whole genome sequencing (lcWGS) has emerged as a powerful and cost-effective approach for population genomic studies in both model and non-model species. However, with read depths too low to confidently call individual genotypes, lcWGS requires specialized analysis tools that explicitly account for genotype uncertainty. A growing number of such tools have become available, but it can be difficult to get an overview of what types of analyses can be performed reliably with lcWGS data, and how the distribution of sequencing effort between the number of samples analyzed and per-sample sequencing depths affects inference accuracy. In this introductory guide to lcWGS, we first illustrate how the per-sample cost for lcWGS is now comparable to RAD-seq and Pool-seq in many systems. We then provide an overview of software packages that explicitly account for genotype uncertainty in different types of population genomic inference. Next, we use both simulated and empirical data to assess the accuracy of allele frequency and genetic diversity estimation, detection of population structure, and selection scans under different sequencing strategies. Our results show that spreading a given amount of sequencing effort across more samples with lower depth per sample consistently improves the accuracy of most types of inference, with a few notable exceptions. Finally, we assess the potential for using imputation to bolster inference from lcWGS data in non-model species, and discuss current limitations and future perspectives for lcWGS-based population genomics research. With this overview, we hope to make lcWGS more approachable and stimulate its broader adoption.

Hosted file

main_text_combined_revision_no_field_code.pdf available at https://authorea.com/users/380682/articles/496574-a-beginner-s-guide-to-low-coverage-whole-genome-sequencing-for-population-genomics

Hosted file

supplementary_materials_revision_no_field_code.pdf available at https://authorea.com/users/380682/articles/496574-a-beginner-s-guide-to-low-coverage-whole-genome-sequencing-for-population-genomics