### Visual Guide on Running SEQSIM on Windows System

### 1. Programs needed on your system:

- Microsoft Visual Studio or similar integrated development environment (IDE)
  - C++ Packages in the IDE

### 2. Preparation of Input File

- SEQSIM accepts ".txt" or ".CSV" formats.
- All sequences need to be the same nucleotide length. In our example, we used a length of 2000 nt.
- 1. SEQSIM will automatically omit any sequences that contain nucleotides other than ATCG from its analysis.

As a guideline, your file should look like the below screenshot. There may also be a tab before the start of each sequence. This makes no difference:

AATCCAAGGCGCTGACACCTGGAGCTACTGCTCGGTTCTCTGAGAGGTTGCAGCACCCCTGCAAACTGAGTCCAGATCTGGTAAGTCACCACCTCCTTAGGGTCATGCCCATCTGATCTGCAG ATAATTGGTTTTTGTGCCCTTAAATTATAGTTCATAGACTTGGTGGCATCGTGATTTTACATGTTAGGCCTTAGTTTTTACTAATACCTTAAAGATATTCCATAGGCCATTCTCAAGAGTGG AGTCTTTGTTCTGAACATGGGACCCATTTTTTTTCCCCTAGATTCCATCCCAAAGTGGGGTCATTGTAGAGTCAGTGTAAGAGCATCAACATTGGAACCTTCATCTGAGAGTTAGTGTTTTTATCCCCTAACTTTCACAGGCTCTAGAACACTGTCATGCTGTTTTTACAATTTCTTGTTACAATTTTTTCAAAATAAAAATAAAGGCGCTGATTCCAAGGAGTCCCTTTAGTCTTCTGCAAGCATGTC AATTGTGGGAACTGAGAATGTAGGCTGTGTGGGGCCACAGGACATTCTCGTTCCCATTGTTTTAGGGTGGTAAGTGACAAGAAATATTTTCTCAAAAAGGTGGAGCTTAGCTTTCAGGATCC AGAATTACAGATTTGTGTCCAGAAAGTGCAGAGTGGAATTGGGGTGAACTAATTACCTTTCCACCTCTACCAGAGCAATGACATTGGCACTAGGAGATGATGATGAAGTGATGTGATTTGCCGTA GTCATGTAATCTAGAGATCTTTCGAGCTGTGCTGGAGGGGCTTGATGCACTGCTTGCCCAGAAGGTTCGCCCAGGTGAGGTGACCCAGCTAACCAGGTGGGGAGGGCTCAGGCATCCAGGG TAGGGTTAGCTGGGTCAGAAGGAGTGGGAGGCCCAAGGGTGGCCCAGAGACTTCTGATGGGGCTGGGGAGAGGCTCAGAGAGGCCTTGGCCATTGTCCAGCTCCTCTGGGAAAGGACTGCT TTGCGGAATGTGGATGAGAACTTCTGGGGCATATGGTCTGGAGCTTCTGCACTCTCCCCAGAGGCCCCTGAGTAAGAGACGAACAGCAGGGAACTGTCCAAGGCCGGGTGGGCAGCAGCCCCTT GATGGTGATCCTAGACCTTTGCTTCAAGAATGGGACGCTGGATGAATG

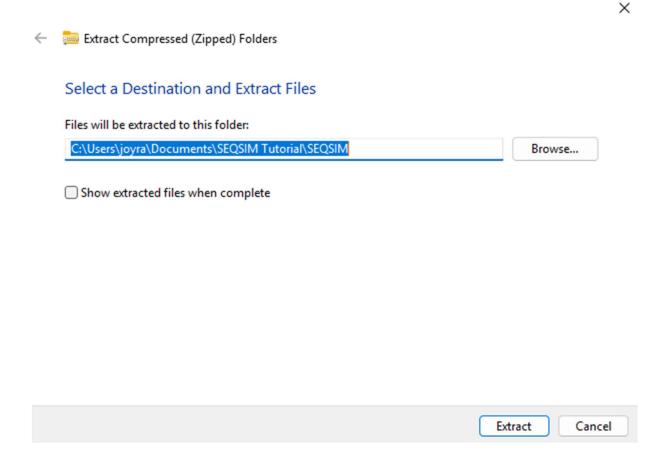
CCTCACCCACTTCTTAGAGTGGGGCAAGCAGAGAAAAGGCTTACTGCACGTGTGTTGCAAGGAGCTGCAGATTTTTGGAATAGCCATCCACAGGATCATAGAGGTCCTGAACACGGTGGAGC TAGACTGTATCCAGGAGGTGGAAGTGTGCTGCCCGTGGGAGCTGTCCATTCTTATAAGGTTCGCCCCTTACCTGGGCCAGATGAGGAATCTCCGCAAACTTGTTCTCTTCAACATCCATGTC TCTGCCTGCATTCCCCTAGACAGGAAGGAGCAGTTTGTCATCCAGTTCACCTCTCAGTTCCCTCAGCTGGACTACTTCCAGAAGCTTTACATGCACTCTGTCTCTTTTCCTCGAAGGCCACCTCATGCTAGGAAGGTACTGCAGGGTTTAGATCTAGTGAGGGTGCATTCGTGAATTCCTCCCGAGGATGTGTCTAAGTTAAGACGATGGGAAATGGGGAGTTGAAGAGGGCACTAAAGAAGA GAATGTCCATCACCCCCTCCATTTTAAAATATAAGGTCTGTCCTCACCTGCCTAGTGAACAGGTAAAATCCTGTGTCTCCCCTCTGTCAGTAAAATGTTTTTGAGCTCCAAAACATGATGGAA TAGAGGGTGAGGAGCAGGAGTGAAGAATGGTGAAAGTGATAGATGGTTTGCTGATGATACGGGCATGTCAGGGTCTCCTGCAGCCTGGCCACCCCAGCTGAGGTTGCAGGATCTTGCCTGG GTTTGTCATTTATGCCTGTGTCTCCATCGGGCTCCTGTGGCCCAGAGATGTGGTTTTCTACCTGACACATGAGGAAAGGGAGCATTAGAGTTCATGGACTTGACCCAGTCACCCGAGTGATG ACCTTGAGGTCTTCCCCACCACTCTCCTCTAACTCCTTCTTGTTCTCCCAGGTGTCTCCAGGCCCCCTTGGAGACAGTCGTAATGACCGAATGCCTGCTGTCAGAGTCGGACCTGAAGCA TCTCTCTTGGTGCCCGAGCATCCGTCAGCTAAAAGAGCTAGACCTGAGGGGCATCACACTGACCCATTTCAGTCCTGAGCCCCCTCTCAGTTCTGCTGGAGCAAGCTGAGGCCACCCTGCAGA CTGGAGAACCTGCTGCGCCACACCGTCGGGCTGAGCAAGCTAAGCCTGGAGCTGTATCCTGCCCCTCTGGAGAGTTATGATGCCCAGGGTGCTCTCTGCTGGGGAAGATTTTCTCAACTTGG TGAATGCCTGCTGTCAGGGTGGATTTATTTAAAGCTTTCTTCTGGTCATTTGGCAACTGAATCCTAGGCCATGAGTGTATGTCAAAGGGAGCACAGACCCATCGTTTCATATGCCTGCTCAA TGTGAACCGGAAAGGAAAGGGATGCAGGAAGGGAGGGACTGGGGGAA

AAGTTGAGTTGGAGTCAATAGGAGCTTTAGAGACCTGTGTCCCAGAGAATCAGAAATGGGAATCTGAATTGCTAGAATCAGGAATCAGGAGAGACACATGAGACAGTTACCCCTGCACG
GATGGTTGTAAAGAAACAGTCAGAAATAAAGGGAAGCTGAGTGGAAACTGTCTGGTGTCCTCCATGATTGCTTCACCTGGCTTAACAATTTTAACCTAAGGAATCTCAAGTTACTCATGGGG

# 3. Downloading and Extracting SEQSIM

As of February 2025, SEQSIM can be downloaded from https://sites.ualberta.ca/~joyramie/SEQSIM.html

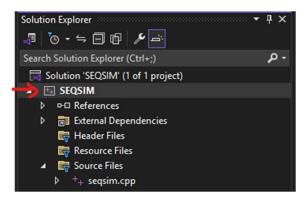
 Once downloaded, extract the SEQSIM folder into the directory of your choice by right-clicking the ZIP file and selecting "Extract All" and selecting your directory of choice. This will be where you are working.



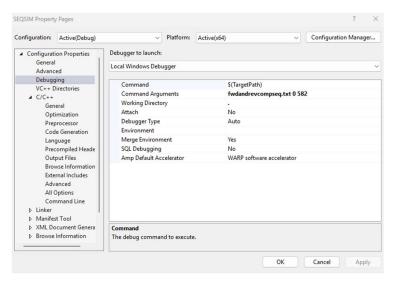
- 2. The extracted document contains 5 files:
  - a. Makefile
  - b. Promoters (Our SEQSIM Input)
  - c. README (Version 1, simpler)
  - d. SEQSIM.cpp
  - e. TopMatches Worksheet (Example of Post-Processing)

### 4. Setting up SEQSIM in Visual Studio

- 1. Open seqsim.cpp in Visual Studio. You can do this by right clicking the file and "Open With" or opening from within Visual Studio itself. Ensure you are working out of the same directory the input files are located.
- 2. Once SEQSIM is open in Visual Studio, right click on the entire program in the "Solution Explorer". Navigate to "Properties."



- 3. In the SEQSIM property pages. Navigate and change the below:
  - a. C/C++>Language>Open MP Support>"Yes (/openmp)
- 4. In the same properties window, navigate to Configuration Properties>Debugging>Command Arguments.
  - a. Here, you will type in 3 things separated by spaces:
    - i. The name of your input file within the directory. For example "fwdandrevcompseq.txt"
    - ii. 0 (the starting sequence index of your analysis)
    - iii. X (the ending sequence index of your analysis. For example, if I am analysing 583 sequences, 0 582 would encompass that range).



Press OK.

# 5. Running SEQSIM

Once your properties have been set up, click the green play button or CTRL+F5 to start the program.



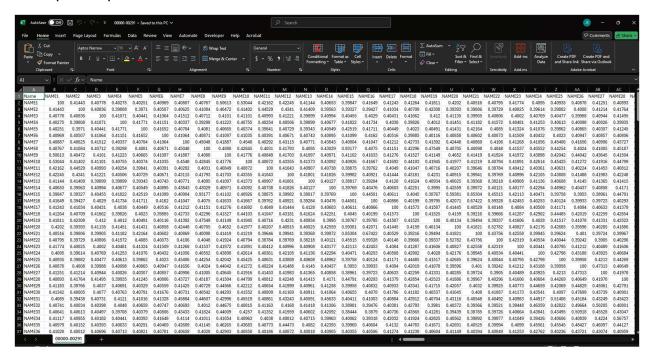
Example screenshot of SEQSIM running:

```
© C:\Users\joyra\source\repos\! ×
[WARN] Skipping NAME116 due to unrecognized letters outside of [ACGT]
[WARN] Skipping NAME91_1 due to unrecognized letters outside of
          Skipping NAME92_1 due to unrecognized letters outside of
                                                                                                      [ACGT]
[WARN] Skipping NAME93_1 due to unrecognized letters outside of
                                                                                                      [ACGT]
[WARN] Skipping NAME94_1 due to unrecognized letters outside of
                                                                                                      [ACGT]
[WARN] Skipping NAME95_1 due to unrecognized letters outside of
                                                                                                      [ACGT]
[WARN] Skipping NAME96_1 due to unrecognized letters outside of 
[WARN] Skipping NAME97_1 due to unrecognized letters outside of 
[WARN] Skipping NAME98_1 due to unrecognized letters outside of 
[WARN] Skipping NAME98_1 due to unrecognized letters outside of 
[WARN] Skipping NAME99_1 due to unrecognized letters outside of
                                                                                                      [ACGT]
                                                                                                      [ACGT]
                                                                                                      [ACGT]
                                                                                                     [ACGT]
[WARN] Skipping NAME100_1 due to unrecognized letters outside of 
[WARN] Skipping NAME101_1 due to unrecognized letters outside of
                                                                                                       [ACGT]
                                                                                                       [ACGT]
[WARN] Skipping NAME102_1 due to unrecognized letters outside of [WARN] Skipping NAME103_1 due to unrecognized letters outside of
                                                                                                       [ACGT]
                                                                                                       [ACGT]
[WARN] Skipping NAME104_1 due to unrecognized letters outside of 
[WARN] Skipping NAME105_1 due to unrecognized letters outside of
                                                                                                       [ACGT]
                                                                                                       [ACGT]
[WARN] Skipping NAME106_1 due to unrecognized letters outside of [WARN] Skipping NAME107_1 due to unrecognized letters outside of [WARN] Skipping NAME108_1 due to unrecognized letters outside of
                                                                                                       [ACGT]
                                                                                                       [ACGT]
                                                                                                       [ACGT]
                                                                                                       [ACGT]
[WARN] Skipping NAME109_1 due to unrecognized letters outside of
[WARN] Skipping NAME110_1 due to unrecognized letters outside of 
[WARN] Skipping NAME111_1 due to unrecognized letters outside of
                                                                                                       [ACGT]
                                                                                                       [ACGT]
[WARN] Skipping NAME112_1 due to unrecognized letters outside of [WARN] Skipping NAME113_1 due to unrecognized letters outside of
                                                                                                       [ACGT]
                                                                                                       [ACGT]
[WARN] Skipping NAME114_1 due to unrecognized letters outside of
                                                                                                       [ACGT]
[WARN] Skipping NAME115_1 due to unrecognized letters outside of
[WARN] Skipping NAME116_1 due to unrecognized letters outside of
Found 584 genes of length 2000
Writing results to 00000-00582.csv
```

Note, the sequences that are omitted from the analysis contain unrecognized letters outside of ACGT and are automatically removed.

Depending on your computer specifications, this analysis may take a significant amount of time. The output is a CSV file that can be opened in Excel at the end of the analysis. SEQSIM will exit on its own or prompt you to press enter to exit.

### **Example Output Screenshot:**



The names of your comparator sequences are found in Row 1 and Column 1. The intersection of each point is SEQSIM's similarity score.