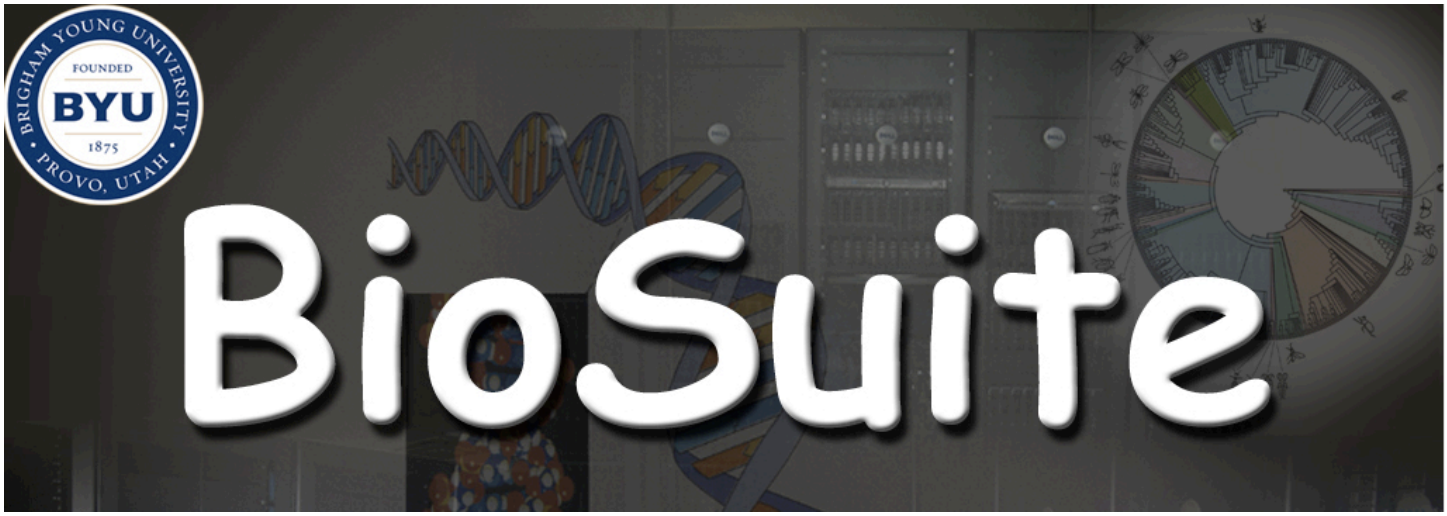


Tutorial

Version 1.0

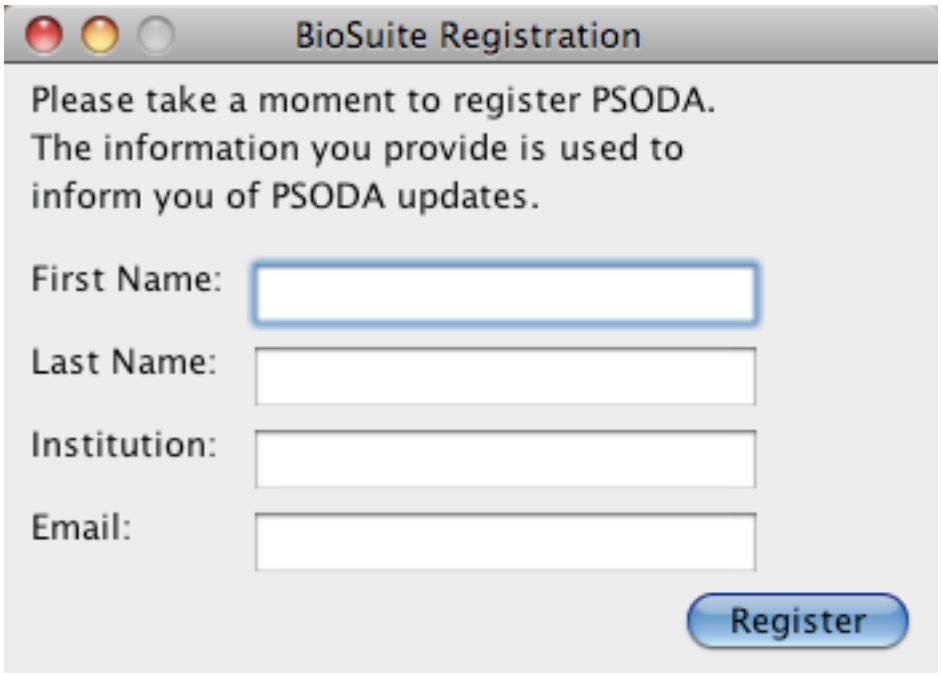
This tutorial will guide you through using BioSuite and the PSODA and TreeSAAP tools. The Phylogenetic Search Open Source Data Analysis (PSODA) tool provides sequence alignment and phylogenetic search using parsimony, likelihood and Bayesian methods. TreeSAAP locates selection sites in sequence data.

To get started with this tutorial, you should download biosuite from http://dna.cs.byu.edu/biosuite

You will find installation instructions there for Windows XP, Macintosh OSX and Linux.

This tutorial will guide you through the alignment, phylogenetic search and selection analysis of sample data.

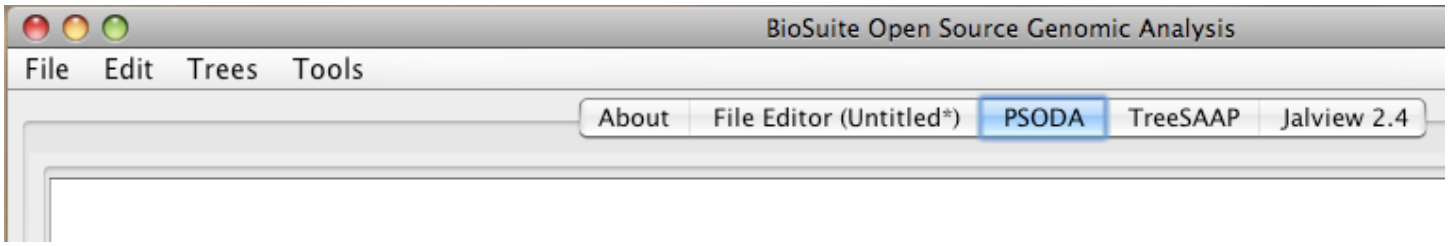
When you start BioSuite, you will see a registration screen. Filling out the information will allow the BioSuite team to contact you when updates are available. You will only have to register the first time you run PSODA.



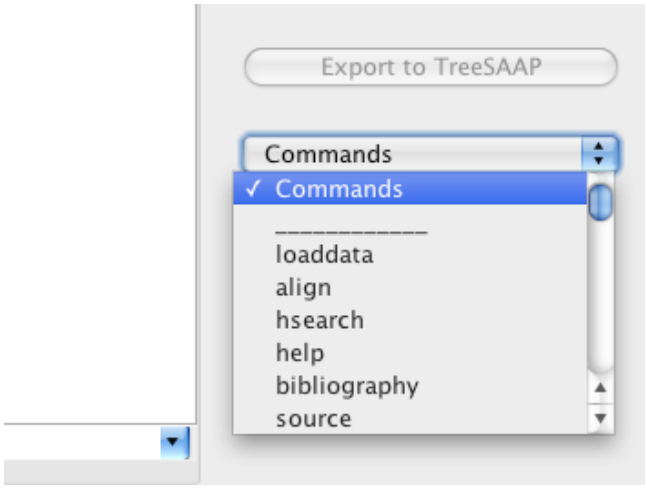
To start the analysis, download the sample data from
http://dna.cs.byu.edu/biosuite/tutorial/alignHIV.fasta

This is some sample HIV data in fasta format. Since we need to perform phylogenetic analysis before doing selection analysis, we will start with PSODA.

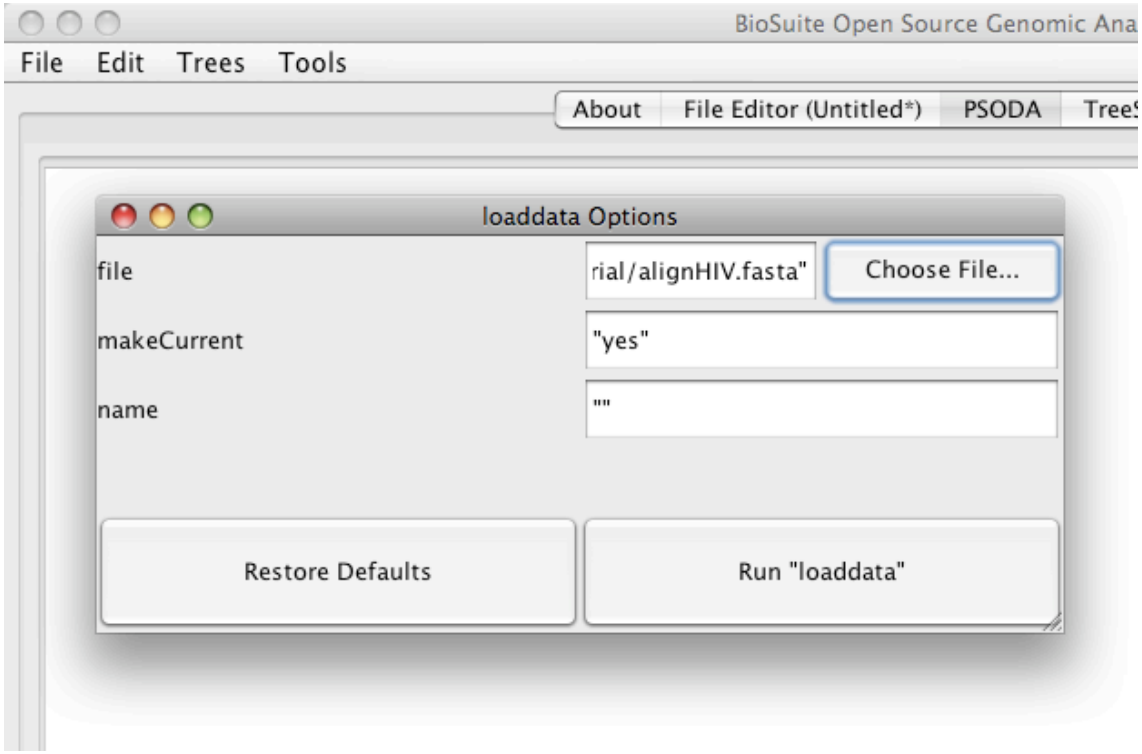
Select the PSODA tab.



Select the loaddata option under PSODA to load the fasta data.

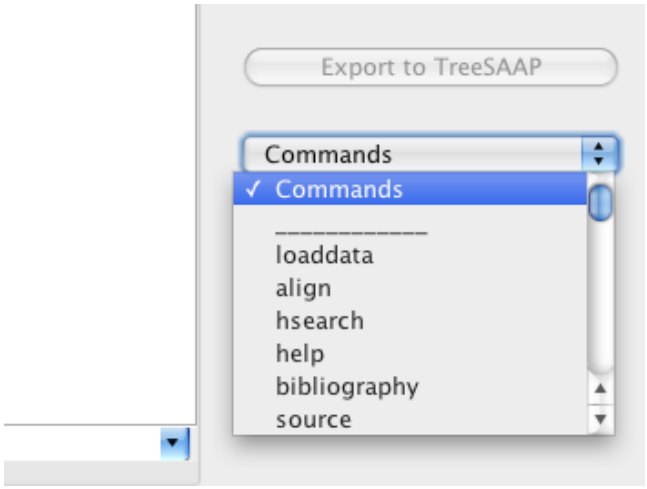


The loaddata command will bring up the following window

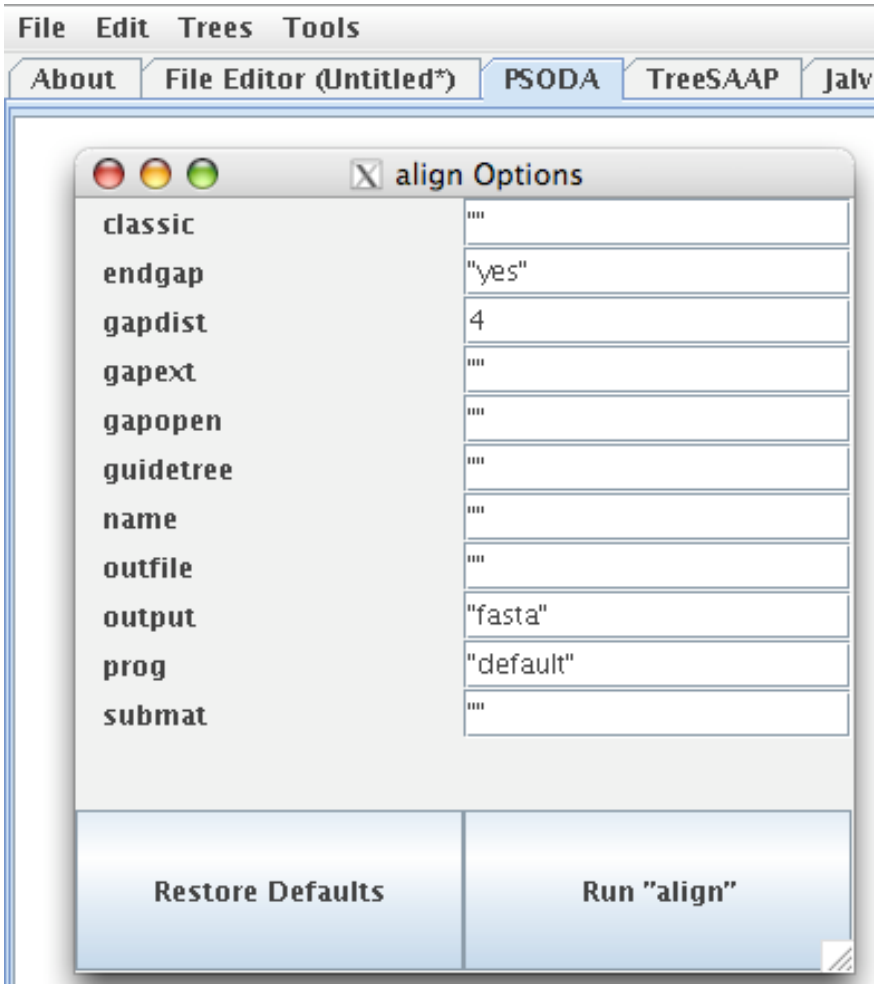


Select the **Run "loaddata"** button when you have the alignHIV.fasta file chosen.

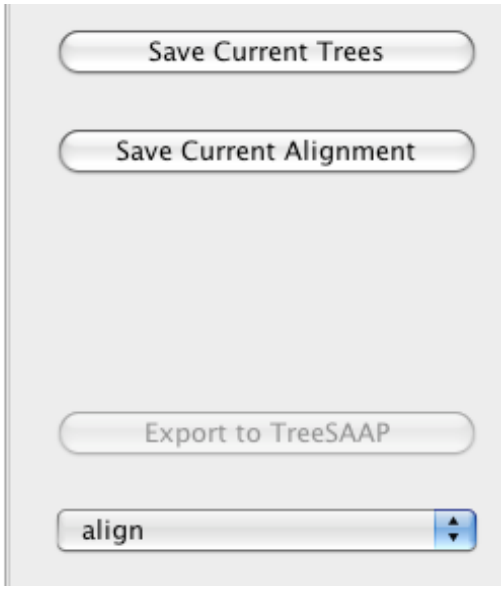
Once the data has been loaded into PSODA, you can run an alignment on the data by selecting the alignment tab from the dropdown menu.



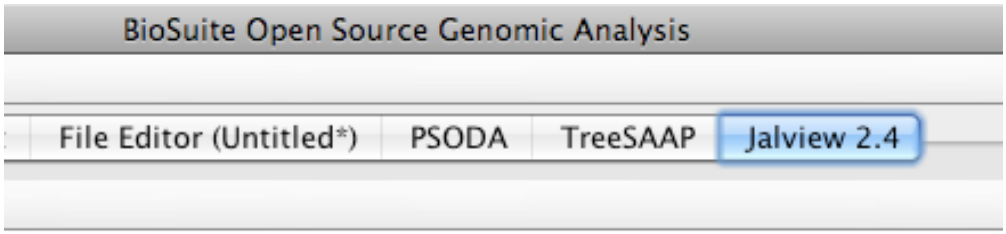
The window will allow you to change the gap and mismatch cost. Click the Run "align" button to run the alignment. All of the commands in psoda can be executed through this menu.



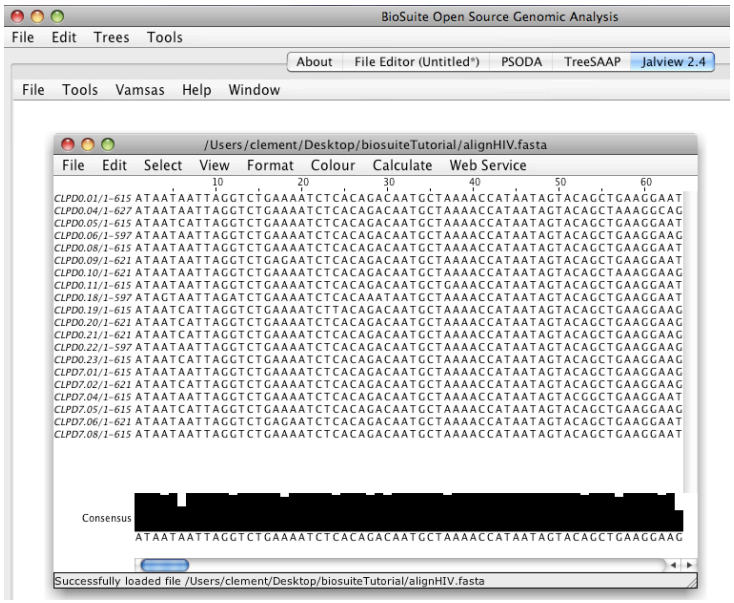
You can save the alignment using the **Save Current Alignment** button to save a copy of the alignment for further analysis.



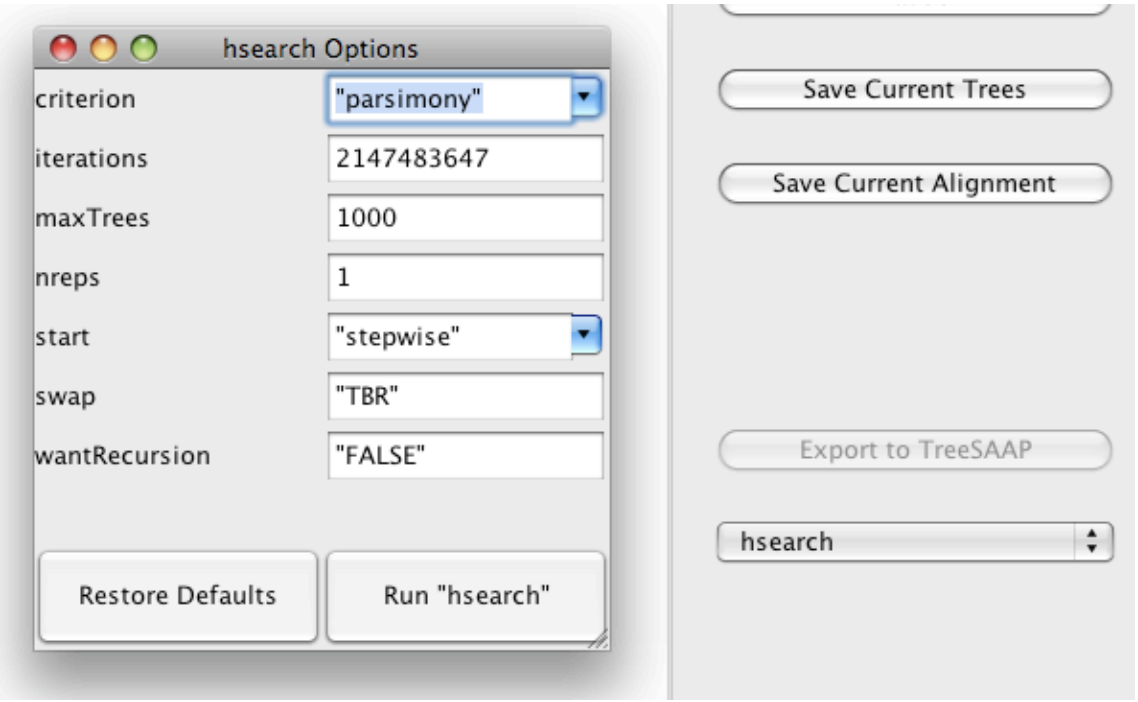
You can use the Jalview tab to view the alignment



You can find more information about Jalview at http://www.jalview.org/



Now that you have an alignment, you can run a phylogenetic analysis using the Heuristic search (hsearch) command.

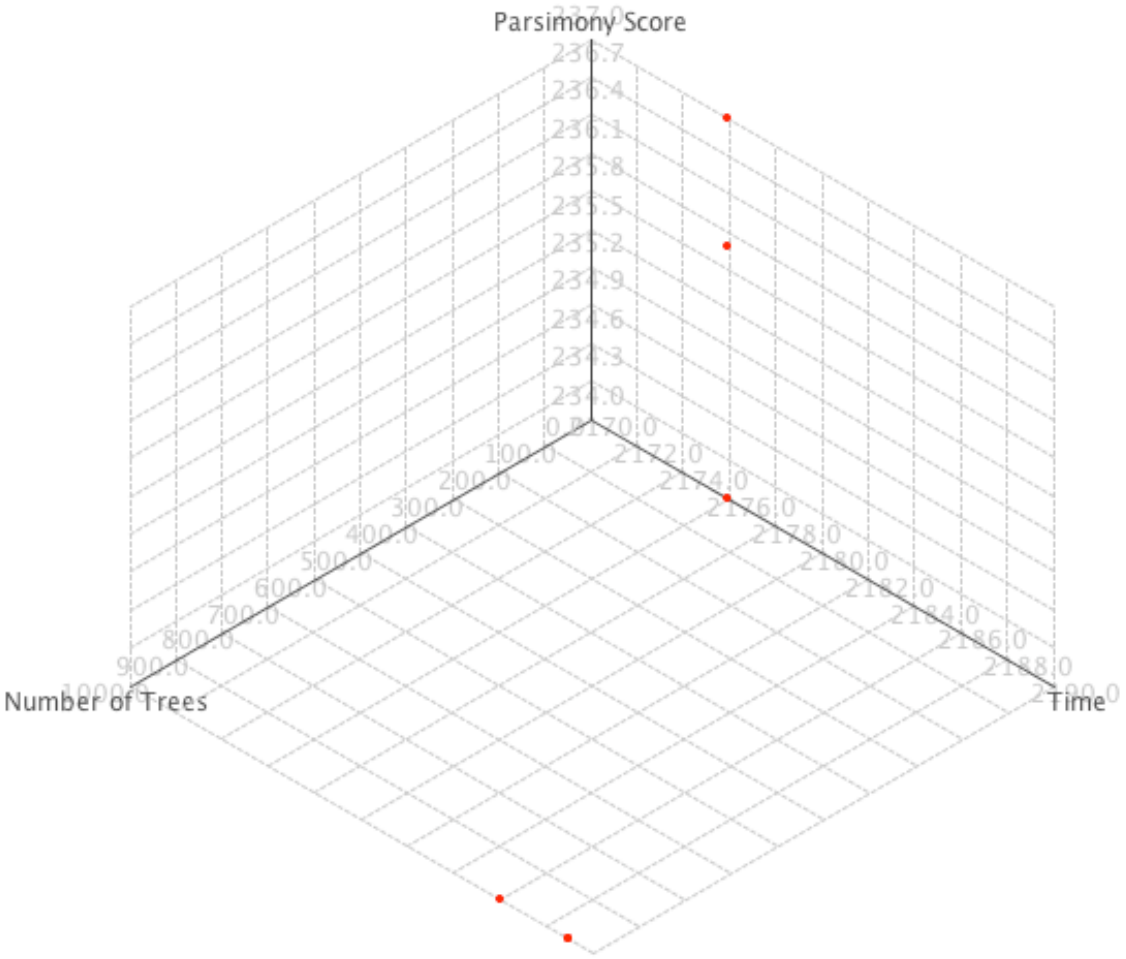


You can select several different optimization criteria including “parsimony” or “likelihood”. You can also select the maximum number of iterations to perform, the maximum number of trees to retain with equally optimal score and the number of replicates to build in the start phase of the search. You can also select SBR or TBR for the swap strategy. PSODA allows for recursion for large trees, but you shouldn’t need this unless you have a large number of taxa in your dataset. When you click Run “hsearch” the output of the hsearch will show up in the PSODA window.

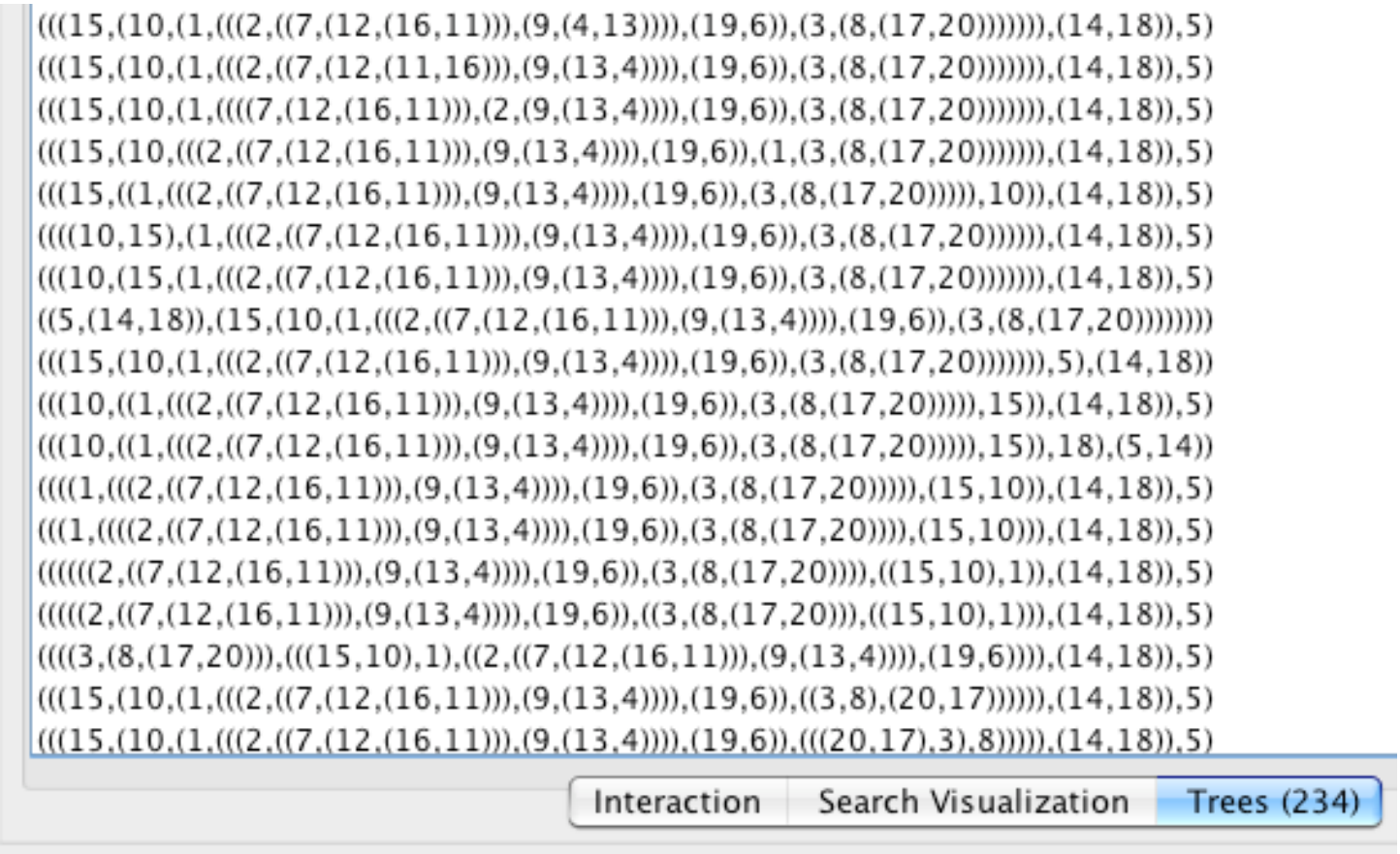
```
> hsearch (criterion="parsimony", iterations=2147483647, maxTrees=1000, nreps=1, start="stepwise", swap="TBR", wantRecursion="F
treestring= (null)
.................
 Stepwise Complete: rearrangements 323.000000 score 237.000000 best 1.797693e+308
Here is the tree we got with score 237.000000
treestring= (((15,(10,(1,((19,6),(3,(8,(17,20)))))),(14,18)),(5,((12,((16,11),(2,7))),(9,(13,4)))));
retainedresults iterations 2147483647

Elapsed         Taxa       Rearr.     No. of trees       Best trees     Search Iter
   time         added      tried   saved left-to-swap
-------------------------------------------------------------------------------
00:00:00          20           35     0     0               237            0
00:00:00          20         2129     1     1               236           30
00:00:00          20         2615     1     1               234           32
00:00:10          20      2903904  1000   219               234        29727
Searched all trees, best score 234.000000
## HSearch Completed Successfully
```

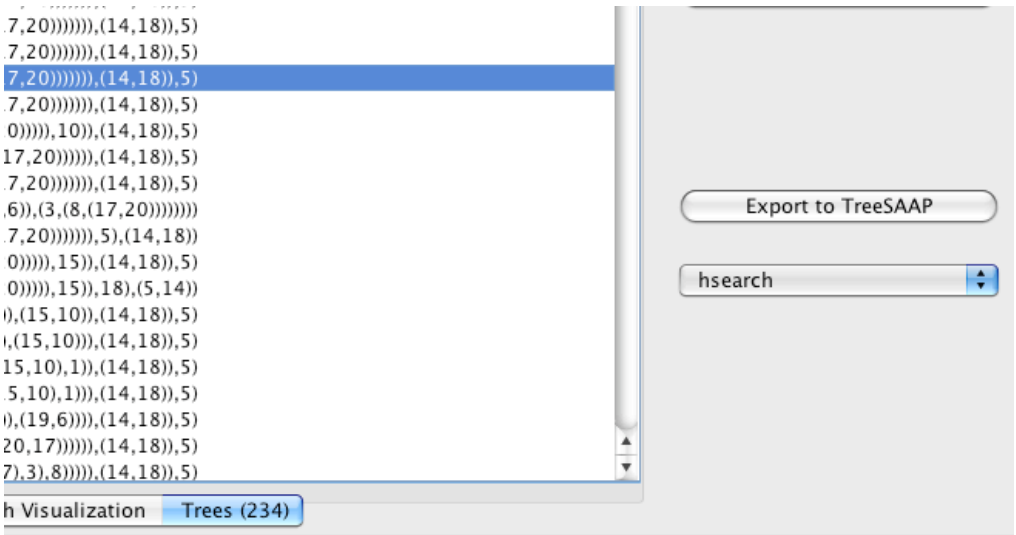
You can visualize the search by selecting the “search visualization” tab at the bottom of the window. You can rotate the search visualization to see how the search is progressing when you have a larger number of taxa.



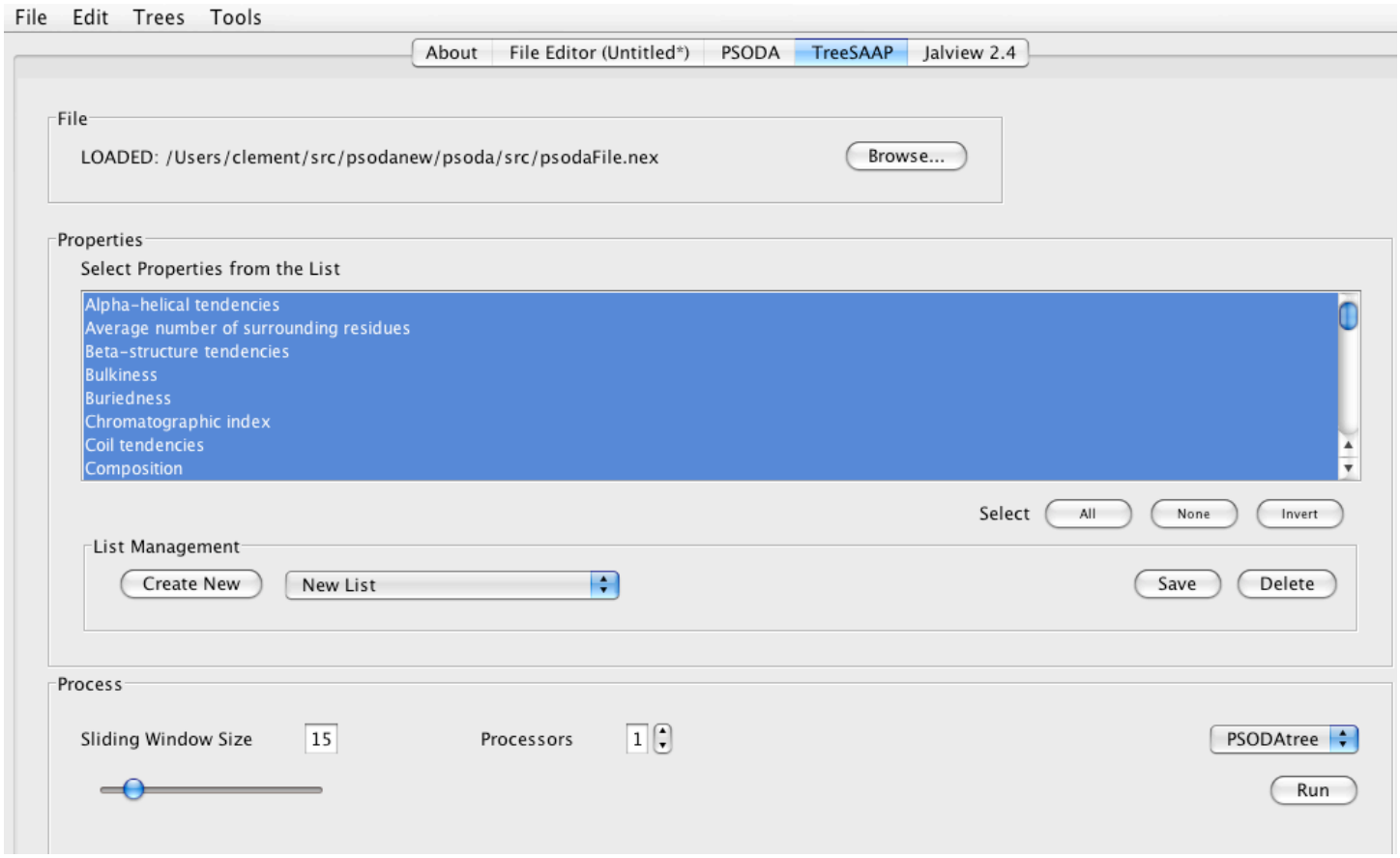
To visualize the trees found by the search, select the "Trees" tab. The tab will show the number of trees found by the search. Double clicking on one of these trees will bring up the "ATV" tree viewer. More information about the viewer can be found at http://www.phylosoft.org/atv/



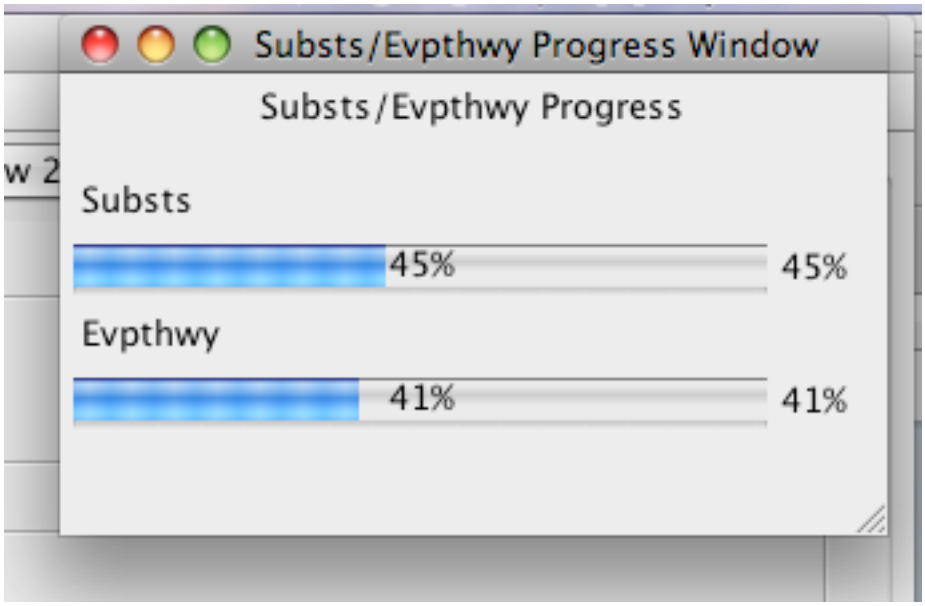
With one of the trees selected, click on the "**Export to TreeSAAP**" button. This will take you to the TreeSAAP window with that tree selected for the selection analysis.



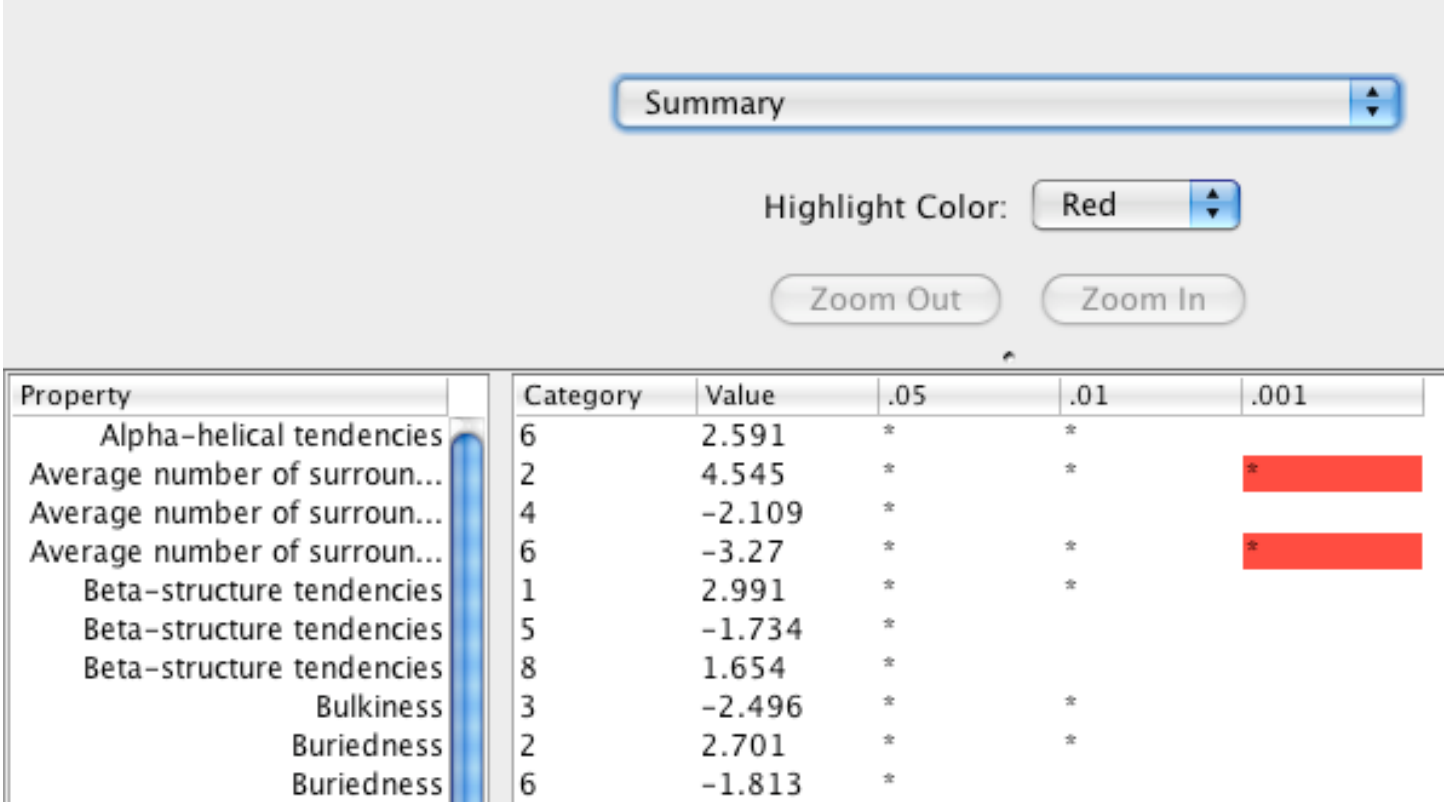
From this window you could also selected another nexus file with aligned data and a phylogenetic tree. To run the analysis, click on the "**Run**" button in the lower right corner.



You should see a progress window showing how the analysis is proceeding.

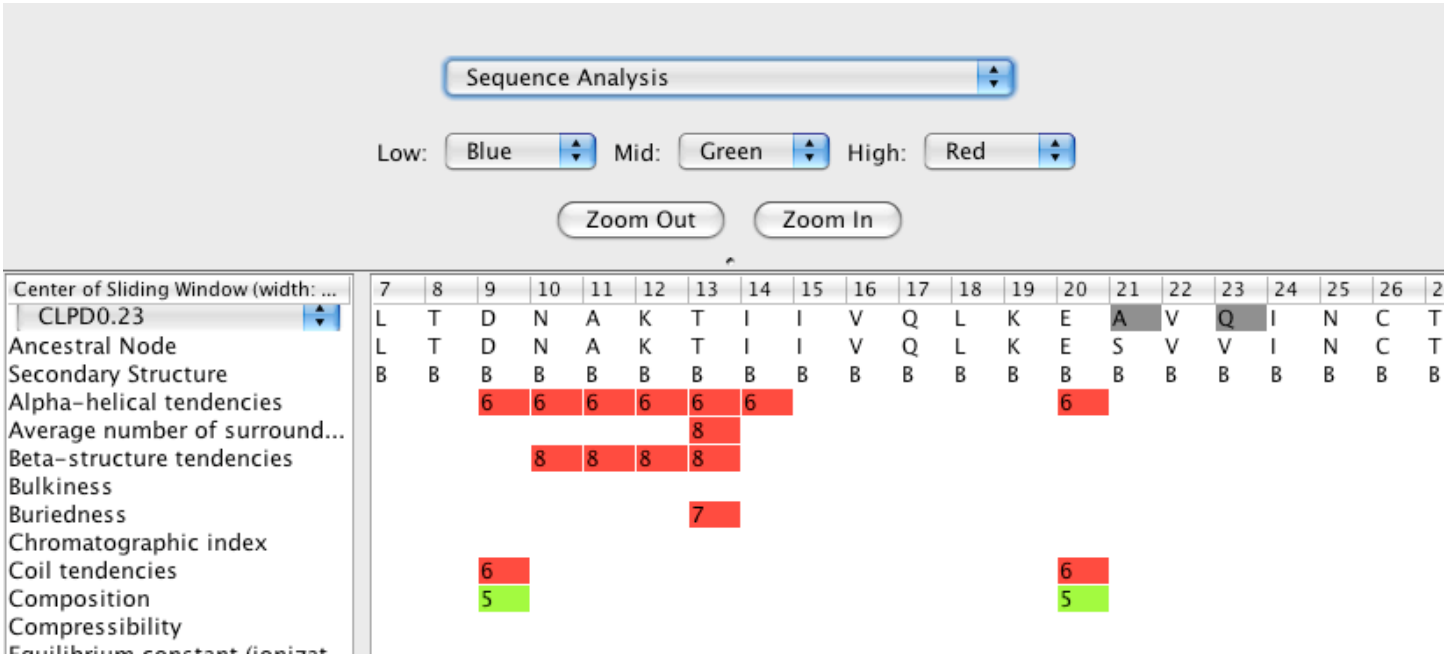


When the analysis is complete, you will see the summary window. This shows the biochemical properties along with their category and P-values. Higher Category values indicate selection for more extreme biochemical changes, lower Category values indicate purifying selection.

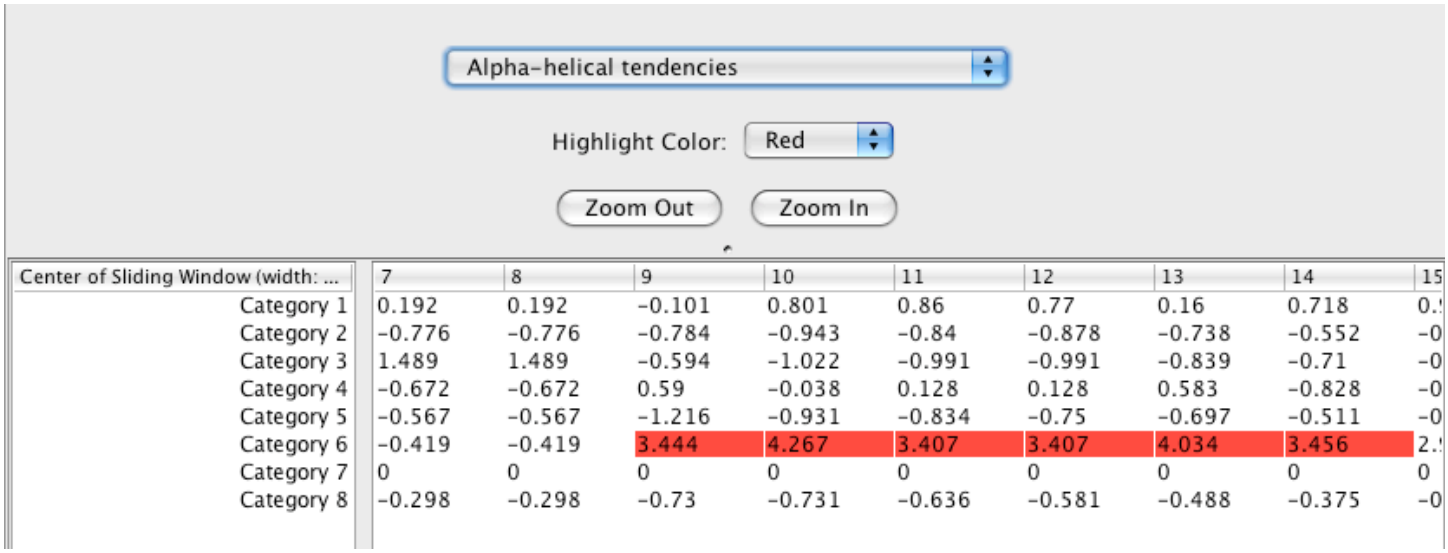


Property	Category	Value	.05	.01	.001
Alpha-helical tendencies	6	2.591	*	*	
Average number of surroun...	2	4.545	*	*	*
Average number of surroun...	4	-2.109	*		
Average number of surroun...	6	-3.27	*	*	*
Beta-structure tendencies	1	2.991	*	*	
Beta-structure tendencies	5	-1.734	*		
Beta-structure tendencies	8	1.654	*		
Bulkiness	3	-2.496	*	*	
Buriedness	2	2.701	*	*	
Buriedness	6	-1.813	*		

By changing to the sequence analysis dropdown, you will be able to see the sequence data (as amino acids) along with the Categories of selection for various chemical properties on the left.



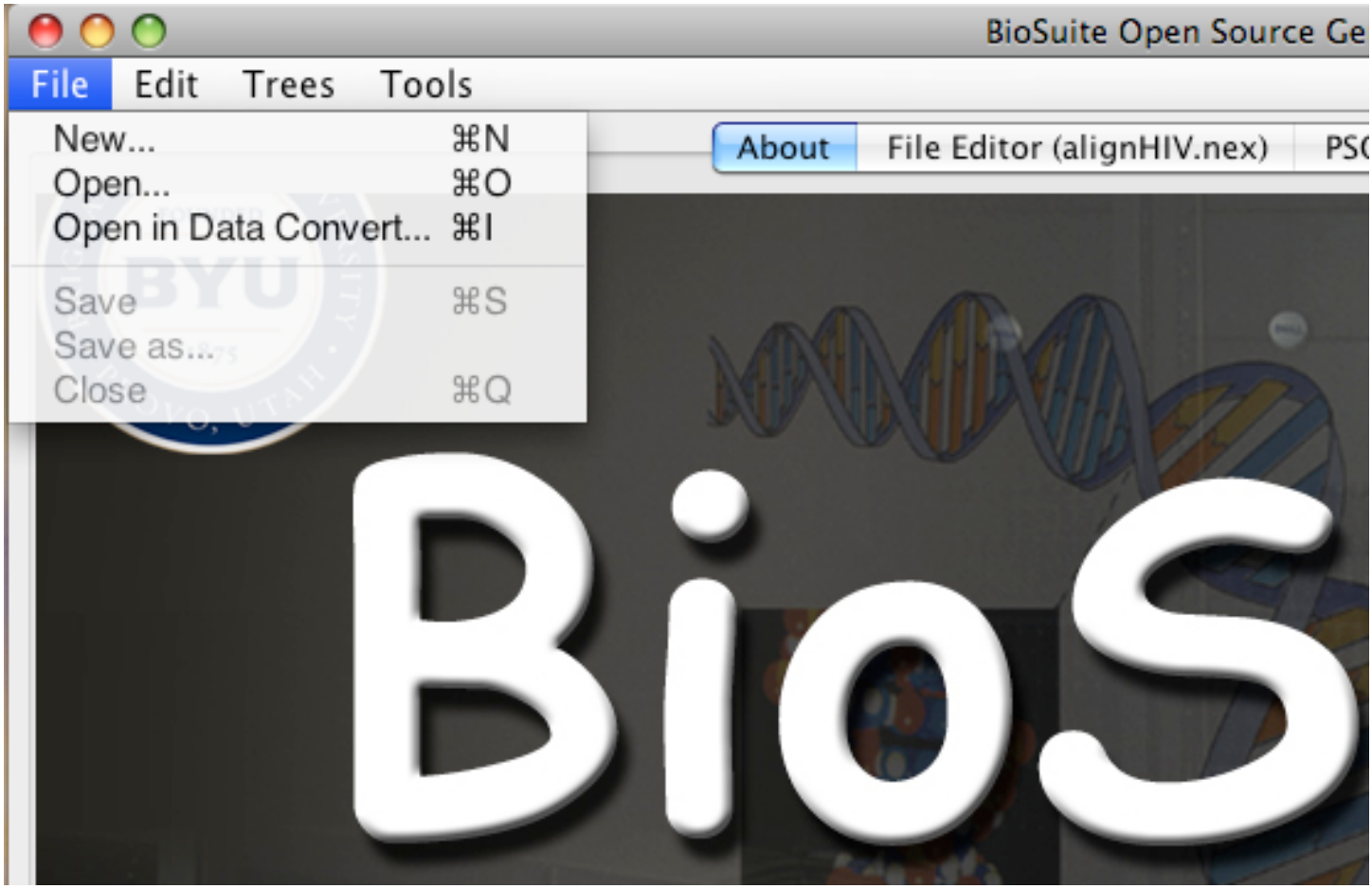
You can also select one chemical property (such as Alpha-helical Tendancies) and see detailed levels of selection for each category.



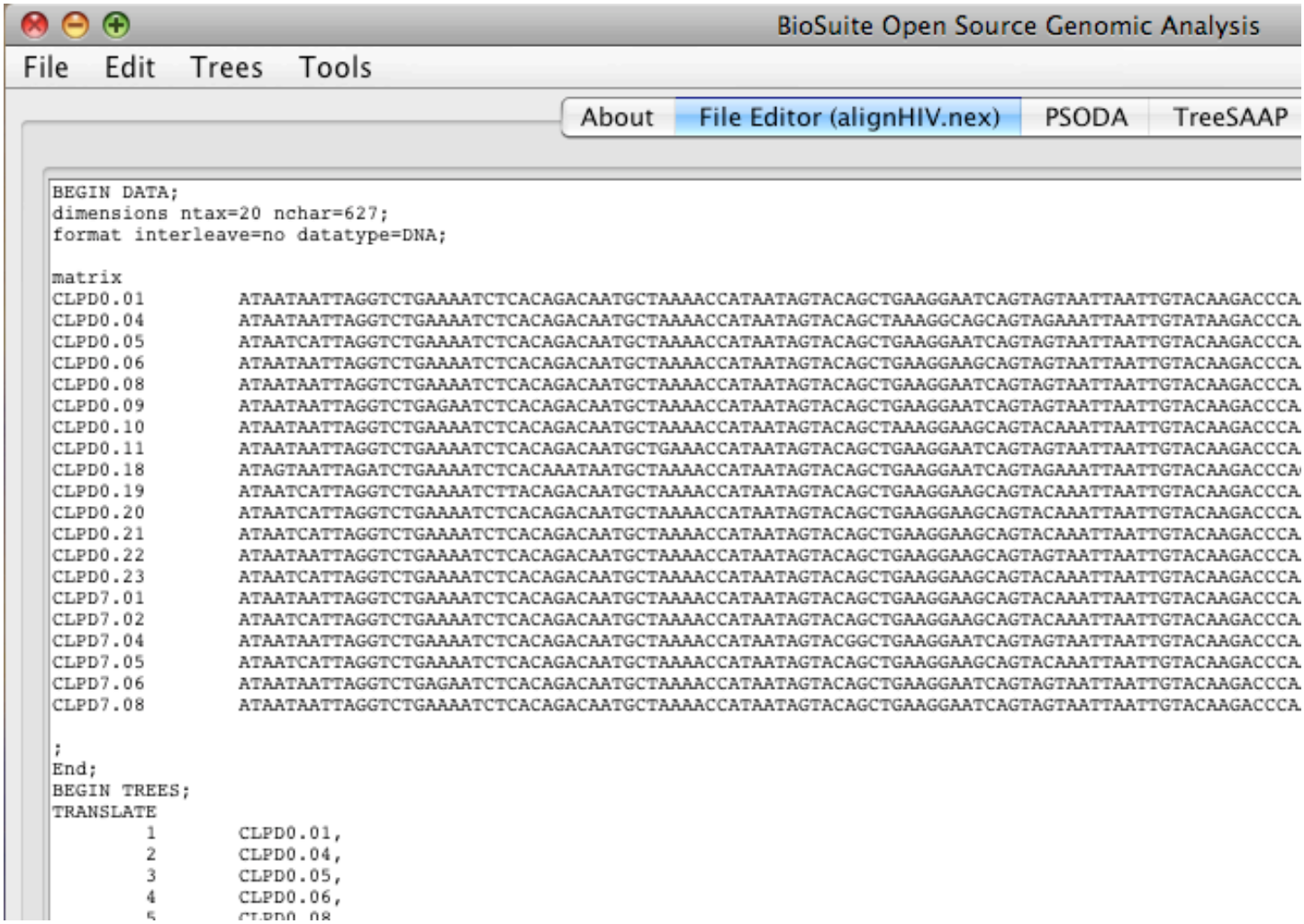
Center of Sliding Window (width: ...	7	8	9	10	11	12	13	14	15
Category 1	0.192	0.192	-0.101	0.801	0.86	0.77	0.16	0.718	0.!
Category 2	-0.776	-0.776	-0.784	-0.943	-0.84	-0.878	-0.738	-0.552	-0
Category 3	1.489	1.489	-0.594	-1.022	-0.991	-0.991	-0.839	-0.71	-0
Category 4	-0.672	-0.672	0.59	-0.038	0.128	0.128	0.583	-0.828	-0
Category 5	-0.567	-0.567	-1.216	-0.931	-0.834	-0.75	-0.697	-0.511	-0
Category 6	-0.419	-0.419	3.444	4.267	3.407	3.407	4.034	3.456	2.!
Category 7	0	0	0	0	0	0	0	0	0
Category 8	-0.298	-0.298	-0.73	-0.731	-0.636	-0.581	-0.488	-0.375	-0

The scroll bar on the bottom of the window allows you to pan through the sequence length.

There are several other ways of using BioSuite. You can open nexus files with Paup style commands using the **File/open** tab in the upper right corner of the window.



Once you have opened the nexus file, you should see the data in the **File Editor** window



You can make changes to the data here and save them out as well. Use **File/Open** to input the file at http://dna.cs.byu.edu/biosuite/tutorial/HIV.nex

Now go to the PSODA tab and click on the **Run "HIV.nex"** button. This should run the same analysis