

Genetic Algorithms are Suitable for Driving Microbial Ecosystems in Desirable Directions

Frederik P.J. Vandecasteele
University of Idaho
Environmental Biotechnology Institute
Moscow, ID 83844-1052, USA
1-208-885-9063
vand8442@uidaho.edu

Thomas F. Hess
University of Idaho
Dept. of Biol. & Agric. Engineering
Moscow, ID 83844-0904, USA
1-208-885-7461
tfhess@uidaho.edu

Ronald L. Crawford
University of Idaho
Environmental Biotechnology Institute
Moscow, ID 83844-1052, USA
1-208-885-6580
crawford@uidaho.edu

ABSTRACT

The behavior of natural, biological ecosystems is for a large part determined by environmental conditions. It should therefore be possible to experimentally manipulate such conditions to drive ecosystems in desirable directions. When a set of environmental conditions can be manipulated to be either present or absent, such an exercise becomes a typical combinatorial optimization problem, and one for which a genetic algorithm should be very suitable. In this work, four exhaustive experimental datasets were assembled, containing growth levels of different natural microbial ecosystem as influenced by all possible combinations of a set of five chemical supplements. The ability of a genetic algorithm to search these datasets for combinations of supplements driving the ecosystems to high levels of growth was compared to that of a random search, a local search, and a hill-climbing algorithm. The results indicate that a genetic algorithm is very suitable for driving microbial ecosystems in desirable directions, which opens opportunities for both applied and fundamental ecology.

Categories and Subject Descriptors

J.3 [Life and Medical Sciences] – *biology and genetics*; I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search – *heuristic methods*.

General Terms

Algorithms, Design, Experimentation, Measurement, Performance, Verification.

Keywords

Biological Application, Ecology, Ecosystem Design, Ecosystem Function, Genetic Algorithm, Microbial Ecology, Optimization.

1. INTRODUCTION

Ecology is the field of biology that studies the interrelationships of living organisms and their environment. Ecosystem dynamics and functions are influenced by environmental conditions, which can cause shifts in organism composition and activity. By purposefully manipulating environmental conditions, a desirable ecosystem function can be obtained. This is for example attempted in biostimulation, the operation of wastewater treatment

plants, and the application of prebiotics. In this work, a genetic algorithm was used to find the combinations of chemical supplements that would result in high levels of growth of natural microbial ecosystems. The performance of a genetic algorithm was compared to that of random search, local search, and hill-climbing, three alternative approaches a human experimenter unaware of genetic algorithms might use. The performance of the algorithms was compared on exhaustive growth datasets of four different microbial ecosystems and results showed that the genetic algorithm outperformed the other optimizers in ten out of twelve instances.

Genetic algorithms have been applied in an experimental microbial context before. They have been used to design mixed microbial cultures from sets of isolated strains [1, 2] and for the optimization of growth media for pure cultures [3]. Our work described here experimentally demonstrates that a genetic algorithm is suitable for driving real-world, undefined microbial ecosystems in desirable directions by manipulating environmental conditions.

2. MATERIALS AND METHODS

Undefined, natural microbial ecosystems were collected from four different sources: human saliva, wastewater treatment plant sludge, soil, and stream water. Five chemical supplements that could influence ecosystem dynamics were selected: glucose (8 g/L), NH₄Cl (10 g/L), HCl (0.0005 M), Na₂EDTA·2H₂O (0.05 g/L), and NaCl (20 g/L). Exhaustive datasets were obtained by growing the environmental samples in a microtiter plate at 30°C in a basal medium (1 g/liter tryptone, 0.5 g/liter yeast extract, and 10 g/liter NaCl) with all 32 combinations of the 5 supplements, in triplicate. Growth was observed by measuring optical density values at 15-minute intervals in an automated plate reader. The growth level of each culture was determined as the maximum optical density value observed in 48 hours.

Four optimization algorithms were run on these four exhaustive datasets to find supplement combinations resulting in high levels of ecosystem growth. Noise was introduced by randomly returning one of three replicate measures to an algorithm for each growth level query. Each algorithm was allowed 48 such queries, with replacement, and returned the average growth level (of three replicates) of the treatment corresponding to the single highest observed growth level. The random search algorithm randomly sampled 48 supplement combinations, with replacement. The local search algorithm repeatedly modified an initially random

Table 1. Statistical comparison of genetic algorithm performance with that of other optimizers

Environmental sample	p-value ^a		
	GA > random search	GA > local search	GA > hill-climbing
saliva	< 0.0001	< 0.0001	< 0.0001
sludge	< 0.0001	< 0.0001	< 0.0001
soil	< 0.0001	0.0130	< 0.0001
stream water	< 0.0001	> 0.9999	> 0.9999

^a significant p-values are in bold

solution by inverting each of the five supplements with a probability of 0.05, until a new solution different from the old one was obtained. The hill-climbing algorithm repeatedly created five new solutions from an initially random solution by sequentially inverting each of the five supplements. The new solution with the highest growth level replaced the old solution if the growth level was improved. The genetic algorithm had the following parameters: binary chromosomes of length five, population size eight, deterministic tournament selection of size two, one-point crossover with probability 0.90, mutation with a bit inversion probability of 0.05 per bit, and elitism of size one. The first population was initialized until no multiple copies of chromosomes were present. New generations were similarly generated until the same criterion was satisfied.

Each of the four optimization algorithms was run 5000 times on each of the four datasets and performances were compared using a permutation test of size 10000. Statistical significance was set at $\alpha = 0.05$ before Bonferroni correction.

3. RESULTS

Exhaustive experimental datasets were obtained that contained growth levels of microorganisms in saliva, sludge, soil, and stream water, as influenced by all possible combinations of five chemical supplements. A random search, local search, hill-climbing, and genetic algorithm were used to find the combination of supplements that would result in optimal growth levels described in the datasets of these different microbial ecosystems. Statistical comparison of 5000 runs of each algorithm revealed that in 10 out of 12 cases, the genetic algorithm outperformed the three competing optimizers (Table 1). The standard genetic algorithm was compared on the saliva dataset to three derivative algorithms that had the selection, crossover, and mutation operators disabled. The standard algorithm statistically outperformed each of the three derivatives. R^2 values were calculated for a linear regression between growth levels and total number of supplements present. The following values were obtained, for saliva, sludge, soil, and stream water, respectively: 0.05, 0.10, 0.26, and 0.57. The stream water sample had a much higher R^2 value than the other samples.

4. DISCUSSION

In this study, a genetic algorithm, random search, local search, and hill-climbing algorithm were used to find combinations of environmental conditions that would lead to optimal levels of growth in a number of natural microbial ecosystems. Statistical comparison of optimization performance revealed that in 10 out of

12 cases, the genetic algorithm outperformed the three competing algorithms. This result experimentally validates the idea that genetic algorithms are suitable for driving microbial ecosystems in desirable directions by optimizing environmental conditions. The three competing algorithms correspond to optimization methods a human experimenter unaware of genetic algorithms might use in the lab to optimize microbial systems: trying random guesses (random search), gradually modifying an existing solution (local search), and changing parameters one at a time (hill-climbing). Even though the genetic algorithm operated with a small population size and low number of generations, comparison of the algorithm to three derivatives demonstrated that selection, crossover, and mutation all contributed to the algorithm's operation. The genetic algorithm performed less well on the stream water dataset, which was the dataset that had the highest correlation between total number of supplements added and overall growth (lower total number of supplements corresponded to higher levels of growth). It is possible that the local search and hill-climbing algorithms exploited an inherent pattern of that dataset by gradually stripping candidate solutions of supplements. This particular result suggests that it may be possible to estimate optimization problem difficulty by a priori analyzing structural properties of the microbial system of interest.

The results obtained here hold promise from both an applied and a fundamental ecological perspective. From a practical point of view, a genetic algorithm could for example be used to design efficient biostimulation treatments, supplements for wastewater treatment plants, or enhanced prebiotics. In fundamental studies, a genetic algorithm could be used to better study ecological functions of interest by selectively inducing them in the lab in samples from a variety of origins. Also, a sample displaying an intermediate level of a function of interest could be driven in radically different functional directions using a genetic algorithm. Molecular genetic analysis of such parallel samples could then help elucidate functionally critical community dynamics or molecular mechanisms.

5. ACKNOWLEDGMENTS

This research was supported by a fellowship from the Inland Northwest Research Alliance (INRA) Subsurface Science Research Institute, which is funded by the Department of Energy under contract DE-FG07-02ID14277.

6. REFERENCES

- [1] Vandecasteele, F. P. J., Hess, T. F., and R. L. Crawford. 2004. Constructing microbial consortia with minimal growth using a genetic algorithm. *In: G.R. Raidl et al. (Eds.), Applications of evolutionary computing - EvoWorkshops 2004. Lecture Notes in Computer Science* Vol. 3005. Springer-Verlag, Berlin, Germany. 123-129.
- [2] Vandecasteele, F. P. J., Hess, T. F., and R. L. Crawford. 2006. (In press). Using genetic algorithms to optimize functions of microbial ecosystems. *In: Hurst, C. J., Crawford, R. L., Garland, J. L., Mills, A. L., Schmidt, S. K., and L. D. Stetzenbach (Eds.), Manual of environmental microbiology*, third edition. ASM Press. Washington, DC.
- [3] Weuster-Botz, D., and C. Wandrey. 1995. Medium optimization by genetic algorithm for continuous production of formate dehydrogenase. *Process Biochemistry*. 30(6):563-571.