**User-Manual for the Sequence Evolver (SeqEvo) Computer Program**

# System Requirements

SeqEvo requires Java 8 or newer to be installed. Java can be downloaded for free (<https://www.java.com/en/download/>).

# Input Files

Running SeqEvo requires 4 input files: (1) A “fixed domains” file, (2) a “variable domains” file, (3) an “oligomers” file, and (4) a “parameters” file.

**Fixed domains file.** A file specifying the base-sequence of domains which will not be modified during optimization. By default, this file is named “se\_in\_domains\_fixed.txt” and is present in the directory where the program is run. A different file can be specified by providing a new value for IN\_FIXED\_DOMAINS in the parameters file. This file must contain one line per domain. Each line should contain (in order) a name for the domain, a tab or space character, and the base-sequence for the domain. Domain names may contain: (1) upper or lower case letters, (2) numbers, (3) dash (-), and (4) underscore(\_) characters. Base-sequences may contain upper or lower case a/t/c/g and should be specified starting from the 5’ end of the molecule.

**Variable domains file.** A file specifying the base-sequence of domains which will be modified during optimization. By default, this file is named “se\_in\_domains\_variable.txt” and is present in the directory where the program is run. A different file can be specified by providing a new value for IN\_VARIABLE\_DOMAINS in the parameters file. This file must contain one line per domain. Each line should contain (in order) a name for the domain, a tab or space character, and the base-sequence for the domain. Domain names may contain: (1) upper or lower case letters, (2) numbers, (3) dash (-), and (4) underscore(\_) characters. Base-sequences may contain upper or lower case a/t/c/g and should be specified starting from the 5’ end of the molecule.

**Oligomers file.** A file specifying the binding domains on each oligomer. By default, this file is named “se\_in\_oligomers.txt” and is present in the directory where the program is run. A different file can be specified by providing a new value for IN\_OLIGOMERS in the parameters file. This file should contain one line per oligomer. Each line should contain (in order) a name for the oligomer, a tab or space character, and a space-separated sequence of domains or domain/complements. Oligomer names may contain: (1) upper or lower case letters, (2) numbers, (3) dash (-), and (4) underscore(\_) characters. The sequence of domains on the oligomer should be specified starting at the 5’ end of the molecule. The sequence of domains may contain any domain specified in the domain files, or the binding complement (i.e., reverse complement) of a domain specified in the domain files. The binding complement of a named binding domain can be indicated using “c.” and the domain name.

**Parameters File.** A file specifying the runtime parameters for the program to use. By default, this file is named “se\_parameters.txt” and is present in the directory where the program is run. A different parameters file can be specified when running the SeqEvo command (i.e., “Java -jar SeqEvo.jar other\_parameters.txt”). The format of the parameters file should contain one line per parameter. Each line should contain (in order) a parameter name, a space or tab character, and a value for the parameter. Example parameters files can be found in the examples folder. An example parameters file containing all available parameters for the program can be created from SeqEvo by using the -ep option.

# Running SeqEvo

The SeqEvo jar file must be compiled before SeqEvo can be run. This can be done by following the instructions in the jar/README.txt file.

To run SeqEvo: (1) open a command line, (2) navigate to the directory containing the input files, and (3) execute the command “java -jar SeqEvo.jar”. If the jar file is not located in the current directory, you will need to specify the correct location for it. For example, when running the example in the 2007\_Zhang-et-al\_Autocatalytic directory, the command would be “java -jar ../../jar/SeqEvo.jar”.

# Available Parameters

The following parameters can be provided to SeqEvo in the parameters file.

### Input File Parameters

**IN\_FIXED\_DOMAINS** - Text file listing the fixed domains for the network. The default value is “se\_in\_domains\_fixed.txt”.

**IN\_VARIABLE\_DOMAINS** - Text file listing the variable domains for the network. The default value is “se\_in\_domains\_variable.txt”.

**IN\_OLIGOMERS** - Text file listing the oligomers for the network. The default value is “se\_in\_oligomers.txt”.

### Heuristic Parameters

**CPL** - Cycles-Per-Lineage. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 100,000.

**GPC** - Generations-Per-Cycle. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 1.

**NDPG** - New-Daughters-Per-Generation. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 1.

**NL** - Number-of-Lineages. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 8.

**NMPC** - New-Mothers-Per-Cycle. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 2.

**FITNESS\_SCORE** - Fitness score to optimize. The default and only accepted value is Wx.

**interSB** – The fitness points contributed by each inter-oligomer duplex are equal to this value raised to the length of the duplex. Must be an integer greater than or equal to 0 and less than 2147483647. The default value is 10.

**interSLC** - Inter-oligomer duplexes with base-pairs less than this value do not contribute to counts or scores. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 1.

**intraSB** – The fitness points contributed by each intra-oligomer duplex are equal to this value raised to the length of the duplex. Must be an integer greater than or equal to 0 and less than 2147483647. The default value is 10.

**intraSLC** - Intra-oligomer duplexes with base-pairs less than this value do not contribute to counts or scores. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 1.

**scoringWeightX** - Wx will be calculated as O times this value plus N. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 10,000.

**maxAA** - Maximum number of consecutive adenosine bases. Any stretch of bases greater than this number will make a network invalid. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 6.

**maxCC** - Maximum number of consecutive cytosine bases. Any stretch of bases greater than this number will make a network invalid. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 3.

**maxGG** - Maximum number of consecutive guanine bases. Any stretch of bases greater than this number will make a network invalid. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 3.

**maxTT** - Maximum number of consecutive thymine bases. Any stretch of bases greater than this number will make a network invalid. Must be an integer greater than or equal to 1 and less than 2147483647. The default value is 6.

### Output Parameters

**OUT\_FILE\_REPORT** - Text file detailing key results and parameters used. Value must be either “false” or end with “.txt”. The default value is “se\_out\_report.txt”.

**OUT\_FILE\_DOMAINS\_VARIABLE** - Text file listing the base-sequence of the variable domains following optimization. Value must be either “false” or end with “.txt”. The default value is “se\_out\_domains\_variable.txt”.

**OUT\_FILE\_OLIGOMERS** - Text file listing the base-sequence of the oligomers following optimization. Value must be either “false” or end with “.txt”. The default value is “se\_out\_oligomers.txt”.

**OUT\_FILE\_SCORES** - Text file listing the scores of the networks in each generation. Value must be either “false” or end with “.csv”. The default value is “se\_out\_score\_trajectories.csv”.

**OUT\_FILE\_LOG\_SCORES** - Text file listing the scores of the networks in logarithmically distributed generations. Value must be either “false” or end with “.csv”. The default value is “se\_out\_score\_trajectories\_log.csv”.

# Output Files

Running SeqEvo can produce the following output files.

1) Report File

2) Score Trajectory File

3) Output Network