April 24th, 2024

Dear Editors for the *Journal of Heredity*,

Please find attached our manuscript titled “Unbiased estimation of the number of segregating sites across unequal sample sizes” for publication as an Original Article in the *Journal of Heredity*.This manuscript provides an approach for standardizing estimates of the number of segregating sites across sample groups or populations. This method will help a broad audience of population geneticists using next-generation sequencing data to answer fundamental ecological and evolutionary questions.

The number of segregating sites is a critical but underutilized measure which conveys an informative aspect of genetic diversity that is different than other commonly used statistics such as observed or expected heterozygosity. Comparing segregating site counts to other statistics can be of great utility as demographic and other population processes do not always impact these estimators identically. Because the probability of identifying rare alleles is higher in larger samples, comparing two sets of samples will introduce bias whenever sample sizes or proportions of missing data differ. The method we present here corrects for these issues, allowing for direct comparisons among populations or sample groups. Our estimator can provide informative estimates of genetic diversity across multiple scales, from genes to chromosomes, thus elucidating how evolutionary forces have shaped portions of a genome in different ways.

We recommend the following internationally recognized experts as potential reviewers (full contact details submitted online):

• Marty Kardos, Northwest Fisheries Science Center, United States, [martin.kardos@noaa.gov](mailto:martin.kardos@noaa.gov)

• Oscar Gaggiotti, University of St. Andrews, United Kingdom, [oeg@st-andrews.ac.uk](mailto:oeg@st-andrews.ac.uk)

• Eric Anderson, Southwest Fisheries Science Center, United States, [eric.anderson@noaa.gov](mailto:eric.anderson@noaa.gov)

• Aaron Shafer, Trent University, Canada, [aaronshafer@trentu.ca](mailto:aaronshafer@trentu.ca)

The attached manuscript is within the specified formatting guidelines for an Original Article and consists of 2858 words and four figures. In the main text, we have also included the link to the software that calculates our method, which can be installed with a single command in R.

Thank you for considering our work.

Sincerely,

A signature on a white background

Description automatically generated

William Hemstrom, PhD

Department of Biological Sciences

Purdue University

West Lafayette, IN 47906

[whemstro@purdue.edu](../../../C:/Users/hemst/Documents/GitHub/filtering_simulation_paper/manuscript/second_submission/whemstro@purdue.edu)