Identifying Necessary Components for Open-Ended Evolution

Anya E. Vostinar[1,2,3,*], Emily L. Dolson[1,2,3,*], Michael J. Wiser[1,3], and Charles Ofria[1,2,3]

[1]BEACON Center for the Study of Evolution in Action
[2]Department of Computer Science and Engineering, Michigan State University
[3]Program in Ecology, Evolutionary Biology, and Behavior, Michigan State University
ofria@msu.edu
[*]Authors contributed equally

Introduction

A central goal of the field of artificial life is to build evolving systems that capture interesting dynamics of natural systems, producing evolutionary outcomes such as sophisticated navigation behaviors, novel cooperative strategies, complex ecosystems, or major evolutionary transitions, to name but a few. Such "open-ended" systems are sought after for a number of reasons: 1) For artificial life researchers, the presence of dynamics that are seen in biology but not artificial life raises the possibility that there is some fundamental and as-of-yet unidentified quality that artificial life systems are missing (Korb and Dorin, 2011). 2) For biologists, access to systems exhibiting complex and nuanced evolutionary processes allows rapid experimentation and facilitates understanding them on a mechanistic level (Tenaillon et al., 2016). 3) For evolutionary computation researchers, insights from open-ended evolutionary systems have the potential to expand the classes of applied engineering problems that we are able to solve with evolutionary algorithms (Hara and Nagao, 1999; Potter and Jong, 2000). Biological evolution is the only process known to have produced general intelligence; replicating this process would provide incredible insights into our own origins, as well as allowing us to harness these dynamics to spur breakthroughs in artificial intelligence. While various artificial life systems have recreated individual dynamics – such as the evolution of complexity, cooperation, and competition (Lenski et al., 2003; Goldsby et al., 2012; Zaman et al., 2011) – these accomplishments have been in highly controlled circumstances. The overarching goal of open-ended evolution research is to create a system where all of these dynamics can emerge more organically, as in the biosphere.

Despite general agreement on open-ended evolution being an important goal in artificial life, the field has yet to come to a consensus on how to define whether systems are making progress towards achieving this goal. A number of dynamics have been proposed as being possibly necessary for open-ended evolution, most notably the continuous production of novelty (Lehman and Stanley, 2011) (also discussed by Wolfgang Banzhaf, among others, in (Taylor *et*
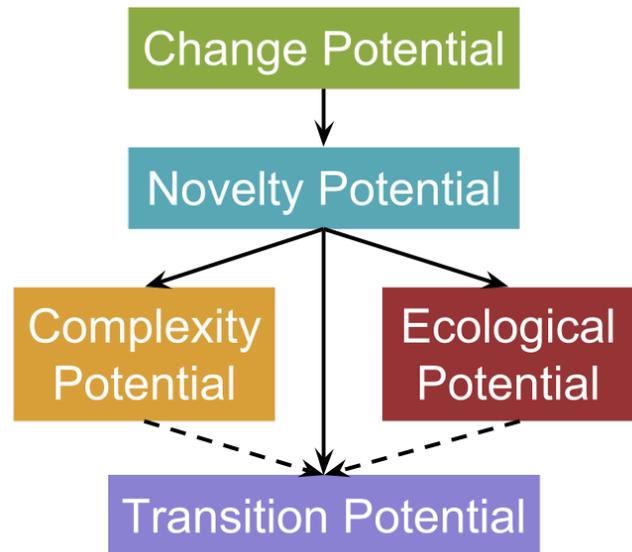

Figure 1: **Relationships between the metrics.** Originally published in (Dolson et al., 2015).

al., in press)), continuous increase in diversity (Bedau and Bahm, 1994), continuous increase in complexity (Lenski et al., 2003), and shifts in individuality such as those often associated with major transitions in evolution (Smith and Szathmary, 1997). We argue that all of these dynamics are important pieces of the open-ended evolution puzzle. In addition, we have previously suggested that there is a fifth necessary and even simpler dynamic: continuous change in the population (Dolson et al., 2015).

These five properties of a system fit into a hierarchy, as shown in Figure 1. For novelty to exist, there must be some degree of change in a population. While this is trivially true, many evolutionary algorithms suffer from premature convergence, which is essentially the absence of change, so it remains an important prerequisite to define and explicitly address. Similarly, complexity and diversity can only increase indefinitely if novel members of the population continue to

be generated. Finally, transitions in individuality typically involve multiple organisms coming together into a single individual, requiring an increase in complexity and a diversity to build from. All of these dynamics capture different subsets of interesting behavior that an evolving system might have and we propose they are all necessary for a system to demonstrate open-ended evolution.

To draw conclusions about what factors of a system promote or inhibit these dynamics, it is critical to first develop a common suite of metrics that are applicable across a wide variety of systems. Some progress has been made toward this end with evolutionary activity statistics (Bedau et al., 1998; Channon, 2001). Evolutionary activity statistics focus on "components" representing the meaningful individual pieces of a system, which the authors admit will need to be defined for each system as appropriate. For instance, in the fossil record, species were used as components. Once these components are decided upon, the diversity and amount of adaptive evolutionary activity of components in a given time step are used to determine into which of three classes of evolution the system falls. These classes of evolutionary activity (none, bounded, unbounded) follow logically from whether diversity and novel evolutionary activity per component is bounded or unbounded. However, to filter out changes that do not represent adaptive evolutionary activity, a neutral "shadow" control with no selective pressure must be run for comparison. While using a shadow population as a null model is quite clever, creating this control treatment can make implementing evolutionary statistics difficult in complex systems. Moreover, the need to select an appropriate component can make results challenging to generalize across systems. Lastly, evolutionary activity statistics may not be able to differentiate dynamics in which a subset of the population is gaining new and interesting traits from less interesting dynamics in which that subset is merely under stabilizing selection.

In this paper, we extend the core ideas behind evolutionary activity statistics. We propose to quantify the idea of open-ended evolution with a suite of four necessary (but not always sufficient) metrics that attempt to balance generalizability across systems with the ability to capture the ideas of the first four dynamics previously discussed: change, novelty, complexity, and ecological interactions. The novelty and ecological metrics track the same dynamics as new components and diversity in evolutionary activity statistics. Complexity provides a direct measure how much information is stored in the genome, and is likely to indicated whether any part of the system is gaining interesting adaptations. We propose an alternative technique for isolating adaptive evolutionary activity, which we hope will be more widely applicable (see Persistent Lineages below). Additionally, we propose a technique for selecting meaningful components that will work for any system in which genomes are composed of elements that collectively determine fitness (see Informative Sites). We present our results when applying these four metrics to an NK system. The fifth metric, transition potential, has proven more difficult to quantify in an intuitive and computationally tractable manner and we are reserving it for future study.

Experimental System

To begin a systematic examination of our metrics, we used a simple NK model (Kauffman and Levin, 1987). An NK model uses two parameters, N and K, to randomly generate a fitness landscape. N specifies the number of sites in the genome, each of which is a 0 or a 1. The fitness landscape specifies the effect of a given value at a given site on the fitness of the bit-string organism. This fitness effect depends on the values at the K subsequent adjacent sites. As such, K tunes the ruggedness of the landscape; low values of K produce smooth landscapes with few peaks, whereas high values produce landscapes with many peaks. We chose to use an NK model because they are a well-understood system for studying general questions about evolutionary dynamics.

Experimental Treatments

Our basic treatment used $N = 20$ (i.e., 20 bits in an individual) and $K = 3$ (the fitness contribution of each bit was influenced by three other bits). We used a population size of 200 and a mutation rate of 3 sites per birth (three bits were randomized in each birth, so there is a 1/8 probability of all three retaining their original values), with tournament selection and a tournament size of 15. On top of this basic treatment, we performed eight experimental treatments: **High K** tests the effect of a highly rugged landscape (K=10) where fitness is effectively randomized whenever a mutation occurs. **High N** tests the effect of longer bit-string genomes (N=100), allowing for a higher potential complexity. **Low Mut** and **High Mut** test the effects of very low and very high mutations rates (1 bit and 6 bit randomizations per birth, respectively); we expect the mutation rate to be important for finding new areas of the fitness landscape and thus our novelty metric. **Small Pop** and **Large Pop** vary the population size (to 20 and 1000 respectively); in small populations we expect more drift in the population, allowing more change, while in a large population we expect stronger selection and consequently that a higher percentage of changes along the line of descent are beneficial. Finally, we included two special treatments: in **Changing Environments**, the fitness function was toggled every 500 generations, allowing us to see the effect of changing selective pressures where the populations couldn't stay on a single peak. In **Fitness Sharing** organisms that were too similar to each other detracted from each others fitness, creating a pressure to explore multiple portions of the landscape at the same time and, ideally, maintain a high diversity (Goldberg and Richardson, 1987).

Persistent Lineages

At any given point in time, the population will contain some maladapted genomes that recently arose via mutation. These genotypes will quickly be purged from the population via natural selection and will add noise to our metrics. To decrease this noise, we must filter out such genomes. We accomplish this by looking backwards in time to see which organisms were the progenitors of lineages that persisted for a substantial number of generations. We mark each organism with a lineage number at a given time point A, as demonstrated in Figure 2 (where color indicates lineage number). The lineage numbers are passed on to offspring throughout the time interval (50 generations in our experiments). At time point A+50, we determine which genomes from the population at A have descendants at A+50. At this point, those genomes are considered **persistent**; in the example in Figure 2 the green and blue lineages are considered persistent at time point A+50. We then compare the green and blue genomes from time point A to the genomes that were persistent previously, either from time point A-50 or from all previous time points, depending on the metric; in the example we would compare to the purple genome at time point A-50. This filtering leads to a delay in counting a genome in a metric until 50 generations later, but enables us to avoid an apparent increase in metrics due to drift via mutation. For example, the red, orange, and blue genomes from time point A-50 would never be considered in our metrics because their lineages do not persist to time point A.

Informative Sites

While a genome may have descendants in 50 generations, this persistent genome may not be phenotypically different than another persistent genome in the population. To ensure we are not counting phenotypically identical genomes separately, we determine which sites in the genome contain information about the environment. In an NK bit-string model, the only information that an organism can have about its environment is whether it is better to have a 1 or a 0 at each site. Thus, informative sites are those for which flipping the corresponding bit would result in a fitness decrease. In calculating all of the following metrics, we first reduce the genome to its informative sites.

This approach could easily be extended to any system in which the genome is made up of a sequence of elements that collectively determine fitness. Whether or not a genomic position is informative can be approximated by measuring the overall fitness effect of changing it to a null alternative that is known to not contribute information. A null version of the element is simply a special element that does nothing but prevents the subsequent elements from shifting positions. A more accurate technique would examine the fitness effect of substituting all possible alternative elements and calculating the potential entropy at that site; when null substitutions are not possible, this technique is an effective method. A caveat
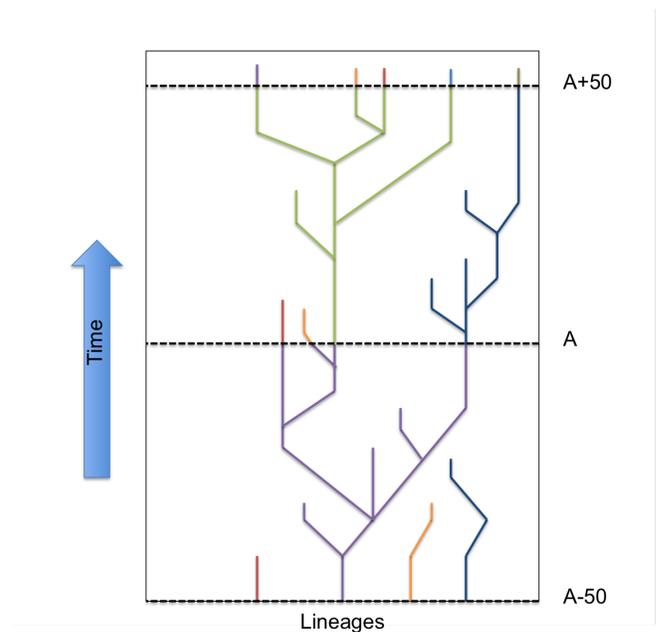

Figure 2: **An illustrative example of how we filter genomes for persistent lineages.** At time point A, the purple lineage has proven to be persistent and therefore the original genome from A-50 will be considered meaningful. Similarly, the green and blue lineages persist to time point A+50 and so the original green and blue genomes will be considered meaningful as they were at time point A.

to this technique is that genomes that achieve a given result in an excessively fragile manner may appear more complex than more robustly built genomes. To mitigate this issue, the combined fitness effect of eliminating multiple genome elements at once can be measured. Fortunately, this is not a risk in an NK model.

Metrics

Change Metric

Our first metric focuses on whether the genetic makeup of the population is changing. This metric should always be above zero unless the population has converged and no beneficial variation is being introduced. We use our method of filtering genomes (explained previously) to ensure that we only record a genome as new compared to the previous time point if its lineage has persisted for one full time point. For this comparison, we first find the genomes from persistent lineages from time point A by determining which genomes have descendants in time point A + 50. In the example shown in Figure 2, this would be the genomes at the root of the green and blue lineages. We then compare these genomes to those found to have been from persistent lineages in time point A-50 because they have descendants in time point A, purple in Figure 2. In this way, we create a

sliding window to observe change in the population.

Novelty Metric

The novelty metric measures how many genomes have evolved in the population that have never been seen previously in the experiment. For this metric we again filter out genomes that do not have descendants in the next time point, enabling us to focus on meaningful novelty. To measure novelty, we simply count how many genomes from persistent lineages (genomes that existed in time point A and have descendants in time point A+50) have never been in a previous time points persistent genome pool. It is possible with this metric for a genome to evolve, but not persist, and therefore not be recorded in the permanent history, but then evolve and persist at a later point and be counted as novel. Once a genome has been counted as novel, however, it is part of the permanent history and will never be counted in the novelty metric again. Thus, while a genome could be delayed in being counted as novel, or not counted if it never persists, it will not be counted twice.

Complexity Metric

The complexity metric measures the maximum complexity (informative sites) of any organism found in the entire population. In the case of the NK landscapes, this is simply the number of sites that, if toggled, would decrease the organism's fitness. While this is easy to saturate for such a simple model system, for more sophisticated encodings the complexity metric counts the number of sites that contribute constructively to an organism's phenotype. Further, complexity can be measured even more accurately by using more sophisticated information-theory techniques where all possible mutations are considered at each site, and ideally some epistatic interactions.

Ecological Metric

The ecological metric measures the amount of information in the population as a whole. While organisms may not contain increasing amounts of information in their individual genomes (as measured by the complexity metric), they could still be increasingly diverse and therefore contain increased information collectively in the population. We can measure such a phenomenon by looking at the diversity of persistent genotypes reduced to informative sites. Complex ecologies in which multiple subsets of the population are using different information about the environment to survive are likely to be characterized by a relatively balanced distribution of individuals across the various successful phenotypes. Thus, we use Shannon entropy, a popular metric of diversity that also measures evenness, to measure the diversity of the persistent genotypes and calculate the ecological metric.
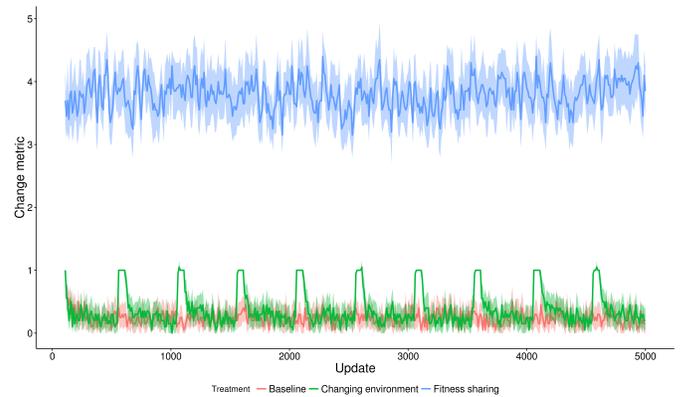

Figure 3: **Amount of change at over time in varying environments.** As measured by the change metric, fitness sharing increases the amount of change in the population over time. Conversely, a routinely changing environment leads to spikes in change that quickly drop as the population converges again. Shaded region represents a bootstrapped 95% confidence interval around the mean.

Results

To ensure that these metrics are capturing the dynamics that we want them to, we tested them on a range of variants of our basic NK model, described above. The preliminary results for each metric are presented here.

Change Metric

For a system to exhibit open-ended evolution, it is clearly necessary for the composition of the population to be changing in some meaningful way, although this change needn't always be novel. We measured the meaningful change in our system under varying environmental conditions to confirm that this metric responds intuitively. As shown in Figure 3, several environmental changes increase the amount of meaningful change found in the populations over time. When organisms are forced to share fitness between others with the same genotype, the amount of change increases and remains higher than the baseline over time. Conversely, when the environment changes frequently, there is an initial spike of increased change that quickly drops back down to the baseline value.

The majority of environmental conditions we tested produced dynamics over time qualitatively similar to the baseline treatment. In Figure 4 we show the amount of meaningful change in populations at the final time point in more environmental conditions. Larger genomes (N) lead to increased meaningful change because they allow for a larger search space that the population can continuously mutate within, leading to more beneficial mutants. A higher mutation rate leads to increased meaningful change because mutations are necessary to create any meaningful change in our system. A smaller population size produces more meaningful change
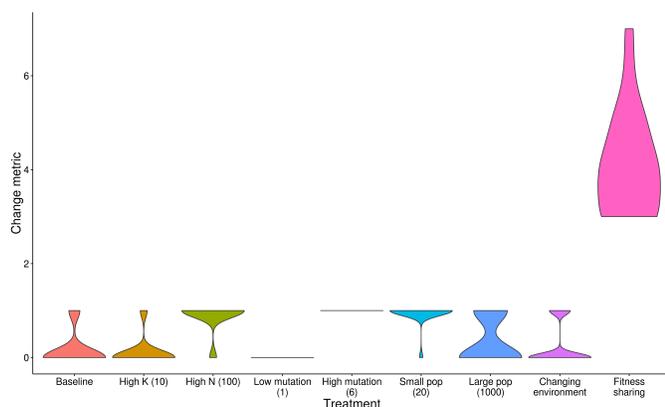
Figure 4: **Amount of change at final generation in varying environments.** As measured by the change metric, environmental conditions that increase the amount of change at the final time point include: increasing the size of the genome (N), increasing the mutation rate, decreasing the population size, and implementing fitness sharing. Shaded region represents a bootstrapped 95% confidence interval around the mean.

Figure 5: **Amount of novelty over time with varying mutation rate.** The novelty metric measures the number of completely new meaningful genomes that have lineages that persisted since the previous time point. As the mutation rate increases, more novelty is continuously produced, however at all mutation rates, the novelty decreases over time. Mutation rate 3 is the baseline treatment in previous graphs. Shaded region represents a bootstrapped 95% confidence interval around the mean.

because a small population cannot hold as many different genomes at one time and therefore there are more genomes that can arise that are different than what is in the previous population. Finally, fitness sharing can produce increased meaningful change because it creates a constant pressure for the population to adapt away from whatever is the current dominant genotype.

While change is a metric often not considered in discussions of open-ended evolution, these results show that the amount of meaningful change can reflect differences in the environment and evolution of the populations and is likely a necessary dynamic for open-ended evolution. Our change metric responds in intuitive ways to variations in parameter settings, suggesting that it is a reliable indicator of the dynamics we designed it to capture.

Novelty Metric

The continuous production of meaningfully new genomes is necessary for what is generally considered open-ended evolution, because without meaningfully novel genomes, nothing "interesting" can emerge. Therefore, our second metric measures the amount of meaningfully novel genomes (i.e., have never been recorded previously) that have persisted in a population since the previous time point. As shown in Fig 5, a higher mutation rate increases the amount of novelty generated by a system. This result is to be expected, because more mutations make it easier to cross fitness valleys and otherwise traverse the fitness landscape. As expected, even at a high mutation rate, novelty does start to decrease over time as the search space is explored.

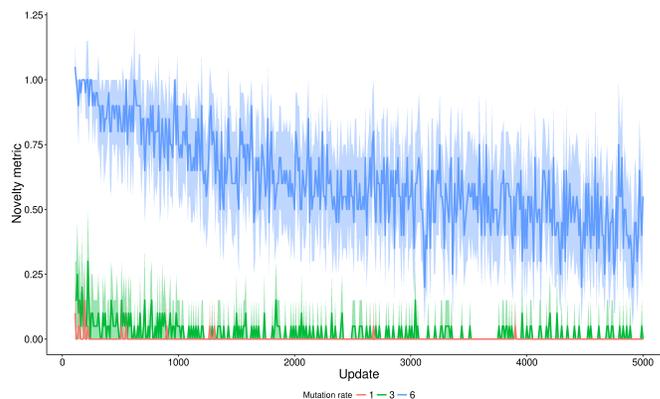We again found that the majority of treatments had a qual-

itatively similar trajectory over time and therefore in Figure 6 we show only the final novelty value. As predicted, the baseline treatment has stopped producing meaningful novelty by the final time point. However, many environments do allow for continuing production of novelty, specifically increased epistasis (K), larger genomes (N), higher mutation rate, differing population size, and fitness sharing. Epistasis increases the value of the novelty metric at the final generation because the population must explore a rugged fitness landscape instead of converging to a single fitness peak, leading to more adaptive novel genomes to discover throughout evolution. Larger genomes allow for more novelty at the final generation because they increase the size of the search space and therefore how many genomes can ever be considered novel. Smaller population sizes increase novelty at the final generation because it takes the population longer to discover many genomes, leaving enough to still be novel at the end of the experiment. Conversely, larger population sizes allow for some increase in novelty over the baseline treatment because the increased number of organisms makes it easier for more of the search space to be explored. Finally, fitness sharing increases the final novelty by creating a selective pressure to explore genotypes that are not currently common in the population; many of these will be novel, as novel genotypes are, by definition, not present in the population.

These results highlight the power of the novelty metric to identify environments and populations that have the potential to be open-ended due to the high number of new genotypes being consistently discovered. Novelty is likely neces-
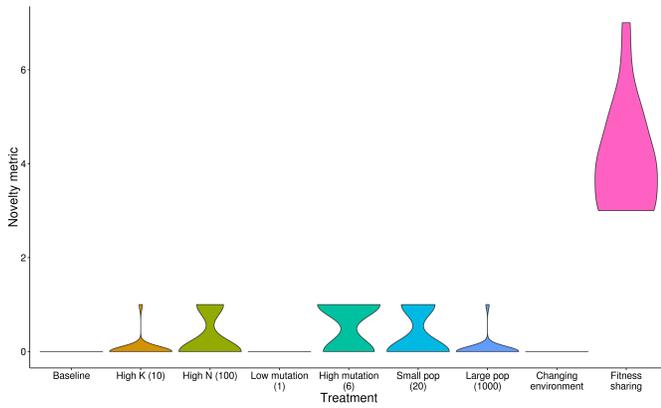
Figure 6: **Amount of novelty at final time point in varying environments.** At the final time point, no meaningful novelty is found in our baseline populations. However, increasing the amount of epistasis (K), increasing genome length (N), increasing mutation rate, decreasing population size, and enabling fitness sharing increase the amount of novelty produced in the final time point.

Figure 7: **Amount of complexity over time with fitness sharing.** The baseline treatment (red) is able to reach the top complexity allowed by the model quickly and remain at that value. When fitness sharing is introduced (blue), the population is not able to attain the top complexity value. Shaded region represents a bootstrapped 95% confidence interval around the mean.

sary, but not sufficient, for open-ended evolution because if nothing new is being produced by a population, neither the complexity nor the ecological metric can be non-zero.

Complexity Metric

The presence of highly complex organisms is a property of biological evolution that is often cited as something that artificial life systems have yet to achieve. Figure 7 shows the trajectories of the complexity metric over time in populations with and without fitness sharing. In our baseline treatment, organisms cannot evolve to be more complex than N. As a result, the complexity increases over time and then saturates. Conversely, fitness sharing leads to lower complexity because it weakens the pressure to be at the very top of a fitness peak (where complexity would be maximized).

Because of the simplicity of our system, the complexity metric stabilizes quickly in most treatments and therefore we show only the final time point values for complexity in Figure 8. The baseline treatment was able to reach the maximum complexity possible for a 20-bit genome and most of the environments did not decrease in the final complexity value. When the genome length was increased to 100 bits, the populations were also able to reach the new maximum complexity value of 100. However, high mutation rate, smaller population size, and fitness sharing all somewhat decreased the final complexity achieved on average. A high mutation rate decreased the final complexity because it introduced more deleterious mutations and therefore increased the percentage of the population that has just mutated away from the fitness peak at any point in time. A smaller population size decreased the complexity somewhat because smaller populations are more susceptible to genetic
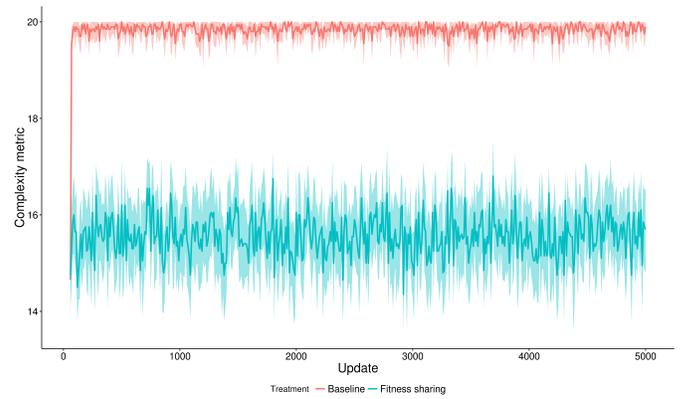
drift. As discussed previously, fitness sharing decreased the overall complexity achieved by the population due to the weakened selection for climbing fitness peaks when there are many organisms on that peak.

These results highlight the strengths of the complexity metric because the NK model is not able to continuously increase the complexity metric over time. Once the maximum complexity allowed by the genome length is reached, no higher value is possible. It is likely that a continuously increasing complexity metric is necessary for an open-ended system, as appears to be the case with the biosphere. This is only possible in a system with a sufficiently complex environment such that there is always new information to be integrated into the genome.

Ecology Metric

Finally, a diverse population of interacting organisms promotes feedback loops that allow for the continuous creation of new and complex adaptations. Even if a system is not evolving more complex individual organisms, it may still be exhibiting interesting dynamics because it allows for a large number of niches that can be occupied by a diversity of organisms. Our ecology metric captures this dimension of open-ended evolution. As shown in Figure 9, if the population is large enough, the ecology metric will remain fairly stable over evolutionary time. Larger population sizes led to higher overall ecology values, while small populations rapidly converged to a single genotype.

The majority of environments led to a complete convergence by the final time point and therefore we show only their final values in Figure 10. The only treatments that did not drop to an ecology value of 0 were the large population
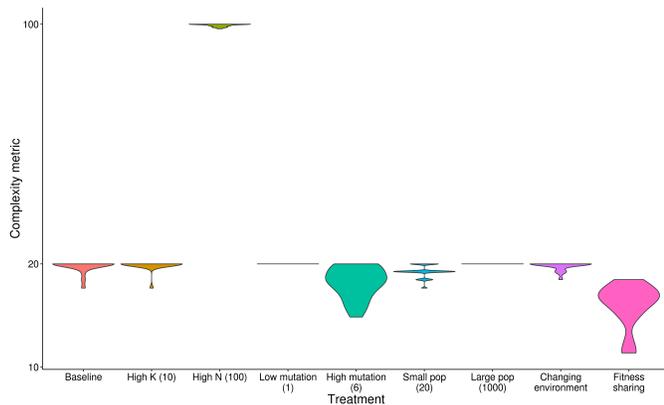
Figure 8: **Amount of complexity at final time point in varying environments.** Most of the treatments reach the maximum complexity allowed by the genome length (20 or 100) and cannot continue to increase. High mutation rate, smaller populations, and fitness sharing decrease the final complexity achieved by the populations on average.
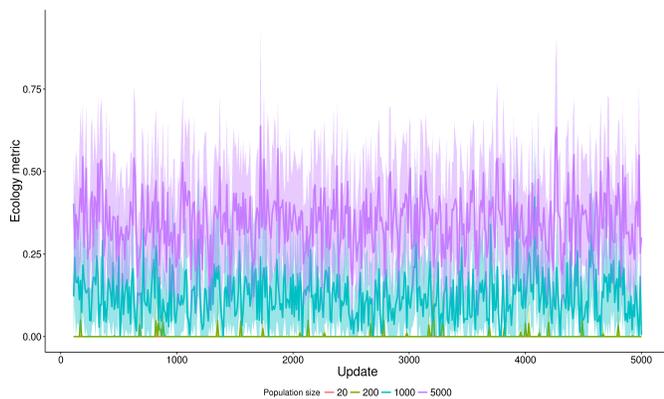

Figure 9: **Amount of ecology over time with varying population size.** At larger population sizes, ecology (measured by Shannon diversity of meaningful genomes) is non-zero throughout evolution. At smaller population sizes, the ecology value drops to 0, indicating there is no diversity. Shaded region represents a bootstrapped 95% confidence interval around the mean.
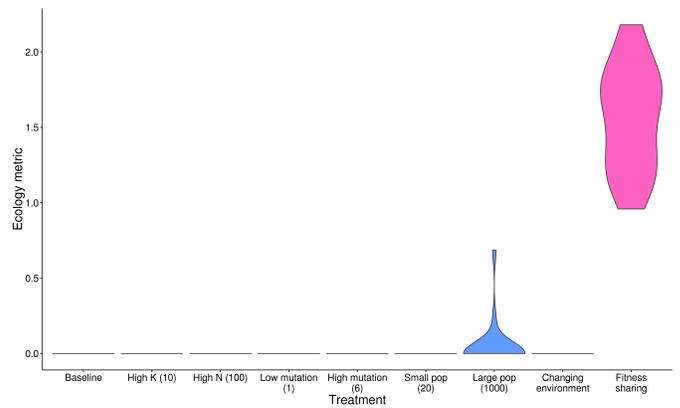

Figure 10: **Amount of ecology at final time point in varying environments.**

size and fitness sharing. Because fitness sharing specifically rewards organisms with less common genotypes, it promotes a stably high ecology value over time. This result shows the trade-off inherent in fitness sharing because it leads to higher ecology at the expense of lower complexity.

We hypothesize that a continuously increasing ecology metric in addition to a continuously increasing complexity metric are necessary for an evolutionary system to truly be open-ended as we observe in the biosphere.

Conclusions

Though our simulation is fairly simple, these initial tests demonstrate that our metrics respond in an intuitive way to the dynamics in an evolutionary system. Thus, they should be able to detect systems that are theoretically capable of achieving open-ended evolution. First, it is clear that for new behaviors to evolve in a system, new genomes need to be evolving. Therefore, both the change and novelty metrics would need to be non-zero – though they could be stable instead of increasing since both metrics measure a change.

Second, continuously increasing maximum complexity with a stable environment would imply increasing biotic interactions, potentially cooperative or complex competitive interactions. If maximum complexity is increasing in a population, at least one organism must be incorporating more information into its genome than any organism had before. This information can be about the environment up to a point. However, if the environment is not changing, maximum complexity will not be able to increase unboundedly without biotic interactions between organisms. Therefore, continuously increasing maximum complexity implies that interesting biotic interactions are evolving in a system.

Finally, a continuously increasing ecological metric in a stable environment implies an ecosystem is forming; if the ecology metric is increasing, the informative diversity in the population is increasing. For diversity to be increasing, new

niches must be emerging – if there were only a single niche, the best genotype for that niche would fill it and take over the population. The only way for diversity to be maintained for a meaningful length of time is for new niches to be created. While new niches can emerge from abiotic factors, if the environment is not changing, those will all eventually fill. Therefore, the only way for an indefinite series of new niches to form (which is required for an unbounded increase of the ecological metric), is via biotic interactions between organisms in the population. Those biotic interactions could fall into a number of categories such as predator and prey, mutualism, parasitism, or commensalism, but as the number of them increase, a complex ecosystem emerges.

We can use these metrics to understand the impact of incremental changes to a system. In order to apply the scientific method to a monolithic problem like designing an open-ended evolutionary system, we need to be able to break that problem down into components that can be addressed in a systematic manner. Moreover, one of the primary goals in building such a system is to understand what components are necessary to do so. By looking at the effects that individual, controlled changes to a system have on this suite of metrics, we can more effectively work towards these goals.

The metrics we propose here are applicable not only to digital systems, but are also relevant to experimental biological systems. The Long-Term Evolution Experiment (LTEE) (Lenski et al., 1991) is an exemplar of experimental evolution, consisting of 12 populations of the bacteria *E. coli*, which have been evolving independently for more than 60,000 generations. As detailed in (Taylor et al., ress), the LTEE exhibits many hallmarks of open-ended evolution, including the criteria we propose here. Because fitness within the LTEE is best described by an unbounded power law function (Wiser et al., 2013; Lenski et al., 2015), the system meets the change metric: populations continue to change in non-trivial ways over time. Further, studies of individual populations within the LTEE have shown numerous examples of the generation of novelty, including exploration of new areas of the fitness landscape (Tenaillon et al., 2016), repeated selective sweeps (Maddamsetti et al., 2015), and new diversity arising after such sweeps (Blount et al., 2012). Towards the ecology metric, several populations within the LTEE demonstrate frequency-dependent fitness dynamics (Ribeck and Lenski, 2015; Rozen and Lenski, 2000; Le Gac et al., 2012; Maddamsetti et al., 2015), which are necessarily cases of ecological interactions. Included in these cases of frequency dependence is a special case (Blount et al., 2008, 2012; Turner et al., 2015a) driven by cross-feeding and specialization on different resources (Turner et al., 2015b). Because all of the populations in the LTEE began as single cells, all ecological complexity in any populations must have arisen during the course of the experiment, and thus satisfies the ecological metric. The complexity metric is inherently harder to quantify in a biological system than in a compu-

tational one, but recent large scale genome sequencing from the LTEE (Tenaillon et al., 2016) offers the promise of being able to measure complexity at the genome level over the course of the experiment. Because our metrics can be applied to a well-studied biological example of open-ended evolution, they can be used to compare dynamics in a broad range of systems and enable the field of artificial life to move forward in quantifiable steps to open-ended evolution.

References

Bedau, M. A. and Bahm, A. (1994). Bifurcation structure in diversity dynamics. In *Artificial life IV*, pages 258–268.

Bedau, M. A., Snyder, E., and Packard, N. H. (1998). A classification of long-term evolutionary dynamics. In *Artificial life VI*, pages 228–237. Cambridge: MIT Press.

Blount, Z. D., Barrick, J. E., Davidson, C. J., and Lenski, R. E. (2012). Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature*, 489(7417):513–518.

Blount, Z. D., Borland, C. Z., and Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proceedings of the National Academy of Sciences*, 105(23):7899–7906.

Channon, A. (2001). Passing the alife test: Activity statistics classify evolution in geb as unbounded. In *Advances in Artificial Life*, pages 417–426. Springer.

Dolson, E., Vostinar, A. E., and Ofria, C. (2015). What's holding artificial life back from open-ended evolution?

Goldberg, D. E. and Richardson, J. (1987). Genetic algorithms with sharing for multimodal function optimization. In *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*, pages 41–49. Hillsdale, NJ: Lawrence Erlbaum.

Goldsby, H. J., Dornhaus, A., Kerr, B., and Ofria, C. (2012). Task-switching costs promote the evolution of division of labor and shifts in individuality. *Proceedings of the National Academy of Sciences*, 109(34):13686–13691.

Hara, A. and Nagao, T. (1999). Emergence of cooperative behavior using ADG; Automatically Defined Groups. In *Proceedings of the 1999 Genetic and Evolutionary . . . .*

Kauffman, S. and Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128(1):11–45.

Korb, K. B. and Dorin, A. (2011). Evolution unbound: releasing the arrow of complexity. *Biology & Philosophy*, 26(3):317–338.

Le Gac, M., Plucain, J., Hindr, T., Lenski, R. E., and Schneider, D. (2012). Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with Escherichia coli. *Proceedings of the National Academy of Sciences*, 109(24):9487–9492.

Lehman, J. and Stanley, K. O. (2011). Abandoning Objectives: Evolution Through the Search for Novelty Alone. *Evol. Comput.*, 19(2):189–223.

Lenski, R. E., Ofria, C., Pennock, R. T., and Adami, C. (2003). The evolutionary origin of complex features. *Nature*, 423:139–144.

Lenski, R. E., Rose, M. R., Simpson, S. C., and Tadler, S. C. (1991). Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. *The American Naturalist*, 138(6):1315–1341.

Lenski, R. E., Wiser, M. J., Ribeck, N., Blount, Z. D., Nahum, J. R., Morris, J. J., Zaman, L., Turner, C. B., Wade, B. D., Maddamsetti, R., Burmeister, A. R., Baird, E. J., Bundy, J., Grant, N. A., Card, K. J., Rowles, M., Weatherspoon, K., Papoulis, S. E., Sullivan, R., Clark, C., Mulka, J. S., and Hajela, N. (2015). Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with Escherichia coli. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1821).

Maddamsetti, R., Lenski, R. E., and Barrick, J. E. (2015). Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli. *Genetics*, 200(2):619–631.

Potter, M. A. and Jong, K. A. D. (2000). Cooperative Coevolution: An Architecture for Evolving Coadapted Subcomponents. *Evolutionary Computation*, 8(1):1–29.

Ribeck, N. and Lenski, R. E. (2015). Modeling and quantifying frequency-dependent fitness in microbial populations with cross-feeding interactions. *Evolution*, 69(5):1313–1320.

Rozen, D. E. and Lenski, R. E. (2000). LongTerm Experimental Evolution in Escherichia coli. VIII. Dynamics of a Balanced Polymorphism. *The American Naturalist*, 155(1):24–35.

Smith, J. M. and Szathmary, E. (1997). *The major transitions in evolution*. Oxford University Press.

Taylor, T., Bedau, M., Channon, A., Ackley, D., Banzhaf, W., Beslon, G., Dolson, E., Froese, T., Hickinbotham, S., Ikegami, T., McMullin, B., Packard, N., Rasmussen, S., Virgo, N., Agmon, E., Clark, E., McGregor, S., Ofria, C., Ropella, G., Spector, L., Stanley, K. O., Stanton, A., Timperley, C., Vostinar, A., and Wiser, M. (In press). Open-ended evolution: Perspectives from the oee1 workshop in york. *Artificial Life*.

Tenaillon, O., Barrick, J. E., Ribeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., Wu, G. C., Wielgoss, S., Cruveiller, S., Medigue, C., Schneider, D., and Lenski, R. E. (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. *bioRxiv*.

Turner, C. B., Blount, Z. D., and Lenski, R. E. (2015a). Replaying Evolution to Test the Cause of Extinction of One Ecotype in an Experimentally Evolved Population. *PLoS ONE*, 10(11):e0142050.

Turner, C. B., Blount, Z. D., Mitchell, D. H., and Lenski, R. E. (2015b). Evolution and coexistence in response to a key innovation in a long-term evolution experiment with Escherichia coli. *bioRxiv*.

Wiser, M. J., Ribeck, N., and Lenski, R. E. (2013). Long-term dynamics of adaptation in asexual populations. *Science*, 342(6164):1364–1367.

Zaman, L., Devangam, S., and Ofria, C. (2011). Rapid host-parasite coevolution drives the production and maintenance of diversity in digital organisms. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 219–226. ACM.