

Artificial ecosystem selection for evolutionary optimisation

Hywel T. P. Williams and Timothy M. Lenton

University of East Anglia, Norwich, NR4 7TJ UK
{h.williams,t.lenton}@uea.ac.uk

Abstract. Artificial selection of microbial ecosystems for their collective function has been shown to be effective in laboratory experiments. In previous work, we used evolutionary simulation models to understand the mechanistic basis of the observed ecosystem-level response to artificial selection. Here we extend this work to consider artificial ecosystem selection as a method for evolutionary optimisation. By allowing solutions involving multiple species, artificial ecosystem selection adds a new class of multi-species solution to the available search space, while retaining all the single-species solutions achievable by lower-level selection methods. We explore the conditions where multi-species solutions (that necessitate higher-level selection) are likely to be found, and discuss the potential advantages of artificial ecosystem selection as an optimisation method.

Key words: artificial ecosystem selection — evolutionary optimisation

1 Introduction

In nature, there are many scenarios where multiple species combine to perform a function that a single species acting alone could not achieve. Many of these occur in the microbial world, at a variety of scales and with varying degrees of integration and cooperation [2]. Global chemical cycles typically involve different species at different stages in the cycle, such as photosynthesisers and respirers in the carbon cycle, or nitrogen-fixers and denitrifiers in the nitrogen cycle. At a smaller scale, biofilms (colonies of microorganisms growing in association on a surface) can also create nutrient cycling loops involving multiple species. Symbiotic associations such as lichens or mycorrhizae are tight, mutually beneficial linkages between species with complementary metabolic processes. These natural collective solutions to problems have been the inspiration for the engineering of ‘designer communities’ in microbial biotechnology, which have been usefully employed for purposes such as bioremediation of pollutants, treatment of sewage, and microbial leaching of metal ores [2]. However, the design of useful microbial communities is not straightforward and new methods are continually being sought.

Laboratory experiments with microbial communities have shown that artificial selection at the ecosystem level can cause a significant adaptive response

[10, 9]. In each experiment, an initial population of microbial ecosystem ‘units’ was created by random sampling of soil or pond water communities to inoculate containers holding neutral growth media. Each successive generation of ecosystems was then created by sampling from the best existing ecosystems (selected on various properties of their environment) to inoculate a new batch of sterile containers. In one set of experiments, pond water communities were selected for their collective effect on the pH level of their liquid environment [9]. In another, soil communities were selected for the growth they supported in above-ground plant biomass [9]. A third set of experiments selected pond water communities for their ability to break down the environmental pollutant 3-chloroaniline [10]. In all experiments, a statistically significant response to selection was observed, but it was unclear whether it resulted in each case from the indirect selection of a single dominant species (which could have been found by lower-level selection methods) or from genuine adaptation of community-level properties (which would necessitate higher-level selection).

Following on from prior population-based modelling of artificial ecosystem selection [4, 6, 5], in previous work [13] we used individual-based evolutionary simulation modelling to re-create the artificial ecosystem selection experiments reported in [10, 9]. We demonstrated a robust response to artificial ecosystem selection and were able to isolate the underlying causal mechanisms, allowing us to determine the level at which adaptation occurred. Here we consider the potential of artificial ecosystem selection as a method for evolutionary optimisation. Evolutionary optimisation methods have traditionally focused on scenarios where candidate solutions are encoded as a single string of values. This ‘single-species’ approach has been successful in many areas, but some harder problems may require more complex encodings. We propose artificial ecosystem selection as a method for designing ‘multi-species’ solutions, where several part-solutions operate in combination to achieve the desired function. We address two key questions. First, is artificial ecosystem selection effective in achieving an arbitrary target ecosystem function? This question has largely been answered in previous work, which we will briefly describe in the next section. Second, under what circumstances is artificial ecosystem selection of benefit over more traditional lower-level selection methods? We approach this question by looking for solutions created by artificial ecosystem selection that could not have been created by lower-level selection methods; we leave efficiency testing, benchmarking, and direct comparison with other optimisation algorithms, for future work.

2 Artificial Ecosystem Selection in the Flask Model

Here we give a brief overview of the ‘Flask’ model of microbial evolutionary ecology [12, 14] and its use in our previous artificial ecosystem selection experiments [13]. More detail and parameter values can be found in online supporting material for this article.¹

¹ <http://researchpages.net/media/resources/2007/04/16/ECALsupportingMethod.pdf>

2.1 The Flask Model

Flask [12, 14, 13] is an individual-based evolutionary simulation model of a flask containing a neutral liquid matrix in which is suspended a microbial population. The composition of the liquid medium determines the environment of the microbes. Some of the chemicals present are ‘nutrients’ that may be consumed as food and converted to biomass, while others are non-consumable and form part of the abiotic environment. The environment is assumed to have properties such as temperature, pH, salinity, etc., that both affect, and can be affected by, microbial activity. Non-consumable chemicals and physical properties of the flask environment are collectively referred to as ‘abiotic factors’, to distinguish them from nutrients. There is a flow of liquid medium through each flask which occurs continuously at a prescribed rate. The inflow brings with it influxes of nutrients at fixed concentrations and steady inputs to abiotic factors, while the outflow removes fixed proportions of stored nutrients and abiotic factors.

Microbes are modelled as simple organisms that consume and excrete nutrients, and affect the levels of abiotic factors in their environment as a by-product of metabolism. The precise ratios in which nutrients are consumed and excreted are genetically encoded for each individual, as are associated effects on abiotic factors, and preferred abiotic conditions (i.e., the state of the abiotic environment in which growth rate is maximised). Microbes grow by converting consumed nutrients to biomass and reproduce by splitting when their biomass reaches a fixed threshold. Mutation may occur during each reproduction event by selecting a new random allele with low probability P_{mut} at each locus, otherwise each offspring microbe receives an identical copy of the parental genotype. Biomass is reduced at a fixed rate to represent the inevitable thermodynamic inefficiency of metabolism and the cost of maintaining cellular machinery. Microbes die if their biomass drops below a fixed threshold, which can happen in sustained periods of nutrient limitation. They may also die ‘from natural causes’ with a low probability at each timestep.

The state of the liquid environment in a flask is a function of the parameterised fluxes of nutrients and abiotic factors, and the activity of the microbial community (i.e., the metabolism of nutrients and associated side-effects on abiotic factors). The shared environment creates individual-level selection pressure on metabolic requirements and environmental preferences, but the nature of this selection pressure changes over time as microbial activity alters the environment. Ecological and (co)evolutionary dynamics of these model ecosystems are discussed elsewhere [14]. Common phenomena include the emergence of nutrient recycling loops and coevolutionary feedbacks between different microbe species and their environment. (The term ‘species’ here refers to genetically identical individuals.)

2.2 Artificial Ecosystem Selection

The ‘phenotypic’ ecosystem trait used for artificial selection is based on the levels of the abiotic factors in the flask environment. Basing the fitness of flask

ecosystems on properties of the environment rather than the biotic population avoids any pre-specification of the type of population that will provide a good solution to the evolutionary problem. An arbitrary target state of the abiotic environment is assigned, with the deviation error of the actual abiotic state of a flask from this target constituting its performance score, $\Phi = \sqrt{\sum_{i=1}^A (\bar{a}_i - \hat{a}_i)^2}$, where \bar{a}_i is the target level for abiotic factor a_i , and \hat{a}_i is the actual level of a_i , in the normalised state vector for the A abiotic factors included in the model. Depending on the direction of artificial selection, the fitness of a flask ecosystem is based on maximising or minimising Φ . In each artificial selection experiment, three lines were selected based on the same initial random population: The ‘high’ line was selected to maximise Φ , the ‘low’ line was selected to minimise Φ , and the ‘random’ line – where the source ecosystem used to create the batch of ecosystems for each iteration was chosen at random – acted as a control. All lines consisted of a number of iterations of directed selection followed by an equivalent number of iterations of random selection, to allow study of the relaxation of the selected response.

Artificial ecosystem selection is an iterative process based on preferentially sampling from successive batches of flask ecosystems in order to create each succeeding batch. Each flask in the initial batch is seeded with a different inoculum, made up of randomly generated microbes. At each subsequent iteration of the selection process a new batch of flask ecosystems is created by inoculating sterile flasks with individuals from the fittest flasks of the previous iteration. A single inoculum of a fixed number of individuals is created by sampling at random from the source flasks, and identical copies of this inoculum are then used to seed the entire new batch of ecosystems. Two sampling methods are used: a ‘propagule’ method where the inoculum is drawn from a single source ecosystem, and a ‘migrant pool’ method where the inoculum takes individuals from several source ecosystems. The propagule method is analogous to asexual reproduction and should preserve ecological interactions between individuals. The migrant pool method is analogous to sexual reproduction and may better represent how new ecosystems form in nature. After inoculation, each ecosystem is propagated for a fixed period before Φ is measured.

2.3 Response to Selection

A similar response to artificial selection (e.g., Fig. 1(a)) is observed in all experiments performed. There is a rapid initial divergence in Φ values for the directed (high and low) selection lines from the control line. When directed selection is stopped and replaced by random selection (after iteration 30) the selected response in each line relaxes towards the non-selected control condition. The response to selection is robust to changes in the target environmental state and the ecosystem sampling method. Inverse relationships exist between the size of the response to selection and mutation rate P_{mut} (Fig. 1(b)), and between the size of response and ecosystem propagation time. Of the high-selected ecosystems, only 3.4% of cases showed adaptation at the community-level, but 36.5% of the

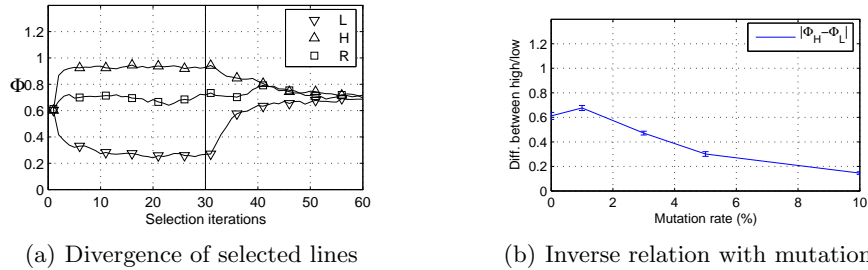


Fig. 1. Artificial ecosystem selection produces a strong adaptive response. An inverse relationship exists between the size of the response and microbe mutation rate. (a) Mean Φ from 49 runs with migrant pool sampling. Data plotted for directed selection for increase (high line) or decrease (low line) in distance from target environmental state, Φ , and for a random selection control line. Directed selection stopped at iteration 30, after which all selection is random. (b) Final deviation between high/low lines in $\{43, 57, 87, 42, 73\}$ runs with P_{mut} set at $\{0, 0.01, 0.03, 0.05, 0.1\}$ respectively.

low-selected ecosystems showed community-level adaptation. Further results and description can be found in [13].

3 Categorisation of Solutions from Artificial Ecosystem Selection

We now focus on showing that artificial ecosystem selection can create solutions that could not be created by standard genetic algorithms. Selecting at the ecosystem level creates the possibility for multiple species to be involved in the solution. Standard genetic algorithms select at the individual level, and even if we allow the case where an individual genotype is cloned in order to be tested in a group context, these algorithms therefore create solutions based on a single genotype. Thus we can reduce our task to categorising the solutions produced by artificial ecosystem selection into those that involve multiple species and those that involve only one species; the latter could have been created by lower-level selection, but the former require selection at a higher level.

The diverse initial seed and ongoing effects of mutation mean that multiple species are almost always present in each flask community. However, not all species in a selected community have a significant involvement in producing the target ecosystem function. If this function only involves a single active species, then any other species present are irrelevant to the response achieved and we have a single-species solution. If the function involves contributions from more than one active species, it is a multi-species solution.

A simple method for finding out if multiple species are involved in producing an observed ecosystem function is to test the converse proposition, that a single

species is alone responsible for that function. We do this for each selected ecosystem by testing each species in the community separately for two observations:

O1: A species exists within the selected community that performs as well or better than the intact community when propagated in isolation as a monoculture population.

O2: A species exists within the selected community that performs as well or better than the intact community when propagated as part of a non-selected wild-type community.

If neither of O1 or O2 is satisfied we can state that the artificially selected ecosystem solution necessarily involves multiple species and that the solution therefore could not have been found by lower-level selection methods. O2 is included to account for the scenario (observed in community selection experiments with flour beetles [1]) where a single species is responsible for an observed ecosystem property, but requires the presence of a non-specific background community for the property to be expressed; this does not count as a genuine multi-species solution since the identity of the background community is not significant. Note that this method is conservative, since satisfaction of O1 or O2 by a species does not necessarily imply that the species is responsible for overall community function.

3.1 Method

We tested O1 and O2 for 1064 artificially selected ecosystems (high and low lines from 532 runs) that were created in the experiments reported in [13]. The performance of each species from each artificially selected ecosystem was tested using a similar method to that used to allocate fitness during the artificial selection process. For O1, a sterile flask was seeded with a clonal inoculum of 100 individuals of the test species. For O2, a sterile flask was seeded with 25 individuals from the test species and 75 individuals randomly sampled from the control line community associated with the relevant artificial selection run. Each seeded flask was then allowed to develop for 2000 timesteps with mutation switched off, after which Φ was measured. The intact evolved community was tested similarly for comparison. When testing O2, and for the intact community, the mean score from 20 repetitions was used, to allow for stochastic effects. In verifying O1 and O2, high scores are ‘better’ in the high line, and low scores are ‘better’ in the low line.

3.2 Results

Overall, the majority (75%) of solutions can be explained by the presence of a single dominant species: 72% of evolved ecosystems satisfy O1, 47% satisfy O2, and 44% satisfy both O1 and O2. This leaves 25% of all the selected ecosystems that rely on contributions from several species. There is a clear divergence in the kinds of solution that are found for each line. Figure 2(a) shows Venn diagrams

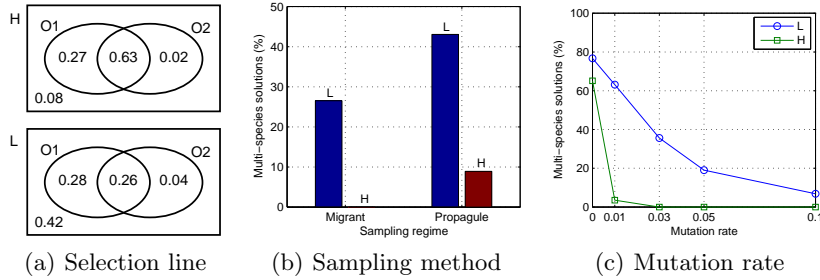


Fig. 2. Proportion of artificially selected ecosystems where multiple species are functionally involved in generating the selected function, found as the percentage of cases where neither O1 nor O2 is satisfied. Each subfigure shows the effect of varying a different parameter of the selection process: (a) selection line, (b) sampling method, (c) mutation rate.

that give the proportion of all the selected ecosystems in each line that satisfy O1 and O2. Multi-species solutions are less likely in high line ecosystems (8% cases, upper diagram) than in low line ecosystems (42% cases, lower diagram). Propagule sampling has a higher likelihood of producing a multi-species solution than migrant pool sampling (Fig. 2(b)). Of the 483 runs performed with the propagule method, 43% of low line and 9% of high line cases are multi-species solutions. Of the 49 runs performed with the migrant pool method (using 4 source ecosystems), 27% of low line and 0% of high line ecosystems were multi-species solutions. A clear inverse relation is observed between P_{mut} and the likelihood of a multi-species solution (Fig. 2(c)). Testing of evolved ecosystems from {43, 57, 87, 42, 73} runs with P_{mut} set at {0, 0.01, 0.03, 0.05, 0.1} respectively showed that when $P_{mut} = 0$ there is a high occurrence of multi-species solutions (77% of cases in the low line, 65% in the high line), but the proportion of multi-species solutions achieved falls dramatically as P_{mut} is increased.

3.3 Interpretation

Since Φ measures the distance from a target environmental state, we may *a priori* say that the low line presents a more difficult problem than the high line, since converging on a target is more difficult than diverging from it; there are many ways to be far from a point in multi-dimensional space, but only one way to hit it. Furthermore, in a complex, dynamic environment, holding an environmental variable close to a particular target level will often require correction in two directions. Since a single species can only push any environmental variable in one direction, at least two species may therefore be needed to provide the necessary opposing influences for the low line (target-seeking) problem. However, a variable can be moved away from a target by pushing in a single direction, so a single dominant species in the community may offer a good solution to the high

line (target-avoiding) problem. Thus the high and low lines offer qualitatively different evolutionary problems that demand qualitatively different ecosystem solutions, and this is reflected in the observed results.

The effect of different sampling methods on the likelihood of multi-species solutions can be understood by noting that the propagule method (which creates inocula using microbes drawn from a single parent ecosystem) transmits the parental community to the offspring ecosystem intact, and thus preserves the interactions between species. The migrant pool method (which creates inocula using microbes drawn from several parent ecosystems) mixes communities and breaks associations between species, thus hindering the formation of collaborative solutions.

At high mutation rates, very few multi-species solutions are found in either the high or low line ecosystems (e.g., 7% of low line and 0% of high line cases when $P_{mut} = 0.1$). This is likely to result from the disruptive effects that high mutation rates cause in any form of evolution; higher mutation rates make adaptation of individual phenotypes more difficult by introducing noise into the search process. In a multi-species community this disruption is amplified, because the interactions between species are also affected. High mutation rates add a large number of non-adaptive mutants to the community, which affect ecological dynamics even though they are likely to die out quickly without reproducing. The disruptive presence of these mutants makes it harder for beneficial associations to reliably form between species. Single-species solutions are not disrupted in this way and are therefore more robust. At the other end of the scale, the zero mutation scenario leads to a much higher incidence of multi-species solutions. With $P_{mut} = 0$, 77% of low line cases and 65% of high line cases were observed to be multi-species solutions, values much higher than the global average over all runs. There are two reasons for this. The first is that the lack of mutation makes ecological dynamics deterministic and stable, making it easier for useful associations between species to form. The second is that without mutation there is less chance that a single species will be found that can alone provide a good ecosystem-level solution, even for the high line problem. Without mutation the search algorithm is limited to those species that are present in the initial seed inoculum, and the indirect individual-level selection pressure applied by ecosystem-level selection cannot create a single ‘super’ species capable of achieving the target function.

4 Discussion

We have seen that artificial ecosystem selection can be effective in designing communities that collectively perform an arbitrary ecosystem function. It can (in theory) generate all the single-species solutions that are available to lower-level selection algorithms, but also allows a new class of multi-species solution which may be beneficial in some problems. Where the problem demands contributions from multiple species (e.g., the low line target-hitting problem) the evolved communities include several complementary species that together per-

form the function. In situations where the problem may be solved by a single species (e.g., the high line target-avoiding problem) the evolved communities reflect this. Where a single-species solution is possible, the search process will usually find it, since these solutions are more robust than multi-species solutions. Increasing mutation rate makes multi-species solutions less likely (Fig. 2(c)). There is a correlated decrease in the size of the response to artificial ecosystem selection (Fig. 1(b)), which may signify the occurrence of situations where multi-species solutions are required by the test problem, but prevented from forming by high mutation rates.

Artificial ecosystem selection is a novel method for evolutionary optimisation. The presence of multiple levels of selection and the isolation of the ecosystem units distinguish the artificial ecosystem selection method from superficially similar methods such as ‘island GAs’ [11]. Within the literature, artificial ecosystem selection is most closely related to work on cooperative coevolution. There are two main forms of cooperative coevolution algorithm in the literature, which differ mainly in how the population is structured. One kind coevolves part-solutions held in separate populations [7], while the other holds all part-solutions in the same population [8, 3]. Artificial ecosystem selection fits neither of these schemes; all species are held in the same community as in [8, 3], but the population of communities adds an additional level of structure. The multi-level population structure creates two levels of selection pressure. The lower level is within-ecosystem selection of microbes, and the higher level is between-ecosystem selection. The coevolutionary nature of the ‘inner GA’ generates new composite solutions that supply variety to the ‘outer GA’, which selects for those composite solutions that provide a good solution overall.

Potter et al [7] describe four issues which cooperative solutions must address: problem decomposition, sub-component interdependency, credit assignment, and diversity maintenance. With artificial ecosystem selection as described in this paper, these issues are addressed adaptively. The coevolutionary and ecological dynamics within each ecosystem prevent problems with sub-component interdependencies, since the persistence of a species in the community demonstrates its stable coexistence with other species. This criterion of survival in the community is the only form of species-level credit assignment. The interplay between internal coevolution and external ecosystem selection leads to adaptive problem decomposition; the niche structure that emerges from within-flask dynamics (steered by between-flask selection) maintains sufficient diversity to solve the problem, but no more. In the example given in this paper, single-species (single niche) solutions were successful in the high line, but multi-species (multiple niche) solutions were necessary in the low line. Within-ecosystem ecological dynamics prevent unnecessary redundancy, since similar species compete for the same ecological niche, while microbial evolution allows the species occupying each niche to adapt and potentially improve. The artificial ecosystem selection process thus allows for a wide variety of solutions with minimal user pre-specification, while also including mechanisms to ensure that the solution achieved is appropriate to the problem.

While the above discussion gives reasonable grounds for optimism, the application of artificial ecosystem selection as an optimisation method is unlikely to be straightforward and further theoretical elaboration will be required before the method can usefully be applied. In particular, one challenge will be in determining what kinds of problem are suitable; the ecological scenario described here is not directly representative of many types of optimisation problem. Artificial ecosystem selection may be most suited to the optimisation or design of ‘collective intelligence’ solutions such as swarms or artificial immune systems, where the community structure is already a feature of the problem. In these scenarios the survival criterion for species (part-solutions) to remain in the community (full solution) will depend on active involvement in solving the target problem; the ecological and coevolutionary selection pressures would be created by competition to contribute and do useful work. Niches would represent different ways of contributing to the whole solution. Future work will apply artificial ecosystem selection to a variety of test optimisation problems.

Acknowledgments. This work was funded by the Leverhulme Trust.

References

1. C.J. Goodnight. Experimental studies of community evolution 2: The ecological basis of the response to community selection. *Evolution*, 44:1625–1636, 1990.
2. M.T. Madigan and J.M. Martinko. *Brock Biology of Microorganisms (11th edition)*. Pearson Prentice Hall, New Jersey, 2006.
3. D.E. Moriarty and R. Miikkulainen. Forming neural networks through efficient and adaptive coevolution. *Evolutionary Computation*, 5(4):373–399, 1997.
4. A. Penn. Modelling artificial ecosystem selection: A preliminary investigation. In W. Banzhaf et al, eds., *Proc. 7th Eur. Conf. Artificial Life (ECAL 2003)*, pages 659–666. Springer, 2003.
5. A. Penn. *Ecosystem selection: Simulation, experiment and theory*. PhD thesis, University of Sussex, Brighton, UK, 2005.
6. A. Penn and I. Harvey. The role of non-genetic change in the heritability, variation, and response to selection of artificially selected ecosystems. In J. Pollack et al, eds., *Proc. 9th Int. Conf. Simulation and Synthesis of Living Systems (ALIFE 9)*, pages 352–357, Cambridge MA, 2004. MIT Press.
7. M.A. Potter and K.A. De Jong. Cooperative coevolution: An architecture for evolving coadapted subcomponents. *Evolutionary Computation*, 8(1):1–29, 2000.
8. R.E. Smith, S. Forrest, and A.S. Perelson. Searching for diverse, cooperative populations with genetic algorithms. *Evolutionary Computation*, 1(2):127–149, 1993.
9. W. Swenson, J. Arendt, and D.S. Wilson. Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation. *Environ Microbiol*, 2(5):564–571, 2000.
10. W. Swenson, D.S. Wilson, and R. Elias. Artificial ecosystem selection. *PNAS*, 97:9110–9114, 2000.
11. D. Whitley, S. Rana, and R.B. Heckendorn. The island model genetic algorithm: on separability, population size and convergence. *J. Comp. Inf. Tech.*, 7(1):33–47, 1999.
12. H. T. P. Williams. *Homeostatic Adaptive Networks*. PhD thesis, University of Leeds, Leeds, UK, 2006.
13. H. T. P. Williams and T.M. Lenton. Artificial selection of simulation microbial ecosystems. *PNAS*, 104(21), 2007.
14. H. T. P. Williams and T.M. Lenton. The Flask model: Emergence of nutrient-recycling microbial ecosystems and their disruption by environment-altering ‘rebel’ organisms. *Oikos*, In press, 2007.