StatAlign Tutorial

November 16, 2012

This is a beginners tutorial on how to align RNA (or DNA) sequences and determine the consensus structure. First you have to download StatAlign. The current snapshot version is available from our download page as a zip archive. To run StatAlign, extract the archive, then double-click StatAlign.jar, or use the following command line syntax to get options:

java -jar StatAlign.jar -help

Once you start StatAlign, a window similar to this one should appear. If it doesn't you probably have to install java.



You start by opening a file which includes RNA sequences in a format similar to this one:

>gi|seq1

GCCTACGGCCATACCACCCTGAAAGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGGGACGGGCCTG GTTAGTACTTGGATGGGAGACCGCCTGGGAATACCTGGTGCTGTAGGCTTT

>gi|seq2

GCCTACGGCCATAGGGCCCTGAAAGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGGGACGGGCCTG GTTAGTAAATGGATGGGAGACCGCCTGGGAATACCTGGTGCTGTAGGCTTT

>gi|seq3

GCCGCGGGGCCATACCACCCTGAAAGCGCCCGATGCTGTCTGATCTCGGAAGCTAAGCAGGGACGGGCCTG GTTAGTACTTGGATGGCGAACCGCCTGGGAATCCATGGTGCTGTAGGCTTT

That can be done by clicking the folder icon which the red arrow points to.



Then you get a pop-up window similar to this one. First you browse for the sequence file, click on it (arrow 1) and then click on *Open* (arrow 2).

Eile MCMC Model Help	
E Sequences 🔀 Alignment 🔆 Tree 🗠 LogLikelihood 🧮 MPD 🏷 Consensus tree 😥 Network	
+ Open ×	
Look In: 🗖 statalign 🔹 🖉 🕼 🛱	
🖬 bin 🖾 src 🚺 LICENSE.txt	
ata Ss.fasta README.txt	
🖆 icons 🔷 5s.fasta.ctree	
lib S.fasta.log	
META-INF Ss.fasta.mpd	
res Ss.fasta.tree	
File Name: 5s.fasta	
Files of Type: All Files	
Open Cancel	
Remove	ove all
Phylogeny Cafe -: Statistical Alignment	

If everything went well you should now see your RNA sequences (green arrow). Next step is to enable the RNA mode. That can be done by clicking the RNA icon (red arrow).



You should now see a pop-up window where you can choose between folding methods. In this tutorial we are just going to use PPFold so you can simply click the *OK* button.

<u>File MCMC Model H</u> elp		
	LogLikelihood 🔠 MPD 🦄 Consensus tree 🔀 Network	
> gi 284925141 ref NR_033271.1 _Xenopus_{Silut GCCTACGGCCATACCACCCTGARAGCGCCCGATCTCGTCTGATCTCGG GGCTTT	ana}_tropicalis_somatic_55_ribosomal_RNA_{LOC100335112},_ribosomal_RNA aascräascrassergesessergestastastastastastastastastastastastastas	
> ailsea2		
GCCTACGGCCATAGGGCCCTGAAAGCGCCCGATCTCGTCTGATCTCGG GGCTTT	▼ RNA options ×	
> qijseq3	Perform sampling and averaging prediction (PPfold).	
GCCGCGGGCCATACCACCCCGAAAGCGCCCGATGCTGTCTGATCTCGG GGCTTT	i 🗌 Perform sampling and averaging prediction (RNAalifold).	
	RNAalifold settings	
	RNAalifold path Browse	
	Temperature (celsius) 37	
	Conformation	
	Covariance term 1	
	Non-compatible penalty	
	Use defaults OK Cancel	
		1
Remove		Remove all

Now you should see 4 new tabs which indicate that RNA mode has been activated. To align and fold the sequences you press the play button which the red arrow points to.

-	StatAlign v2.0	- + ×
<u>File MCMC Model H</u> elp		
🔆 Conservus tree 🛛 🔁 Networ	k 🔚 Base-pairing matrix 📎 Consensus Structure 🖓 Entropy 🖓 Similarity	
📃 📜 quences	Alignment	MPD
GGCTTT	opus {Silurana} tropicalis somatic 55 ribosomal RNA {LOC100335112}, ribosomal RNA reneRterceGRAGCTAACCAGEGREGEGECTEGETARTACTEGEAGEAGECECCEGGEARTRCCEGETECTETA	
GGCTTT	TCTGRTCTCGGRRGCTRGCRGGGRCGGGGCCTGGTTRGTRRTGGRTGG	
> gi seq3 GCCGCGGGGCCATACCACCCTGAAAGCGCCCCGATGCTG GGCTTT	TCTGRTCTC6GRA6CTAB6CR66GRC66GCCT6GTTA6TRCTT6GRT66CGARCC6CCT66GARCCAT6GT6CT6TA	
Remove		Remove all
Phylogeny Cafe :: Statistical Alignme	nt	

If that went well you should see "*Burn-in:* some number which increases rapidly" in the lower left corner. To see the consensus structure you can click on the tab which the red arrow points to.



When the burn-in has finished the folding starts and you should see the current consensus structure from the samples we have already taken . The number of samples taken can be seen in the lower left corner.



This is the get started tutorial. For a more detailed usage of StatAlign, please visit this website.