Thesis Experimental Plan

6-13-19

Introduction

The general goal of these experiments is to examine the effects of evolution under different circumstances. Through the use of the *in silico* evolutionary tool **aevol**, we can simulate the evolution of artificial organisms over hundreds, thousands, or millions of generations, changing the evolutionary conditions at will. This allows us to change evolutionary properties such as the mutation, duplication, and rearrangement rates in a controlled manner. This control allows us to isolate one variable at a time through careful experimental design; by changing one factor at a time, we thus hope through simulated evolution to better understand some of the phenomena which occur in natural evolution which might lead to reduced genomes in the real world.

Experiment Preliminaries and Strategy Outline

1. Start with 3 populations, each being a clonal population of one of three wild types which have each evolved for 10 million generations. These are our *control populations*:
   a. control_wt_1_yyy
   b. control_wt_2_yyy
   c. control_wt_3_yyy

      Where yyy is the *run number*, i.e. how many iterations of this experiment have been run (see 3. below).

2. For each experiment, we begin with these control populations and clone them again to get *experimental populations*. With each *experimental population* we change one variable in **aevol** and let all populations (the control and experimental) run for another 100,000 generations. The *experimental populations* evolve with their new variable setting (e.g. point mutation rate increased) and the *control population* continues to evolve with the variable at its original value. At the end, the two population types are compared.

3. For each experiment, repeat 2) with different random seeds (***SEED***) for 100 runs to gain statistical significance.

The resulting populations are of the form exp_[variable]_[x]_[yyy]_[zzz] where [variable] is some variable being changed (e.g. point mutation rate, small insertion rate, etc.), and x ∈ {1, 2, 3} is one of the three wild type strains, [yyy] encodes the value of the parameter, and [zzz] encodes the run number (e.g. 001, 075, 100, etc.).

Experiment Group 1 – Mutation Rates

Experiment 1-1 – Point Mutation Rate

Parameter Changed: ***POINT_MUTATION_RATE***

Resulting Populations: exp_point_mutation_up_x_yyy, exp_point_mutation_down_x_yyy

Experiment 1-2 – Small Insertion Rate

Parameter Changed: ***SMALL_INSERTION_RATE***

Resulting Populations: exp_small_insertion_up_x_yyy, exp_small_insertion_down_x_yyy

Experiment 1-3 – Small Deletion Rate

Parameter Changed: ***SMALL_DELETION_RATE***

Resulting Populations: exp_small_deletion_up_x_yyy, exp_small_deletion_down_x_yyy

Experiment 1-4 – Maximum Insertion & Deletion Size

Parameter Changed: ***MAX_INDEL_SIZE***

Resulting Populations: exp_max_indel_up_x_yyy, exp_max_indel_down_x_yyy

Experiment Group 2 – Rearrangement Rates

Experiment 2-1 – Duplication Rate

Parameter Changed: ***DUPLICATION_RATE***

Resulting Populations: exp_duplication_up_x_yyy, exp_duplication_down_x_yyy

Experiment 2-2 – Deletion Rate

Parameter Changed: ***DELETION_RATE***

Resulting Populations: exp_deletion_up_x_yyy, exp_deletion_down_x_yyy

Experiment 2-3 – Translocation Rate

Parameter Changed: ***TRANSLOCATION_RATE***

Resulting Populations: exp_translocation_up_x_yyy, exp_translocation_down_x_yyy

Experiment 2-4 – Inversion Rate

Parameter Changed: ***INVERSION_RATE***

Resulting Populations: exp_inversion_up_x_yyy, exp_inversion_down_x_yyy

Analysis

After all runs of all experiments are completed, we will analyze the data to determine which conditions lead to a change in:

- overall fitness
    - aevol_misc_create_eps
- the number of genes (coding and non-coding)
    - aevol_misc_create_eps
- the average gene length
    - aevol_misc_extract
- Protein profile
    - aevol_misc_protein_map
- robustness
    - aevol_misc_ancestor_robustness