

# Installation and Usage Instructions for Tagd Plug-in for Mesquite

by T. Eck, S. Angelov, S. Fisher, and J. Kim  
Biology Department, University of Pennsylvania

---

## Introduction

This document provides manual installation and usage instructions for a Mesquite plug-in module that integrates Mesquite's UI with the Tagd.exe application that generates discriminating tags from biological sequence data. This plug-in enables a four-step process for generating and selecting discriminating tags:

1. Extraction of discriminating tags from the raw sequence files
2. Computation of several statistics on each tag (melting temperature, folding potential, GC content, information content)
3. Visualization of the statistics in the form of histograms
4. Exporting tags that meet user-supplied filter criteria to FASTA format

The plug-in employs a "project-based" approach to managing the taxa, trees, sequence data, and results files. The project itself is implemented in a NEXUS file, which is a standard and extensible format utilized in phylogenetics research.

## Installation

### Pre-requisites

The basic Mesquite software must first be installed and functional. This software can be freely obtained from <http://mesquiteproject.org>. Before proceeding with this installation procedure, please confirm that Mesquite executes properly on your system. For the remainder of this document, the root of the Mesquite installation (e.g. c:\program files\Mesquite\_Folder) shall be represented by <Mesquite Root>.

## Manual Installation Steps

The included Zip file contains four subdirectories. Installation merely requires extracting these files into the appropriate directories as specified here.

1. Copy the contents of the *CopyToJarsFolder* directory to <Mesquite Root>\jars
2. Copy the entire *ngramEntropy* directory to <Mesquite Root>\ (thus creating the directory <Mesquite Root>\ngramEntropy )
3. Copy the entire *tagd* directory (and its subdirectories) to <Mesquite Root>\Mesquite\ (thus creating the directory <Mesquite Root>\Mesquite\tagd)

## Test Steps

To confirm that the installation has succeeded, launch the Mesquite application. For Windows platforms, the application is located in <Mesquite Root>\Mesquite.exe When Mesquite is launched, a series of diagnostic messages is written to the "Mesquite Log" window. Once the Mesquite "Projects and Files" window appears, inspect the "Mesquite Log" window and search for the following text:

548 modules installed.

**4 prerelease substantive modules installed.**

This plug-in registers four modules with Mesquite. If you see this message then Mesquite is recognizing the plug-in's modules.

## Usage

### Project File

Now that the plug-in has been successfully registered with Mesquite we can load our first sample project. The Zip file contains a directory named "cf-small" which contains simulated data. We will use this to step through the plug-in's features.

Copy the entire cf-small directory to a location on the system that is writable (the plug-in will generate files and update the cf.nex file)

Open the file cf.nex in a text editor. The first section (TAXA) defines the taxa contained in the project.

The second section (TREES) defines the guide tree that this tagd project will utilize. The plug-in supports one tree per project file. Note that the structure of the tree is provided in Newick format.

The next section (TAGD) provides the mapping between each taxon in the tree and its associated sequence file. Depending on where you copied the cf-small directory, you may need to change the paths in this section to point to the location of the sequence files (\*.fasta). After making any necessary changes save the .nex file in your text editor.

## Loading a Project

Launch Mesquite (as described above) and select the File menu from the “Mesquite Projects and Files” window. Select “Open File” from the menu and then open the cf-small.nex file that you edited in the prior step. Mesquite and the tagd plug-in will cooperate to read and parse the Nexus file.

*NB: If you receive an error while loading the Nexus file, check the syntax in the file and confirm that the paths to the sequence files in the TAGD section are correct.*

Mesquite will display a new Taxa window that reflects the TAXA section in the project file. Open the Taxa&Trees menu from the Taxa window and select the “New Tree Window” option. Select the “Stored Trees” option and click OK. Mesquite will display a Tree Window containing an interactive model of the tree that it parsed from the project file.

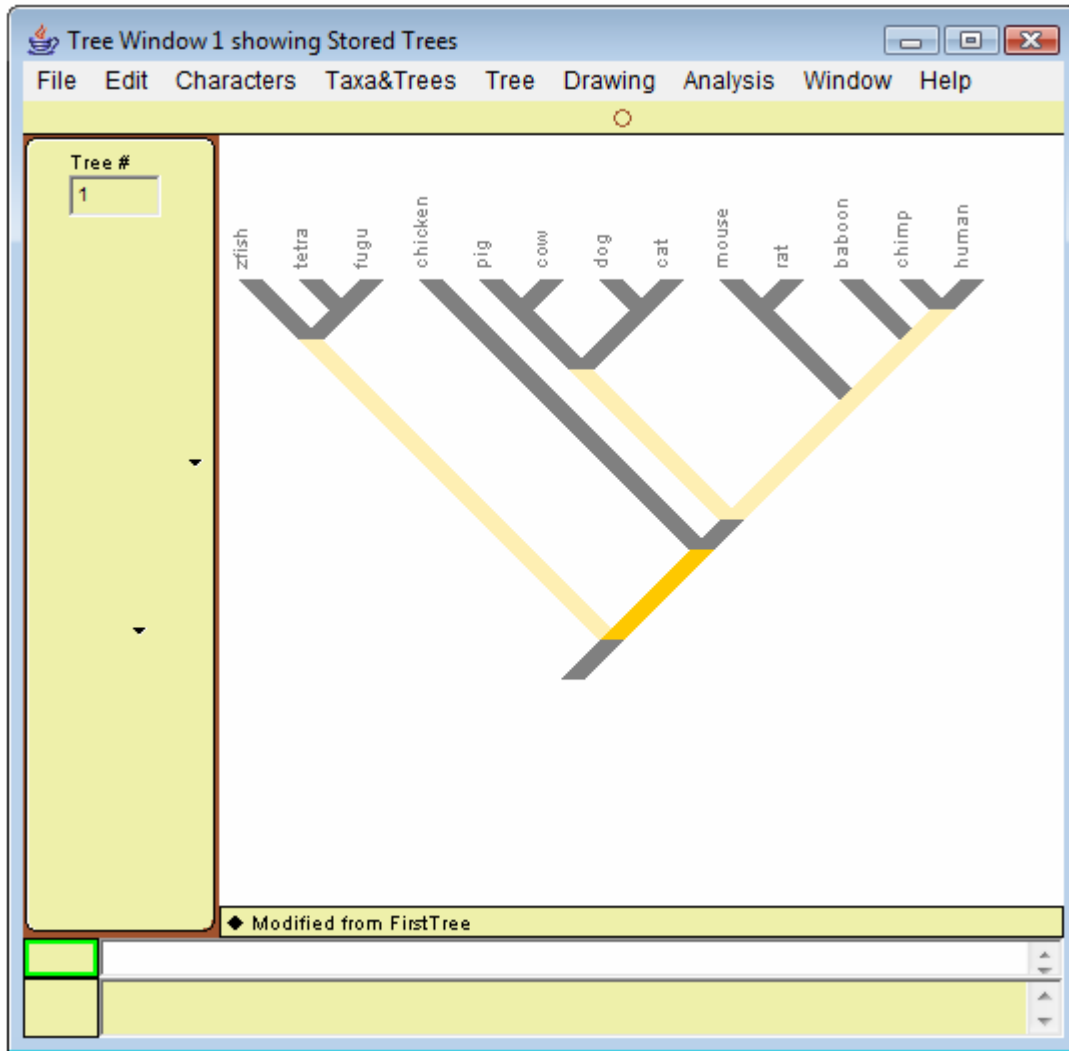


Figure 1 - Mesquite Tree Window

## Performing Tagd Analysis on the Project

All tagd operations are performed on the tree that has been loaded into the Mesquite Tree Window in the prior step. The tagd operations are accessed through menu options that are only available within the Tree Window. All tagd operations are performed on a single branch at a time.

### Selecting a Branch

To select a branch on which tagd operations will be performed, you must activate the Mesquite “Select Branch” tool which appears in the “Tree Window’s” tool palette (the yellow section on the left of the window). Click the icon labeled “Select Branch” to activate it. Now click one of the edges on the tree to select it (its color will change when selected). All taxa to the right of the selected branch away from the root are

considered the “Right” taxa for tagd operations, and all taxa to the left and away from the root are considered the “Left” taxa.

### Extracting Tags

The first tagd operation that must be performed on a branch is the extraction of discriminating tags. With the desired branch selected, open the Tree menu and then open the Utilities submenu within it. Then select the “Tagd-1::Extract Discriminating Tags” menu option. The plug-in will launch tagd.exe externally and capture its output stream. While tagd.exe is executing, a dialog window will be displayed which allows you to cancel the operation if you want. Depending on the size of the tree, the size of the sequence files, and the edge that you selected in the tree, this processing can take minutes or hours to complete.

When tagd.exe has completed its processing, a dialog window will appear containing a message similar to the following:

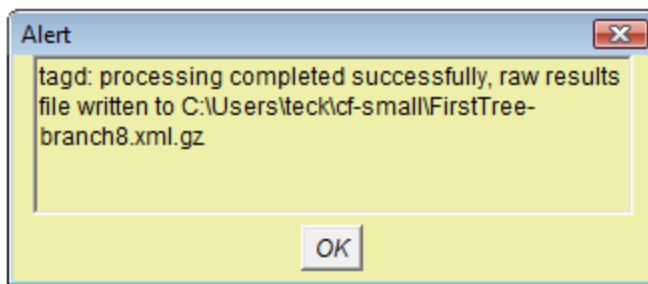


Figure 2 -- Successful extraction of tags

The tagd results are encoded in an Xml file and subsequently compressed with Gzip. The plug-in manages this output file and keeps it synchronized with the project (Nexus) file. The plug-in automatically decompresses the file when additional tagd operations are performed on that branch. If you wish to view the results directly, you can use any gzip-compatible tool to extract the Xml file and then view the Xml file in any text or Xml editor.

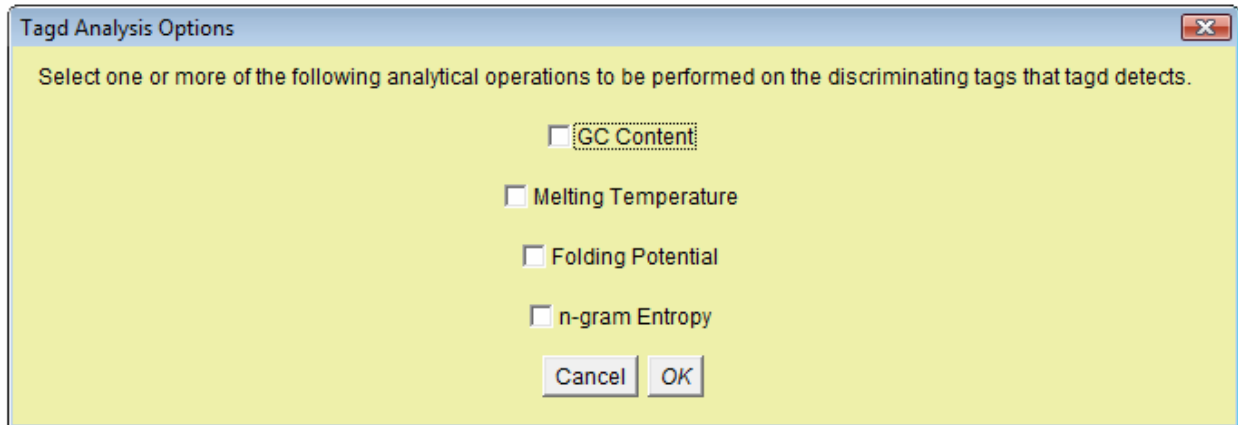
### Analyzing Tags

The tagd plug-in supports the computation of the following analytical methods to each tag:

- GC Content
- Melting Temperature

- Folding Potential (free energy)
- N-gram Entropy (information content)

To apply one or more of these methods to the tags extracted for a particular branch, select the branch (using the Mesquite Branch Selection tool), open the Tree menu and the Utilities submenu within it, then select the "Tagd-2::Analyze Extracted Tags" menu option. Select one or more operations from the "Tagd Analysis Options" dialog window and click OK to initiate the processing.



**Figure 3 - Tagd Analysis Options dialog**

The "n-gram Entropy" operation requires a one-time construction of an index for each project. So the first time that you select this operation for a given project, you will be notified that the index will first be constructed.

When the analysis operation(s) are complete, a dialog window will be displayed indicating the successful completion of the operation(s).

You may apply subsequent analysis operations to a branch that has been previously analyzed and you can re-apply analysis operations to the same branch at any time.

### Visualizing Tags

The plug-in allows you to visualize the results of the analysis operation(s) that you performed on the selected branch. Results are displayed in histogram format.

To visualize these values for the selected branch, open the Tree menu and the Utilities submenu within it, then select the "Tagd-3::Visualize Extracted Tags" menu option. A dialog window will be presented that allows you to select one of the analysis operations that you applied to the branch. Select one operation that you wish to visualize and click OK.

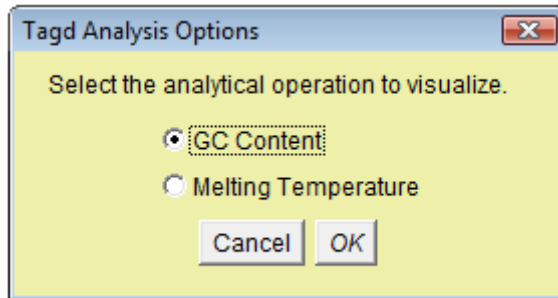


Figure 4 - Tagd Analysis Options dialog

The histogram will be computed and displayed in a window. You may open multiple histogram windows simultaneously for the same or different branches.

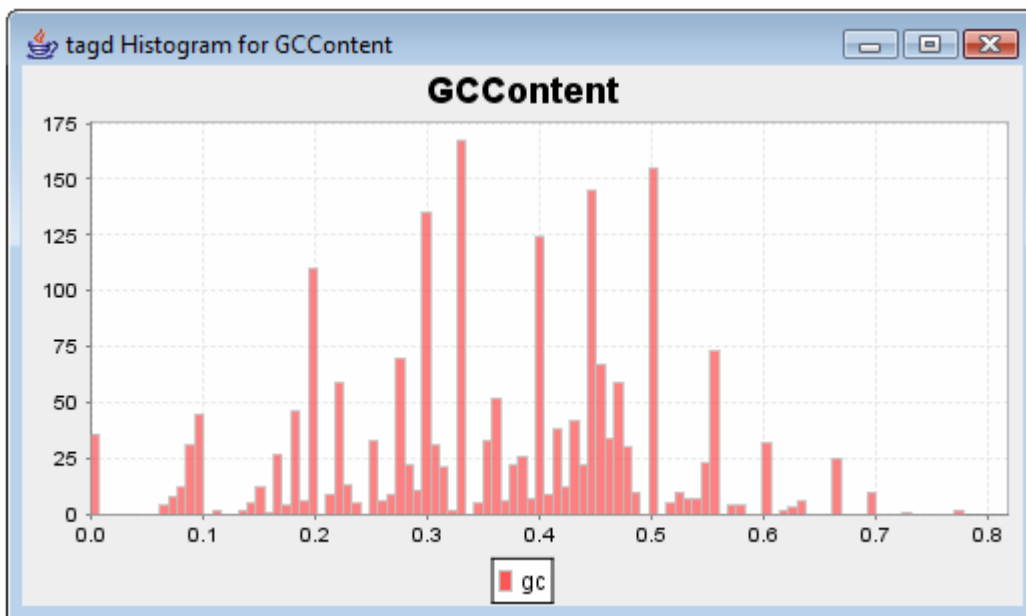


Figure 5 - Histogram produced by the plug-in

### Exporting Tags

The plug-in allows you to export a branch's tags to a FASTA file. You are given the option to filter the tags based on the values of the four analysis operations supported by the plug-in. Note that you are not required to perform the Analysis and Visualization steps in order to export the tags, although doing so facilitates your selection of the filter criteria.

To export tags for a selected branch, open the Tree menu and the Utilities submenu within it, then select the "Tagd-4::Filter and Export Tags" menu option. The "Tagd Export Options" is displayed.

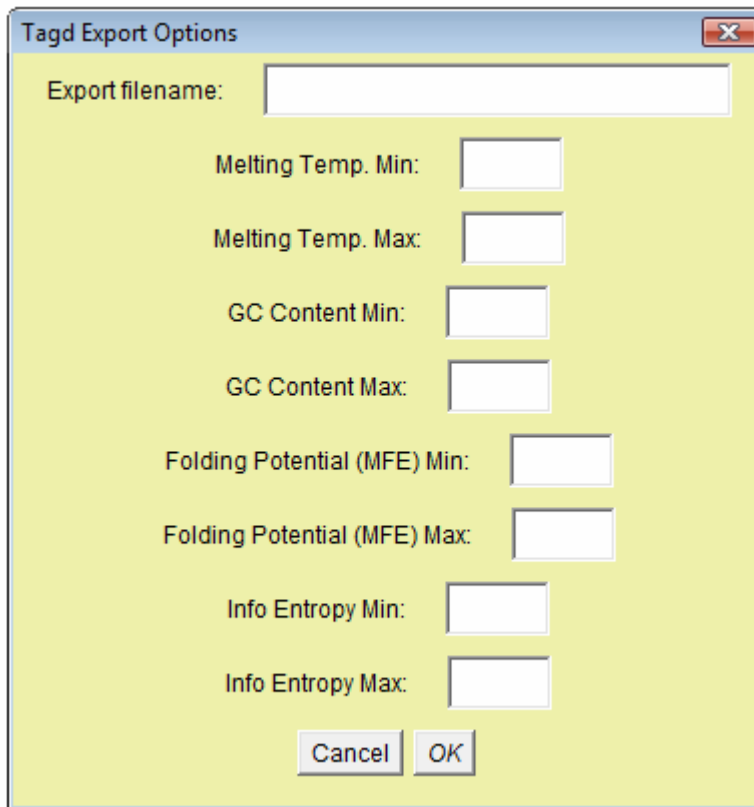


Figure 6 - Tagd Export Options dialog

The only mandatory field in the dialog is the “Export filename”. Provide a path and filename for the FASTA file that the plug-in will produce.

The remaining fields are pairs of minimum and maximum values for each of the four analysis operations supported by the plug-in. You can supply any combination of values. The plug-in treats any blank minimum field as negative infinity and any blank maximum field as positive infinity. Thus if both fields for a particular statistic are empty, then the filter range is interpreted to be +/- infinity.

Once you have provided the desired filter criteria, click OK to perform the export operation. Upon completion a dialog window will be displayed indicating successful execution of the operation.