Techniques from Evolutionary Computation to Implement as Experimental Approaches in Synthetic Biology: Tests in Silico

Alexander Spirov
Institute of Scientific Information for Social Sciences RAS
Moscow
Russia
alexander.spirov@gmail.com

Ekaterina Myasnikova
Peter the Great St. Petersburg Polytechnical University
St. Petersburg
Russia
myasnikova@spbcas.ru

ABSTRACT

Evolution of biological macromolecules in a tube (in vitro evolution) of modern synthetic biology can be reasonably interpreted as the implementation of genetic algorithms (GA) in a biochemical experiment. This area of modern biology and bioengineering needs both new experimental approaches and new mathematical tools. In our report, we simulate how evolution occurs in vitro using the example of selection of RNA control devices (or RNA-based sensors). We demonstrate that heuristic recombination algorithms are significantly more efficient in a test tube evolution model than the standard mutation and crossover operators. We believe that the implementation of new biochemical methods, based on such heuristic algorithms, can significantly improve the efficiency of in vitro evolution.

CCS CONCEPTS

• Computing methodologies → Artificial life; Discrete-event simulation; • Applied computing → Synthetic and systems biology

KEYWORDS

Building blocks, crossover, heuristic recombination, evolution in test tube, biological macromolecules, RNA-devices, RNA-modules, modular design, approaches transfer

ACM Reference format:

A. Spirov, and E. Myasnikova. 2019. Techniques from Evolutionary Computation to Implement as Experimental Approaches in Synthetic Biology: Tests in Silico. In *Proceedings of ACM GECCO conference, Prague, Czech Republic, July 2019 (GECCO'19)*, 2 pages.
DOI: 10.1145/3319619.3326787

1 INTRODUCTION

Since the 60s of the XX century, the transfer of ideas from biology to applied mathematics has served to form the vast and thriving field of Evolutionary Computation, EC [1,2]. This is an impressive example of the formation of a new field at the intersection of two sciences.

At the last couple of decades a clear reverse transfer of concepts and approaches from EC to the synthetic and systems biology has become apparent (e.g. [3,4]). For the most part, it laid the foundation for modern areas of modeling the evolution of gene and protein networks and evolutionary design of genes, RNA devices and proteins. Moreover, the approaches of such a rapidly developing field of modern biology as the directed (forced) evolution of macromolecules can be reasonably treated as the experimental implementation of Genetic Algorithms (GA) [4].

In our report we consider the transfer of GA approaches to one of the promising areas of the modern synthetic biology and bioengineering. We will focus on the problems of directed evolution (evolution in silico) using an example of the engineering of artificial RNA molecules with the desired properties and functions. These are known as RNA devices (RNA control devices, RNA-based sensors) and destined for a great future in biotechnology and medical biology [5]. At the same time, this area suffers from a lack of effective mathematical and computer approaches necessary to reduce the cost of experimental molecular evolution and improve its efficiency [5,6]. This makes the purposeful transfer of algorithms and approaches from EC to this biological field extremely promising.

2 PROBLEM STATEMENT

Evolutionary design of RNA devices. RNA-devices are multimodular RNA molecules, including sensory and effector blocks, connected by transmitter modules. They are designed to be able to sense a highly specific signal from environment and turn on or switch the corresponding function [5,6]. RNA-devices are typically produced by the methods of in vitro evolution, followed by the in vivo evolution methods [5,6].

Modularity of RNA devices. Not only the methodology, but also the character of the selected RNA devices (their modularity) bring this experimental area closer to EC. The modularity is critical for

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07
https://doi.org/10.1145/3319619.3326787

the design of RNA devices and dictates the appropriate approaches to their development: the plug and play strategy [3,6].

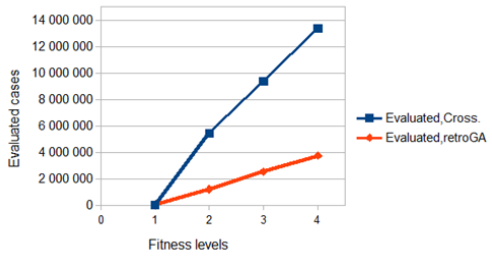

Figure 1: Average total number of fitness function evaluations to achieve nth fitness level for the four-modular RNA sequence: standard 2-point crossover + mutation vs retroGA operator. See text and Table 1 for details.

The modularity of RNA devices has direct analogies to the concept of Building Blocks (BB) in EC. In EC, we can find extensive studies of the role of BB in evolutionary search and ways of manipulating BB in order to improve the efficiency of evolutionary search (see e.g. [7]). It is reasonable to try to transfer and implement these approaches in these areas of synthetic biology and biotechnology (Cf [4,8]).

3 RESULTS AND DISCUSSION

We here demonstrate how the implementation of GA approaches to the BB preservation could significantly increase the efficiency of finding of RNA device sequences. Specifically, we test the retroGA approach [3,9], in comparison with the standard mutation operator, and standard crossover & mutation. retroGA operator generates a child string from a given "multi-parent set", combining the function of reproduction and crossover [3]. Crossover points are determined by regions of local homology in the parent strings.

Namely we consider the task to find a four-modular RNA device (RNA molecule with two sensors, effector and transmitter) from scratch. We begin with the initial population of random sequences, selected for the domain of the first sensor. Next, the population of the selected sequences is subjected to the selection for the second module – the second sensor. Then, the population is subjected to the selection for the third module – the effector-actuator. Finally, we select the obtained population of three-domain sequences for a fully functional device by finding the last domain – the transmitter.

For our tests, we performed a series of 100 runs with point mutation only (test 1), with standard single- and two-point crossover operators (test 2 and test 3, respectively), and with the retroGA operator (test 4). The sizes of populations (as well as other parameters) were chosen such as to achieve a high efficiency of evolutionary search (preferably 100%). For test 1, we had to stay on the population size of 480 thousand individuals, for test 2, 3, 160 thousand individuals, for test 4 the population size of 80 thousand was sufficient.

Table 2: The effectiveness of tests for different operators

Test	Efficacy,%	Evaluated cases
1	15	19364316±4746742
2	62	12376104±4851423
3	65	11768089±2807074
4	100	3845264±612328

It is noteworthy that the average total number of fitness function evaluations to achieve nth fitness level (nth module) for the four-modular RNA sequence rises nearly linearly (Fig.1).

As a result, we found that the recombination algorithm that preserve BB is significantly more efficient than the mutation only and more efficient than the standard mutation + crossover (Table 1). Namely, the retroGA procedure is approximately three times more efficient than the standard crossover operator (Table 1; Fig.1). retroGA finds the desired sequence of the four-modular device with high efficiency for a population of 80 thousand, while the standard GA shows low success rate even for population of 480 thousand (Table 1).

retroGA, along with other similar approaches, demonstrated its ability to preserve already found BB on well-known benchmark tests [3,9]. This property was explained by the fact that this operator performs recombinations only in sites of local homology of the parent sequences. It is for the tasks of searching for multi-modular RNA devices from scratch that the effectiveness of non-trivial heuristic recombination algorithms is most impressive.

ACKNOWLEDGMENTS

The research was supported by the Russian Science Foundation (grant 17-18-01536).

REFERENCES

[1] J. Holland. 1975. Adaptation in Natural and Artificial Systems MIT Press.
[2] J. R. Koza, F.H. Bennett, D. Andre, and M. A. Keane. 1999. Genetic Programming III: Darwinian Invention and Problem Solving. San Francisco, CA: Morgan Kaufmann.
[3] A. V. Spirov, and D. Holloway. 2012. New approaches to designing genes by evolution in the computer. Real-World Applications Of Genetic Algorithms, Olympia Roeva ed., IntechOpen, 235-260.
[4] C. A. Voigt, C. Martinez, Z. G. Wang, S. L. Mayo, and F. H. Arnold. 2002. Protein building blocks preserved by recombination. *Nat. Struct. Biol*. 9:553-558.
[5] C. M. Kim, and C. D. Smolke. 2017. Biomedical Applications of RNA-Based Devices. *Curr. Opin. Biomed. Eng*. 4:106-115.
[6] J. Chappell, K. E. Watters, and M. K. Takahashi. 2015. A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. *Curr. Opin. Chem. Biol*. 28:47–56.
[7] K. A. De Jong. 2006. Evolutionary Computation – A Unified Approach. MIT press.
[8] A. V. Spirov, and E. M. Myasnikova. 2018. Evolutionary Computations and Modular Organization of the Gene Regulatory Regions. Proceedings of 11th International Multiconference Bioinformatics of Genome Regulation and Structure \ Systems Biology, Novosibirsk, IEEE, 94-99.
[9] D. Holloway, and A. V. Spirov. 2011. Retroviral genetic algorithms: implementation with tags and validation against benchmark functions. Proceedings of The International Conference on Evolutionary Computation Theory and Applications, 233-238.