# Geneious Microsatellite Plugin

Biomatters Ltd

November 24, 2018

# Introduction

This plugin imports ABI fragment analysis files and allows you to visualize traces, fit ladders, call peaks, predict bins, display alleles in tabular format and export your data for further analysis.

It requires a valid license for Geneious R7 or later. If your copy of Geneious is older you should purchase an upgrade.

The Geneious Microsatellite Plugin is intended for research purposes only and has not undergone validation for human identification analysis or diagnostic procedures.

# Installing the Plugin

The Geneious Microsatellite Plugin can be downloaded from https://www.geneious.com/plugins/microsatellite-plugin/. Once it is downloaded simply drag it into Geneious to install, or select "Install plugin from a gplugin file" from the "Tools" → "Plugins" menu.

Once it is installed, the Geneious Plugin Manager will let you know about updates if the "Automatically check for updates to installed plugins" options is checked.

# Importing your Data

Once you have the plugin installed, simply select the .fsa files you want to import and drag them into Geneious. The importer will automatically recognize them and store them as Geneious documents, which you can then select to visualise and analyse your data.

# Traces Viewer

When one or more traces documents are selected a number of viewers will be shown. The left most tab is the "Traces'' tab which looks like Figure 1.

In the main panel the selected documents are stacked vertically in the same order that they are in the document table. This view can be highly customised using the controls on the right:

- **Y Scale** – Adjusts the vertical amplitude of the traces

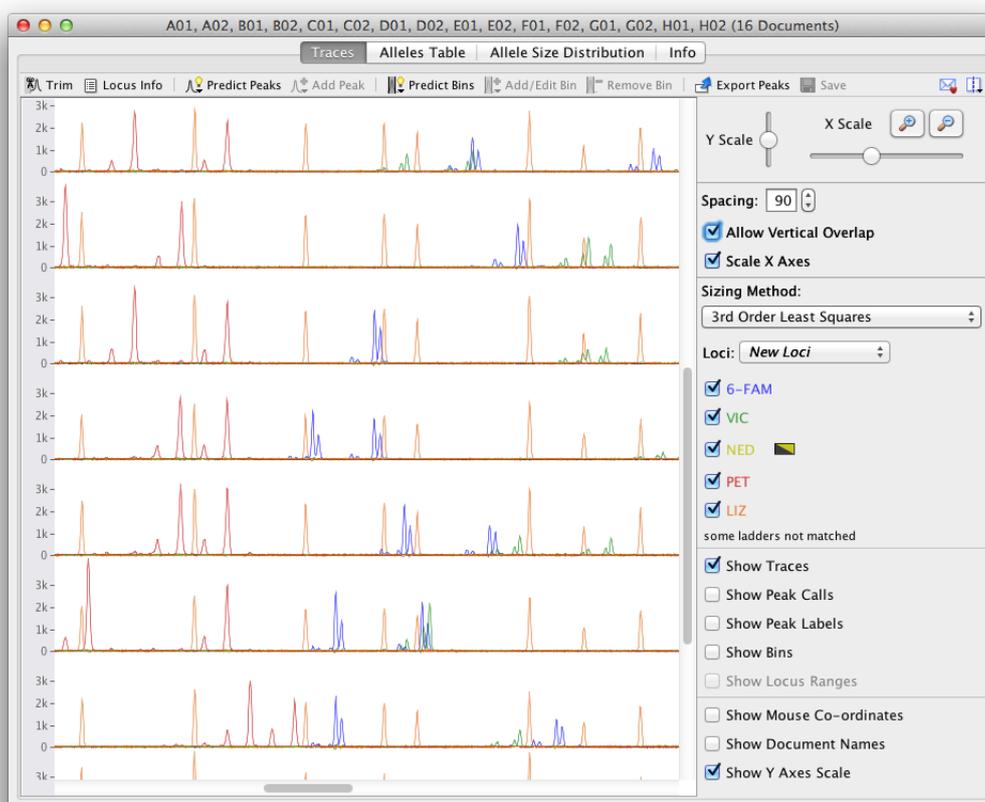- **X Scale and Zoom buttons** – Adjusts the X axes zoom

Figure 1: The Traces view

- **Spacing** – Adjusts the vertical separation between rows. This option is not available when viewing a single document.

- **Allow Vertical Overlap** – Allows traces from different documents to overlap vertically, disabling this option often gives better drawing performance. This option is not available when viewing a single document.

- **Scale X Axes** – Whether to use raw coordinates from the electropherogram machine on the x axes or to scale them into base pair (nucleotide) coordinates. This option is not available when viewing a single document.

- **Sizing Method** – The method that is used to do the sizing conversion from raw machine coordinates to base pairs

- **Loci** – The document that provides the locus information and bins

- **Trace buttons** – Show or hide drawing of specific trace dyes. *Note:* the yellow dye can be switched to black for enhanced visibility

- **Show Traces** – Toggles drawing trace lines

- **Show Peak calls** – Toggles drawing peak calls

- **Show Peak Labels** – Toggles drawing peak call labels

- **Show Bins** – Toggles drawing bins

- **Show Locus Ranges** – Toggles drawing the size ranges of the allele regions. *Note:* this is only enabled when viewing a single microsatellite dye.

- **Show Mouse Co-ordinates** – Show the mouse X coordinate in base pair units and the amplitude of the largest trace under the cursor in relative fluorescence units (RFU)

- **Show Document Names** – Toggle showing the document names on the left

- **Show Y Axes Scale** – Toggle showing a Y axis scale for traces on the left in relative fluorescence units (RFU)

Along the top of the main panels are buttons that allow you to analyse your data. These are ordered left-to-right in approximate order of use:

- **Set Trim** – This allows you to set the trim for all documents. The trim is greyed out area at the start of each trace which represents the area of noise around fragment length 0-50bp that should be ignored. You can also set this per document by dragging the right hand edge of the trace, if there is no trim visible just drag from the far left of the trace window to create one.

- **Set Locus Info** – This allows you to name your loci and specify their repeat unit, number of expected peaks and the size range they occur over, shown in Figure 2. Multiple loci can be specified for a single dye by clicking the "+" button on the right. To visualise the size range see the "Show Locus Range" option in the view controls on the right.
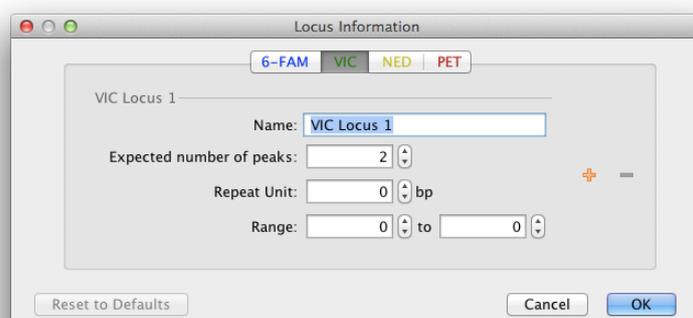


Figure 2: Setting your locus info

- **Predict Peaks** – Re-runs the automated peak calling algorithms and allows setting the minimum height threshold in RFU below which peaks will not be called. When data that could lead to a change in peak calls (e.g. trim, ladder peaks) is changed these are run automatically until manual changes are made by the user, at which point they stop in order to not overwrite the user's changes. This option allows re-running them in this case.

- **Add Peak** – You can manually add a peak call by clicking a point on a trace and clicking this button. You can hold the CTRL key (⌘-Command on Mac OS) and click to select multiple points. You can also add peaks by right clicking.

- **Remove Peak** – You can remove peak calls by selecting them in the trace view and clicking this button, pushing the DELETE key on your keyboard or right clicking. To select a peak click on it, multiple peaks can be selected by holding the CTRL key (⌘-Command on Mac OS) and/or clicking and dragging to select all peaks in a region.

- **Predict Bins** – Predict bins based on the present peak calls and locus information. For each locus, you can optionally specify an "example allele" as a fixed point for the bin labels. There will not be a bin at that fixed point if no peaks are called there, but all the bin labels for that locus will be forced to be consistent with that fixed point.

- **Add/Edit Bin** – This allows adding a bin or editing an existing a bin. Select a point to add a Bin, select an existing Bin to edit it. You can also add bins by right clicking.

- **Remove Bin** – You can remove bins by selecting them in the trace view and clicking this button or pushing the DELETE key on your keyboard or right clicking. To select a bin

click on it, multiple bins can be selected by holding the CTRL key (⌘-Command on Mac OS).

- **Export Peaks** – This allows exporting the calculated peak sizes in base pairs (nucleotides) as a comma separated values (CSV) file for import into other programs, such as Microsoft Excel.

- **Save** – This saves any changes made to the files. *Note:* Locus Info and Bins are saved to a separate Geneious document to allow flexibility in applying them to different trace files.

There is also sliding notification that pops out from time to time to offer helpful suggestions (e.g. Figure 3)
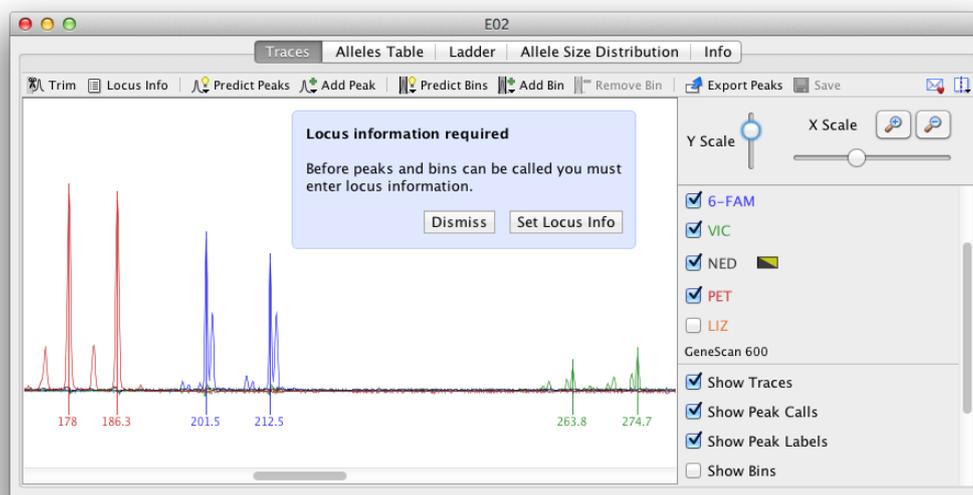


Figure 3: Sliding notification visible

## Ladder Viewer

The ladder viewer is only available when viewing a single document and only works when a ladder is fitted. It displays a graph showing the raw X coordinates from the electropherogram machine and the corresponding matched length in base pairs for the ladder fragments. A line of best fit is plotted which corresponds to the 1st order linear least squares sizing method (Figure 4.) For more information on sizing methods see the below section.
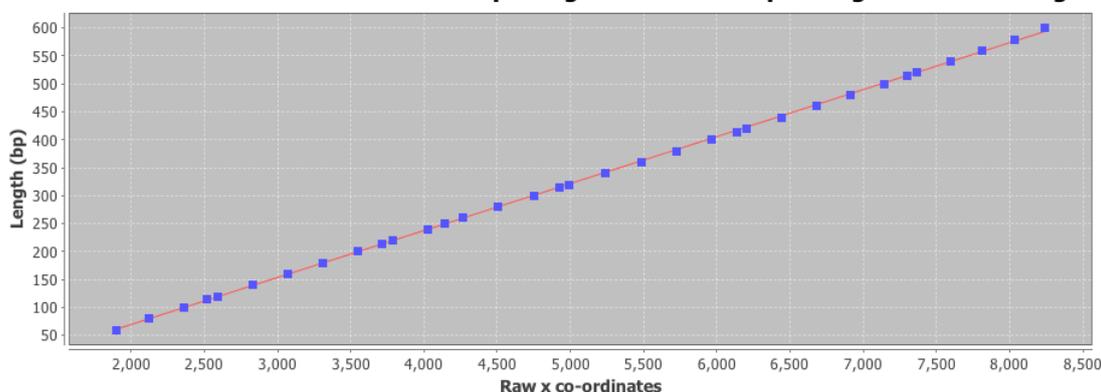
Figure 4: Ladder with all the peaks for the GeneScan 600 ladder

# Sizing Methods

In order to convert the X coordinates from the electropherogram machine to base pair coordinates, a ladder of known fragment sizes is run alongside the microsatellite fragments. Once the ladder fragment peaks are identified the size of the microsatellite fragment peaks can be inferred. The simplest approach is to look at the size of the ladder fragment peaks on either side of the microsatellite peak and work out how far between them it is, e.g. if the unknown size microsatellite fragment peak is three quarters of the way between the 100bp and 140bp ladder fragment peaks it is going to be about 130bp in size. This is actually what the "Two Surrounding Peaks" sizing method does.

Similarly the least squares algorithms fit a first, second (quadratic) and third (cubic) order polynomial function to the all ladder peaks. These functions are able to convert between sizes that are outside the range of the ladder and are hence known as global functions. On the other hand the "Two Surrounding Peaks", "Local Southern" and "Cubic Spline" methods require a certain number of peaks on each side of the microsatellite peak in order to size it and are hence known as local methods. The "Two Surrounding Peaks" method is described above, the "Cubic Spline" method fits a series of cubic functions between the ladder peaks and also requires a ladder fragment peak above and below, finally the "Local Southern" method takes the average of two curves as described by E.M. Southern in http://www.ncbi.nlm.nih.gov/pubmed/525781 and hence requires two ladder peaks above and below.

The recommended algorithms to use for sizing are Third Order Least Squares or Local Southern.

## Adding a Custom Ladder

As of plugin version 1.2 the ladders are stored in a text file which can be manually edited to add new ladders.

This file can be found in the Geneious plugin installation directory. The location of this varies by computer operating system and permissions, the easiest way to find it is search the filesystem for the following directory name:

`com.biomatters.plugins.microsatellite.MicrosatellitePlugin` (Mac or PC)

or `com.biomatters.plugins.microsats.MicrosatsPlugin` (Linux)

Within this directory are a series of sub-directories

`com/biomatters/plugins/microsatellite/data`

and within this directory is a `ladders.txt` file.

This file contains a tab-separated list of ladders where the first field is the name of the ladder and the rest are the ladder fragment peak sizes in bp. Geneious must be restarted for changes to this file to take effect.

## Allele Size Distribution Viewer

This viewer displays the allele sizes for all peaks in the selected documents and dyes sorted by ascending size. Once peaks are correctly sized alleles should be clearly visible as steps as in Figure 5.

## Alleles Table Viewer

The alleles table shows the allele calls that are the result of combing the microsatellite peak calls with the defined bins. Any peak that falls inside a bin is considered an allele and is displayed in the table with a label corresponding to the name of the bin it is in. If there are error conditions such as peaks outside bins or multiple peaks falling inside a single bin warnings are displayed in the table. This can be disabled by deselecting the "show warning text" option on the right.

The "Duplicate homozygote alleles" checkbox toggles whether homozygote alleles (ie single peaks) are repeated across all columns or not.

The Export button in the main toolbar allows exporting the table in Comma Separated Value (CSV) format for further analysis in other programs such as Microsoft Excel.
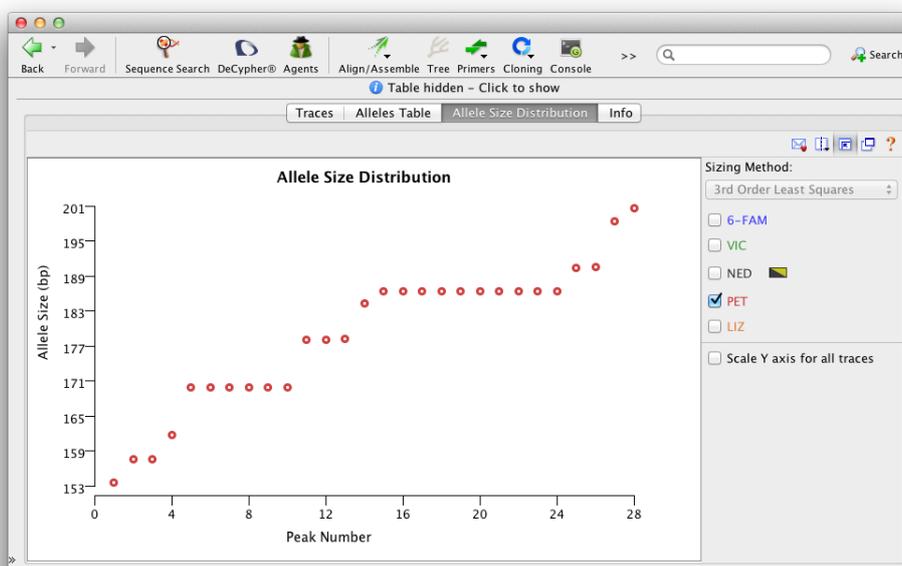
Figure 5: Checking the allele sizes



Figure 6: Showing alleles table

## Locus Document Viewer

The "Locus Document" (Figure 7) is a regular Geneious document. It stores the Locus Information and Bins that you create when analyzing traces in the Trace Viewer to allow flexibility in applying them to different trace files.

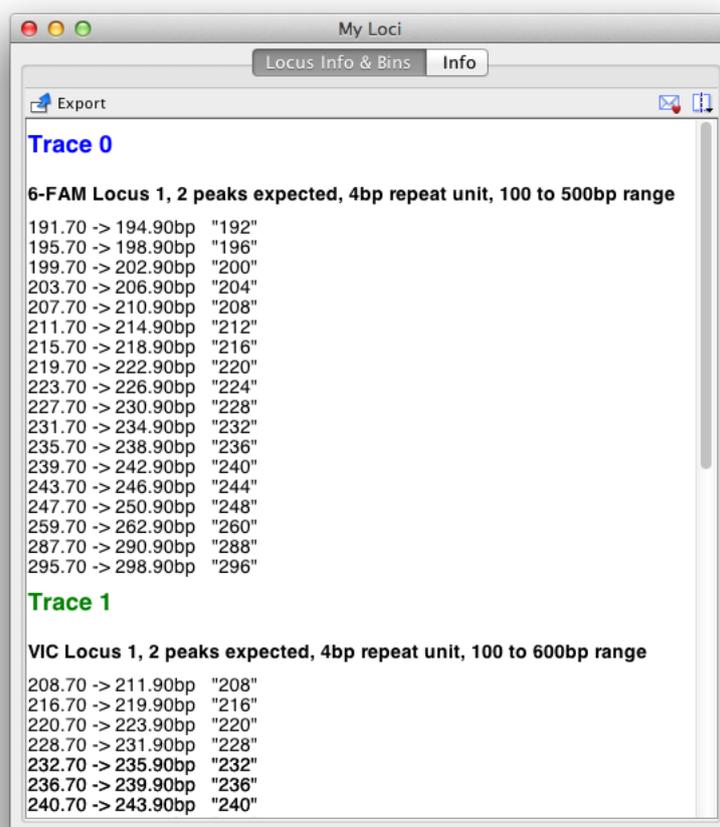You can select a single locus document to view the information it contains.



Figure 7: Locus Document

If you select a Locus Document along with some Trace Documents Geneious will use it as the Locus Information for those traces, prompting you if you have previously been using a different Locus Document with those files.

The "Loci:" combo box option also allows you to select any locus document that you have stored in your Geneious database.