

Analysis of neutral networks based on global and local sampling of RNA sequences

Sarah J. Berkemer^{1,2,+,*}, Kofi K. Khamit-Kush^{3,+}, Jacob P. Rutten^{4,+}, and Alice C. Schwarze^{5,+}

¹Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

²Bioinformatics Group, Department of Computer Science, Universität Leipzig, Germany

³Department of Biology, Clark Atlanta University, Atlanta, GA

⁴Theoretical Biology group, Utrecht University, Netherlands

⁵Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom

*bsarah@bioinf.uni-leipzig.de

+these authors contributed equally to this work

ABSTRACT

Fitness landscapes are used to explore the phenotype space based on fitness values. Neutral networks are used to explore fitness landscapes for RNA sequences (genotype) and corresponding secondary structures (phenotype). A neutral network is constructed based on genotypes showing the same phenotype. Sequence being the nodes of the network are connected by an edge if they differ by a single point mutation. Due to the exponential increase in sequence number, it is impossible to explore complete sequence spaces, even for relatively small sequences. Therefore, sampling of biological sequences based on a given secondary structure is used to explore parts of the space.

In this work, we use a new sampling approach where we first uniformly sample sequences from different distant parts of the sequence space using *RNAredPrint*. We then increase the size of the network by locally sampling sequences around each of the sequences sampled at first. In order to keep the network in a reasonable size, we additionally constrain the neutrality criterion such that not only the secondary structure but also the value of its folding energy is considered, using only the lowest 5% of folding energy values. This is done for three different type of structures, an artificial short hairpin structure, a Hammerhead ribozyme and a tRNA. We analyze the structure of the resulting networks and compare with already existing work. We are able to confirm earlier results and show that our sampling technique helps to explore distinct parts of the fitness landscape. We also show that local minima consist of more than one sequence and low folding energy sequences are spread throughout the network and separated by sequences with higher folding energy.

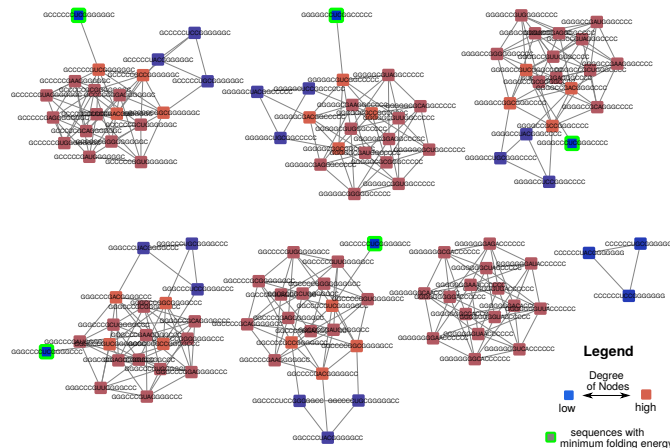


Figure 1. Neutral Network for the artificial small hairpin structure. Nodes represent sequences in the neutral network and are connected by an edge if sequences differ with Hamming distance 1. Node colors indicate the degree of a node which nicely shows strongly connected areas. Sequences folding with minimal folding energy are marked in green. Network visualized using Cytoscape 3.6.