



UNITED STATES DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center
Fisheries Ecology Division
110 McAllister Way
Santa Cruz, CA 95060

July 15, 2024

Editorial Board
Molecular Ecology Resources

To the Editor:

I am delighted to submit the accompanying manuscript, “GSCRAMBLE: Simulation of admixed individuals without reuse of genetic material,” to *Molecular Ecology Resources*. While it has been understood since the exposition in Anderson, Waples, & Kalinowski (2008; *Can. J. Fish. Aquat. Sci.*) that care must be taken when simulating data sets to assess the power of a marker panel for population assignment and/or genetic stock identification, it has been less widely appreciated that the same holds true for assessing how much power is available for estimating admixture fractions and identifying hybrids between closely related groups. With the advent of data sets with thousands, or even hundreds of thousands, of genetic markers, this problem becomes severe. At the same time, with so many markers, it is critical to account for physical linkage when assessing how accurately different hybrid categories can be resolved.

In our paper, we start with several simple simulations to show how dramatic the bias of power estimation can be with modern data sets, a phenomenon that we refer to as resampling-induced spurious power inflation, or RISPI. We then describe a novel, pedigree-based approach to sampling alleles to create hybrid and admixed individuals to properly estimate the power for admixture analysis and hybrid identification that avoids RISPI, and takes full account of physical linkage. We are confident that this approach provides state-of-the-art methodology for rigorously comparing the outcome of genetic clustering approaches, such as STRUCTURE, and ADMIXTURE, to results on simulated data sets that are commensurate with the original data.

In the remainder of the paper, we provide two examples of how the approach can be used to assess the power for hybrid identification and admixture proportion estimation in case studies involving Pacific trouts and feral pigs.

The approach we have developed is implemented in the R package ‘gscramble’ which is

currently available for download from the Comprehensive R Archive Network (CRAN), and includes full documentation and vignettes. Members of our team have used 'gscramble' already in one published project and the package is currently in use in several more. We predict that the package will be widely used in the future, and as a consequence, the paper will be highly cited.

Thank you for the opportunity to submit our work to *Molecular Ecology Resources*.

Sincerely,

A handwritten signature in black ink, appearing to read "Eric C. Anderson". The signature is fluid and cursive, with the first name "Eric" being more prominent.

Eric C. Anderson

eric.anderson@noaa.gov

ph: 831-227-4239