

SELVa: Simulator of Evolution with Landscape Variation

Elena Nabieva

*Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute),
Bolshoy Karetny pereulok 19, Moscow 127994, Russia, enabieva@gmail.com*

Georgii A. Bazykin

*Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute),
Bolshoy Karetny pereulok 19, Moscow 127994, Russia; Skolkovo Institute of Science and Technology, Skolkovo
Innovation Center, Building 3, Moscow 143026, Russia, gbazykin@iitp.ru*

Computer simulation methods have established themselves as invaluable tools in the study of molecular evolution. Since evolution is a tremendously complicated process, simulation methods vary in which of its aspects they can model and, therefore, which questions about evolutionary processes they can help answer (for a review, see Arenas 2012). An important feature of evolution is the change in fitness landscapes over time, whether reflecting environmental changes or, in the case of single-position fitness landscapes, epistatic interactions with other sites that are also undergoing change (Bazykin 2015). Studying the effect of changing fitness landscapes therefore calls for an evolutionary simulator capable of modeling landscape changes; to the best of our knowledge, this niche has not yet been filled. We present SELVa, the Simulator of Evolution with Landscape Variation, a novel simulator that allows the user to model molecular evolution with changing fitness landscapes. Written in Java and distributed as an executable *jar* file, SELVa provides a flexible framework that allows the user to choose from a number of implemented rules governing landscape change. In the course of one run, SELVa can simulate multiple parallel “landscape histories”, i.e., independent simulations governed by the same rules, with the option of multi-threading them. The general simulation framework of SELVa follows Chapter 9.5.1.3 of Yang 2006. The user provides the program with a rooted phylogenetic tree in Newick format and a configuration file detailing the options of the simulation. The simulation starts with an initial fitness landscape encoded as a fitness vector and either provided by the user in a file or generated from one of the supported probability distributions. As a first step, SELVa generates the root sequence by sampling from the stationary distribution corresponding to the

fitness vector (Yang 2006, Yang and Nielsen 2008). Thereupon, an event-based simulation takes place along the phylogenetic tree, where at any point, either of the following events can occur: a substitution in the sequence (governed by the current fitness landscape), or an instantaneous change in the fitness landscape. The mutation events are modeled as a Markov chain with inter-event times described by an exponential distribution with the mean parameter derived from the current fitness landscape; once a mutation event occurs, the new allele is chosen probabilistically according to the stationary distribution of the current fitness vector.

The rules governing the landscape change process can be specified along several dimensions: whether the landscape-change events occur deterministically at equal intervals or stochastically; whether the landscape history is the same for all branches that exist at a particular time point or are independent for “parallel” branches; and how the new landscape is selected. The values of the new fitness vector can be sampled from the same distribution as the initial fitness vector, obtained from the previous vector by randomly permuting its values, or derived from the previous vector by either increasing or decreasing the fitness of the current allele at the site. The last option permits the modeling of processes in which the current allele becomes more or less desirable with time, e.g., as a result of epistatic interactions with other sites. Other rules for selecting the new landscape are in the planning. Upon completion, SELVa prints out the values of the sequence(s) generated at each node of the phylogenetic tree in the course of the simulation, and, optionally, the times of landscape changes and the fitness vectors generated at those times, thus providing the user with detailed information about the course of the simulation.

With its flexibility in the rules governing landscape changes and the option of printing out detailed intermediate data, SELVa promises to be an invaluable tool for simulating evolutionary processes in which the change of fitness landscapes plays an important role. .

Acknowledgements:

The authors are grateful to Anastasia Neverova for invaluable comments and being the first user of SELVa and to Alexey Neverov for helpful discussions.

References:

1. M. Arenas (2012) Simulation of Molecular Data under Diverse Evolutionary Scenarios, *PLoS Comput Biol* **8**(5): e1002495 .
2. G.A. Bazykin (2015) Changing preferences: deformation of single position amino acid fitness landscapes and evolution of proteins, *Biol. Lett.* **11** 20150315.
3. Z. Yang Z and R. Nielsen (2008) Mutation-Selection Models of Codon Substitution and Their Use to Estimate Selective Strengths on Codon Usage, *Mol. Biol. Evol.* **25**(3):568–579. 2008
4. Z. Yang (2006) Computational Molecular Evolution. Oxford University Press 2006