Dear Editor
GENETICS

We would like to submit our manuscript

"CLEAR: Composition of Likelihoods for Evolve And Resequence Experiments"

for possible publication in GENETICS. Our manuscript develops a new method identifying selection in experimental evolution studies, which provide an important platform for understanding evolution in real-time. Specifically, by evolving-and-resequencing a population of an organism, such studies can help explain how organisms genetically adapt to selection pressures such as pesticide, drug, antibiotic resistance. Over past decade, there has been a substantial effort has focused on analyzing longitudinal studies of sexually reproducing organisms with short generation-time, esp. *D. melanogaster* . However, there are many unresolved issues regarding computational/statistical methods:

- Only a few tools can make full use of time-series points. Most just handle two time points.
- As individuals of population are pooled and sequenced together, coverage of a variant changes among different replicates and generations, which can have a critical effect on the power of different methods.
- A majority of existing tests, fail when the initial frequency of the favored allele is small, which is very likely in real word scenarios.
- Except for multi-locus Gaussian process which is computationally intractable, all the existing tools provide tests for single locus.

In this manuscript, we address these issues and develop a novel tool, CLEAR (available as open source) that can analyze time-series data, displays better power in detection of selection for a wide range of experimental evolution and selection parameters. We carefully examine the role of different experimental evolution parameters, including initial frequency of the favored allele, sampling time relative to the onset of selection, sequencing coverage, and the span of sampling on the power of selection.

We also found that existing tools for handling single time point data (often based on analysis of the site-frequency spectrum) and those for handling time-series data (based on a modeling of the favored allele frequency over time) are based on completely different principles, and no serious effort has been made to reconcile the two approaches.

In our experiments, we show that the problem of detecting a genomic region under selection is distinct from identifying the favored site that is functionally responding to selection constraint, and evolve-and-resequence strategies that work for detection may not be optimal for site-selection. In this sense, our analysis will guide future design of experimental evolution experiments. For all of these reasons, we hope that our manuscript will be of interest to GENETICS readers. We thank you in advance for your consideration.

In terms of potential editors/reviewers, our work is of course most closely related to GENETICS's Theoretical Population Genetics section. In particular, Dr. Yun Song and Dr Rasmus Nielsen has previously published on the same topic.

Sincerely,
Arya Iranmehr
Vineet Bafna