# Evolution through Programming

Lecturer: Prof. Yitzhak Pilpel  
TA: Omer Kerner – [omer.kerner@weizmann.ac.il](mailto:omer.kerner@weizmann.ac.il)

## Assignment 6 – Genetic Code

Choose ONE question to answer.

### Question 1 - Genetic-Code Robustness Landscape

**Goal**

Quantify how “special” the Standard Genetic Code (SGC) is by comparing it to randomly generated genetic codes.

**Tasks**

1. Generate 10,000 random genetic codes:
   * Choose a method of randomization. For example:
     1. Shuffle codon-to-amino acid assignments randomly while keeping stop codons fixed.
     2. Maintain the “box” structure and shuffle the amino acids between boxes.
   * Explain why you chose this specific randomization process. Is it what would be implied by a specific hypothesis (Frozen accident, Coevolution, etc.)?
2. Define an error-cost function:
   * Choose a distance matrix representing physiochemical differences between amino acids - [Amino acid replacement - Wikipedia](https://en.wikipedia.org/wiki/Amino_acid_replacement) (Grantham, Sneath, Epstein, Miyata, Experimental Exchangeability)
   * Explain how the distance matrix you chose was created and what it measures.
3. Analyze:
   * For each generated genetic code and for the SGC:
     1. For each codon, calculate the sum of distances to each of its one-base neighbors.
     2. Sum the distances over all codons to get a robustness score for the generated genome.
   * Plot the distribution of random-code costs as a histogram.
   * Clearly mark the SGC’s error-cost value on the histogram.
   * Determine the percentile ranking of the SGC.
   * Discuss how often you find codes as robust (low error-cost) or more robust than the SGC.

**Deliverables**

* Script/notebook containing visualizations and a brief report summarizing findings and analysis.

### Question 2 - Variant Code Robustness Comparison

**Goal**

Compare the robustness of real variant genetic codes to the Standard Genetic Code (SGC) by examining their local evolutionary landscapes.

**Tasks**

1. Hard-code two variant genetic codes:
   * Vertebrate mitochondrial code
   * CTG→Ser variant code in Candida
2. Compute error costs:
   * Apply the error-cost function described previously (Question 1) to these variant codes and the SGC.
3. Simulate evolutionary neighborhood:
   * Generate 1,000 random one-swap neighbors for each code (variant and SGC).
     1. One-swap means to swap the designated amino acids of 2 codons
   * Calculate error-costs for each neighbor.
4. Visualization:
   * Create boxplots to compare neighbor error-cost distributions for each genetic code.
5. Analysis:
   * Identify which genetic code resides on a "flatter" local optimum (i.e., more neighbor swaps have similar or improved robustness).

**Deliverables**

* Script/notebook with visualizations (boxplots and heatmap)
* Brief report summarizing key insights.