A Correlated Fitness Landscape Describes Growth in Experimental Microbial Ecosystems: Initial Results

Frederik P.J. Vandecasteele[1,2], Thomas F. Hess[1], Ronald L. Crawford[2]

[1] Dept. of Biol. and Agric. Engineering, University of Idaho, Moscow, ID 83844-0904
[2] Environmental Biotechnology Institute, University of Idaho, Moscow, ID 83844-1052
{vand8442, tfhess, crawford}@uidaho.edu

Abstract. This paper describes experimental tests of the biological notion that correlated fitness landscapes underlie the functioning of ecosystems. We have demonstrated that a correlated fitness landscape topology describes overall growth in microbial ecosystems as a function of species composition. This confirms the intuitive notion that the more similar the structures of ecosystems are, the more similar their functions will be. These methods and results provide opportunities for better understanding how ecosystems function. They also provide a rationale for using evolutionary computation in optimizing microbial ecosystems.

1 Introduction

Fitness landscapes are a common conceptual framework in evolutionary computation. They provide a way of thinking about solutions to a problem as being points in a topological space. In this way, fitness landscapes represent the fitness of a particular solution and its relatedness to other solutions.

A common intuitive notion in ecology is that the more similar ecosystems are in composition, the more similar they will behave. In other words, the more similar the structures of ecosystems are, the more similar their functions will be. If this ecological notion holds true then it would mean that the behavior of ecosystems can be described by correlated fitness landscapes.

In this work, it is tested experimentally if correlated fitness landscapes underlie the functioning of microbial ecosystems. Species composition was chosen as a measure of ecosystem structure and overall growth as a measure of ecosystem function.

2 Materials and Methods

2.1 Isolation of Strains

Microbial isolates with morphologically distinct colony features were obtained by incubating a dilution of a surface layer soil sample on R2A agar and incubating the plates at different temperatures. From these isolates, 20 different fast growing strains were selected that had grown well in an overnight Luria Bertani (LB) broth at 37 °C, as judged by the turbidity of the culture medium. An equal volume of 40% glycerol was added to the cultures (final concentration 20% glycerol), after which they were aliquoted in cryovials and stored at -80 °C.

2.2 Ecosystem Assembly and Fitness Determination

Experimental ecosystems were constructed by combining particular subsets of a main set of N isolated microbial strains. For each assembled ecosystem, total growth after 24 h was determined.

Every ecosystem was represented as a binary string of length N, encoding for the presence or absence of the corresponding organism in the main set. The separate strains were diluted from their stock vials by adding 13.3 µL of each vial to 6 mL of LB. Each ecosystem was then assembled in a standard glass test tube by transferring 100 µL volumes of the right dilution tubes. After this, LB was added to each tube to a final volume of 2 mL. The test tubes were incubated at 37 °C and 200 rpm for 24 hours. Growth after 24 hours was assessed by transferring 200 µL of each test tube to a 96 well microtiter plate and measuring the optical density in each well using an automated spectrophotometric plate reader. Optical density (OD) (dimensionless units) is a measure of turbidity, which is proportional to the amount of microbial cells present in a sample and thus OD constitutes a measure of microbial growth.

3 Results and Discussion

3.1 Correlated Fitness Landscape Test

A set of 60 (= n) random bitstrings of length 20 was generated and the overall growth of the corresponding ecosystems was assessed as described above, with N = 20.
A test was developed to determine if a correlated fitness landscape was associated with this experimental dataset of random ecosystems. All possible unique pairwise combinations of the 60 data points were considered, amounting to a total of 1770 (= $(n^2 - n) / 2$). For each pair, two normalized distance metrics were calculated. The first

distance metric (d_1) was a measure of structural distance and described the dissimilarity between the species composition of two ecosystems. The second distance metric (d_2) was a measure of functional distance and described the dissimilarity in overall growth between two ecosystems:

$$d_1(i, j) = \frac{H(b_i, b_j)}{N}$$

$$d_2(i, j) = \frac{|OD_i - OD_j|}{OD_i + OD_j} .$$

with $H(b_i, b_j)$ the Hamming distance between bitstrings b_i and b_j
N length of the bitstrings
OD_i optical density measurement for ecosystem i

Linear regression was performed on the scatter plot resulting from these distance measures and a slope of 0.00844 was obtained (Figure 1). For visualization purposes, a moving averages plot was also generated by averaging the values of d_2 for each level of d_1 (Figure 2):

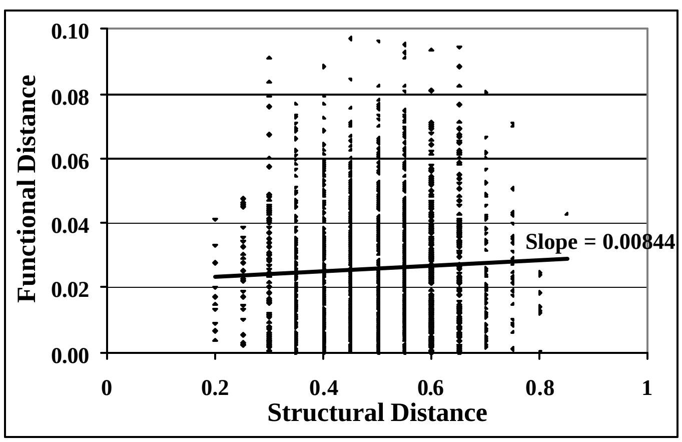

Fig. 1. Scatter plot of structural (d_1) versus functional (d_2) distance values of all possible pairs of ecosystems

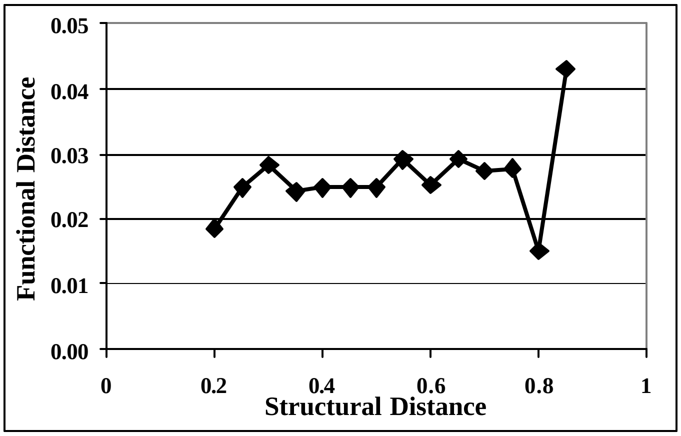

Fig. 2. Moving averages plot of structural (d_1) versus functional (d_2) distance values

To determine the statistical significance of the slope obtained, a randomization test was used [1], with the null hypothesis being that there is no correlation between d_1 and d_2. A random permutation procedure of size 10^5 was used on the list of pairs of d_1 and d_2 values. This resulted in a rejection of the null hypothesis with $p = 0.0327$. This indicated that there was indeed a correlation between d_1 and d_2, which means that a correlated fitness landscape topology underlies the behavior of the ecosystems. We have demonstrated for this set of 60 random ecosystems that overall, the more similar the ecosystems are with regards to structure (species composition), the more similar they are with regards to function (overall growth).

3.2 Estimating Required Sample Size for the Correlated Fitness Landscape Test

Experimentally assessing fitness values for sets of bitstrings is labor intensive. Therefore, it is important to know the minimal number of experiments required to be able to demonstrate a correlation between ecosystem composition and ecosystem function with statistical significance, for a given number of microbial strains N. In other words, what is the expected p-value for a correlation between structure and function for experimental datasets of various sizes? This question was addressed with the following test, presented in pseudo-code (with n = 60, the size of the random experimental dataset):

```
for k is 3 to n

   do 1000 times

      choose k data points from the set of n, without
      replacement
```

```
    perform correlated fitness landscape test on this
    subset, as described above

    determine significance of the correlation obtained
    for this subset (permutation, size 100)

  out of the 1000 correlation significance levels,
  report the value at the 95 percentile level
```

For a dataset of n points, this test 1000 times chose a random subset, for subset sizes of 3 through n. For each subset size level, a correlated fitness landscape analysis was performed on each of the 1000 random subsets and a significance level (p-value) for each of these 1000 correlations was determined by permutation. For each subset size level, the 95 percentile p-value was reported. This resulted in a 95 % upper limit estimate of p-values for correlated fitness landscape tests on sets of sizes 3 through n.

The result of this test obtained for the random dataset of size n = 60 described above is given in Figure 3:

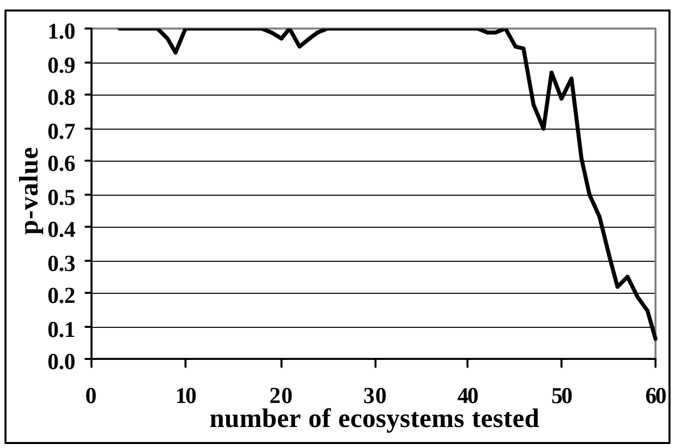

Fig. 3. Estimated 95% upper limits of p-values of structure-function correlation as a function of number of ecosystems tested

These results suggest that the current number of 60 ecosystems tested for a number of organisms of 20 could probably not have been lowered without losing significance on the observation.

3.3 Random Walk

In addition to the random dataset described above, a random walk experiment was performed to further investigate correlated fitness landscapes underlying the behavior of microbial ecosystems. A random bit string of length 20 was generated and a walk of 60 steps was determined by randomly flipping the value of one bit in the string per step. Total growth in each ecosystem was determined with all 60 experiments of the random walk performed in a single batch to avoid between batch variability. The result of this experiment is given in Figure 4:

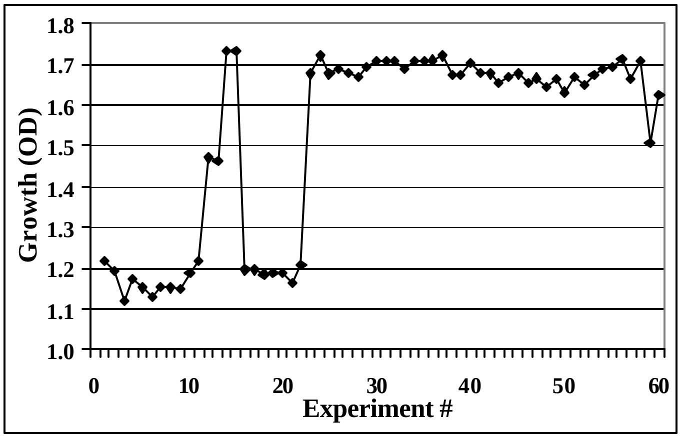

Fig. 4. Growth of mixed microbial ecosystems in a 60 step random walk

The correlated fitness landscape test was performed on this dataset and here too, a correlation between similarity in ecosystem composition and similarity in ecosystem function was demonstrated ($p<0.000001$).

3.4 Autocorrelation Function and Correlation Length

Using the data of the random walk, an autocorrelation function [2] was calculated. The autocorrelation function was estimated according to:

$$r_i = \frac{\sum_{t=1}^{T-i} \left(y_t - \overline{y}\right)\left(y_{t+i} - \overline{y}\right)}{\sum_{t=1}^{T} \left(y_t - \overline{y}\right)^2} .$$

with $\bar{y} = \frac{\sum_{t=1}^{T} y_t}{T}$ [3] and y the experimental measurement (OD) of overall growth.

The significance interval was calculated as $\left(-2/\sqrt{T}, +2/\sqrt{T}\right)$ with T = 60 [3]. The autocorrelation function is given in Figure 5:

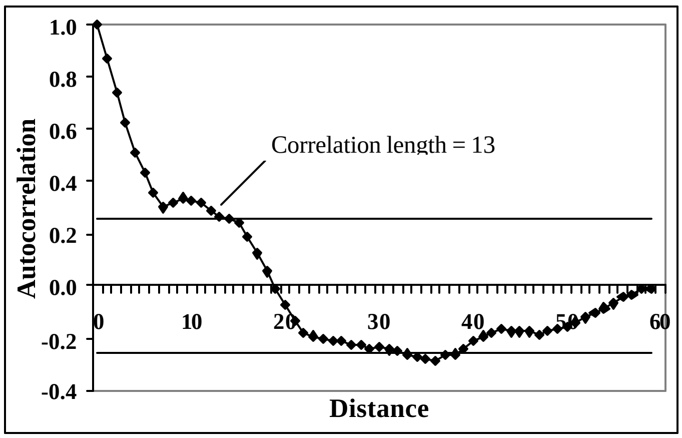

Fig. 5. Autocorrelation function of the 60 step random walk with confidence intervals

Based on the significance interval, a correlation length of 13 was estimated. Since this value is much larger than zero, it again indicates the presence of a correlated fitness landscape, in at least a section of the solution space.

4 Future Work

Further work should include performing correlated fitness landscape tests on more random sets of ecosystems, performing multiple random walks, determining statistical isotropy and possibly inferring properties of the entire solution space. This could result in opportunities for modeling and predicting ecosystem behavior. The approach described here should also be applied to various other ecosystems, composed of different organisms and performing different functions.

5 Conclusions

A common intuitive notion in ecology holds that the more similar the structures of ecosystems are, the more similar their functions will be. Here, this notion was tested and proven valid experimentally. A microbial system was set up with ecosystem structure determined by species composition and overall growth as a measure of ecosystem function. A new method was introduced and successfully used to demonstrate that a correlated fitness landscape topology underlies the behavior of this system. An initial attempt was made at estimating minimal dataset sizes for successfully applying this test. Finally, the autocorrelation function and correlation length of a section of the landscape were determined. These methods and results can lead to a better understanding of ecosystem functioning. In particular, correlation lengths and slopes of structure-function graphs could be used as a measure of relatedness between ecosystem structure and ecosystem function and as such provide a way of comparing that characteristic between various types of isolate sets or various types of environmental conditions. These results also provide a rationale for using evolutionary computation techniques to construct optimally functioning microbial ecosystems (see [4], [5], [6] and [7]) since such techniques can perform well on correlated fitness landscapes.

Acknowledgement

F.V. is 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.

References

1. Edington, E.S.: Randomization Tests. 3rd edn. Marcel Dekker, inc, New York (1995)
2. Weinberger, E.: Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference. Biological Cybernetics 63 (1990) 325-336
3. Hordijk, W.: A Measure of Landscapes. Evolutionary Computation 4(4) (1996) 335-360
4. Vandecasteele, F.P.J.: Constructing Efficient Microbial Consortia Using a Genetic Algorithm. Biological Applications for Genetic and Evolutionary Computation (BioGEC) – 2003 Genetic and Evolutionary Computation Conference Workshop Program (2003) 69-713
5. Vandecasteele, F.P.J., Hess, T.F., Crawford, R.L.: Constructing Microbial Consortia with Optimal Biomass Production Using a Genetic Algorithm. 2003 Genetic and Evolutionary Computation Conference – Late-Breaking Papers (2003) 299-302
6. Vandecasteele, F.P.J., Hess, T.F., Crawford, R.L.: Constructing Microbial Consortia with Minimal Growth Using a Genetic Algorithm. EvoBIO2004, 2nd European Workshop on

Evolutionary Bioinformatics, Lecture Notes in Computer Science, Springer-Verlag, Berlin Heidelberg New York (2004) 123-129

7. Vandecasteele, F.P.J., Hess, T.F., Crawford, R.L.: Thoughts on Using Evolutionary Computation to Assemble Efficient Ecosystems. Biological Applications for Genetic and Evolutionary Computation (BioGEC) – 2004 Genetic and Evolutionary Computation Conference Workshop Program (2004) in press