Selection schemes from evolutionary computing show promise for directed evolution of microbes

Alexander Lalejini
lalejini@umich.edu
University of Michigan
Ann Arbor, Michigan, USA

Emily Dolson
dolsonem@msu.edu
Michigan State University
East Lansing, Michigan, USA

Anya E. Vostinar
vostinar@carleton.edu
Carleton College
Northfield, Minnesota, USA

Luis Zaman
zamanlh@umich.edu
University of Michigan
Ann Arbor, Michigan, USA

ABSTRACT

Directed microbial evolution harnesses evolutionary processes in the laboratory to construct microorganisms with enhanced or novel functional traits. Directing evolutionary processes for applied goals is fundamental to evolutionary computation, which harnesses the principles of Darwinian evolution as a general purpose search engine for solutions to computational problems. Despite overlapping aims, artificial selection methods from evolutionary computing are not commonly applied to living systems in the laboratory. Here, we summarize recent work wherein we ask if parent selection algorithms from evolutionary computation might be useful for directing the evolution of microbial populations when selecting for multiple functional traits. To do so, we developed an agent-based model of directed microbial evolution, which we used to evaluate how well three selection schemes from evolutionary computing (tournament selection, lexicase selection, and non-dominated elite selection) performed relative to schemes used in the laboratory (elite and top-10% selection). We found that lexicase selection and non-dominated elite selection generally outperformed the commonly used directed evolution approaches. Our results are informing ongoing work to transfer these techniques into the laboratory and motivate future work testing more sophisticated selection schemes from evolutionary computation in a directed evolution context.

CCS CONCEPTS

• **Applied computing** → **Computational biology**; • **Computing methodologies** → *Model development and analysis.*

KEYWORDS

directed evolution, artificial selection, digital organisms, selection schemes, agent-based modeling

ACM Reference Format:
Alexander Lalejini, Emily Dolson, Anya E. Vostinar, and Luis Zaman. 2022. Selection schemes from evolutionary computing show promise for directed evolution of microbes. In *Genetic and Evolutionary Computation Conference Companion (GECCO '22 Companion), July 9–13, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3520304.3528900

GECCO '22 Companion, July 9–13, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9268-6/22/07.
https://doi.org/10.1145/3520304.3528900

1 INTRODUCTION

Directed evolution harnesses artificial selection in the laboratory to generate biomolecules or organisms with desirable functional traits [2, 22]. The scale and specificity of artificial selection has been revolutionized by a deeper understanding of evolutionary and molecular biology in combination with technological innovations in sequencing, data processing, laboratory techniques, and culturing devices. These advances have cultivated growing interest in directing the evolution of whole microbial communities with functions that can be harnessed in medical, biotech, and agricultural domains [22].

Of course, attempting to direct evolutionary processes for applied goals has not been limited to biological systems. Evolutionary computing harnesses the principles of Darwinian evolution as a general-purpose search algorithm. In this brief communication, we overview recent work, "Artificial selection methods from evolutionary computing show promise for directed evolution of microbes" ([14]), wherein we use agent-based modeling to investigate the whether parent selection algorithms developed for evolutionary computing might be effective for directing the evolution of microbial populations.

As in evolutionary computing, directed evolution in the laboratory begins with a library—or population—of variants (*e.g.*, communities, genomes, or molecules). Variants are scored based on a phenotypic trait (or set of traits) of interest, and the variants with the "best" traits are chosen to produce the next generation. Such approaches to picking progenitors are known as elitist selection algorithms in evolutionary computing [3]. Evolutionary computing research has shown that these elitist approaches to artificial selection can be sub-optimal in complex search spaces. On their own, elitist selection schemes fail to maintain diversity, which can lead to premature convergence [12, 15], and they lack mechanisms to balance multiple objectives. Artificial selection routines (*i.e.*, parent selection algorithms or selection schemes) are intensely studied in evolutionary computing, and many *in silico* selection techniques have been developed that improve the quality and diversity of evolved solutions (*e.g.*, [7, 8, 13, 16, 17, 21]).

Given their success, we expect that artificial selection methods developed for evolutionary computing will improve the efficacy of directed microbial evolution in the laboratory, especially when simultaneously selecting for more than one trait (a common goal in evolutionary computation). However, directed microbial evolution differs from evolutionary computing in ways that may inhibit our ability to predict which techniques are most appropriate to apply in the laboratory. For example, candidate solutions in evolutionary computing are evaluated individually, resulting in high-resolution genotypic and phenotypic information that can be used for selecting parents, which are then copied, recombined, and mutated to produce offspring. In directed microbial evolution, individual-level evaluation is often intractable at the scale required for directed evolution; as such, evaluation often occurs at the population-level, and the highest performing populations are partitioned (instead of copied) to create "offspring" populations. Moreover, when traits of interest do not benefit individuals' reproductive success, population-level artificial selection may work against individual-level selection, which increases the difficulty of steering evolution. Additionally, conducting directed evolution experiments in the laboratory can be slow and labor intensive, making it difficult to evaluate and tune new approaches to artificial selection *in vitro*. To address these issues, we developed an agent-based model of directed evolution of microbes for evaluating which techniques from evolutionary computing might be most applicable in the laboratory.

In [14], we ask if artificial selection schemes developed for evolutionary computing might be useful for directing the evolution of microbial populations when selecting for multiple traits of interest: both for enhancing multiple traits in a single microbial strain and for producing a set diverse strains that specialize on different traits. To do so, we use our agent-based model of laboratory directed evolution wherein we evolve populations of self-replicating computer programs ("digital microbes") that perform computation that contributes either to the phenotype of the individual or the phenotype of the population. We evaluated how well three selection mechanisms from evolutionary computing (tournament, lexicase, and non-dominated elite selection) performed in a setting that mimics directed evolution on functions measurable at the population-level. Overall, we found that lexicase selection and non-dominated elite selection generally outperformed the selection schemes commonly applied to directed microbial evolution (elite and top-10% selection). In particular, our findings suggest that lexicase selection is a good candidate technique to transfer into the laboratory, especially when aiming to evolve a diverse set of specialist microbial populations. We also found that a selection scheme's performance in a conventional evolutionary computing context does not necessarily predict its performance in our model of laboratory directed microbial evolution, which indicates the value of more directly modeling laboratory setups for predicting which techniques will be effective.

2 RELATED WORK

For brevity, we limit this discussion of related work to relevant biological applications of evolutionary computing techniques. See [14, 22, 24] for an overview of current methods of directed evolution in the laboratory.

To the best of our knowledge, sophisticated methods of choosing progenitors from evolutionary computing have not been applied to directed microbial evolution. However, artificial selection techniques from evolutionary computing have been applied in a range of other biological applications. For example, multi-objective evolutionary algorithms have been applied to DNA sequence design [4, 20]; however, these applications are treated as computational optimization problems. A range of selection schemes from evolutionary computing have also been proposed for both biomolecule engineering [5, 9] and agricultural selective breeding, especially for scenarios where genetic data can be exploited [19]. For example, using an NK landscape model, O'Hagan et al. evaluated the potential of elite selection, tournament selection, fitness sharing, and two rule-based learning selection schemes for selective breeding applications [18]. Inspired by genetic algorithms, island model approaches [23] have been proposed for improving plant and animal breeding programs [19, 25], and Akdemir et al. applied multi-objective selection algorithms like non-dominated selection to plant and animal breeding [1]. In each of these applications, however, artificial selection acted at the level of *individuals* (*e.g.*, individual genetic sequences). Therefore, our work focuses on applying artificial selection at the population-level in order to test the applicability of evolutionary computing selection algorithms for directed microbial evolution.

3 MODEL OVERVIEW

Here, we provide an abbreviated overview of our model of laboratory directed evolution; a more detailed description can be found in [14]. Our model contains a population of populations (*i.e.*, a "metapopulation"). We initialize each population with a digital organism capable only of self-replication. After initialization, directed evolution proceeds in cycles. During a cycle, we allow each population to evolve for a "maturation period" that comprises a fixed number of time steps. After the maturation period, we then evaluate each population's performance on a set of objectives, and we select performant populations as "parental" populations. To create an "offspring" population, we use a random sample of digital organisms from the chosen parental population; in this work, we used 1% of the maximum population size.

Digital organisms are self-replicating computer programs that reproduce asexually by copying their genome instruction-by-instruction and then dividing. In addition to self-replicating, digital organisms can perform functions by acquiring inputs from the environment, performing computations on those inputs, and producing outputs. In this work, a population was evaluated based on the number of designated functions (18 possible) that digital organisms performed during the population's maturation period, just as we might screen for the production of different biomolecules in a laboratory population.

4 LEXICASE AND NON-DOMINATED ELITE SELECTION SHOW PROMISE FOR DIRECTED EVOLUTION

Using our model of laboratory directed evolution, we investigated if selection schemes from evolutionary computing might be useful for

directed evolution of microbes. Specifically, we compared two selection schemes used in directed evolution (elite and top-10% selection) with three other methods used in evolutionary computing (tournament, lexicase, and non-dominated elite selection). Additionally, we ran two controls that ignored population-level performance: random and no selection. We describe each of these artificial selection methods in [14].

For each selection scheme, we ran 50 independent replicates of our model of directed evolution for 2,000 cycles of population maturation, evaluation, and propagation. Within each replicate, the metapopulation comprised 96 populations (following the number of samples held by a standard microtiter plate used in laboratory experiments), each with a maximum carrying capacity of 1,000 digital organisms. After 2,000 cycles of directed evolution, we measured the task profiles of each population in the metapopulation. A population's task profile is the set of functions that individuals within that population are capable of performing during a maturation period. We define a *metapopulation's* task profile as the union of all population task profiles within the given metapopulation. We measured directed evolution success in two ways: "best-population task coverage" and "metapopulation task coverage". Best-population task coverage is the size of the largest population task profile in the metapopulation, and metapopulation task coverage is the size of the metapopulation's task profile.

Figures 1a and 1b show the best-population and metapopulation task coverages, respectively. All selection schemes resulted in greater single-population task coverage than both random and no selection controls (Bonferroni-corrected Wilcoxon rank-sum test, $p < 10^{-4}$). Metapopulation coverage under tournament selection was not significantly different than coverage under the no selection control, but all other selection schemes resulted in significantly better metapopulation coverage than both controls (Bonferroni-corrected Wilcoxon rank-sum, $p < 0.03$). Overall, lexicase and non-dominated elite selection scored the greatest population and metapopulation task coverage out of all selection schemes, and lexicase was the overall best selection scheme according to both metrics of performance.

While differences were significant on the best-population task coverage, they were not necessarily substantial. However, other measures had more substantial differences. Both multi-objective selection schemes—lexicase and non-dominated elite—had the greatest metapopulation task coverage (Figure 1b), and the greatest diversity of task profiles in the final metapopulations (Figure 1c; Bonferroni-corrected Wilcoxon rank-sum test, $p < 10^{-4}$). Lexicase selection in particular also had the greatest task profile *spread* (Figure 1d; Bonferroni-corrected Wilcoxon rank-sum test, $p < 10^{-4}$), which is a measure of *how* distinct task profiles are. Lexicase's ability to produce diverse metapopulations are consistent with previous results demonstrating that lexicase excels at maintaining diverse specialists [6, 10–12].

We hypothesized that lexicase and non-dominated elite selection's mechanisms for selecting different *types* of parental populations underpinned their improved performance over elite, top-10%, and tournament selection. This, however, is confounded by each selection scheme's varying capacity to select a greater number of different populations (regardless of differences in those selected). As such, we asked whether lexicase and non-dominated elite's success could be explained by a capacity to select a greater number of different parental populations. Elite selection selected exactly one population per cycle, top-10% selected 10, lexicase selected an average of 12, tournament selected an average of 50, and non-dominated elite selected an average of 83 different populations. Thus, we can rule out the number of populations selected per cycle as the sole explanation for lexicase selection's success; we argue that this, in combination with our diversity data, suggests that directed evolution practitioners should consider incorporating mechanisms for selecting phenotypically diverse parental populations into their artificial selection approaches.

These results are also informative when compared to our genetic programming control experiment (reported in [14]). While results across these contexts are not directly comparable, we argue that, taken together, our experiments suggest that steering evolution at the population-level is more challenging than steering at the individual-level. For example, across all treatments, no single population in our model of directed evolution performed all 18 population-level functions. Yet, after a similar number of organism-level generations (~55, 000), both elite and lexicase selection produced programs capable of all functions in a genetic programming context; even after only 2,000 generations (the number of rounds of artificial selection in our directed evolution experiments), we found that conventional genetic programming produced more performant programs than those evolved under our model of laboratory directed evolution (see [14]). We also observed differences in the rank order of selection schemes between experiments. For example, non-dominated elite selection performed poorly in a genetic programming context relative to the other non-control selection schemes; however, non-dominated elite outperformed all selection schemes except lexicase selection in our model of laboratory directed evolution. On its own, non-dominated elite's difference in performance is not surprising, as it is not conventionally used for evolving computer programs where evaluation criteria are evaluated on a pass-fail basis. More broadly, however, we argue that this result highlights modeling as an important intermediate step when evaluating which techniques from evolutionary computing are likely to be effective in a laboratory setting.

5 CONCLUSION

We see digital experiments like those reported here as a critical step for transferring techniques developed for evolutionary computing into the laboratory. Indeed, our results are currently informing the design of laboratory experiments that apply evolutionary computing techniques to the directed evolution of *E. coli*. Our model of directed microbial evolution provides a testbed for rigorously evaluating different artificial selection methods with different laboratory setups (*e.g.*, metapopulation size, maturation period, *etc.*) before embarking on costly or timing consuming laboratory experiments.

ACKNOWLEDGMENTS

This research was supported through computational resources provided by the University of Michigan's ARC and by Michigan State University's ICER. Additionally, this research was supported by the

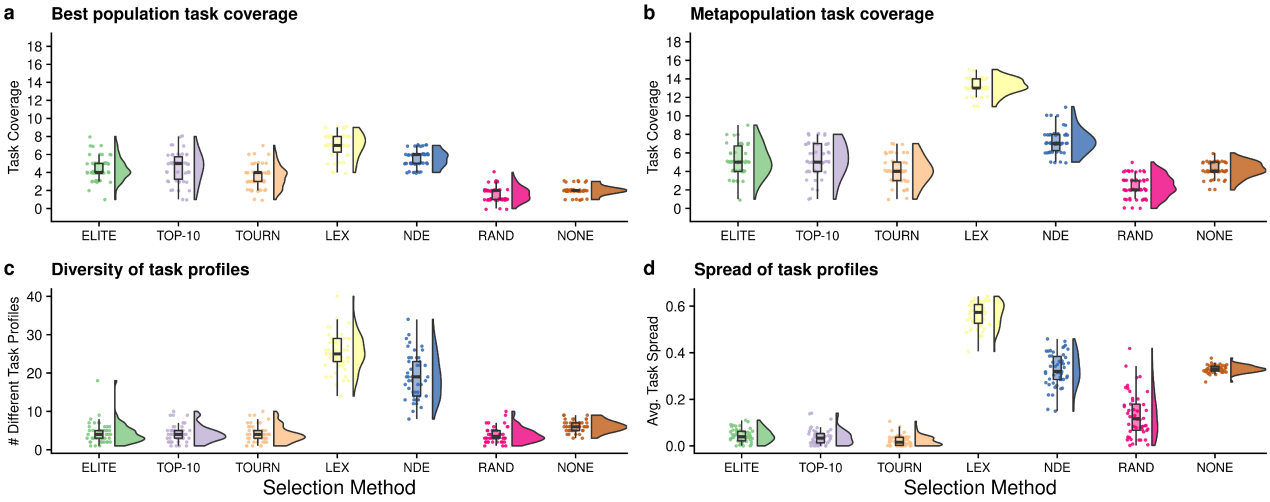

Figure 1: Digital directed evolution results. Differences among treatments were statistically significant for each panel (Kruskall-Wallis, $p < 10^{-4}$).

National Science Foundation (NSF) through support to LZ (DEB-1813069) and a sub-award to AEV (MCB-1750125).

REFERENCES

[1] Deniz Akdemir, William Beavis, Roberto Fritsche-Neto, Asheesh K. Singh, and Julio Isidro-Sánchez. 2019. Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity* 122, 5 (May 2019), 672–683. https://doi.org/10.1038/s41437-018-0147-1

[2] Frances H. Arnold. 1998. Design by Directed Evolution. *Accounts of Chemical Research* 31, 3 (March 1998), 125–131. https://doi.org/10.1021/ar960017f

[3] Thomas Bäck, D.B Fogel, and Z Michalewicz (Eds.). 1997. *Handbook of Evolutionary Computation* (1 ed.). https://doi.org/10.1201/9780367802486

[4] Jose M. Chaves-González. 2015. Hybrid multiobjective metaheuristics for the design of reliable DNA libraries. *Journal of Heuristics* 21, 6 (Dec. 2015), 751–788. https://doi.org/10.1007/s10732-015-9298-x

[5] Andrew Currin, Neil Swainston, Philip J. Day, and Douglas B. Kell. 2015. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society Reviews* 44, 5 (2015), 1172–1239. https://doi.org/10.1039/C4CS00351A

[6] Emily L Dolson, Wolfgang Banzhaf, and Charles Ofria. 2018. *Ecological theory provides insights about evolutionary computation.* preprint. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.27315v1

[7] Sherri Goings, Heather Goldsby, Betty H.C. Cheng, and Charles Ofria. 2012. An ecology-based evolutionary algorithm to evolve solutions to complex problems. In *Artificial Life 13*. MIT Press, 171–177. https://doi.org/10.7551/978-0-262-31050-5-ch024

[8] David E Goldberg, Jon Richardson, et al. 1987. Genetic algorithms with sharing for multimodal function optimization. In *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*. Hillsdale, NJ: Lawrence Erlbaum, 41–49.

[9] Julia Handl, Douglas B. Kell, and Joshua Knowles. 2007. Multiobjective Optimization in Bioinformatics and Computational Biology. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 4, 2 (April 2007), 279–292. https://doi.org/10.1109/TCBB.2007.070203

[10] Thomas Helmuth, Nicholas Freitag McPhee, and Lee Spector. 2016. Lexicase Selection for Program Synthesis: A Diversity Analysis. In *Genetic Programming Theory and Practice XIII*, Rick Riolo, W.P. Worzel, Mark Kotanchek, and Arthur Kordon (Eds.). Springer International Publishing, Cham, 151–167. https://doi.org/10.1007/978-3-319-34223-8_9

[11] Thomas Helmuth, Edward Pantridge, and Lee Spector. 2019. Lexicase selection of specialists. In *Proceedings of the Genetic and Evolutionary Computation Conference on - GECCO '19*. ACM Press, Prague, Czech Republic, 1030–1038. https://doi.org/10.1145/3321707.3321875

[12] Jose Guadalupe Hernandez, Alexander Lalejini, and Charles Ofria. 2022. *An Exploration of Exploration: Measuring the Ability of Lexicase Selection to Find Obscure Pathways to Optimality*. Springer Singapore, Singapore, 83–107. https://doi.org/10.1007/978-981-16-8113-4_5

[13] Gregory S. Hornby. 2006. ALPS: the age-layered population structure for reducing the problem of premature convergence. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation - GECCO '06*. ACM Press, Seattle, Washington, USA, 815. https://doi.org/10.1145/1143997.1144142

[14] Alexander Lalejini, Emily Dolson, Anya E. Vostinar, and Luis Zaman. 2022. Artificial selection methods from evolutionary computing show promise for directed evolution of microbes. *bioRxiv* (2022). https://doi.org/10.1101/2022.04.01.486727

[15] Joel Lehman and Kenneth O. Stanley. 2011. Abandoning Objectives: Evolution Through the Search for Novelty Alone. *Evolutionary Computation* 19, 2 (June 2011), 189–223. https://doi.org/10.1162/EVCO_a_00025

[16] Joel Lehman and Kenneth O. Stanley. 2011. Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation - GECCO '11*. ACM Press, Dublin, Ireland, 211. https://doi.org/10.1145/2001576.2001606

[17] Jean-Baptiste Mouret and Jeff Clune. 2015. Illuminating search spaces by mapping elites. *arXiv:1504.04909 [cs, q-bio]* (April 2015). http://arxiv.org/abs/1504.04909 arXiv: 1504.04909.

[18] Steve O'Hagan, Joshua Knowles, and Douglas B. Kell. 2012. Exploiting Genomic Knowledge in Optimising Molecular Breeding Programmes: Algorithms from Evolutionary Computing. *PLoS ONE* 7, 11 (Nov. 2012), e48862. https://doi.org/10.1371/journal.pone.0048862

[19] Vishnu Ramasubramanian and William D. Beavis. 2021. Strategies to Assure Optimal Trade-Offs Among Competing Objectives for the Genetic Improvement of Soybean. *Frontiers in Genetics* 12 (Sept. 2021), 675500. https://doi.org/10.3389/fgene.2021.675500

[20] S.-Y. Shin, I.-H. Lee, D. Kim, and B.-T. Zhang. 2005. Multiobjective Evolutionary Optimization of DNA Sequences for Reliable DNA Computing. *IEEE Transactions on Evolutionary Computation* 9, 2 (April 2005), 143–158. https://doi.org/10.1109/TEVC.2005.844166

[21] Lee Spector. 2012. Assessment of problem modality by differential performance of lexicase selection in genetic programming: a preliminary report. In *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference companion - GECCO Companion '12*. ACM Press, Philadelphia, Pennsylvania, USA, 401. https://doi.org/10.1145/2330784.2330846

[22] Álvaro Sánchez, Jean C.C. Vila, Chang-Yu Chang, Juan Diaz-Colunga, Sylvie Estrela, and María Rebolleda-Gomez. 2021. Directed Evolution of Microbial Communities. *Annual Review of Biophysics* 50, 1 (May 2021), annurev–biophys–101220–072829. https://doi.org/10.1146/annurev-biophys-101220-072829

[23] Reiko Tanese. 1989. *Distributed genetic algorithms for function optimization.* PhD Thesis. https://proxy.lib.umich.edu/login?url=https://www.proquest.com/dissertations-theses/distributed-genetic-algorithms-function/docview/303814520/se-2?accountid=14667

[24] Yajie Wang, Pu Xue, Mingfeng Cao, Tianhao Yu, Stephan T. Lane, and Huimin Zhao. 2021. Directed Evolution: Methodologies and Applications. *Chemical Reviews* (July 2021), acs.chemrev.1c00260. https://doi.org/10.1021/acs.chemrev.1c00260

[25] Shiori Yabe, Masanori Yamasaki, Kaworu Ebana, Takeshi Hayashi, and Hiroyoshi Iwata. 2016. Island-Model Genomic Selection for Long-Term Genetic Improvement of Autogamous Crops. *PLOS ONE* 11, 4 (April 2016), e0153945. https://doi.org/10.1371/journal.pone.0153945