Detecting the genetic signature of mutualisms from lineages

Nadine Strasser[1], Alexander Lalejini[2] and Charles Ofria[2]

[1]University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria
[2]Michigan State University, East Lansing, MI 48824
nadine.strasser@gmail.com

Abstract

We study genetic signatures of tightly co-evolving populations. Specifically, we want to measure the effect of co-evolution on mutational changes along lineages under ideal conditions to determine if lineage-based genetic data can be used to identify such relationships. To reach this goal, we created a population of higher-level organisms that each consist of two different types of lower-level cells. These cells are tightly linked and jointly determine the fitness of the overall organisms. When the fitness contribution of one cell type is dependent on the state of the other cell type, we expect that one will always change in response to the other. We built a simple computational model that depicts various idealized scenarios in which we expect evolving organisms to exhibit co-evolution. We track the mutational changes along lineages and analyze the patterns with an accumulated mutations metric.

We find that a lockstep pattern can successfully be measured with our metric. In the next phase, we will explore a broader range of symbiotic behavior by introducing migration and disentangling the two populations of lower-level individuals. Cells that have the possibility of reproducing individually can display either antagonistic interactions (e.g., predators and prey or parasite and hosts), or mututalistic interactions, as occur in cooperating groups or after major evolutionary transitions.

Introduction

Major evolutionary transitions in individuality redefine what it means to be an individual. Such transitions occur when formerly distinct individuals unite to form a more complex lifeform capable of reproducing as a single, higher-level entity.

Transitions in individuality generally fall into two categories: fraternal and egalitarian transitions (Queller, 1997). Fraternal transitions occur when genetically identical lower-level individuals stay together to form higher-level organisms that subsequently reproduce as one (*e.g.*, the evolution of multicellularity or eusocial insect colonies). Egalitarian transitions occur when different types of lower-level individuals come together as a higher-level organism to fulfill a united goal (Smith and Szathmary, 1995; Szathmary, 2015). For example the origin of the eukaryotic cell, where a prokaryotic cell ingested another type of prokaryotic cell and the latter became a component of the first.

Mutualistic relationships are present in many types of organisms, such as lichens, which are composite organisms consisting of algae or cyanobacteria and fungi species; the Buchnera–aphid symbiosis; and most plant-pollinator pairs. (Queller and Strassmann, 2009; Szathmary, 2015)

Even microbiomes might be considered as egalitarian transitions: Plants and animals live in close association with microbial organisms as a synergistic unit (Guerrero et al., 2013; Lopez-Garcia et al., 2017). It can be hard to identify tightly coupled relationships, but we know that *e.g.*, the human gut microbiome is important for the health of its host since it helps with the degradation of non-digestible polysaccharides (Thursby and Juge, 2017).

The flexibility of artificial life systems make them an ideal testbed for exploring useful metrics that help detecting the traits of tight evolutionary couplings, which will extend the understanding on how egalitarian transitions and mutualistic behavior arose through evolution.

Can the genetic signatures of an egalitarian transition in individuality be detected in the evolutionary histories of constituent species without taking into account the specific genetic architectures of the participants? As transitioning species become increasingly coupled, we might expect genetic changes in one lineage to drive genetic changes in the other. As a result, we might also expect that the amount of mutation accumulation in one lineage at a given point in time will correspond with the amount of mutation accumulation in a tightly coupled lineage.

We propose a simple metric that looks at mutation accumulation in tightly coupled lineages. We are using this metric to detect genetic signatures of cooperation between two lower-level individuals that together form a higher-level organism. We assume that a genetic signature is present when genomes are being modified in some sort of systematic way. Concretely, we are looking at the variance of beneficial mutation accumulation along two coupled lineages from lower-level individuals.

At the moment, it can be challenging to apply our metric

to biological data since it is hard to collect sequence data for whole lineages at a useful resolution. But as sequencing technology improves, it will get easier to produce detailed mutation accumulation data for natural systems. In the meantime, this metric gives valuable insights into how lineage data from biological systems could be processed in the near future, and provides an initial foray into a broad range of possible analyses.

Methods

In order to study the evolution of cooperation and specialization, we use evolutionary algorithms (Bäck, 1996). Evolutionary algorithms imitate natural evolution with the objective of generating efficient solutions for computational problems. To implement and evaluate the populations of evolving digital organisms, we use MABE (Modular Agent Based Evolution platform) (Bohm et al., 2017). For data analyses, Python (Van Rossum and Drake, 2009) and R (Team, 2017) are used.

Metric

We developed a metric that looks for simultaneous (or close to simultaneous) genetic changes in tightly coupled lineages. Therefore, the metric tracks mutation accumulation over time (based on (Dolson et al., 2020)). The beneficial mutations are counted over time in both types of the lower-level individuals that form one higher-level organism since we expect the beneficial mutations to occur closely aligned in the lockstep patterns. By measuring the variance of those differentials, we expect to identify if a genetic signature is present.

Experimental Design

We incorporated individuals into a simulated environment and observed how evolution shapes the individual's phenotype over a long period of time. For the formation of an egalitarian organism, we used two populations of individuals such that the organism represents a tight link of one individual from the first population with one individual from the second population. These higher-level organisms each replicate as an individual, copying both lower-level individuals with it. We used tournament selection to manage the evolutionary process, with a tournament size of seven and per-site mutation rates for the lower-level individuals. Selection is always done on the higher-level organisms, whereas the mutations, which are simple bit flips, are performed on the lower-level individuals.

To demonstrate that our algorithm is robust to population size and mutation rate, we used different combinations of per-site mutation rates (0.03, 0.01, 0.003, 0.001) and populations sizes (10000, 1000, 100, 10) for our experiments. A population size of 1000 and a mutation rate of 0.01 served as our pivot. We ran 5000 generations per replicate.

Organism and Fitness Evaluation

Each lower-level individual consists of a bit string, and we utilized a counting-one problem for the fitness evaluation. The number of leading ones (*i.e.*, the number of ones that are in the genome before the first zero occurs) in each of those bit strings determines the fitness contribution of that lower-level individual. Initially, the bit strings started off with all zeroes and a fixed genome size of 100. Every leading one counted as a fitness benefit of 1. Ones are only desirable at the beginning of the genomes but not after the first zero occurred. Specifically, we penalized any ones after the first zero in the genome with a fitness deduction of 0.005, which is the fitness benefit that a leading one generates (1) divided by two times the genome length ($2 \cdot 100$). We chose to multiply the genome length by two since each higher-level organism consists of two lower-level individuals and we did not want a leading one to ever be outperformed by its tail.

Genetic Signature and Scenarios

In egalitarian populations three different patterns of genetic signature are possible and we want to look at each of them: (1) Simultaneous genetic changes between species. Thus, both sides are mutating in a lock-step coordinated fashion. (2) Alternating mutations, where changes in one partner trigger changes in the other partner. (3) A control where no mutational patterns should be observed at all. In such a case, each lower-level individual would be evolving in a similar genetic pattern as it would have evolved in non-cooperative circumstances.

In order to see different genetic signatures, we implemented five different idealized scenarios that foreordain how the fitness scores of the two lower-level individuals must interact to be a successful higher-level organism. Those five scenarios are: (1) Random Drift on both lower-level individuals; (2) Independent Evolution, where the fitness contributions are simply added; (3) Lockstep, where both lower-level individuals must have the same number of leading ones in order to have their fitness contributions added. If the number of leading ones differs, their combined fitness is negative; (4) One-Off Lockstep, which is like Lockstep, but the number of leading ones can be equal or differ by at most one for the fitness contributions to be added together; (5) One-Follows, where one lower-level individual is allowed to be behind the second lower-level individual by as many leading ones as it wants, but it can never be ahead. If it gets ahead, the the higher-level organism gets a negative fitness score assigned.

Results

Our metric successfully detects simultaneous genetic changes between species. Specifically, we can see a genetic signature in the Lockstep and the One-Off Lockstep scenarios. In both of these cases, we are effectively able to see those lockstep-like patterns.

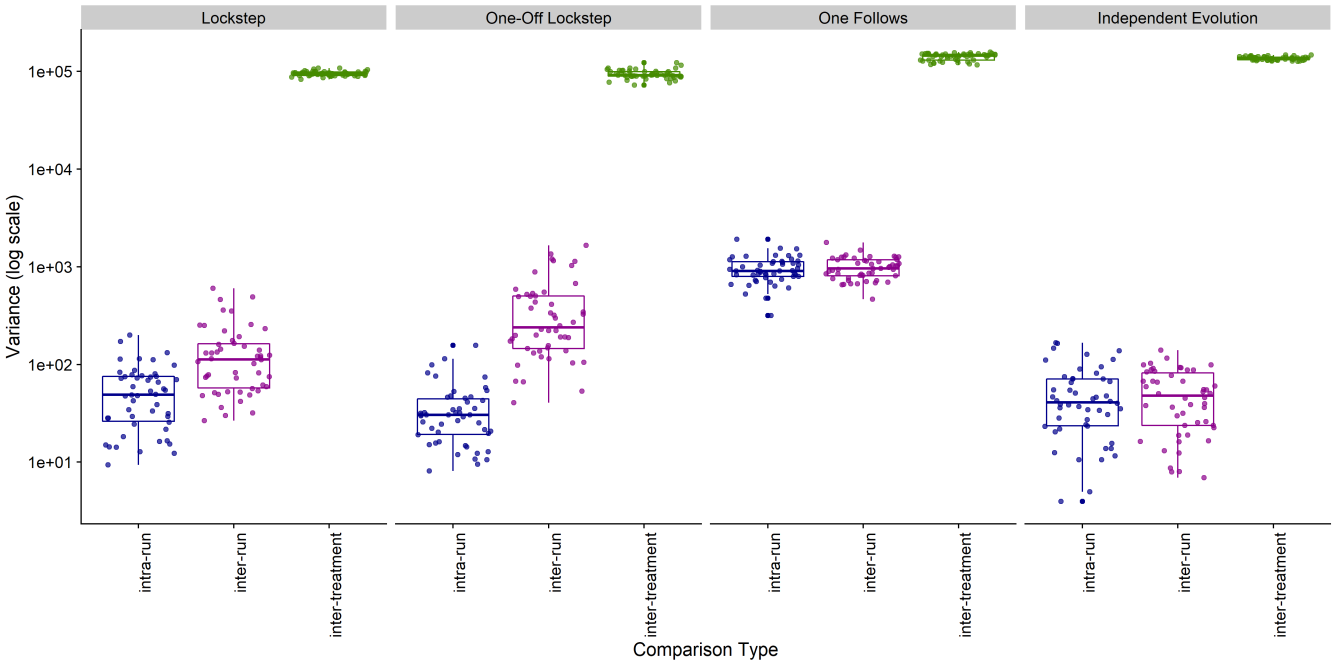

Figure 1: This figure shows the intra-run versus inter-run versus inter-treatment comparison on the x-axis for different scenarios (Lockstep, One-Off Lockstep, One Follows, Independent Evolution). The per-site mutation rate is 0.01 with a population size of 1000. The y-axis depicts the variance that was computed with the mutation accumulation metric on a log scale.

We use three different comparisons with different replicates to ensure that the signal for a lockstep pattern really is there: Intra-run, inter-run and inter-treatment comparison. In the intra-run comparison, both lower-level individuals actually evolved together, in the inter-run the lower-level individuals are from different replicates, but from the same scenario. The inter-treatment comparison takes one lower-level individual from a scenario-replicate and the other lower-level individual from the Random Drift-scenario. For each comparison, we generated 50 independent data points and plotted them. In the lockstep-like scenarios we found that the lockstep was detected by our mutation-accumulation-over-time metric since the intra-run, inter-run and inter-treatment comparisons were significantly different.

Figure 1 shows that our metric detects lockstep patterns. All scenarios in figure 1 were analyzed using Wilcoxon rank sum test. Lockstep and One-Off Lockstep were significant with p-values of 2.4e-06 (intra-run versus inter-run), <2e-16 (inter-run versus inter-treatment) and 4.3e-16 (intra-run versus inter-run), <2e-16 (inter-run versus inter-treatment), respectively.

Those results are robust in regards of almost all mutation rate and population size combinations we have run. Only when the mutation rate is very high (0.03) or the population size very small (10), there is no evidence of a lockstep-like co-evolution between the two lower-level individuals that form the higher-level organism. In the One-Follows scenario, we detected no difference between intra- and inter-run lineage pairs. Because one of the lower-level individuals can fall behind the other by many leading ones, the lineages are less tightly linked, allowing 'unanswered' mutations to accumulate in one of the lineages.

Discussion and Conclusion

Within an evolutionary process of a multi-level population, selection mechanisms may act at each level. When a system with two levels exists and those levels aren't perfectly aligned, one level of selection will determine which higher-level organisms (or groups) move on to the next generation, which a second level of selection will determine which individuals are used in the propagule that forms the offspring group in that next generation. At the higher level, cooperation to replicate the whole organism is going to be most important, whereas at the individual cell level, it will be most important that one's own genetic material is passed on into the propagule that forms the next generation. (Peck, 1992; Leigh Jr, 1977) Hence, individual interests as well as group interests need to be satisfied and the selection mechanisms play a crucial role in doing so.

In the study presented here, we focused only on the high-level selective pressures, mandating that both cells in the higher-level organism would always be pass on to the offspring. Under these idealized conditions, where mutualisms were forced, we were able to detect a genetic signature, al-

beit only when beneficial mutations in one cell type needed to be closely linked to beneficial mutations in the other cell type for the higher-level organism to benefit.

After looking at those preliminary baseline-models to get a feeling for what we should expect, we plan to look at a broader range of symbiotic behavior where the relative strengths of higher-level and lower-level selective pressures can be adjusted. As of yet, the two types of lower-level individuals have been measured with the same fitness function and the fitness of the higher-level organism is determined by how well the lower-level individuals contribute to the current scenario. In the future, we will have three distinct fitness functions and allow multi-level selection with conflicting pressures. Each of the two lower-level populations will have their own fitness function plus a different one for the higher-level population. To ensure egalitarian behaviour, we will additionally allow horizontal gene transfer (migration) while keeping the vertical gene transfer (mutation). By varying the probabilities for migration and mutation, we expect to see interesting patterns in the interaction of the individual-level and group-level selection mechanism.

Our metric has given us a promising direction for looking more into lockstep-like patterns since we were able to see co-evolution happen when analyzing the lineages. Therefore, we plan to use fitness functions that trigger the lockstep pattern. The individuals that form the higher-level organism will be disentangled and evaluated separately. On the group level, we will determine fitness by the number of matching bits between the two lower-level individuals. Therefore, it is also a manifestation of the lockstep pattern. We will no longer use the leading-ones component since it appears to overcomplicate the evolutionary process. On the individual level, the fitness functions will discourage matching bits by selecting for the number of zeroes and the number of ones, respectively, in the two lower-level populations.

We will introduce migration as follows: A migration rate of *e.g.*, ten percent means that ninety percent of the lower-level individual-pairs from the current generation stay together and move on to the next generation as pairs. Then, ten percent are picked from the first population of lower-level individuals and ten from the second population. Those lower-level individuals are randomly paired and move on to the next generation. In that way, we hope to produce a trade-off between group-level fitness goals and individual-level fitness goals when varying the migration rate. We hope to see loose and tight bindings between groups and individuals with migration rates ranging form zero to hundred percent.

Acknowledgements

This work was supported by the Austrian Marshall Plan Foundation and by Michigan State University through computational resources provided by the Institute of Cyber-Enabled Research (ICER) and through the BEACON Center for the Study of Evolution in Action. NS would like to particularly thank members of MSU's Digital Evolution Lab and Dr. Anya Vostinar for their overwhelming support, the many ideas, the thoughtful discussions and for sharing their enthusiasm.

References

Bohm, C., G., N., and Hintze, A. (2017). Mabe (modular agent based evolver): A framework for digital evolution research. In *Artificial Life Conference Proceedings 14*, pages 76–83. MIT Press.

Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York.

Dolson, E., Lalejini, A., Jorgensen, S., and Ofria, C. (2020). Interpreting the tape of life: Ancestry-based analyses provide insights and intuition about evolutionary dynamics. *Artificial Life*, 26:58–79.

Guerrero, R., Margulis, L., and Berlanga, M. (2013). Symbiogenesis: the holobiont as a unit of evolution. *International Microbiology*, 16:133–143.

Leigh Jr, E. G. (1977). How does selection reconcile individual advantage with the good of the group? *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 74(10):4542–4546.

Lopez-Garcia, P., Eme, L., and Moreira, D. (2017). Symbiosis in eukaryotic evolution. *Journal of Theoretical Biology*, 434:20–33.

Peck, J. R. (1992). Group selection, individual selection, and the evolution of genetic drift. *Journal of Theoretical Biology*, 159(2):163–187.

Queller, D. C. (1997). Cooperators since life began. *The Quarterly Review of Biology*, 72(2):184–188.

Queller, D. C. and Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philosophical Transactions of the Royal Society B*, 364:3143–3155.

Smith, J. M. and Szathmary, E. (1995). *The Major Transitions in Evolution*. Oxford University Press, New York.

Szathmary, E. (2015). Toward major evolutionary transitions theory 2.0. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 112(33):10104–10111.

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

Thursby, E. and Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal*, 474:1823–1836.

Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA.