Dear Editor
PLoS Genetics

We would like to submit our manuscript

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

for possible publication in *PLoS Genetics*. There are many reasons why we think the manuscript is appropriate for the journal. Experimental Evolution of organisms is an important technique for understanding evolution in action, and in particular, how they genetically adapt to selection pressures, e.g. pesticide, drug, antibiotic resistance, etc. With new sequencing technologies, we have the opportunity for deep sampling of the evolving population at different time points.

Increasingly, scientists are looking at sexually reproducing organisms, esp. *Drosophila,* and acquiring time series data. However, there are many questions about the computational/statistical tools for analyzing such data in order to detect regions evolving under selection.

1. Only a few tools can make full use of time-series points. Most just handle two time points. A recently published tool, Gaussian Process (GP), published in *PLoS Genetics* 2015, can analyze time series through full likelihood calculations, but its assumptions do not hold in real world scenarios. Moreover, it is computationally intensive and cannot scale to handling genome scale data. In our paper we deployed an improved model to improve the power.
2. As individuals of population are pooled and sequenced together, sequencing coverage can have a critical effect on the power of different methods. We evaluate different methods under various sequencing coverages and show that CLEAR is robust to change of coverage.
3. A majority of existing techniques, including GP, fail when the initial frequency of the favored allele is small, which is very likely in real word scenarios. We show that CLEAR is not sensitive to this parameter and performs well over many different starting frequencies..
4. We also found that 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 efforts.

In this manuscript, we address these questions and develop an open source tool that can analyze time-series data, displays better power in detection of selection for a wide range of experimental evolution and selection parameters, and is considerably faster than GP.

It also resolves questions regarding the performance of site-frequency based methods by extending them to handle time-series data. We also carefully examine the role 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 show that different experimental evolution regimes might be optimal for the problems of detecting region under selection while some of them it is quite difficult to identifying the favored site itself mainly due to intrinsic structure of the population. In this sense, our calculations will also guide future design of experimental evolution experiments. For all of these reasons, we hope that our manuscript will be of interest to PLoS Genetics readers. We thank you in advance for your consideration.

In terms of potential reviewers, our work is of course most closely related to Dr. Yun Song's work which was published in PLoS Genetics last year. We did discuss an early version of the manuscript with him, and even with competition, it would be OK to ask him to be a referee. We share one of the co-authors from that paper. Other potential reviewers include Dr. Dmitrii Petrov who is doing a lot of work on experimental evolution, and some of the authors of the competing tools:
- Hande Topa
- Robert Kofler
- Sergey Kryazhimskiy
- Alison Feder

Sincerely,

Arya Iranmehr
Vineet Bafna