Start Using SuiteMSA

Download programs:

1. All programs are available from **SuiteMSA** website:
   http://bioinfolab.unl.edu/~canderson/SuiteMSA/

2. For Mac OS X 10.5 or higher, download **"SuiteMSA_Package_forMacOSX.dmg"**. Double-click on the dmg file, and drag **"SuiteMSA_Package"** folder to your computer.

3. For other OS, you need to download each package separately.

   • Download **"SuiteMSA package"**, and double-click the zip file to uncompress.

   • Download **"indel-seq-gen-2.1.06.tar.gz"** or **"indel-seq-gen-2.1.06.zip"** file. Depending on the OS, you can double-click the file to uncompress or on Linux-like OS, type the following two commands:
   ```
   gunzip indel-seq-gen-2.1.06.tar.gz
   tar -xvf indel-seq-gen-2.1.06.tar
   ```

   • To use ClustalW2 with SuiteMSA GUI, you need to download **"ClustalW2 v2.1"** from ClustalW2 website (http://www.clustal.org/download/current/). Depending on the OS, follow their instructions. Note that you need to download ClustalW, not ClustalX.

   • To use MUSCLE with SuiteMSA GUI, you need to download **"MUSCLE v3.8"** from MUSCLE website (http://www.drive5.com/muscle/downloads.htm).

Set up each program (If you downloaded the **"SuiteMSA_Package_forMacOSX.dmg"**, you can skip these steps):

1. (Optional) Move iSGv2.1, ClustalW2, and MUSCLE folders to **"SuiteMSA_Package"** folder.

2. To set up iSGv2.1 in Linux-like environment (use Terminal in MacOS X), change the directory to **"indel-seq-gen-2.1.06"** directory and type the following two commands:
   ```
   ./configure
   make
   ```

   This should create **"indel-seq-gen"** executable file in **"src"** directory. Try to run it by typing:
   ```
   ./indel-seq-gen
   ```

   You should see the message:
   ```
   You must specify the substitution model (-m).
   ```

   If you see any error message, consult the iSG website for trouble-shooting (http://bioinfolab.unl.edu/~cstrope/iSG/).

   NOTE: If `./configure` does not succeed (you see error messages and indel-seq-gen executable cannot be generated), it is possible that you have file/directory names with irregular characters (', ", etc.). Please check the directory path and try to remove any irregular characters.

3. To confirm if clustalw2 is properly installed in Linux-like environment, change the directory to **"clustalw-2.1-macosx"** (or similarly named) directory and type the following command:
   ```
   ./clustalw2
   ```

   You should see the ClustalW menu. If you see any error message, consult the Clustal website for trouble-shooting (http://www.clustal.org/).

Do the same for MUSCLE by typing the name of the executable such as the following command:

```
./muscle3.8.31_i86darwin64
```

You should see the MUSCLE help. If you see any error message, consult the MUSCLE website (http://www.drive5.com/muscle/).

4. For SuiteMSA, nothing needs to be done. The executable is **"SuiteMSA-1.3.22B.jar"** located in **"SuiteMSA"** folder. Do not move or delete this and other files in this folder. You can create an alias of the jar file and put the alias wherever you like for your convenience.

Run SuiteMSA:

1. Run **SuiteMSA** by double-clicking on **SuiteMSA-1.3.22B.jar** in **"SuiteMSA"** folder.

2. The SuiteMSA main window lists the six tools available. Click on a button to start any tool.

3. **MSA Viewer**, **MSA Comparator**, and **Pixel Plot** require alignments in fasta format. **Phylogeny Viewer** requires a phylogeny in Newick format. **ClustalW2** and **MUSCLE** use sequences in fasta format.

Sample files (available in "sample" folder):

lipocalinSuperfamily_seqs.fasta (unaligned sequences in fasta format)
lipocalinSuperfamily_alignment.fasta (manually aligned sequences in fasta format)
lipocalinSuperfamilyPhylogeny.tree (a phylogeny in Newick format)
lipocalinSuperfamily.tree (a guide tree file used for iSGv2 simulation)
lipocalinSuperfamily.spec (a lineage specification file used for iSGv2 simulation)
lipocalinSuperfamily_template.maroot (a MSA root file used for iSGv2 simulation)
lipocalin.freq (an amino acid frequency file used for iSGv2 simulation)
lipocalin.idlen (an indel length-distribution file used for iSGv2 simulation)

MSAs generated by ClustalW2, Muscle, and MAFFT (in fasta format) are also available in this folder. More sample files are also available in **"additionalSamples"**.

Run iSG simulation:

* This brief tutorial uses the sample files listed above. See Anderson *et al.* (submitted) for the description of parameters included in these files.

1. Click on **"iSG Simulation"** button in the SuiteMSA main window. It brings up the main iSG simulation screen.

2. From the **File** menu, choose **"Set folder locations…"**.

  • **Input file folder:** Choose "**SuitMSA/sample**" folder where the sample files are located.

  • **Output file folder:** Automatically filled. You can change to a new location.

  • **Log file folder:** Automatically filled. You can change to a new location. Fill in (or choose) the log file name. This may be a new file, which will be created when this setting up is done.

  • **iSG executable folder:** Choose **"SuiteMSA_package/indel-seq-gen-2.1.06/src"** or wherever indel-seq-gen executable exists.

  Click on **"Set"** button.

3. Within **"Basic parameters"** panel:

- Click on **"Choose guide tree file…"** button and choose the **lipocalinSuperfamily.tree** file.

- From **Simulation options**, choose **"WAG"** from the **"Substitution model"** pop-up menu.

4. Within **"Advanced parameters"** panel:

  - Check **"Use lineage file"**, then click on **"Choose linage file..."** button. Choose the **lipocalinSuperfamily.spec** file.

  - Check **"Gamma-distribution rate heterogeneity"** and type in 3.88 for the alpha value.

  - Check **"Amino acid or nucleotide frequencies"**, and type in the amino acid frequencies as follows: A: 0.064, R: 0.046, N: 0.048, D: 0.058, C: 0.02, Q: 0.04, E: 0.079, G: 0.057, H:, 0.018, I: 0.047, L: 0.083, K: 0.072, M: 0.022, F: 0.045, P: 0.045, S: 0.064, T: 0.063, W: 0.016, Y: 0.05, V: 0.064.

5. The guide tree and lineage specification files contain all other necessary parameter information. **"Edit guide tree file"** and **"Edit lineage file"** panels show you the parameters specified in these files and any other support files.

6. Click on **"View/update command line"** button at the top to see the iSG command line.

7. Click on **"Run iSG"** button at the top to run the simulation. When the simulation is done, **"iSG simulation summary"** window comes up showing that the simulation is successful.

8. Click on **"Display MSA"** or **"Display event tree"** to examine the true MSA of simulated sequences or the phylogeny with indel-event mapping.

9. You can find all expected output files in the **"iSG_sample_output"** folder. To reproduce the same result, type in 404 for **"Random number seed"** in **"Advanced parameter"** panel.

Run ClustalW2:

* This brief tutorial uses the sample fasta sequence file available in the **"sample"** folder.

1. Click on **"MSA Reconstrucion"** button in the SuiteMSA main window and choose **"ClustalW2"**. It brings up the ClustalW2 screen.

2. From the **"Setup"** menu, click on **"ClustalW2 location"**. Navigate to the folder where the **clustalw2** executable is installed. Select this folder.

3. For **"Input sequence file"**, choose the **lipocalinSuperfamily_seqs.fasta** file in the **"sample"** folder.

4. Click on **"Run ClustalW2"** button at the top.

5. When alignment is done, **"Run complete"** message shows up at the top. Two default output files are: **lipocalinSuperfamily_seqs_CLTW2** (the alignment output in fasta format) and **lipocalinSuperfamily_seqs.dnd** (the guide tree used for progressive alignment)

6. From the **"Alignments"** menu, choose **"MSA Viewer", "MSA Comparator"**, or **"Pixel Plot"** to view the alignment or compare it with other alignments.

Run Muscle:

* This brief tutorial uses the sample fasta sequence file available in the **"sample"** folder.

1. Click on **"MSA Reconstrucion"** button in the SuiteMSA main window and choose **"Muscle"**. It brings up the Muscle screen.

2. From the **"Setup"** menu, click on **"Muscle executable"**. Navigate to the folder where the muscle executable is installed. Select this executable.

3. For **"Input sequence file"**, choose the **lipocalinSuperfamily_seqs.fasta** file in the **"sample"** folder.

4. Click on **"Run Muscle"** button at the top.

5. When alignment is done, **"Run complete"** message shows up at the top. The default output file is: **lipocalinSuperfamily_seqs_Muscle** (the alignment output in fasta format).

6. From the **"Alignments"** menu, choose **"MSA Viewer"**, **"MSA Comparator"**, or **"Pixel Plot"** to view the alignment or compare it with other alignments.

Use MSA Comparator or Pixel Plot:

* This brief tutorial uses the sample alignments available in the **"sample"** folder.

1. Click on **"MSA Comparator"** (or **"Pixcel Plot"**) button in the SuiteMSA main window. Choose **lipocalinSuperfamily_alignment.fasta** and **lipocalinSuperfamily_seqs_CLTW2**. In **Pixel Plot**, you can add up to four alignment (in fasta format) using the **"Add"** button.

2. To move the region to be compared, click on the scale bar for the reference alignment (MSA 1). To shift the region to be compared, use **">>"** or **"<<"** button.

3. To change the shifting step size or the length of the selection bar (blue bar), use the **"Display options"** menu at the top.

4. Click on **"Show consistency"** button (only in **MSA Comparator**) to show all positions consistent between the alignments in blue. Click on **"Show"** button to highlight only the selected region. The alignment positions in agreement (consistent) are colored in blue, or else in red.

Note on memory usage with SuiteMSA:

MSA Viewer and **MSA Comparator** use a large amount of memory when big alignments (e.g., > 2000 characters) need to be opened and compared. If you experience slowness or difficulty in opening a new window, you need to increase the memory allocated to the program.

• On Linux or Mac OS X (in Terminal), you can do this by using the following command line to start the program:

```
java –Xms350m –Xmx350m –jar SuiteMSA–1.1a.jar
```

This increases the program's memory (heap) size to 350mb. How much memory you can allocate to the program depends on how much free memory is available on your computer.

(last updated on February 17, 2013)