

Evolution through Programming

Lecturer: Prof. Yitzhak Pilpel

TA: Omer Kerner – omer.kerner@weizmann.ac.il

Assignment 4 – Fitness Landscapes and Sequence Spaces

CHOOSE ONE of the following questions.

Attach your well-documented simulation code (in any language of your choice) along with your report that includes visualizations, calculations, and explanations.

Question 1 - Quasispecies Simulation on Binary Genomes

Objective:

Using a binary genome of length L , define a fitness landscape in sequence space and simulate a quasispecies of size N evolving under the quasispecies equation.

Tasks:

- **Simulation:**
 - Represent each genome as a binary string of length L .
 - Define a fitness value for each sequence to create your fitness landscape.
 - Evolve the quasispecies population using the quasispecies equation, incorporating mutation events.
- **Calculations:**
 - Compute the average fitness of the population over time.
 - Show that the equilibrium state of the quasispecies does not necessarily correspond to the maximum possible average fitness.
 - Experiment with different mutation rates and analyze their impact on the equilibrium distribution.
- **Explanation:**
 - Write a brief discussion on how mutation rate influences selection and why the population may settle at an equilibrium that doesn't maximize average fitness.

Question 2 - Exploring Fitness Landscapes with the NK Model

Objective:

Create fitness landscapes for binary genomes using the NK model, where you can tune the ruggedness of the landscape. In this assignment, you will also explain the concept of epistasis and demonstrate how varying levels of gene interaction (epistasis) shape the landscape.

Tasks:

- **Simulation:**
 - **Implement the NK Model:**
 - Represent each genome as a binary string of length N .
 - Use the NK model framework: N is the number of genes, and K is the number of other genes each gene interacts with.
 - For each gene, assign a fitness contribution that depends on its state and the states of K other genes chosen at random.
 - Compute the overall fitness of a genome as the average (or sum) of all individual fitness contributions.
 - **Tunable Ruggedness:**
 - Vary the parameter K from 0 (no epistatic interactions, smooth landscape) to $N-1$ (maximal interactions, rugged landscape).
 - Generate and visualize different fitness landscapes (e.g., using heatmaps or 3D plots) for several values of K .
 - Analyze the number of local fitness maxima for each landscape to quantify ruggedness.
- **Calculations:**
 - Calculate the average fitness of randomly sampled genomes for different K values.
 - Analyze how the number and distribution of local optima change as K increases.
- **Explanation:**
 - **Define Epistasis:**
 - Shortly explain epistasis as the phenomenon where the effect of one gene mutation is influenced by the presence or absence of mutations in other genes. How does increased epistasis (higher K) lead to a more rugged landscape with many local peaks and valleys?