Rank epistasis: A new model for analyzing epistatic interactions in the absence of quantifiable fitness interactions

Acacia L. Ackles[1,†], Austin J. Ferguson[2,†], Connor Grady[3,†], and Charles Ofria[2]

[1]Integrative Biology / Ecology, Evolutionary Biology, and Behavior
[2]Computer Science and Engineering / Ecology, Evolutionary Biology, and Behavior
[3]Biomedical Engineering / Institute for Quantitative Health Science and Engineering
BEACON Center for the Study of Evolution in Action, Michigan State University
[†] These authors contributed equally to this work.
alackles@msu.edu

Introduction

Epistasis refers to multiple genes interacting to create a single phenotype, and it is a key component of complex genetic architectures. If loci are epistatic, the fitness effects of individual mutations at each loci may not fully determine their combined fitness effect. Determining how strong epistatic interactions are at individual sites or for individual genes is therefore an important problem in evolutionary biology if we are to understand fitness landscapes. However, challenges arise when measuring epistatic load both because absolute fitness is difficult to measure in biological organisms, and because we do not have a perfect baseline expectation against which to compare epistatic interactions.

Fitness in biological systems is measured in relation to the other organisms in the population (Elena and Lenski, 1997; Trindade et al., 2009); the organisms with highest fitness are those which produce the most offspring in comparison to others in the population. However, these comparative fitnesses are inexact and can be confounded by organismal interactions and environmental changes.

Furthermore, even when exact fitness values are known, it can be challenging to predict how the fitness effects of two non-interacting mutations should be combined to provide a baseline for epistasis. For example, we might model fitness contributions from non-epistatic sites as either additive or multiplicative; both models are commonly used as baselines, but provide very different results (Puniyani et al., 2004).

In order to address these challenges, we propose a rank-based epistasis metric that avoids the requirement of absolute fitness measurements of organisms, focusing instead on how perturbations to one site affect the ordering of fitness contributions from other sites.

Methods

Rank Epistasis Metric

Evaluating a target locus with rank epistasis consists of:

1. *Measure and rank the fitness of organisms in a wild type population.* The means by which fitness is determined depends on the target system.
2. *Apply a mutation to the target locus and repeat step 1.* We will now have a new ranking of the same population, but in combination with the target mutation.
3. *Calculate the edit distance between the two rankings.* Using the organism rankings from steps *1* and *2* to create ranked lists of individuals, calculate the minimum number of adjustments that need to be made to put the lists in the same order (edit distance).
4. *Repeat step 2 and 3 for all distinct mutations at the target locus.* We can then average these results to measure how meaningfully the target locus interacts with other loci in the genome.

This procedure allows us to calculate the epistasis at a single locus, without needing to first identify baseline interactions among loci. This per-locus epistasis can be aggregated to measure epistasis across a whole organism (as is common in other epistatic measurements (Elena and Lenski, 1997; Franklin et al., 2019)), or for a single locus across a population at a point in time.

Fitness Landscapes

To evaluate the efficacy of this metric, we used three distinct classes of fitness landscapes based on the NK-landscape model (Kauffman and Levin, 1987). Here, N is the number of sites, and K is the number of interaction each bit has with other genes. K is thus one way to quantify epistasis.

We chose this model because the degree of interaction between sites in the genome can be adjusted, allowing for an easy comparison across different levels of epistatic interaction. We use a classic NK landscape, an oscillatory NK landscape ("NK treadmill"), and an NK landscape with narrow, high peaks and broad, flat peaks ("fit and flat") as in Wilke et al. (2001).

Comparison Metrics

We compare our metric to existing measures of epistasis discussed in Elena and Lenski (1997). A strong correlation between our rank-based metric and their interaction-based metric, β, would provide evidence that we are directly measuring epistatic activity and not a related signal.

Evolutionary System and Analysis Pipeline

We conducted experiments in this paper using the Modular Agent Based Evolver framework (MABE) (Bohm et al., 2017). All analyses were done using the R statistical computing language (R Core Team, 2017). Plots were created using the ggplot2 R package (Wickham, 2009). Source code, data analyses, and additional figures are availble as supplemental material at `https://github.com/alackles/ALIFE-2020-Rank-Epistasis`.

Results

Convergence muddles rank-based metric The fit and flat fixed NK landscape provides little challenge to evolving organisms. The extreme case sees a population converging to a single genotype, and indeed this is often observed with the lowest mutation rate. In this case, mutating a single locus may cause a shift in fitness, but that shift is seen in all organisms and thus the ranking does not change. This yields an edit distance of zero. Future work must be conducted to disentangle these effects.

Edit distance varies as expected with K Unlike fixed NK landscapes, the classic, randomly-generated landscapes provide considerable levels of noise and are more of a challenge for the evolving populations. These attributes make examining the rank-based metric more feasible in this domain, as populations are less likely to converge en masse.

Edit distance is stratified according the value of K, as expected (Fig 1). This result provides early evidence that the metric measures some facet of epistasis.

Rapidly changing environments increase epistasis measurement The NK treadmill landscape requires a velocity parameter, which determines how fast the landscape changes. While an increase in velocity results in an increase in rank epistasis across a population, we hypothesize this is *not* a result of increased epistasis, but rather a decrease in convergence. As the landscape quickly shifts, the population must continually adapt. This continuous adaptation thus maintains diversity.

Rank epistasis correlates with traditional measures when not converging We also compared the rank epistasis metric to the calculation described in (Elena and Lenski, 1997). At higher mutation rates we observe a positive correlation between the rank-based metric and the more traditional epistasis measurement. These correlations are weakly positive for at low mutation rates and moderately to strongly positive for high mutations rate.

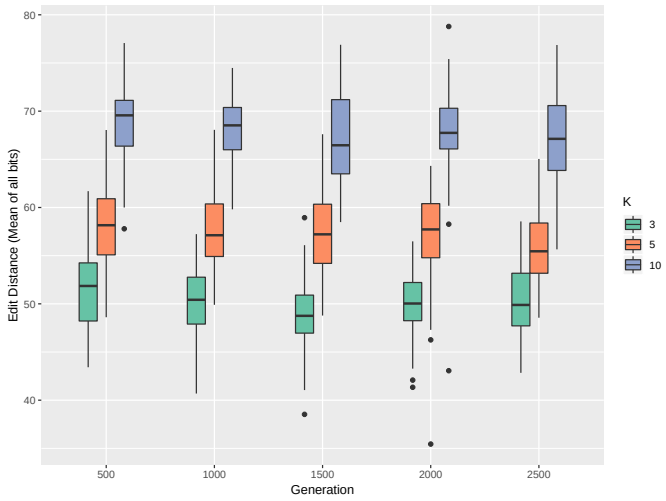

Figure 1: Boxplots showing the edit distance of the rank-based metric at several points in the evolutionary run and at three levels of K (3, 5, 10). Each boxplot represents the 50 replicates of that treatment at each time point.

Conclusion

Here we present preliminary results for a binary genome model. We found this model to closely track existing models of epistatic interactions, including K on an NK landscape and the parameter β from Elena and Lenski (1997). This shows promising results for potential future developments of this rank-based epistasis metric.

A model that allows for more single-step mutations per locus is under development. Such a model would be applicable to many questions in biological systems, especially as generating landscapes of single-step mutations is coming within reach of wet-lab systems. Applying our rank epistasis metric to rapidly adapting biological populations can help us understand which sites are biologically important, and how these sites interconnect with the rest of the genome.

Acknowledgements

The authors thank Christoph Adami, Wolfgang Banzhaf, and Clifford Bohm for assistance in designing this study. This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Michigan State University provided computational resources through the Institute for Cyber-Enabled Research. Michigan State University resides on Anishinaabeg land ceded in the 1819 Treaty of Saginaw.

References

Bohm, C., C G, N., and Hintze, A. (2017). MABE (Modular Agent Based Evolver): A framework for digital evolution research. In *Artificial Life Conference Proceedings 2017*, pages 76–83. MIT Press, Cambridge, MA.

Elena, S. and Lenski, R. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature*, 390:395–398.

Franklin, J., Labar, T., and Adami, C. (2019). Mapping the peaks: Fitness landscapes of the fittest and the flattest. *Artificial Life*, 25:250–256.

Kauffman, S. and Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128:11–45.

Puniyani, A., Liberman, U., and Feldman, M. W. (2004). On the meaning of non-epistatic selection. *Theoretical Population Biology*, 66:317–321.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Trindade, S., Sousa, A., Xavier, K., Dionisio, F., Ferreira, M., and Gordo, I. (2009). Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genetics*, 5:e1000578.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilke, C., Wang, J., Ofria, C., Lenski, R., and Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412:331–333.