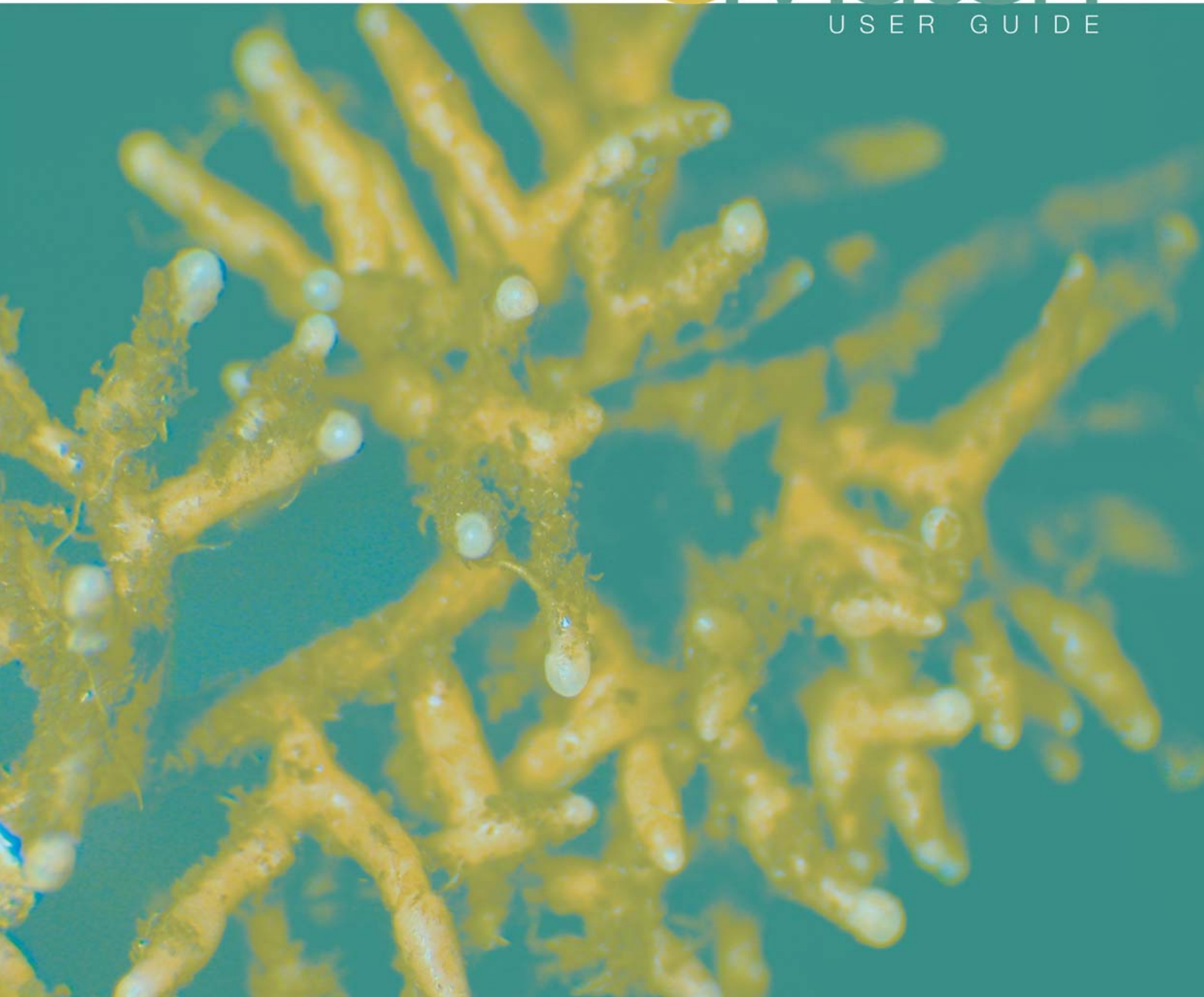


FragMatch

USER GUIDE



Index

<i>Overview</i>	3
<i>Quickstart</i>	4
1. Installing and System Requirements	5
2. Running the Program	5
3. Main Window	6
3.1 Menu	6
3.2 Toolbar	7
3.3 Document View	8
4. Reference Set Window	8
4.1 Input (Value) Areas	8
4.2 Buttons	9
5. Result Window	9
5.1 Result Text Area	9
5.2 Buttons	10
6. Settings	10
7. Sample Run	12

The program and documentation herein is copyright of the authors Tuomas Saari, Susanna Saari and The Macaulay Institute.

Overview:

This program is designed to match sample (=unidentified) DNA fragment patterns to known reference species/individuals. Comparison is based on DNA fragment patterns created using T-RFLP, microsatellite PCR or similar procedures that produce DNA fragment data. Fragment values from sample files are compared to a set of reference fragment patterns for known individuals/species. For example, T-RFLP fragment patterns derived from ectomycorrhizal fungal communities (tips or soil) can be compared to a set of reference fragment patterns derived from identified fruit bodies, or ectomycorrhizal root tips providing a means for species identification.

As output, the program reports all full matches found within a sample file (i.e. all reference species for which all reference fragments were found within the sample file according to the set error limits). If no full matches are found, a number (determined using the settings option) of close matches are shown (see 6. Settings). The output can be viewed in two different formats. The detailed view reports, for each full/partial match, the matched reference fragments along with the found match from the sample. This form of output also includes other information, such as the error limits used, any weighting used and the date and time of the analysis. Alternatively, an overview of the results can be viewed by checking the “view overview” box in the results window. The overview shows a column of all the sample files analysed on the left and each subsequent column to the right represents the individuals in the reference set. For each sample, the number of reference fragments detected per reference unit is shown (e.g. 1_4, means that a match has been detected for 1 out of a possible 4 fragments for the reference unit).

The program accepts sample files in Tab delimited text file format (*.txt). Fragment data from ABI sequence analysis software can be directly exported in this format, although it may be necessary to combine/modify files in Notepad or Microsoft Excel. The use of other file types should be avoided as they may result in abnormal results, program run-time errors, etc.



Quick Start:

First of all, refer to the instructions for installing the program “Installing and System Requirements” on page 5. There is also a pictorial sample FragMatch run demonstrated from page 12 onwards in this guide.

To prepare the sample files, export the DNA fragment data from the sequencer analysis software as Tab delimited text files. If required, modify the files in either Excel or Notepad to ensure that all fragment data relating to a sample are contained within a single file. For example, make sure that all data for each enzyme/primer are included within one file and that each primer/enzyme is clearly identified. Within each sample file the first column should include the enzyme/primer identifiers.

To start running the program, double click on the FragMatch.jar file in the “FragMatch PROGRAM” folder. Open the settings menu where the order of the column containing the fragment size data can be set from the settings menu. From the main program window, open the settings panel from “view” menu and set the column that contains the DNA fragment information. Then, set the name for value sets (=primer/enzyme name) and the indicator for each of the value sets (=either the full enzyme/primer name or preferably a single identifying letter, to be found in the first column from left in your sample files). To save the settings you must close the program and then open it again.

In the program main window, open a sample file you wish to analyse by clicking “File” menu – Open. Several sample files can be analysed simultaneously. Load or create a new Reference set by clicking “Reference” menu – Load/New.

In the reference set window, create the reference units (i.e. known species, strains etc.) and type in the fragment data for each of the primers/enzymes, and set the error limits for analysis. Remember to save each reference unit individually, before creating another one. Finally, save the reference set before running the analysis, remember to type “.rss” after the file name, e.g. “Sample-REFERENCE.rss”.

To run the analysis click the “Compare” menu – Compare (for a single sample) or Compare all (for multiple samples).



1. Installing and System Requirements

Required System specs.

Java2 1.5 Runtime Environment (jre) installed.

(FragMatch has been tested on Sun Microsystems Java2 1.5 Runtime Environment) and Operating System and Computer able to run the jre. The Java2 1.5 Runtime Environment is available as a free download from Sun Microsystems (<http://java.com/en/download/>).

Recommended System specs.

System Memory: 32 MB of free system memory (memory footprint on WindowsXP Operating system is roughly 20MB but dependent on the Java Runtime Environment used and the number of files open).

Installing:

At the moment all of the program is contained in a folder called “**FragMatch**”, this folder contains a further two subfolders “FragMatch PROGRAM” and “EXAMPLE FILES”. Copy all files from the “**FragMatch**” to a folder that you wish to run the program from (the only required program file is **FragMatch.jar**, but the other files are needed to correctly draw the Toolbar etc.). “EXAMPLE FILES” contains sample T-RFLP community data files from an ABI sequencer (.txt files) for practice purposes and a sample database of reference units (.rss file).

To start running the program once installed, double click on the FragMatch.jar file.

2. Running The Program

Microsoft Windows 9X, 2000, XP

Depending on what Java Runtime Environment has been installed on your computer, the program can be executed by double clicking the FragMatch.jar file. If this doesn't work, the following command has to be typed in Command Prompt (Start menu – run – type “cmd” in the text field – press Enter)

```
java -jar [full installation path]FragMatch.jar
```

Where [full installation path] is the directory the program was installed to (see section 1.Installation and System Requirements). For example, if the program was installed to Directory “program” on drive C, the [full installation path] would be “C:\program\ “ (without the quotes) and the full command would be:

```
java -jar C:\program\FragMatch.jar
```

[full installation path] can be left out, if the command is typed in the folder that the program was installed to.

For these commands to work, java has to be set to your path. If this hasn't been done, the command has to be typed in the folder which contains file "java.exe" (most likely the folder that the Java Runtime Environment was installed to).

(NOTE! FragMatch has NOT been tested on following Operating Systems, it should, however, work correctly on any system supporting Java.)

UNIX/Linux

To run the program type the following command:

```
java -jar [full installation path]FragMatch.jar
```

Where [full installation path] is the directory program was installed to (see section 1.Installation and System Requirements). For example, if the program was installed to Directory "program" in Home directory, the [full installation path] would be "~\program\" (without the quotes) and the full command would be

```
java -jar ~\program\FragMatch.jar
```

[full installation path] can be left out, if the command is typed in the folder that the program was installed to.

For these commands to work, java has to be set to your path. If this hasn't been done, the command has to be typed in the folder which contains file "java.exe" (most likely the folder that Java Runtime Environment was installed to).

(Running the program on other operating systems should be similar)

3. Main Window

3.1 Menu

"File" menu is used to load and close sample files (*.txt). Successfully opening a file will load the contents of the file into a new tab in the Document View (see section 3..3 Document View). Closing a file will close and remove the tab created by a successful Open operation. Closing all tabs at once is also possible.

"View" menu is used to set the "Look & Feel" and settings of the program. Either Windows or Java Look & Feel can be used. However Windows Look & Feel might not be available on all systems, if this is the case, the program will default to Java Look & Feel. "Settings" menu item will open settings window.



In this window settings used in any comparison can be modified. To change a value, type a new value and press enter (only one value can be changed at a time). Care should be taken when changing these values as they have a direct impact on the results shown. For more information on settings, see 6. Settings.

[NOTE!] Changing the weight used will increase/decrease (depending on whether the weight is set to above or below 1.0) the weighting of primers/enzymes that have full matches in the sample file].

“Reference” menu is used to load, edit, save and create new reference sets (see section 4. Reference Set Window). Clicking Load/New menu item opens a window in which an existing reference set file can be selected (*.rss) or a new one created (to create a new file, just type the name of the file you wish to create in the “File Name” field, not forgetting to add the “.rss” at the end of the name). Also a track is kept of the 5 most recently used reference set files, clicking any of these file names in menu loads the set. Edit button reactivates the reference set window.

“Compare” menu is used to compare samples. Sample files can be analysed individually or all of the opened files can be analysed simultaneously. Only the active (open during analysis) reference set will be used in the analysis. At least one sample file must exist and a reference set must be open (active) for the analysis to work. After a successful analysis a result window will open (see section 5. Result Window).

“Results” menu is used to browse all open result windows. There are two styles to viewing the results. Firstly a detailed window shows the full or partial matches found to reference units, including each of the reference fragments and their matches in the sample file. Secondly an overview shows a column of all the sample files analysed on the left and each subsequent column represents the individuals in the reference set. For each sample analysed, the number of matching fragments per reference unit is reported. For example, 1_4, denotes a match has been found in the sample for 1 out of a possible 4 fragments for the reference individual in that column. Clicking any of the items will reactivate the result window.

“Help” menu is used to access manual (if manual exists).

3.2 ToolBar

Each toolbar item has an action similar to some menu item.

From left to right toolbar comprises of:

Open file button: See 3.1 Menu under the “File”.

Close file button: See 3.1 Menu under the “File”.

Load/New button: See 3.1 Menu under the “Reference”.

Save button: See 3.1 Menu under the “Reference”.

Edit button: See 3.1 Menu under the “Reference”.



Compare button: See 3.1 Menu under the “Compare”.

Comapre all button: See 3.1 Menu under the “Compare”.

3.3 Document View

“Document View” holds the contents of all loaded sample files in separate tabs. Only one of these tabs is visible at time. The contents of these files can’t be edited in the program.

4. Reference Set Window

4.1 Input (Value) Areas

Name field is used to set the name of a reference unit, name has no limits.

Primer/enzyme value fields are used to set and store values identifying unique reference units (i.e. species, strains etc.). All values can be either floating-point or integer values (integer values will be automatically converted into floating-point values). Values should be separated by “, “(comma followed by space). Name & number of primer/enzyme fields can be set from the “settings” panel under the “view” menu.

Accuracy bound value field is used to set error bounds for the current reference unit. The following format is to be used when setting these bounds:

[Max Value]/[Error Bound]

Where [Max Value] is the highest value (largest DNA fragment) that this error bound applies to, and [Error Bound] is the maximum error (+/-) value that can occur. Both of these values should be positive integer or floating-point values. Each [Max Value]/[Error Bound] pair should be separated by “, “(comma followed by space). For example “250/2, 300/3” would mean that all values of 250 and below would have error bounds of +/- 2 and all values between (and including) 251 and 300 would have error bounds of +/-3.

(NOTE! If a higher value than the highest [Max Value] is found in any of the primer/enzyme value fields, [Error Bound] of the highest [Max Value] is used. Setting too small [Max Value] should be avoided.)

Any of these fields can be left empty, however, it should be noted that setting all fields empty will result in full matches regardless of the actual matches included in the sample file.

4.2 Buttons

The reference set window has 3 buttons and a drop down list. The drop down list is used to select a reference unit in this set. Simply click the



drop down list and choose one of the units (name of the unit is displayed in the list entry).

Save button is used to save all changes made to the current unit selected. Only the currently selected unit will be saved and selecting a new unit before saving will result in loss of all changes made to this unit. This command will NOT save the set to file. Menu or toolbar commands should be used for this purpose.

(NOTE! All saves are final and irreversible!)

New button is used to add a new reference unit into the set. An empty unit is added to the set and activated. No empty units should be left inside a set as they will result in full match regardless of the Sample file being compared.

(NOTE! If the previously selected unit had unsaved changes, all of them will be lost when a new unit is created.)

(NOTE! When a new set is created one empty unit is added automatically.)

Delete button is used to remove any unwanted units. All empty units should be removed before any comparisons are performed.

(NOTE! All deletions are final and irreversible!)

5. **Result Window**

5.1 Result Text Area

Result text area is used to display the comparison results. From top to bottom the Result text comprises of:

Date: Date of analysis.

Sample file: Name of the sample file analysed.

Matching Reference: Name of the reference unit that matched the sample file.

Match type: Whether the match was full or partial.

Weight Used In Comparison: Weight value used in this comparison (see 6. Settings)

(NOTE! Weight will only be shown in the result text area if the weight used was different to the default 1.0.)

“Primer/enzyme name” Value Tested: One of the fragment values identifying the “Matching Reference”. All values in the reference units are tested.

Match Found: The value found in the sample file analysed that matches (within the error limits) the “primer/enzyme” Value Tested.



(if one of the primers/enzymes had no inserted values in the reference set, no tested values should appear in this window)

(NOTE! If the result is a partial match, those values that didn't match, but were the closest matches, have "*" (asterisk) prefix.)

Error Bounds Used: A list of the error limits used (error limits of the matching reference unit).

5.2 Buttons

Checkboxes are used to choose what parts of the result are visible and to be saved in the result text file. The created result file will always only contain the parts selected at the time the result file was created.

Next and Previous buttons are used to cycle through result texts (only available if more than one sample file was tested or more than one match was found)

Delete button is used to remove any unwanted results. Only the currently visible result will be removed.

(NOTE: All deletions are final and irreversible!)

Save as button is used to save the results into a file (.txt). All the results in this window (accessible by consequent Next or Previous button clicks) are saved in the same file. The format of the file created is the same as result text area's (see section 5.1 Result Text Area). Different results are separated by a new line.

(NOTE: Result file is saved as a plain text file, but it is designed to show correctly on Microsoft Windows Operating Systems. Any other operating system may display the file contents incorrectly or distorted.)

6. Settings

Three different settings are currently available; the weight to be used in the comparison, the number (maximum) of partial matches to be shown in the results, and the column in sample files that contains the fragment data. In addition to these three settings, the name and indicator for the value sets (= primers/enzymes) can be set in this window.

The weight used in the comparison works as following:

If one of the primers/enzymes fully matches, in other words, reference set values for some primer/enzyme all have a matching value in the sample file, this primer/enzyme gets more weight when results are computed.

For example if two reference units both have 9 values in common with the sample file being analysed, but only one of the units has a full match for 2 of the three primers/enzymes. If these 2 primers/enzymes have 3 values each, and weight used was 2.0, then one of the reference units (the one that had 2 primers/enzymes fully matching) is considered to



have $3*2 + 3*2 + 3 = 15$ matches instead of 9 and is more likely to be shown as the best partial match. Using a weight higher than 1.0 might be useful if the “most likely” best result of many partial matches needs to be found.

Care should be taken when using a weight higher or lower than 1.0, as this will prevent some partial matches from showing in the result window (see 5. Result Window). The set weight should never be much higher than 1.0 and in most cases should be 1.1 or below.

The number of partial matches selected determines how many (maximum) partial matches will be shown in the result window (see 5. Result window).

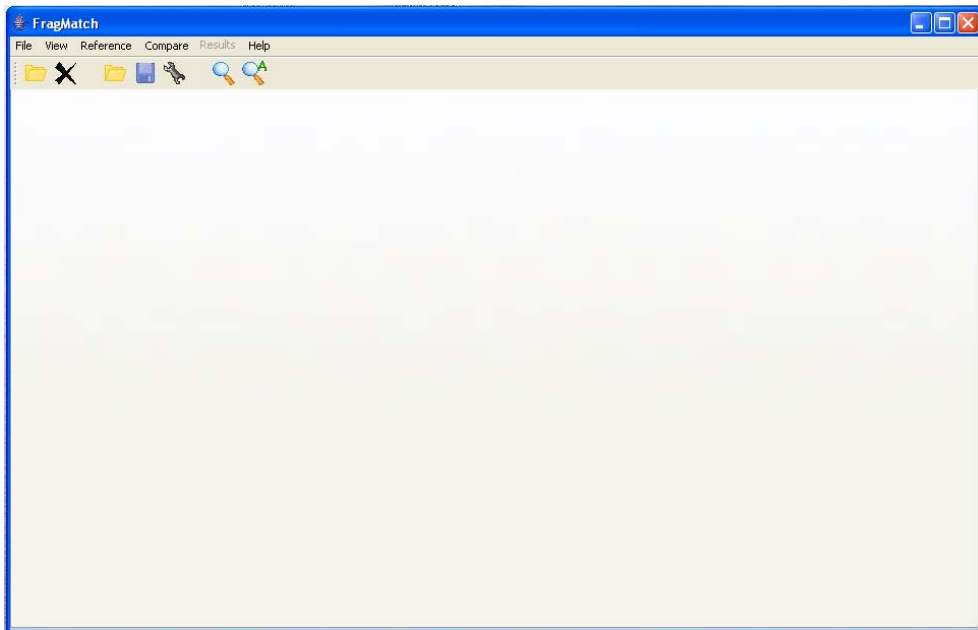
Setting this to a small number will prevent some partial matches from showing in the result window. If you wish to see partial matches to all the reference units, the number of partial matches should be set to equal the number of reference units in the reference set.

The column to be used in comparison should be set to match the order (from left) of the column that contains the fragment data in the sample files.

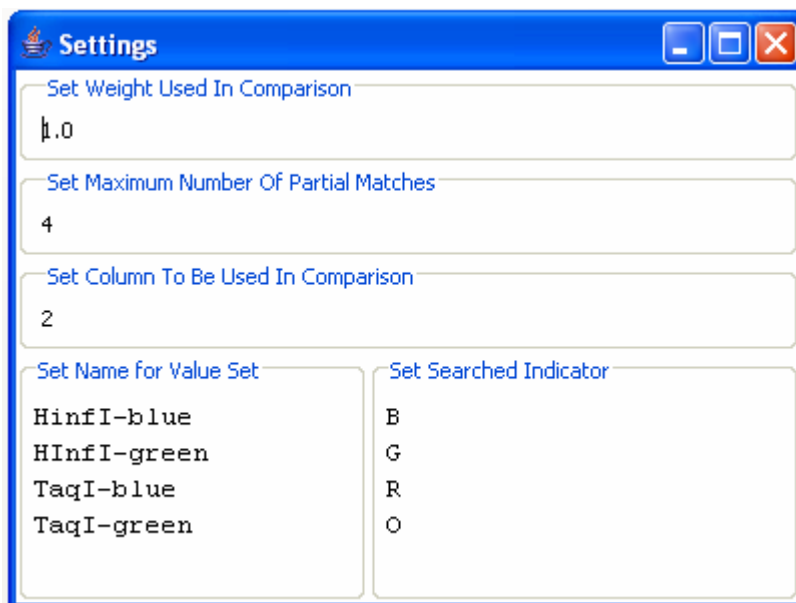
The name for the value sets (=primer/enzyme) should be entered along with the indicator for each of these in the two bottom columns of the settings window. At the moment, the number of value sets has no upper limit. The indicators should be kept short, however the program can handle indicators up to 20 characters long.



7. Sample Run

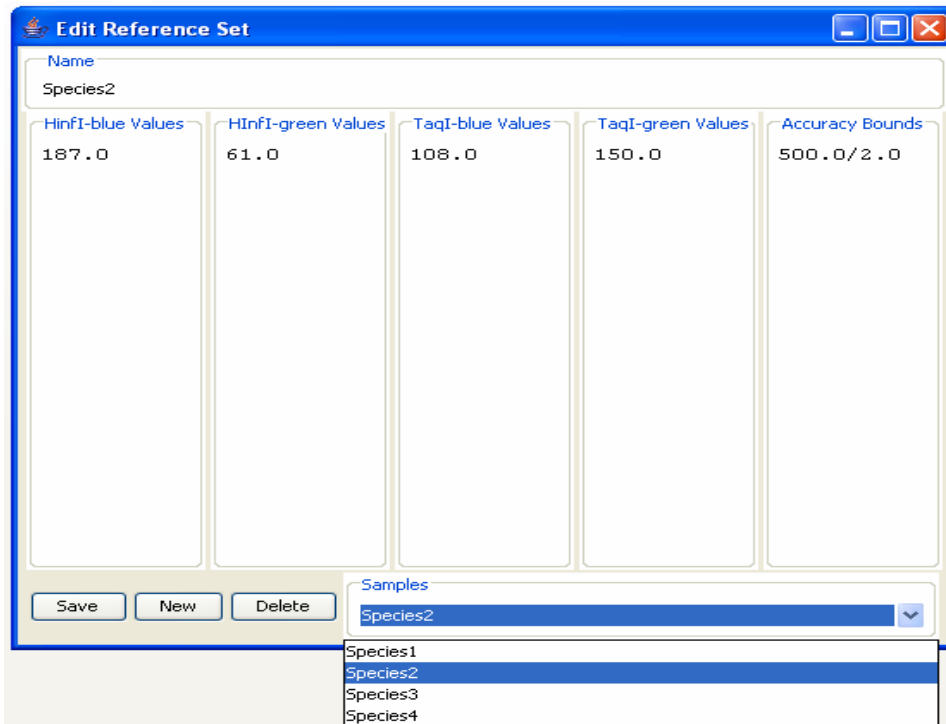


Program main window (appears when you have double-clicked on the FragMatch.jar file in the "FragMatch PROGRAM" folder).

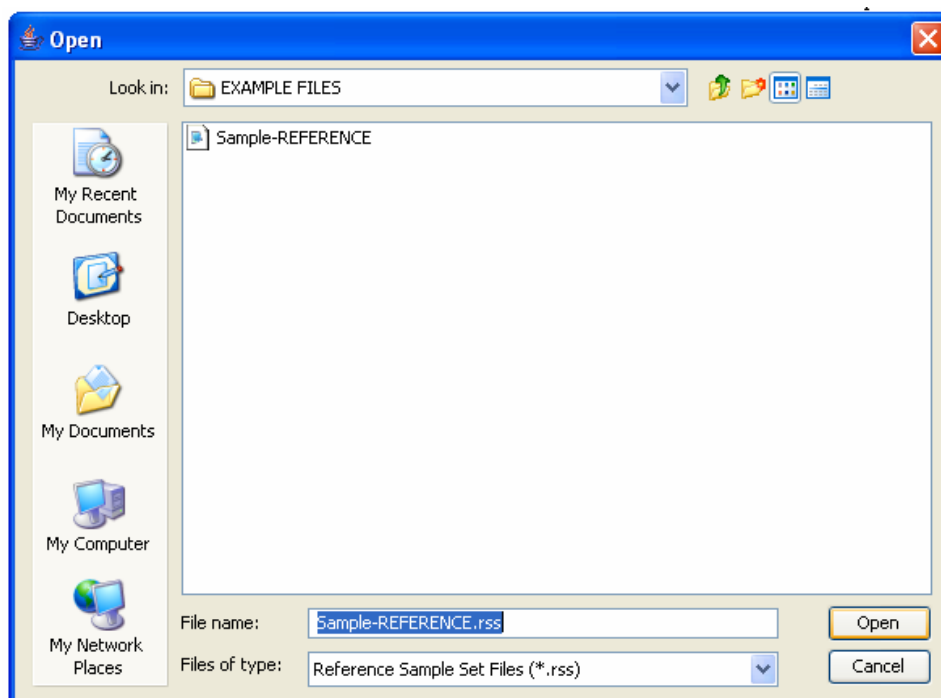


Setting primer/enzyme indicators in the "settings" panel (from "view" menu)

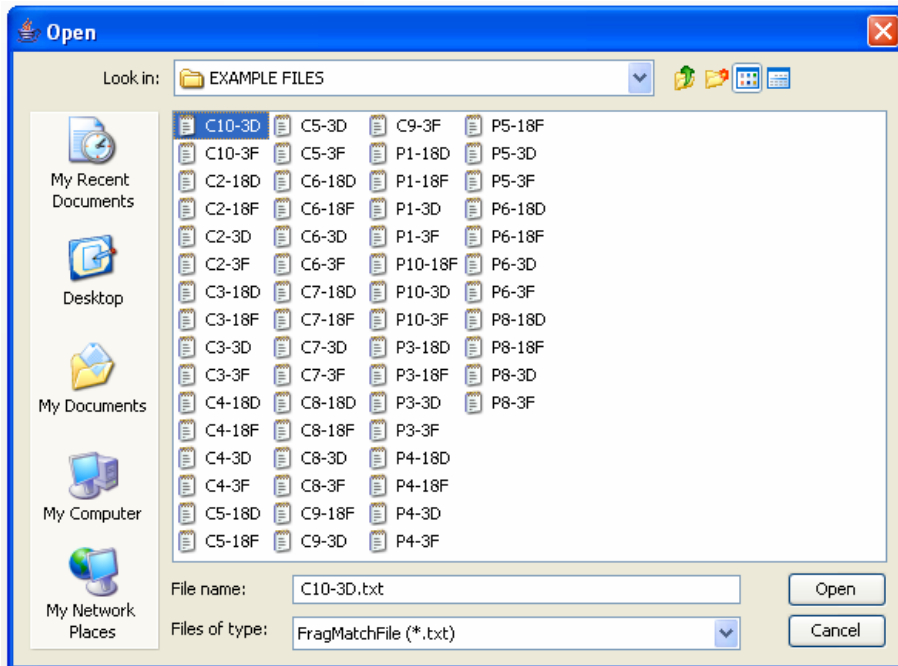




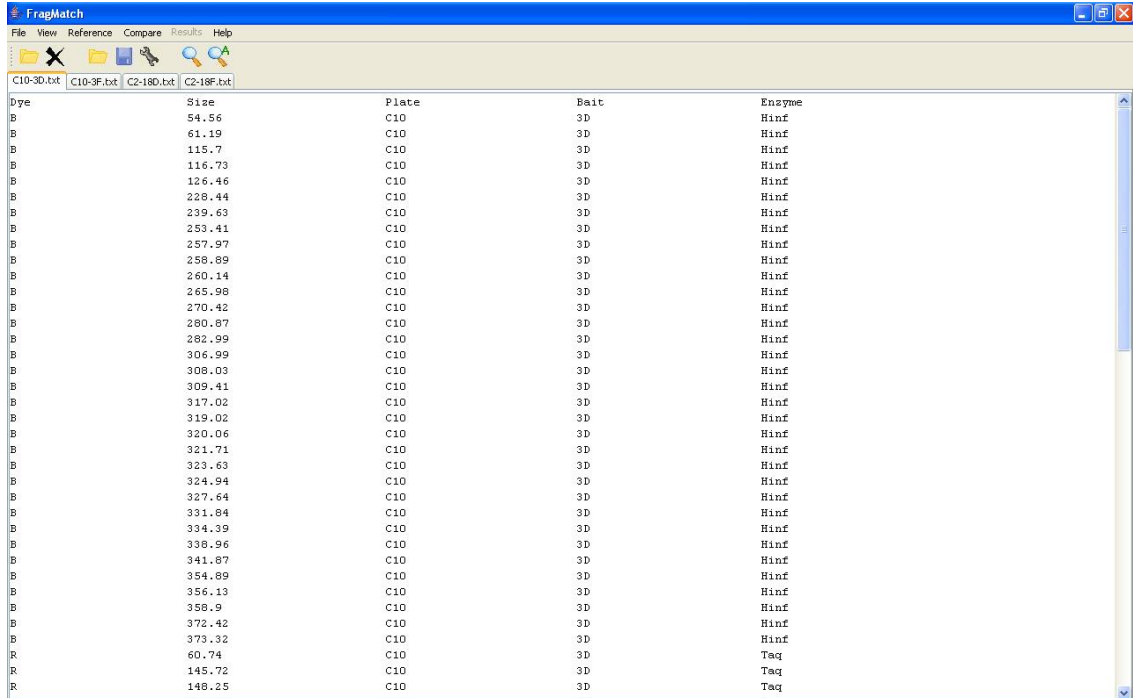
Constructing a new reference data set from “load/new” in the “reference” menu.



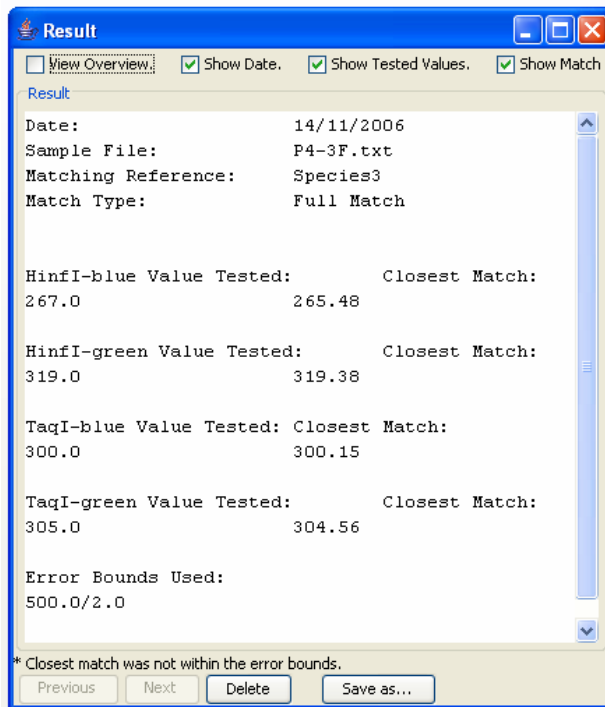
Loading a pre-existing reference set file - “Sample-REFERENCE.rss” set to be loaded (this particular file is provided for you in the “EXAMPLE FILES” folder).



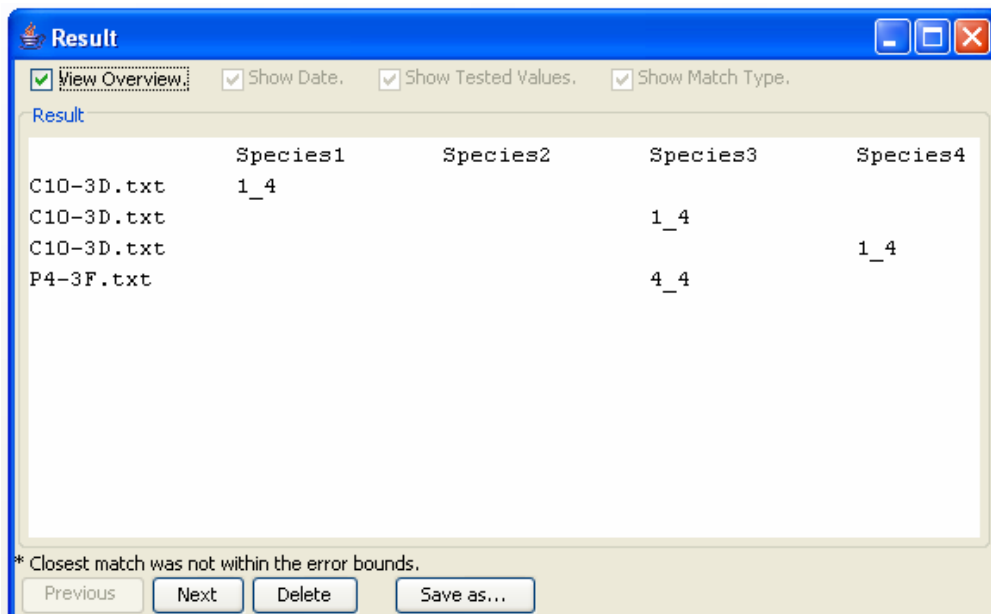
File opening window. File C10-3D.txt set to be loaded as Sample File (these files are provided for you in the “EXAMPLE FILES” folder).



Main program window with four sample files opened ready for analysis. Once you have all your sample files and the reference set (.rss) file open, all you have to do is select either the “compare” or “compare all” function, depending on whether you want to analyse all the sample files separately or if you wish to analyse all of them together.



Detailed results window. A full match to "Species3" of the active reference set is reported.



Overview results window showing a column of sample files on the left and each of the reference units in subsequent columns. For each of the reference units the number of matching fragments over the total number of reference fragments is shown, e.g. 1_4 means that one fragment in the sample file matched to the four reference fragments of that reference unit (e.g. "Species1"). 4_4 denotes a full match (see above for a detailed view of this result).

