IMPACT_S TUTORIAL

Integrated
Multiprogram
Platform to
Analyze and
Combine
Tests of
_SELECTION

IMPACT_S Version: 1.0.0

Starting...5

Open Source GNU GPL

2011/2012

Emanuel Maldonado[1,#], Kartik Sunagar[1,2,#], Daniela Almeida[1,2,#], Vitor Vasconcelos[1,2] and Agostinho Antunes[1,2,*]

1- CIIMAR/CIMAR - Centro Interdisciplinar de Investigação Marinha e Ambiental,
Universidade do Porto, Universidade do Porto, Rua dos Bragas 289, 4050-123 Porto, Portugal.
2- Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal.

# - These authors contributed equally to this work.
* - Corresponding author. Email: aantunes@ciimar.up.pt

NOTICE OF USE:

Before running this program the user should be aware of all the methods included and the statistical tests to be performed. We recommend the user to select the methods that are more appropriate for the dataset at hand. Please see the following resume and references therein for more and detailed information:

Codeml from the PAML package (Yang, 1997; Yang, 2007; Yang et al, 2005) implements several models to detect natural selection. The site specific models addressed in our software (M2a and M8) include Bayes Empirical Bayes (Yang et al, 2005) for identifying positively selected sites. This is the only type of selection identifiable by Codeml, contrasting with other models available in the Datamonkey web-server (http://www.datamonkey.org): Single-Likelihood Ancestor Counting (SLAC) (Kosakovsky Pond & Frost, 2005), Fixed Effects Likelihood (FEL) (Kosakovsky Pond & Frost, 2005), Random Effects Likelihood (REL) (Kosakovsky Pond & Frost, 2005) and Fast Unconstrained Bayesian AppRoximation (FUBAR) (Murrell et al, 2013). The recently proposed Mixed Effects Model of Evolution (MEME) (Murrell et al, 2012) is a state-of-the-art method for detecting sites that evolve under episodic selection pressures, which are often difficult to identify using traditional site-specific methods. According to Kosakovsky Pond & Frost (2005) there is a model more adequate to each dataset and choosing the appropriate method is depending on the question to be addressed as well as the size of the user's dataset. These methods are able to detect negative selection (except MEME) and all should be considered based on the author's recommendations. Thus, the appropriate statistical interpretation follows the tests recommended by the authors of each one of the packages here integrated (for further details please see Kosakovsky Pond & Frost, 2005; Murrell et al, 2013; Murrell et al, 2012; Yang, 1997; Yang, 2007; Yang et al, 2005). Moreover, it has been suggested that the application of all the major selection assessment methods in HyPhy (SLAC, FEL, REL, FUBAR and MEME), followed by the classification of sites detected in consensus as positively selected can minimize false positives (Kosakovsky Pond & Frost, 2005). This feature is available in the Datamonkey webserver named as "Integrative Analysis". Results from codon-level analysis/analyses could be intersected with the sites pointed by TreeSAAP (Wooley et al, 2003) in order to map the sites with amino acid properties under positive or negative selection.

REFERENCES

1. Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22: 1208-1222
2. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genetics 8: e1002764
3. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol 30: 1196-1205
4. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19: 671-672
5. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences : CABIOS 13: 555-556
6. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591
7. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22: 1107-1118
8. *The reader should not be limited to and consult other references. Other can be found at the end of this document.*

Index

Part I – Getting Started

In order to run IMPACT_S you will need the following:

- ✓ Java (version 1.6.0 minimum installed);
- ✓ **TreeSAAP** (curr. version 3.2);
- ✓ **PAML** (Codeml) (curr. version 4.7);
- ✓ **PhyML** (curr. version 2.4.4);
- ✓ **Archaeopteryx** (curr. version 0.9812) jar archive file (formerly known as ATV);
- ✓ **Jmol** (curr. version 13.0) jar archive file;
- ✓ **Gnuplot** (curr. version 4.6).

Unzip the archive (*IMPACT_Svx.x.x.zip*). Note: folder names should not contain spaces to comply with CodeML and PhyML's requirements. Replace spaces with "_". The user only needs to install Java (http://www.oracle.com/technetwork/java/javase/overview/index.html).

After the contents are extracted, you are ready to run IMPACT_S under:

- **Windows OS –** open IMPACT_S folder and click to run **impact_srun_win.bat** file **OR** open a cmd.exe terminal and type **chdir** *location of the IMPACT_S* in the DOS prompt, type "**impact_srun_win.bat"** and hit the Enter key;

- **Linux/UNIX OS –** open IMPACT_S folder and click to run **impact_srun_linux.sh** file **or** open a terminal/console and type **cd** *location of the IMPACT_S directory* in your computer and run by typing "**sh ./impact_srun_linux.sh"** or just **"./impact_srun_linux.sh";**

- **Mac OS –** open IMPACT_S folder and click to run **"impact_srun_mac.command"** file **or** open a terminal and type **cd** *location of the IMPACT_S directory* in your computer and run by entering "sh ./**impact_srun_mac.command"** or just "./**impact_srun_mac.command"**.

Once IMPACT_S is running for the first time, you will be greeted by the **Preferences window** (see section 3.1). Please make sure that all the fields are entered appropriately.

1.1. *Shortcuts and quick access* (main window)

Quit (Menu File>Quit) or Ctrl+Q.

Preferences (Menu Edit>Preferences) or Ctrl+P.

Alignment Editor / PhyML (Menu Tools>Alignment Editor) or Ctrl+Shift+E.
PAML tab; MSA file “View” button.

Alignment Filter (Menu Tools>Alignment Filter) or Ctrl+Shift+F.
TreeSAAP Substs tab; Datamonkey tab; PAML tab; “Filter Alignment” button.

MSA Format Converter (Menu Tools>MSA Format Converter) or Ctrl+Shift+C.

Protein Modeler (Menu Tools>Protein Modeler) or Ctrl+Shift+M.
Results & 3D tab; “Model...” button.

PDB Downloader (Menu Tools>PDB Downloader) or Ctrl+Shift+D.
Results & 3D tab; “Get PDB File...” button.

CSV Viewer (Menu Tools>CSV Viewer) or Ctrl+Shift+V.
Some or Most of the “View” buttons.

Archaeopteryx (Menu Run>Archaeopteryx) or Ctrl+A.
PAML tab; Tree file “View” button.

Codeml (Menu Run>PAML) or Ctrl+C.
PAML tab; “Run PAML” button.

Gnuplot (Menu Run>Gnuplot) or Ctrl+G.
TreeSAAP tab; Property “View” button (needs Plot selected).

Jmol (Menu Run>Jmol) or Ctrl+J.
Results & 3D tab; “View in Jmol” button.

TreeSAAP (Menu Run>TreeSAAP) or Ctrl+T.

Color Schemes (Help>Color Schemes).

About (Help>About).

Part II – Quick Tutorial

IMPACT_S makes use of a **specific directory hierarchy**. The base directory contains **three directories** named:

- **ISystem** – containing IMPACT_S jar archive or executable file and configuration files;
- **IPrograms** – containing all executables/software supplied with the IMPACT_S;
- **IDataSets** – allows the user to create multiple datasets for various analyses. Here data is organized in four directories per dataset (defined by the user), corresponding to each tab in the main window of the program.

At first, when the program starts, the user is presented with the **Preferences window** (available by accessing **Edit>Preferences menu**) where the user is asked to set or reset configuration for further utilization:

- **GUI tab** – the user defines the **dataset and its location**;
- **Programs tab** – the user defines programs for the **specific OS** under which IMPACT_S is running **and their locations.**

In IMPACT_S **main window**, the user is presented **with four principal tabs**:

- **TreeSAAP** (in green);
- **Datamonkey** (in orange);
- **PAML** (in blue);
- **Results & 3D** (in grey).

Provided that the user has the **outputs from TreeSAAP**, **Datamonkey** (CSV files obtained from the website for any or all of SLAC, FEL, REL, MEME and FUBAR methods) **and PAML**, it is possible **to choose/select, in each one of the above-mentioned colored tabs**, the corresponding data in order to be used for further analyses in IMPACT_S.

❖ IMPACT_S TreeSAAP tab

Under the **TreeSAAP tab**, the user can analyze **data from:**

- **Evpthwy directory –** by selecting the **IMPACT_S Evpthwy tab:**

1) **Browse** the directory **Evpthwy –** IMPACT_S will automatically load all files needed from this directory (*Z-SigProp file* and the *SlidingWindow directory*);

2) **Adjust the options** (by default the three categories are selected in each case):

- **Conservative option** – categories 1, 2 and 3;

- **Radical option** – categories 6, 7 and 8.

In case of selecting the option **☑Categories 1,2** or **☑Categories 7,8** IMPACT_S will be limited to these categories. Also it is possible to select **positive ☑(+)** and/or **negative ☑(-)** Z-score values, in order to obtain sites with properties under positive and/or negative selection;

3) **Click "View Most Significant"** near the **Z-SigProp field** to be presented with a list of the most significant properties [P≤0.001(Z-score > |3.09|)] according to the options chosen in step 2;

4) **Select a property from the list** and click **"View"** to see complete data file from this property and the corresponding Gnuplot Options window *(see section 3.7)* from which a graph can be produced;

5) **Click "Retrieve Properties by Range" to produce a table (PBR table)** saved to a file to be used in the final Results & 3D tab for 3D mapping. This file consists of all properties found for each sliding window range, according to the options selected in step 2). **Click "View" to see this table.**

- **Substs directory** – by selecting the **IMPACT_S Substs tab:**

1) **Browse** the directory **Substs –** IMPACT_S will automatically load all files needed from this directory (*SynSubs.txt* and *NSynSubs.txt file*);

2) **Adjust the options** (by default the three categories are selected in each case):

- **Conservative option** – categories 1, 2 and 3;
- **Radical option** – categories 6, 7 and 8.

In case of selecting the option ☑**Categories 1,2** or ☑**Categories 7,8** these will be defined as the limiting categories. Also it is possible to select **positive ☑(+)** and/or **negative ☑(-)** Z-score values (P≤0.001), in order to obtain sites with properties under positive and/or negative selection;

3) **Select "Branches" and/or "Condon & AA"** if branches and/or codon information is necessary in final table. **Click "Retrieve Properties by Site" to produce** (or rebuild, according to selected options) **the PBS table** saved to a file to be used in the final Results & 3D tab for 3D mapping. This table consists of all properties found for each site, according to the options selected in step 2). **Click "View" to see this table**;

4) **Click "Statistics" to have access to property counts and total counts. Click "Branches Statistics"** in case of the table was built with option "Branches" selected, to access branches statistics. Two files are produced one for Node-Species (differentiated by having "bns" in its name) and another for Node-Node (differentiated by having "bnn" in its name). **To see these files select which file to be displayed in the CSV Viewer window in the "View File" box**;

5) **Click "Filter Alignment" to generate an alignment/image consisting of only the sites (columns) retained as most significant from the 3rd step**. A new window will appear to select the alignment file and other options that allow for instance, to place gaps among non-consecutive sites.

Both the tables/files (PBR and PBS) are automatically placed in the Results & 3D tab, as they are generated. The user can choose to use both, but at least one is required.

❖ IMPACT_S Datamonkey tab

To use Datamonkey tab, the user must **run all analyses in Datamonkey web-server** and **download all the corresponding CSV files for each model needed** (SLAC, FEL, REL, MEME, FUBAR). The user can now generate new tables related to positively and negatively selected sites using IMPACT_S, without relying on the Datamonkey web-server:

1) Load all the CSV files for any of the models (not necessarily all) in each corresponding tab using the **"Browse" button;**

2) Verify the significance value below editing or selecting from the box for each model;

3) Click "Retrieve" (or **"Retrieve All"**, at once) in order to generate all the positive and negative selection tables for each browsed model (CSV file) and chosen significance value. Click the wide **"View" button** to access both tables. **To see these files select which file to be displayed in the CSV Viewer window in the "View File" box;**

4) Choose from "Selection" field, if common sites table is meant to consist of positive or negative selection;

5) Verify which models are selected in "Choose Models" field. This will allow the user **to generate a table of all common sites** from only the selected (and existing CSV files) models;

6) Click "Combine" to generate the common sites table, at any time (and to update if previous options have changed). **Click "View" to see this table**;

7) In the **"Select Data (P/N):" field/board** select which data table will be used **to generate an alignment/image consisting of only the sites (columns) retained as most significant** by clicking **"Filter Alignment"** button and/or which data table will be used for Results & 3D tab for 3D mapping by clicking **"Send to Results & 3D tab"**. Feel free to change step 4) option in case a positive or negative table from any model is of interest. In the case of a different common sites table is required, repeat steps from either 2), 4) or 5) to 6), depending on the needed change.

❖ PAML tab

If the user has already run Codeml analyses externally, this tab can be used for processing the results. Alternately, the user can easily run these site-specific analyses using IMPACT_S.

1) Click "Browse" to select the codeml results file.

2) Select which models to retrieve information, **either ⊙"Options(M7,M8)"**, **⊙"Options(M1,M2)"**, **⊙"Options(M0,M3)"**, **or ⊙"All(M0-M8)"**.

3) Click the wide "View" button to view the Likelihood and Ratio tables, built from the extracted information regarding the selected option in step 2) – each result is saved to a file. **To see these files select which file to be displayed in the CSV Viewer window in the "View File" box.**

4) Select the significance to apply as filter from the "Select" field. **Either P>95% (*) or P>99% (**)**. The first option encompasses both single star (*) and double star values (**), and the second option only values marked with double star (**).

5) Click "Extract BEB" to extract the Bayes Empirical Bayes table to a file according to step 2). Click the near **"View"** button to view data. This file is automatically placed in the Results & 3D tab. The user can create BEB tables for both models M2 and M8, which will be automatically placed in Results & 3D tab. **Change "BEB Results" file name** (*this can be done quickly by using the* **"GFN" small button**) and repeat step 2), 4) and 5).

6) Click the "Filter Alignment" to generate an alignment consisting of only the sites (columns) retained as most significant from step 5).

❖ Results & 3D tab

The results of various selection analyses (PAML and/or TreeSAAP and/or Datamonkey) can be integrated under this tab. Sites detected in common by various analyses as positively selected can be integrated under this tab. These sites can be further mapped onto the three-dimensional structure of the protein.

The integration of results from TreeSAAP, Datamonkey and/or PAML can be accomplished under the "Merge Results" section.

- **⊙Positive selection – results of TreeSAAP, Datamonkey and/or PAML are required (**at least two methods required): Select by ticking the checkboxes ☑TreeSAAP and ☑Datamonkey or ☑TreeSAAP and ☑PAML;
- **⊙Negative selection – only TreeSAAP and Datamonkey**.

Under the **3D View section**, the user can **retrieve a PDB structure** in two ways. Either through homology modeling using Swiss-Model, when an already solved crystal structure for the protein under study is absent, by using the **"Model..." button.** Alternately, the user can retrieve PDB files from RSCB-PDB or Swiss-Model Template Library (SMTL) using the **"Get PDB File..." button**. **This file is automatically placed in the PDB field below.**

1) Click "Browse" to select a (one-chain) PDB file located in your computer. IMPACT_S will automatically retrieve information from this file. Drag/Click the mouse over the **"[?]"** mark. Use the **Modeled Range** to change or limit the range of residues mapped in the **3D structure**.

2) In the "Select Color Scheme" field, select the color scheme to be applied. Each scheme represents a specific configuration applied to each specific case:

- **TreeSAAP** – 5 options;
- **Datamonkey** – 1 option (were any file can be used, provided from the Datamonkey tab – step 7));
- **PAML** – 3 options. The user will need two BEB tables in order to be possible to use options "PAML BEB (M2+8,2)" and "PAML BEB (M2+8,3)" for both Model 8 and Model 2. See a quick reminder from the menu **Help->Color Schemes**.

3) Click "View Scheme" to view or modify the selected color scheme *(see section 3.5).*

4) Select the "Shape" of the molecule to be presented:

☑ **"Only" –** with or **without** amino-acids depicted;

☑ **"Restrict" –** with all or **just** the significant residues;

☑ **"Small" –** with large or **small** size.

5) Finally, **click the "View in Jmol" button to map** and view the sites from the data associated to the selected color scheme, mapped onto the 3D structure from the selected PDB file. Use this button to update the 3D view, whenever options are changed from previous steps.

❖ Other Tools

IMPACT_S provides other helpful tools:

- **Change dataset or IMPACT_S configuration: (Edit->Preferences)** (see section 3.1);

- **Edit or view Multiple Sequence Alignments and/or run PhyML:** MSA Editor/Viewer **(Tools->Alignment Editor)** (see section 3.9);

- **Produce or highlight alignments with only significant sites:** Alignment Filter **(Tools->Alignment Filter)** (see section 3.6);

- **Convert Multiple Sequence Alignments:** MSA Format Converter **(Tools->MSA Format Converter)** (see section 3.10);

- **Homology modeling:** Protein Modeler **(Tools->Protein Modeler)** (see section 3.3);

- **PDB retriever:** PDB Downloader **(Tools->PDB Downloader)** (see section 3.4);

- **View CSV files:** CSV Viewer **(Tools->CSV Viewer)** (see section 3.8);

- **Plotting TreeSAAP Sliding Window properties:** Gnuplot (Options) (**Run->Gnuplot)** (see section 3.7).

For further detailed information about each option available in IMPACT_S, please refer to the following sections, which are accompanied with illustrative figures.

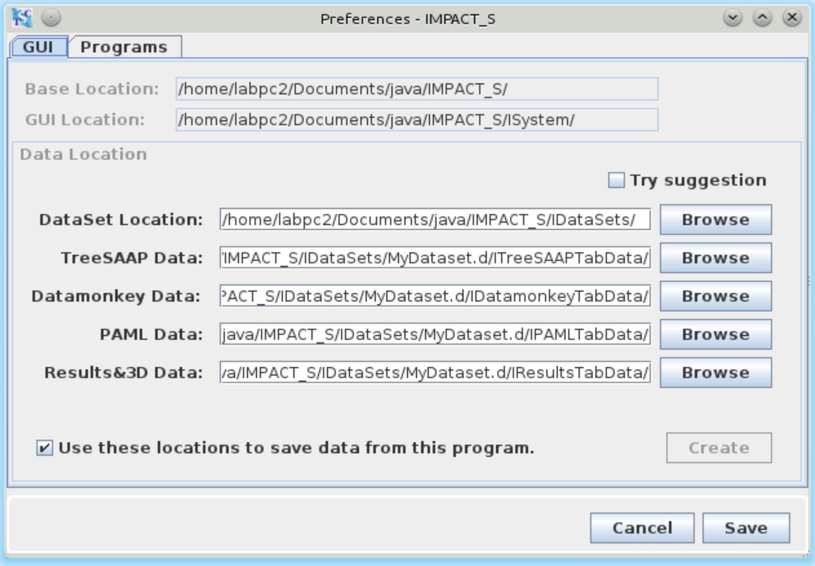
Part III – IMPACT_S Manual

3.1. Preferences window.

When IMPACT_S is initialized for the first time, the user will find the preferences window open, which indicates that the location(s) of program(s) are not configured. Only if these directories are modified, you will see this this window again during the subsequent sessions.

This window presents two tabs:

- **"GUI" tab** – here you find the path to programs that are currently being used. These are seen in non-editable fields**.** Below there is a section that allows the user to choose a working directory (preferably inside the ***IDataSets*** directory), which will serve as a container for all the files resulting from this program and also files needed as input for this program. The user can create as many DataSets as needed. This, however, requires care on the user behalf, because one directory is being used at a time by the program, so the user must always come to the preferences window to change to the required DataSet.



Preferences - IMPACT_S

Base Location: /home/labpc2/Documents/java/IMPACT_S/

GUI Location: /home/labpc2/Documents/java/IMPACT_S/ISystem/

Data Location

□ Try suggestion

DataSet Location: /home/labpc2/Documents/java/IMPACT_S/IDataSets/

TreeSAAP Data: 'IMPACT_S/IDataSets/MyDataset.d/ITreeSAAPTabData/

Datamonkey Data: ⁾ACT_S/IDataSets/MyDataset.d/IDatamonkeyTabData/

PAML Data: java/IMPACT_S/IDataSets/MyDataset.d/IPAMLTabData/

Results&3D Data: /a/IMPACT_S/IDataSets/MyDataset.d/IResultsTabData/

☑ Use these locations to save data from this program.

The *IDataSets* structure consists of the following:

IMPACT_S/IDataSets/<mydataset>/

IMPACT_S/IDataSets/<mydataset>/ITreeSAAPTabData/

IMPACT_S/IDataSets/<mydataset>/IDatamonkeyTabData/

IMPACT_S/IDataSets/<mydataset>/IPAMLTabData/

IMPACT_S/IDataSets/<mydataset>/IResultsTabData/

Each subdirectory of <mydataset> is always relative to each (name related) tab, of the main window; hence for instance, the ***ITreeSAAPTabData* directory** is relative to **“TreeSAAP” tab**. This means, that whenever the user browses for some file, IMPACT_S will readily open the correct location from the tab related directory.

Use:

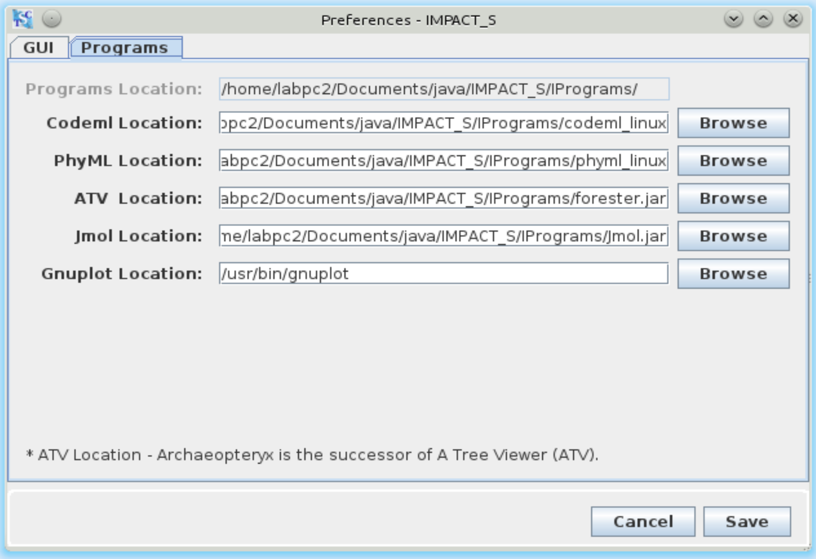
1) The **“DataSet Location”** field (of the base data sets directory) – to automatically adjust all paths below;

2) **☑“Try suggestion”** – to see or use a suggestion of a datasets directory structure;

3) **“Create” button** – to create these directories, once you are sure that you need them;

4) **☑“Use these locations to save data from this program”** – to enable or disable the use of these directories. If this is disabled, by default, the program will save all produced files in IMPACT_S base directory. In this case, when browsing files, IMPACT_S will remember the location of the last opened file. Use the **“Browse” buttons** to choose directories.

- **“Programs” tab** – the user will have to verify/select all the programs. All the IMPACT_S **executables** are located in the ***IPrograms* directory**.



Programs Location: /home/labpc2/Documents/java/IMPACT_S/IPrograms/

Codeml Location: ɔpc2/Documents/java/IMPACT_S/IPrograms/codeml_linux

PhyML Location: abpc2/Documents/java/IMPACT_S/IPrograms/phyml_linux

ATV Location: abpc2/Documents/java/IMPACT_S/IPrograms/forester.jar

Jmol Location: ne/labpc2/Documents/java/IMPACT_S/IPrograms/Jmol.jar

Gnuplot Location: /usr/bin/gnuplot

* ATV Location - Archaeopteryx is the successor of A Tree Viewer (ATV).

Table 1- Operating Systems

	OPERATING SYSTEMS		
IMPACT_S	**Windows OS**	**Linux/UNIX**	**Mac OS**
CODEML	codeml47.exe	codeml47_linux	codeml47_macosxIntel
PHYML	phyml_win32.exe	phyml_linux	phyml_macOSX
ARCHAEOPTERYX	forester.jar		
JMOL	Jmol.jar		
GNUPLOT	gnuplot46.exe	gnuplot46_linux	gnuplot46_macosxIntel

1) The user must refer to the table above and **browse** to select the appropriate **executable**.
2) The java archives (".jar") can be replaced/updated in order to upgrate to newer/different versions.
3) Once the locations are configured appropriately, press the **"Save" button** to save changes. **Or otherwise**, select **"Cancel"** to discard any changes and close window.

3.2. IMPACT_S Main window

There are four tabs in the main window: **"TreeSAAP"**, **"Datamonkey"**, **"PAML"** and **"Results & 3D"**.

3.2.1. TreeSAAP tab

In order to use this tab the user needs to first run the TreeSAAP software. When the run is over, the user is able to use both underlying tabs:

Adjust the IMPACT_S default options:

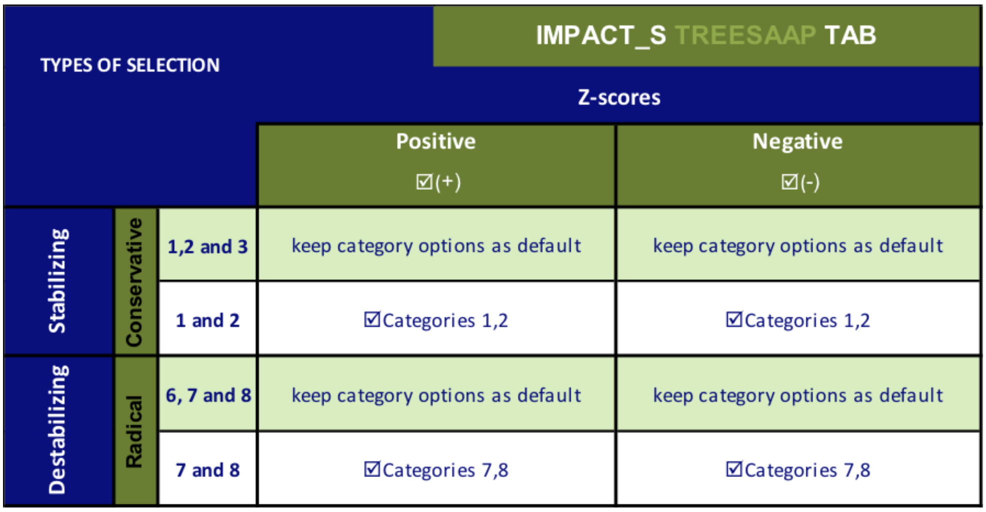
- Conservative option – includes the categories 1, 2 and 3 are assumed by default. The user can limit this to the categories (1 and 2) by selecting the option **☑Categories 1,2.**

- Radical option – includes the categories 6, 7 and 8 and are assumed by default. The user can limit this to the categories (7 and 8) by selecting the option **☑Categories 7,8**.

In both cases, it is possible to limit or extend the information retrieved using the options **positive ☑(+)** and/or **negative ☑(-)** that are relative to the Z-scores. The value underlying these options is Z-score > |3.09| for a P≤0.001. Hence, when selecting these options **☑(+)** and/or **☑(-)** the user is requiring positive (Z-score > 3.09) and/or negative (Z-score < -3.09) Z-score values, corresponding to sites with properties under positive and/or negative selection, respectively.

Synthesizing, to infer about positive and/or negative, stabilizing and/or destabilizing selection **select the following options,** according to the table:

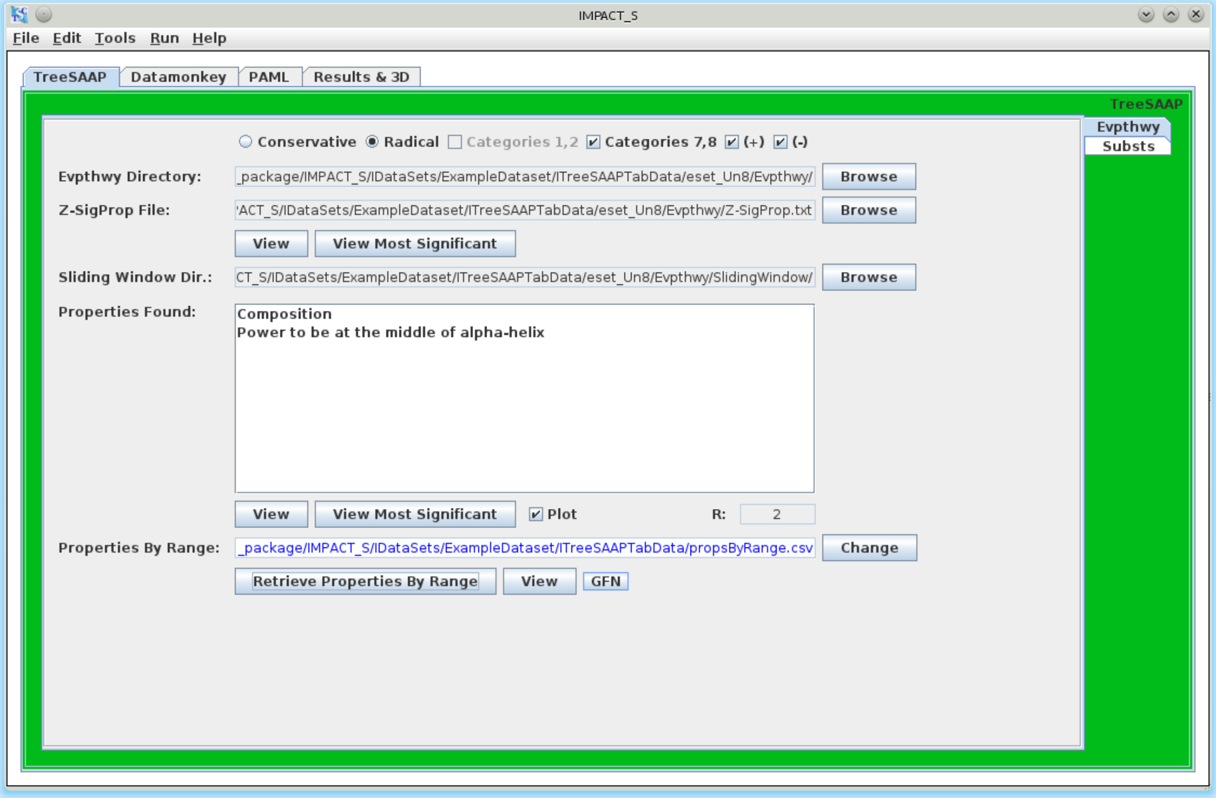
Table 2 – Types of Selection in TreeSAAP



TYPES OF SELECTION			IMPACT_S TREESAAP TAB	
			Z-scores	
			Positive ☑(+)	Negative ☑(-)
Stabilizing	Conservative	1,2 and 3	keep category options as default	keep category options as default
		1 and 2	☑Categories 1,2	☑Categories 1,2
Destabilizing	Radical	6, 7 and 8	keep category options as default	keep category options as default
		7 and 8	☑Categories 7,8	☑Categories 7,8

- **Evpthwy tab –** for *Evpthwy* directory from TreeSAAP software.

1) Select the Evpthwy tab;



○ Conservative ● Radical □ Categories 1,2 ☑ Categories 7,8 ☑ (+) ☑ (-)

Evpthwy Directory: _package/IMPACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/eset_Un8/Evpthwy/

Z-SigProp File: 'ACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/eset_Un8/Evpthwy/Z-SigProp.txt

Sliding Window Dir.: CT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/eset_Un8/Evpthwy/SlidingWindow/

Properties Found:
Composition
Power to be at the middle of alpha-helix

☑ Plot R: 2

Properties By Range: _package/IMPACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/propsByRange.csv

2) Browse the **directory "Evpthwy" –** IMPACT_S will automatically load all files needed from this directory (*Z-SigProp file* and the *SlidingWindow directory* containing the biochemical properties used);

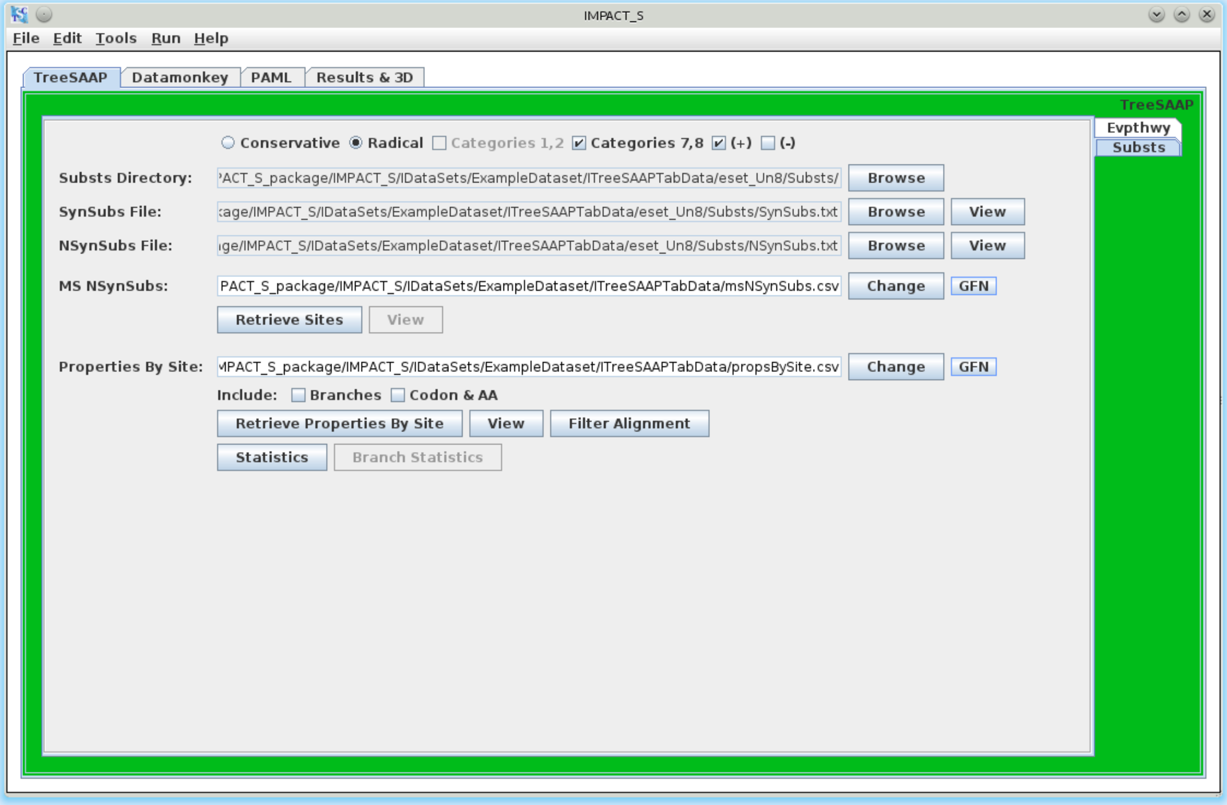
3) Adjust the IMPACT_S default options: see orange text box above. This tab comes by default with **⊙"Radical"** option selected and options **positive ☑(+)** and **negative ☑(-)** selected.

4) The Properties By Range (PBR) table can be generated using the **"Retrieve Properties By Range" button**. This table contains all the ranges produced by TreeSAAP, which are most significant. These ranges are followed by a column of all the biochemical properties that fall or are found in each range, and is preceded by a column ("Total") that presents the number of total properties found in each range. The user can **view this table** using the **"View" button** next to the previous button.

5) Across all fields in this tab the user can visualize the contents of the tables using the corresponding **"View"** buttons, or the **"View Most Significant"** buttons, for the most significant data relative to the same files or fields. **"Save As..." option from the File menu** can be used to save this information to file. The files produced for viewing are only temporary and will not be saved and hence the user can save them, if required.

- **Substs tab –** for *Substs* directory from TreeSAAP software.

1) Select the Substs tab;



○ Conservative ● Radical □ Categories 1,2 ☑ Categories 7,8 ☑ (+) □ (-)

Substs Directory: ACT_S_package/IMPACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/eset_Un8/Substs/

SynSubs File: age/IMPACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/eset_Un8/Substs/SynSubs.txt

NSynSubs File: ge/IMPACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/eset_Un8/Substs/NSynSubs.txt

MS NSynSubs: PACT_S_package/IMPACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/msNSynSubs.csv

Properties By Site: MPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/propsBySite.csv

Include: □ Branches □ Codon & AA

2) Browse the **directory "Substs" –** IMPACT_S will automatically load all files needed from this directory (*SynSubs.txt and NSynSubs.txt files*). In this tab all operations are relative to the *NSynSubs.txt* file. The *SynSubs.txt* file is available only for viewing.

3) Adjust the IMPACT_S default options: see text box above. This tab comes by default with ⊙**"Radical"** option selected and option **positive ☑(+)** selected.

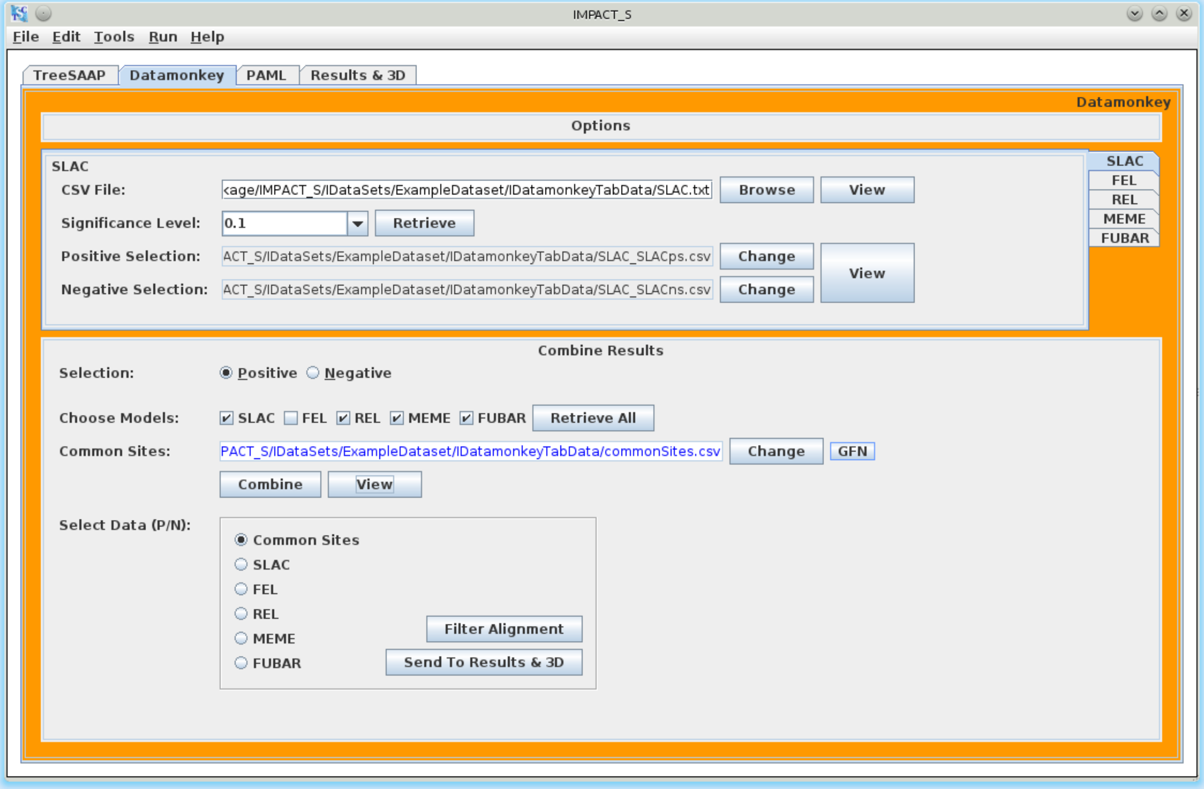
4) It is possible to obtain an extraction of the most significant rows from this file, by clicking the **"Retrieve Sites" button**. In this case all the properties columns and Z-scores are verified to get all the rows that have at least one significant property. This is made according to the options selected;

5) Most importantly and finally **IMPACT_S is able to create a** Properties By Site (PBS) **table** containing all the most significant sites and all the corresponding properties per site. **This table is obtained through the "Retrieve Properties By Site" button**. Use this button whenever this table needs to be updated or reconfigured. **This table can have several configurations.** The most simple is comprised of **three columns (Codon, Total and Properties)**. Other configurations are available in case the user wants to include branches information with the **☑"Branches" option** and/or to include the codons and amino acids information with the **☑"Codon & AA" option (four additional columns).** This additional information is placed between the Codon and the Total columns. In these cases, the sites presented are not unique, instead they are repeated, presenting differential information, as opposite to the three column configuration. **This table can then be used for the** Alignment Filter **functionality (see section 3.6).**

6) **Along with the** PBS table**, statistics are produced, were counts of the number of same properties, the number of unique properties and total properties are presented in a separate table.** This can be accessed through the **"Statistics" button**. Also for the case of branches, statistics are presented to count the number of properties per node. **This is divided into two tables** one for **Node-Node cases** and another for the **Node-Species cases**. Access to this is provided through the **"Branch Statistics" button**, which presents both contents in the **CSV Viewer (see section 3.8).** To do this, the user needs to select the **☑"Branches" option** and click the **"Retrieve Properties By Site"** to update the **PBS table** and hence retrieve the statistics.

3.2.2. Datamonkey tab

To use this tab, the user first needs to run the required analyses on the Datamonkey web-server at (www.datamonkey.org): SLAC, FEL, REL, MEME and/or FUBAR and save all the CSV resultant files.



SLAC

CSV File: <age/IMPACT_S/IDataSets/ExampleDataset/IDatamonkeyTabData/SLAC.txt

Significance Level: 0.1

Positive Selection: ACT_S/IDataSets/ExampleDataset/IDatamonkeyTabData/SLAC_SLACps.csv

Negative Selection: ACT_S/IDataSets/ExampleDataset/IDatamonkeyTabData/SLAC_SLACns.csv

Combine Results

Selection: ● Positive ○ Negative

Choose Models: ☑ SLAC □ FEL ☑ REL ☑ MEME ☑ FUBAR

Common Sites: PACT_S/IDataSets/ExampleDataset/IDatamonkeyTabData/commonSites.csv

Select Data (P/N):
● Common Sites
○ SLAC
○ FEL
○ REL
○ MEME
○ FUBAR

1) There are five sub-tabs referring to each Datamonkey method: SLAC, FEL, REL, MEME and FUBAR. The user selects (by using the **"Browse"** buttons) several CSV files generated and downloaded from the Datamonkey web-server, corresponding to each one. The user can then choose different significance levels for each method, even by typing in any valid value.

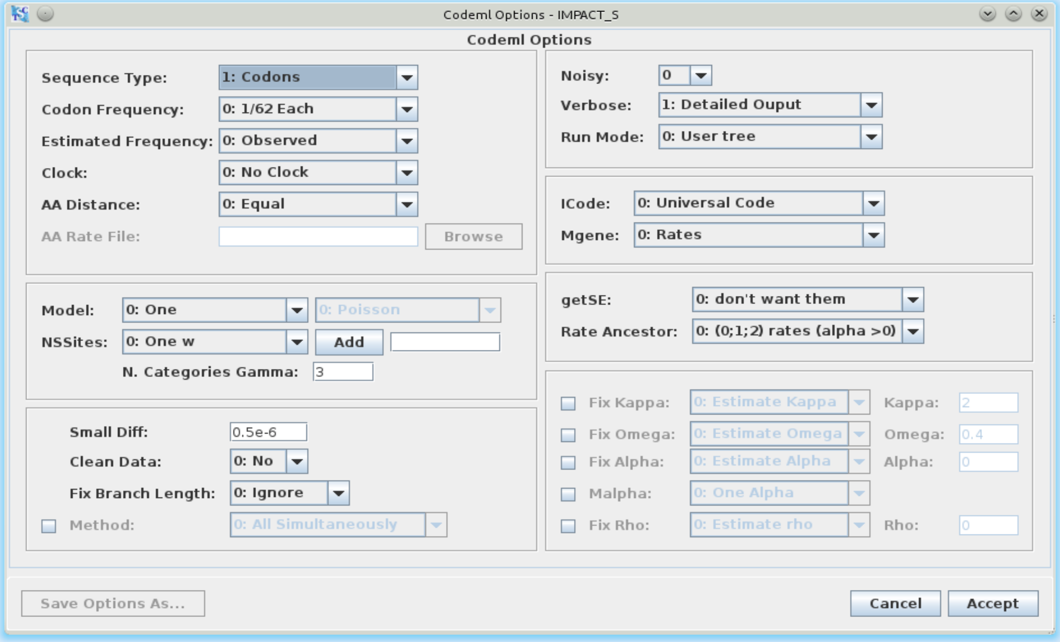
2) The user can retrieve positive and negative selection tables by using the **"Retrieve"** button, or using the **"Retrieve All"** button for all available CSV files and selected methods. This information can be inspected (and saved) by using the **"View"** buttons available from the respective method tabs.

3) From the selected methods the user creates the **Common Sites** table for either "Positive" or "Negative" selection, depending on the user selection and on the existence of results. To this, the user must first have retrieved all tables and then press the **"Combine"** button. Results from this operation can be viewed (and saved) through the **"View"** button. Finally, in the **"Select Data (P/N)"** board the user can perform an **Alignment Filter** (see section

3.6) on the selected data table and choose which table to be used in the **Results & 3D tab** (see section 3.2.4.). Both options have into account, if the required data is for "Positive" or "Negative" selection.

3.2.3. PAML tab

This tab consists of two sections: i) the section for running Codeml; and ii) the section for using the Codeml results, when these are already available.



Codeml Options - IMPACT_S

Codeml Options

Sequence Type: 1: Codons
Codon Frequency: 0: 1/62 Each
Estimated Frequency: 0: Observed
Clock: 0: No Clock
AA Distance: 0: Equal
AA Rate File:

Model: 0: One 0: Poisson
NSSites: 0: One w
N. Categories Gamma: 3

Small Diff: 0.5e-6
Clean Data: 0: No
Fix Branch Length: 0: Ignore
□ Method: 0: All Simultaneously

Noisy: 0
Verbose: 1: Detailed Ouput
Run Mode: 0: User tree

ICode: 0: Universal Code
Mgene: 0: Rates

getSE: 0: don't want them
Rate Ancestor: 0: (0;1;2) rates (alpha >0)

□ Fix Kappa: 0: Estimate Kappa Kappa: 2
□ Fix Omega: 0: Estimate Omega Omega: 0.4
□ Fix Alpha: 0: Estimate Alpha Alpha: 0
□ Malpha: 0: One Alpha
□ Fix Rho: 0: Estimate rho Rho: 0

- **Running Codeml –** *Options* section from PAML tab.

1) The user selects the Multiple Sequence Alignment (**MSA**) and the **Tree** file corresponding to the MSA. Use the **"View"** button in line with the **MSA** file to view or edit the alignment (by opening the Alignment Editor (see section 3.9)) and produce a corresponding Tree file (by running PhyML). Use the **"View"** button in line with the **Tree** file field to view or edit the tree file which opens in Archaeopteryx.

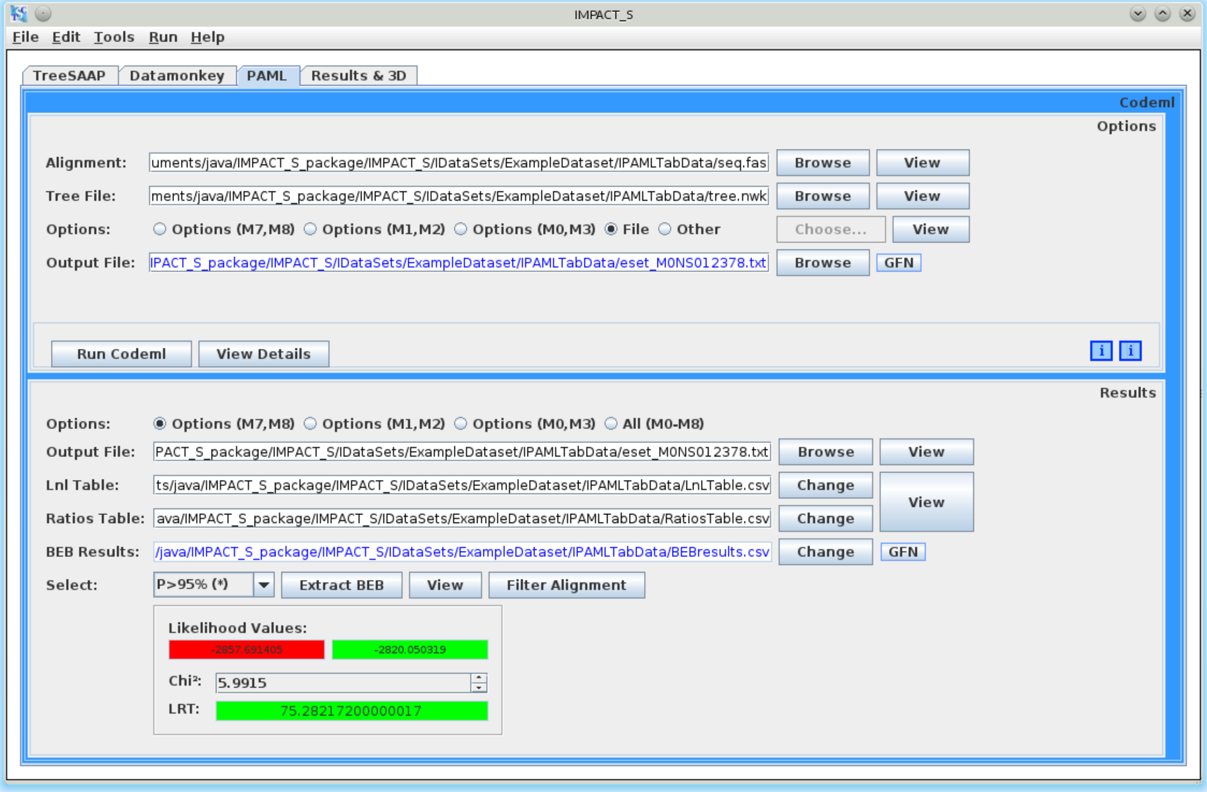
2) Select the option for running Codeml. The options introduced for running

Codeml consist of 3 default sets of parameters and two extra possibilities.

1. **⦿"Options (M7,M8)"** Codeml options for running with Codeml NSsites 7 and 8.
2. **⦿"Options (M1,M2)"** Codeml options for running with Codeml NSsites 1 and 2.
3. **⦿"Options (M0,M3)"** Codeml options for running with Codeml NSsites 0 and 3.
4. **⦿"File"** Options contained in the Codeml control file near the executable placed by the user.
5. **⦿"Other"** Which allows a user to select his own options from a panel exhibited from the **"Choose..."** button (see figure above).

These sets of options can be viewed with the **"View"** button next to the **"Choose..."** button.

3) Select the output file for resulting data. Then, for running Codeml just press the **"Run Codeml"** button. If there is some problem with the execution use the **"View Details"** button for viewing any errors or information from or reported by Codeml.



Alignment: uments/java/IMPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/seq.fas

Tree File: ments/java/IMPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/tree.nwk

Options: ○ Options (M7,M8) ○ Options (M1,M2) ○ Options (M0,M3) ● File ○ Other

Output File: IPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/eset_M0NS012378.txt

Options: ● Options (M7,M8) ○ Options (M1,M2) ○ Options (M0,M3) ○ All (M0-M8)

Output File: PACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/eset_M0NS012378.txt

Lnl Table: ts/java/IMPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/LnLTable.csv

Ratios Table: ava/IMPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/RatiosTable.csv

BEB Results: /java/IMPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/BEBresults.csv

Select: P>95% (*)

Likelihood Values: -2857.691405 -2820.050319

Chi²: 5.9915

LRT: 75.28217200000017

- **Analyzing Results –** *Results* section from PAML tab.

1) The resulting file from Codeml execution is automatically presented in the corresponding field in the "Results" section. Alternately, the user can also select any Codeml result file generated previously.

2) In the Results section, the same set of options from the panel above is maintained except the two options on the right, which are replaced with **⊙"All (M0-M8)"** option (see Figure above). These options are meant for the **BEB extraction** and for the creation of the **Likelihood and Ratios tables** from the available models in the results file. The resulting amount of information is only dependent of the options selected. This last option as opposite to the remaining, considers all the models M(0,1,2,3,7,8).

3) The user is able to easily extract the Bayes Empirical Bayes (BEB) section (provided that exists) from the results file selected/browsed in the field above. In order to extract the BEB section from the Model 8 or Model 2, the user should have selected the corresponding option from the options provided in the same section (namely **⊙"Options (M7,M8)"** or **⊙"Options (M1,M2)"**, respectively). If a BEB section is not found, a message will be presented, accordingly. In case the BEB section exists, it will be written to the file selected in the bottom field, named **"BEB results"**. The user must remember to change the file name in order to save different BEB sections. The lines extracted from the BEB section, match the option selected in combo box to the left of the **"Extract BEB"** button. By default, the user can extract all the significant lines (**P>95% (*)**) which encompasses both values maked with star (*) and double star (**); or otherwise, just the values marked with double star (**) as the most significant (**P>99% (**)**).

4) IMPACT_S also provides an **LRT** (**Likelihood Ratio Test**), which is automatically performed when extracting the BEB section. This LRT takes into account the degrees of freedom based in the difference of number of parameters from each pair of models and the Chi-square values matching significance values regarding p-values of p=0.05 (selected by default), p =0.01 and p=0.001. This is possible for any of the three first options, **⊙"Options(M7,M8)"**, **⊙"Options(M1,M2)"** and **⊙"Options(M0,M3)"**, although, the latter does not provide a BEB section.

5) Any of these four options allow the extraction and creation of **Likelihood and Ratios tables** for every one of these methods, depending, if available in the results file. This is obtained by using the wide **"View"** button, which will

exhibit the final result in the CSV Viewer window (see section 3.8). These tables gather for each model its relative information **Model**, number of parameters (**np**) and Likelihood (**lnL**) for the first case; and **Model**, proportion values (**p:**) and omega values (**w:**) (in paired lines; same column), followed by all possible classes from K=1 to K=11 from the mentioned models M(0,1,2,3,7,8).

3.2.4. Results & 3D tab

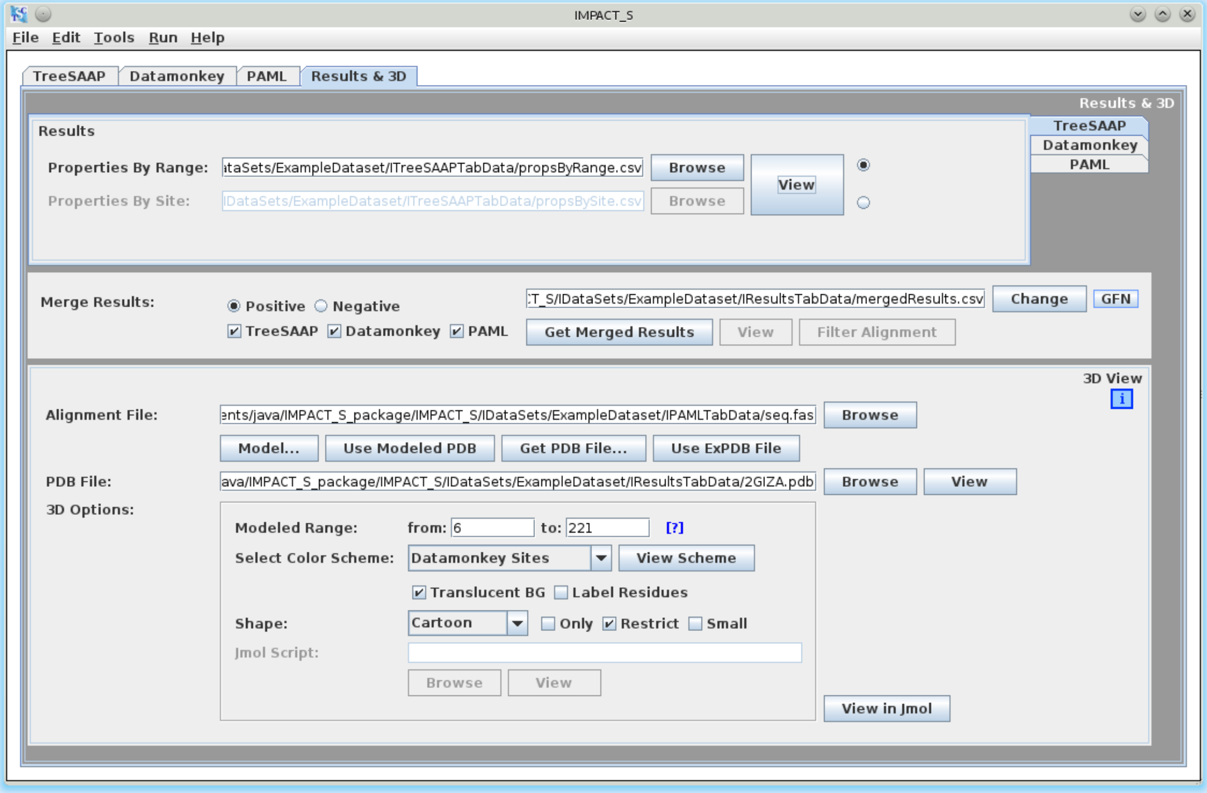
This tab consists of 2 sections, the "Results" section and the "3D View" section.

- **Analyzing Results –** *Results* section from Results & 3D tab.

1) In the **"Results"** section the user browses the corresponding files needed. These files are automatically inscribed in these fields as long as the user makes use of the previous tabs. In this section, the user can once again view the files contents using the corresponding **"View"** buttons. The user can also select a file, where the resulting data from the Merge Results (MR) procedure will be saved.
2) The MR procedure takes always the **PBR** or **PBS** tables generated from the **TreeSAAP** tab, the chosen **Datamonkey** table and/or one of the **BEB results** table from **PAML** (in case of positive selection), selected using the associated radio buttons. The user is also free to select if the MR procedure will be produced with only data from PAML or Datamonkey or both, by using the corresponding checkboxes. This procedure produces a table much similar to the PBR or PBS table extended with the Datamonkey and/or BEB information, namely the site numbers. An additional column **"Common Sites"** is placed before the "Total" column containing only the common sites to all the two/three programs. Matched sites under PAML BEB and/or Datamonkey are placed between the TreeSAAP ranges (PBR) or codons (PBS) and the column "Total" in columns identified by each program name. The PAML column additionally identifies the Model from which information was obtained (e.g.: M8Site or M2Site). Each range (PBR) or codon (PBS) will

have any number of sites identified by Codeml and Datamonkey. The sites are separated by "/" character. The **"Common Sites"** column will further be used for mapping in the 3D structure, according to the Merged Results scheme.

- **3D Structure Mapping –** *3D View* section from Results & 3D tab.



Results

Properties By Range: ıtaSets/ExampleDataset/ITreeSAAPTabData/propsByRange.csv ●

Properties By Site: IDataSets/ExampleDataset/ITreeSAAPTabData/propsBySite.csv ○

Merge Results: ● Positive ○ Negative ☑ TreeSAAP ☑ Datamonkey ☑ PAML :T_S/IDataSets/ExampleDataset/IResultsTabData/mergedResults.csv

Alignment File: ents/java/IMPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/seq.fas

PDB File: ava/IMPACT_S_package/IMPACT_S/IDataSets/ExampleDataset/IResultsTabData/2GIZA.pdb

3D Options:

Modeled Range: from: 6 to: 221

Select Color Scheme: Datamonkey Sites

☑ Translucent BG ☐ Label Residues

Shape: Cartoon ☐ Only ☑ Restrict ☐ Small

Jmol Script:

1) In the section **"3D View"**, the user will make use of the tables available from the section above. Here the user has the possibility to map the sites detected by any of the programs (PAML (Codeml), TreeSAAP and Datamonkey). If there are tables missing from the section above, some of these options cannot be used. In this section, the user will browse for the **MSA** file (if not loaded from "PAML" tab) and the **PDB** file if exists. Any PDB file can be used for viewing in Jmol, but we recommend to use a PDB file containing only one chain in order to better visualize the sites detected by any of the programs. For this, the user has the possibility to produce a Protein homology model

using the Swiss-Model interface from one of the sequences in the alignment. To open this window press the **“Model...”** button which will bring the **Protein Modeler window** (see section 3.3) with the alignment already loaded. After producing or obtaining the model use the **“Use Modeled PDB”** button which will load the modeled PDB file and the modeled range values in its fields. Another way of obtaining the PDB file is through the **“Get PDB File...”** button. This will bring the **PDB Downloader window** (see section 3.4). After downloading the file use the **“Use ExPDB File”** button to load the path of this newly downloaded file.

2) It is always possible to use the **“View”** button next to the PDB field to view the entire PDB file in text format. Underneath the PDB field, a box is presented where the user can choose options for visualization in Jmol. This box consists of the **“Modeled Range”** field which allows the user to limit the number or the extent to which sites will be presented painted in Jmol window. The user must specify a **“From”** value lower or equal than the **“To”** value. In case, these are not specified, no limits are used and thus all possible sites are visualized.

3) Below this field, the user finds the **color schemes** in a selection box, which can be selected/applied for mapping sites (see color schemes table below). These color schemes depend on the availability of result files browsed in the **“Results”** section. Next to the color scheme selection box, the button **“View Scheme”**, presents a window showing the color scheme for the scheme option previously selected, making possible to view the colors used or even change the colors according to the user preferences (see section 3.5) and thus creating new color schemes.

4) The fields regarding to the tag **“Jmol Script”** are activated for the last scheme option **“Use my own...”**, which allows the user to browse for any Jmol script provided by the user.

5) The check box **☑“Translucent BG”**, allows the user to remove translucency from the color associated to the 'Other' field, found in the **Color Scheme window** (see section 3.5). By default, this color is white across almost all the schemes. The check box **☑“Label Residues”** is useful to insert labels next

to the residues, allowing the user to better understand the molecule, the residues location and to export for publication purposes.

6) The field **“Shape”** allows the user to specify a different rendering for the protein to be visualized. The box of options allows 11 rendering options plus the “None” option. These options belong to Jmol set of options.

Table 3 – Jmol Shape Options

Shape Options	
None	**Dots**
Cartoon	**Spacefill**
Rockets	**Geosurface**
Ribbons	**Wireframe**
Strands	**Trace**
Meshribbon	**Backbone**

7) The checkbox ☑**“Only”** associated, can be used to make the selected rendering option the only one visible and thus overwhelms other related options.

8) The check box ☑**“Restrict”** allows the user to view only the residues that are important or under selection. Selecting this option will enable the check box ☑**“Small”** that can be used to view these residues in small size.

9) After all options are chosen the user presses the **“View in Jmol”** button to open Jmol and view the chosen/modeled PDB file with the selected color scheme. Always use this button to update the 3D view of the PDB file for newly applied color schemes.

Table 4 – IMPACT_S Color Schemes

Color Scheme	**Description**

None	No scheme is applied. Regular PDB file visualization.
Manual Coloring...	This option allows free application of colors to any sites needed by the user, in the PDB Sequence window (see section 3.5).
TreeSAAP (Type I)	This option applies to PBR and PBS tables. Selects all ranges/sites having at most one property under selection. $-\infty < NP < 2$.
TreeSAAP (Type II)	This option applies to PBR and PBS tables. Selects all ranges/sites having between 2 to 5 properties under selection, inclusive. $2 \leq NP \leq 5$.
TreeSAAP (Type III)	This option applies to PBR and PBS tables. Selects all ranges/sites having more than 5 properties under selection. $5 < NP < +\infty$.
TreeSAAP (User)	This option applies to PBR and PBS tables. Selects all ranges/sites having more than X properties under selection. X is chosen by the user. $X < NP < +\infty$.
TreeSAAP (Type I;II;III)	This option applies to PBR and PBS tables. It applies all previous schemes (Types I, II, III) in the same 3D structure.
Datamonkey Sites	This option enables visualization of any table from Datamonkey, either from single method alone or using the Common Sites table.
Codeml (M2 or M8)	This option applies to one BEB table and can be used for either Model 2 or 8, just depending on the BEB table/field selected. No restrictions.
Codeml (M2+8;2)	This option applies to both BEB tables. This scheme applies 2 colors: one for the common sites found among both models and another for the remaining uncommon sites aggregating both models.
Codeml (M2+8;3)	This option applies to both BEB tables. This scheme applies 3 colors: one for the common sites found among both models; another for the remaining uncommon sites only belonging to Model 2 and another for the remaining uncommon sites only belonging to Model 8.
Merged Results	This option applies to the Merged Results table. All the Datamonkey Sites and/or PAML BEB sites that fall in the TreeSAAP ranges (PBS) or codons (PBS). It maps the Common Sites column found among these methods, in this named column.
Side Chains	Based on the characteristics of the side chains. It consists of 7 groups or colors as follows: AGILPV,FWY,ED,RHK,ST,CM,NQ Corresponding to *Aliphatic, Aromatic, Acidic, Basic, Hydroxylic, Sulphur-*

	containing, Amidic in http://www.cryst.bbk.ac.uk/education/AminoAcid/the_twenty.html
Seaview	Based on SeaView program Amino Acid colors for the alignment. It consists of 8 groups or colors as follows: KR,AFILMVW,NQST,HY,C,DE,P,G in http://pbil.univ-lyon1.fr/software/seaview
Amino (Jmol)	Color scheme taken from Jmol website. This option can be used to visualize residues differently by choosing different colors for each residue. See http://jmol.sourceforge.net
Shapely (Jmol)	Color scheme taken from Jmol website. This option can be used to visualize residues differently by choosing different colors for each residue. See http://jmol.sourceforge.net
Use my own...	Allow a user to select or browse his own Jmol script. This selection will activate fields located at the bottom.

3.3. *Protein Modeler (Swiss-Model) window*

This window consists of two tabs: **“Automated Mode”** and the **“Alignment Mode”**. Each tab handles the respective Swiss-Model mode being employed for homology modeling.

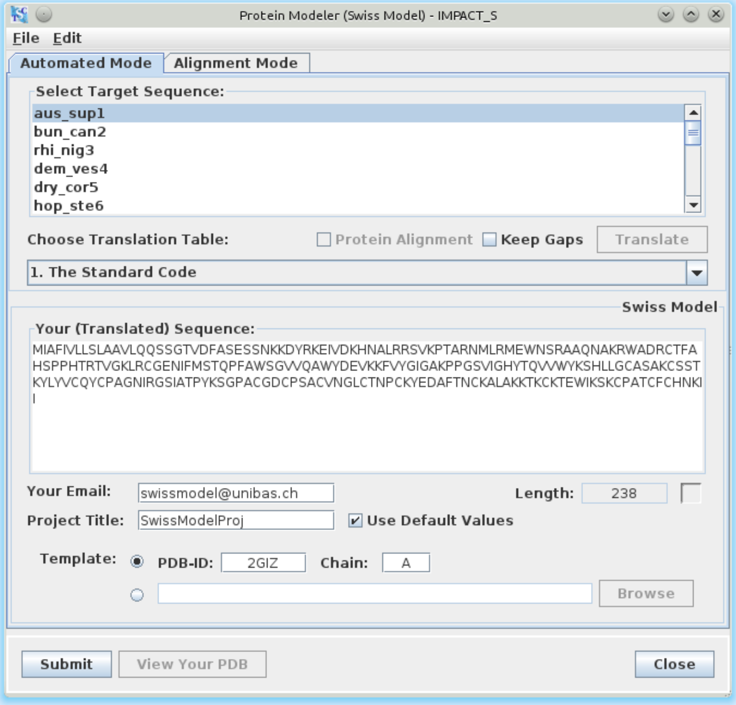
When opening this window from the main window **Results & 3D tab**, through the **“Model...”** button, the MSA file comes automatically loaded ready for further use in the Protein Modeler, provided it was browsed in the **“Alignment File”** field above.

- **Automated Mode tab –** from Protein Modeler window (see figure below).

1) In the **“Automated Mode” tab**, the user can select, from the top list box, a sequence to use as a target for submission. If eventually, the alignment is a protein alignment the user should tick the ☑**“Protein Alignment”** check box.
2) Otherwise, the user needs to translate the sequence into protein. For that, make sure to select the appropriate translation table and use the **“Translate”** button to translate. The sequence selected and translated will appear in the text area in the section below.

3) In this section, named **Swiss-Model**, other options are provided for Swiss-Model submission. Namely, the possibility to select a **template PDB-ID and CHAIN-ID code** or otherwise, a **template file**. Also the user can choose to enter an email address and project title to be used in submission.

4) All these options are provided in the same way as in Swiss-Model website. After all is chosen, the user can then press the **"Submit"** button.



Select Target Sequence:
aus_sup1
bun_can2
rhi_nig3
dem_ves4
dry_cor5
hop_ste6

Choose Translation Table: □ Protein Alignment □ Keep Gaps

1. The Standard Code

Your (Translated) Sequence:

```
MIAFIVLLSLAAVLQQSSGTVDFASESSNKKDYRKEIVDKHNALRRSVKPTARNMLRMEWNSRAAQNAKRWADRCTFA
HSPPHTRTVGKLRCGENIFMSTQPFAWSGVVQAWYDEVKKFVYGIGAKPPGSVIGHYTQVVWYKSHLLGCASAKCSST
KYLYVCQYCPAGNIRGSIATPYKSGPACGDCPSACVNGLCTNPCKYEDAFTNCKALAKKTKCKTEWIKSKCPATCFCHNKI
I
```

Your Email: swissmodel@unibas.ch Length: 238 □

Project Title: SwissModelProj ☑ Use Default Values

Template: ● PDB-ID: 2GIZ Chain: A

○

- **Alignment Mode tab –** from Protein Modeler window (see figure below).

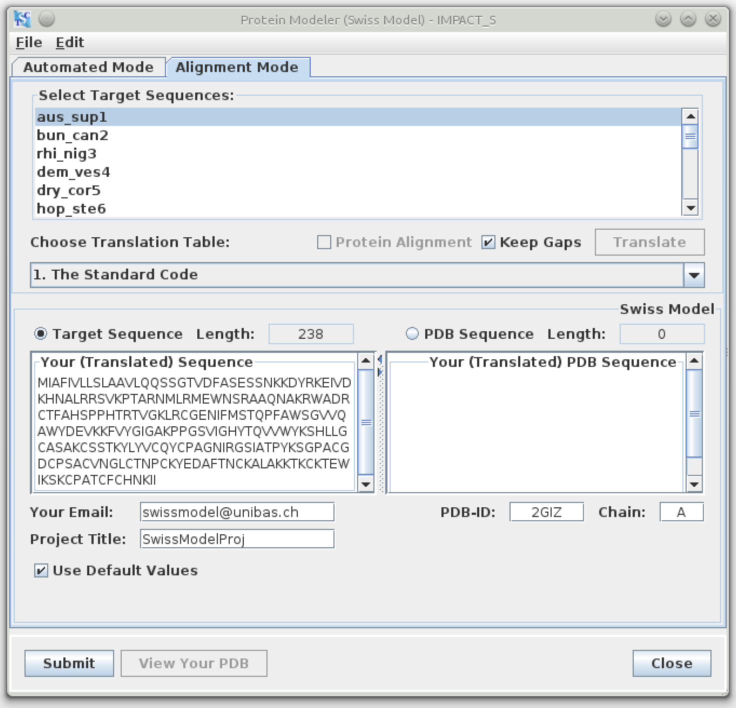
1) In the **"Alignment Mode" tab**, the user has access to the same top section, allowing to select sequences and translate each one or make sure to tick the **☑"Protein Alignment"** check box, if it is the case. In this mode, the user will need to select two sequences from the top section.

2) First, the user selects a target sequence and translates (the **☑"Target Sequence"** check box is selected), and then after selecting the **☑"PDB sequence"** check box, the user can select the template sequence and

translate. This PDB or template sequence has an associated PDB-ID code that needs to be specified in the text fields below. Likewise, the user can optionally introduce an email address and a project title.

3) After the user has everything selected, it is possible to submit the project to Swiss-Model. This mode comprises a few more steps during submission, which are fully handled by the submission process implemented. Hence, the user just needs to be sure of the data being submitted.

4) After the user presses the **“Submit”** button regardless of mode being used, a new window will appear showing the progress of submission. IMPACT_S will wait until the submitted project is complete. Once it is complete, the PDB file is automatically downloaded and the modeled range retrieved. This window, exhibiting a black background, shows the results link obtained from Swiss-Model, which can be used for obtaining other information, for instance, accessing the website. In this window, it is possible to also find the final location where the modeled PDB file was saved. The button **“View Your PDB”** near the **“Submit”** button is enabled allowing the user to display the downloaded PDB file. Any errors that can arise from the submission, are shown in this window or through dialogs.



Select Target Sequences:
aus_sup1
bun_can2
rhi_nig3
dem_ves4
dry_cor5
hop_ste6

Choose Translation Table: □ Protein Alignment ☑ Keep Gaps

1. The Standard Code

● Target Sequence Length: 238
○ PDB Sequence Length: 0

Your (Translated) Sequence

```
MIAFIVLLSLAAVLQQSSGTVDFASESSNKKDYRKEIVD
KHNALRRSVKPTARNMLRMEWNSRAAQNAKRWADR
CTFAHSPPHTRTVGKLRCGENIFMSTQPFAWSGVVQ
AWYDEVKKFVYGIGAKPPGSVIGHYTQVVWYKSHLLG
CASAKCSSTKYLYVCQYCPAGNIRGSIATPYKSGPACG
DCPSACVNGLCTNPCKYEDAFTNCKALAKKTKCKTEW
IKSKCPATCFCHNKII
```

Your (Translated) PDB Sequence

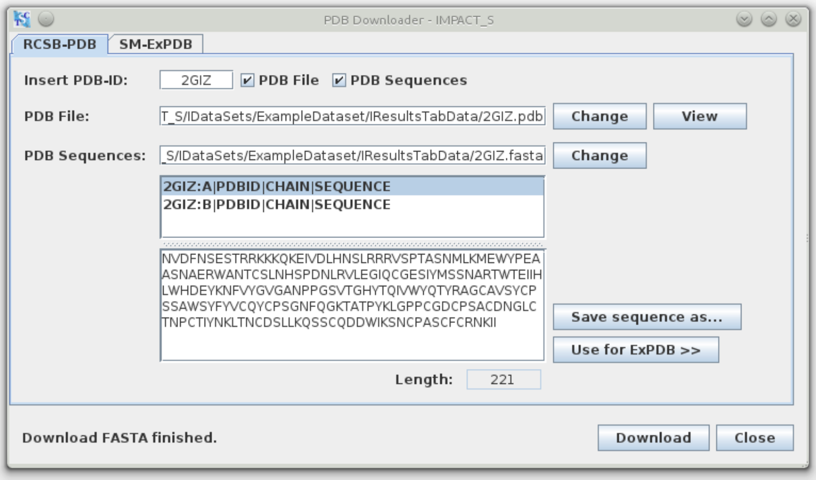
Your Email: swissmodel@unibas.ch
Project Title: SwissModelProj
PDB-ID: 2GIZ
Chain: A

☑ Use Default Values

3.4. PDB Downloader window

This window comprises a tool for downloading PDB files and/or PDB FASTA files from RCSB website. Or one chain PDB files, from the Template Library from Swiss-Model. It is composed by two tabs, one named **“RCSB-PDB”** and the other named **“SM-ExPDB”**.

- **RCSB-PDB tab –** from PDB Downloader window (see figure below).



Insert PDB-ID: 2GIZ ☑ PDB File ☑ PDB Sequences

PDB File: T_S/IDataSets/ExampleDataset/IResultsTabData/2GIZ.pdb

PDB Sequences: _S/IDataSets/ExampleDataset/IResultsTabData/2GIZ.fasta

```
2GIZ:A|PDBID|CHAIN|SEQUENCE
2GIZ:B|PDBID|CHAIN|SEQUENCE
```

```
NVDFNSESTRRKKKQKEIVDLHNSLRRRVSPTASNMLKMEWYPEA
ASNAERWANTCSLNHSPDNLRVLEGIQCGESIYMSSNARTWTEIIH
LWHDEYKNFVYGVGANPPGSVTGHYTQIVWYQTYRAGCAVSYCP
SSAWSYFYVCQYCPSGNFQGKTATPYKLGPPCGDCPSACDNGLC
TNPCTIYNKLTNCDSLLKQSSCQDDWIKSNCPASCFCRNKII
```

Length: 221

Download FASTA finished.

1) In the former tab, the user enters the **PDB-ID code** (no CHAIN-ID required). Near this code entry, 2 check boxes are selected, meaning that both files will be downloaded. Deselect, if not necessary. These will allow the download of the **☑“PDB File”** and/or the **☑“PDB Sequences”** FASTA file containing the PDB sequences in FASTA format.

2) As the user enters the PDB-ID code, the fields where the files are to be saved, are automatically adjusted, showing file names based on the code inserted.

3) Near these fields it is possible to select a different directory where to save files, using the **“Change”** buttons.

4) After options are correctly chosen, the user can press the **“Download”** button on the bottom of the window to download. If the code is correct and exists, the download will succeed.

5) Below the **“PDB Sequences”** field, a divided box will present, the **sequence**

descriptions found in the FASTA file on the **top division** and the **sequence** itself in the **bottom division**. Once a sequence description is selected above the corresponding sequence becomes displayed below. It is possible to save the selected sequence to file using the **"Save sequence as..."** button.

6) The button **"Use for ExPDB >>"**, will pass the selected sequence information (**PDB-ID and CHAIN-ID codes**) to the "SM-ExPDB" tab in order to download the PDB file containing the specified chain only.

- **SM-ExPDB tab –** from Protein Modeler window.

1) In this tab, the user can directly insert the **PDB-ID and CHAIN-ID code** to download from the Template Library of Swiss-Model (SMTL or ExPDB).

2) Additionally, the user can choose the directory where to save the PDB file.

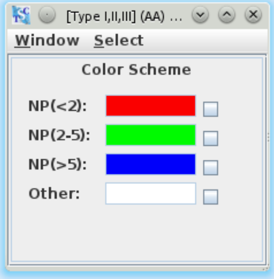
3) After options are correctly chosen, the user can press the **"Download"** button on the bottom of the window to download. If the code is correct and exists, the download will succeed.

3.5. Color Schemes *and PDB Sequence windows*

This is a very simple window, showing a list of fields exhibiting the colors for the residues or the options in the selected scheme.

1) Bring this window after selecting the scheme in main window **Results & 3D tab** and by clicking the **"View Scheme"** button. To view a complete listing of color schemes see the color schemes table in section 3.2.4.

2) In case of options like *Side Chains*, *Seaview*, *Amino and Shapely*, the residues names are presented, each one associated to a color. For these schemes the 'Other' field color, can be used to "hide"



[Type I,II,III] (AA) ...

Window Select

Color Scheme

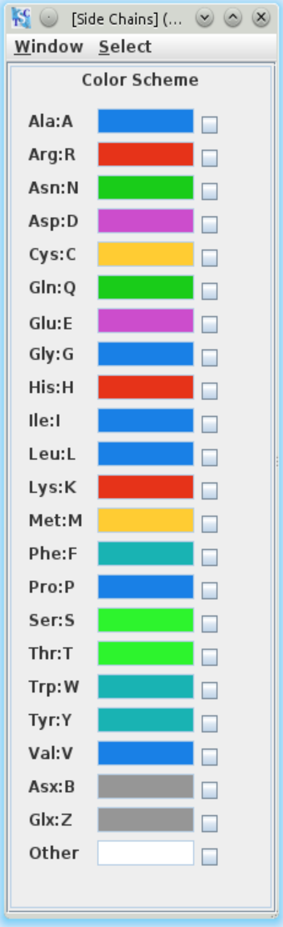
NP(<2): 🟥 □

NP(2-5): 🟩 □

NP(>5): 🟦 □

Other: ⬜ □

residues, if the user just needs to focus on a small subset. Also, this makes possible to just label these wanted residues. For this, the user just chooses the same color for all the residues to “hide”, which must be the same as for the 'Other' color.



Ala:A	🟦	□
Arg:R	🟥	□
Asn:N	🟩	□
Asp:D	🟪	□
Cys:C	🟨	□
Gln:Q	🟩	□
Glu:E	🟪	□
Gly:G	🟦	□
His:H	🟥	□
Ile:I	🟦	□
Leu:L	🟦	□
Lys:K	🟥	□
Met:M	🟨	□
Phe:F	🟦	□
Pro:P	🟦	□
Ser:S	🟩	□
Thr:T	🟩	□
Trp:W	🟦	□
Tyr:Y	🟦	□
Val:V	🟦	□
Asx:B	▨	□
Glx:Z	▨	□
Other	⬜	□

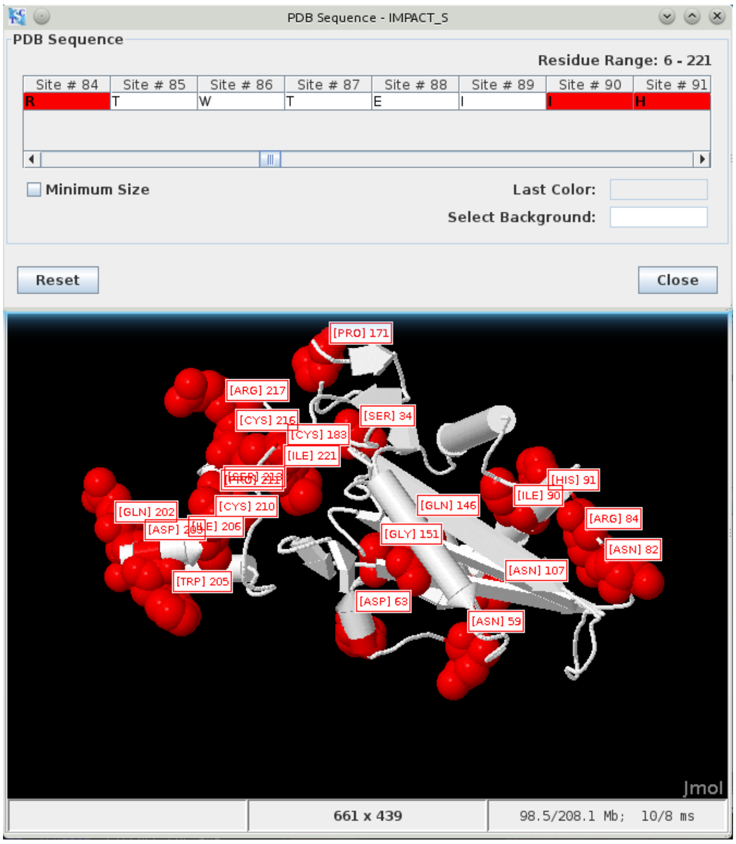
3) In the remaining options, except for the **“None”** and **“Use my own...”**, the scheme options are presented with the default colors associated.

4) In every case, the user has the possibility, to change each color associated. For this, the user clicks on the color he wants to change and a **color chooser window** will appear allowing the user to select the wanted color. Additionally, the checkboxes near the colors can be used to apply a certain color to a set of selected color fields (see right figures). Which means, that the user can select any entries in the window and after clicking on one of the selected entries, the **color chooser** will appear to allow color selection that will be applied to all the selected entries.

5) These sets of colors will be used in the main window to display the PDB in Jmol, according to the currently selected scheme and chosen colors.

6) The small menu on this window provides useful shortcuts like **“Select->All”** that selects all entries or **“Select->None”** that deselects any selected entry. The **“Window->Reset”** to reload the default colors and the **“Window->Close”** to close this window.

Note: This window does not need to be opened for the scheme to be applied. The only need for opening this window is to be aware of the scheme of colors and eventually make modifications to the set of colors that comprises the scheme to be applied.



PDB Sequence - IMPACT_S

PDB Sequence

Residue Range: 6 - 221

Site # 84	Site # 85	Site # 86	Site # 87	Site # 88	Site # 89	Site # 90	Site # 91
R	T	W	T	E	I	I	H

□ Minimum Size

Last Color:

Select Background:

661 x 439

98.5/208.1 Mb; 10/8 ms

Jmol figure obtained from IMPACT_S for a modeled PDB file. This is for illustration purposes only.

Applied scheme: **PAML BEB(M2 or M8) Sites**; **Label Residues** selected; Shape **Rockets**; **Restrict** selected; **Translucent BG**, deselected.

1) In the above **"PDB Sequence"** window, the user can get understanding of which sites are colored in which positions from a sequence perspective, for the selected scheme relative to the PDB file used; or can color any other sites by free will.

2) The **"Manual Coloring..."** option allows the user to color the 3D structure by selecting which sites and colors to use, finally mapping the choices in the 3D structure, by clicking the **"View in Jmol"** button.

    a. To apply a color, click the sequence amino-acids positions or click and drag the mouse pointer to select a contiguous set of amino-acids.

    b. Then the color chooser will appear to apply the preferred the color.

3) It is possible for a user to **switch from any other scheme** that was previously applied, to this option and thus modify the colors of the distinguished sites or add new colored sites.

4) This window also allows the user to choose the background color for every other non-distinguished sites (white in this case), by clicking the **“Select Background”** color field.

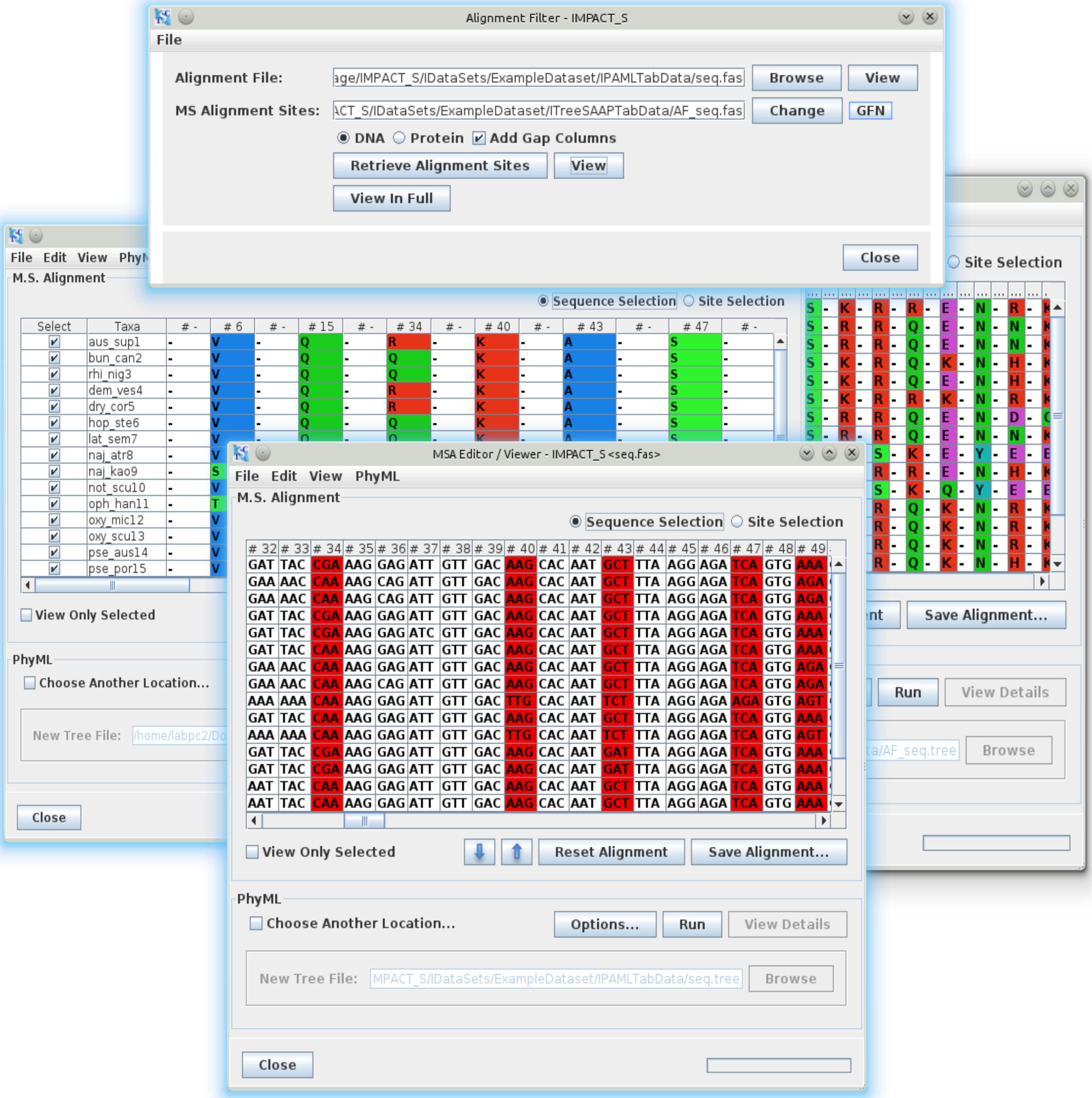
3.6. Alignment Filter window

This window provides a new functionality that enables the user to highlight sites that are considered to be most important. These are the significant selection sites obtained directly from each tab or any sites in a CSV file (which in contrast, requires the user to open specifically this file). Any CSV file will work either from IMPACT_S or created by the user.

1) This can be achieved in two ways, either by creating a new alignment file where only the most important sites are displayed (see background pictures) or just by coloring sites in the full alignment (see foreground bottom picture). This way the user can create a view where only the positive or negatively selected sites are shown on the alignment.

2) The user must provide a **MSA file,** which must be either **⊙“DNA”** or **⊙“Protein”** in **FASTA** format**.** The user can use this tool independently, by accessing the Tools menu in main window. By going to its **“File”** menu, it is possible to open a file containing the needed codons, provided that the file is **CSV** formatted, with the first column named “Codon”, containing all the codons to be considered. Remaining columns are discarded. The result is shown in the **Alignment Editor/Viewer** (see section 3.9) by clicking the **“View”** button.

3) The option **☑“Add Gap Columns”** inserts columns of gaps between non-consecutive sites, this way providing the idea of proximity between sites. These views are widely used in publications.



Alignment Filter - IMPACT_S

File

Alignment File: age/IMPACT_S/IDataSets/ExampleDataset/IPAMLTabData/seq.fas [Browse] [View]

MS Alignment Sites: ACT_S/IDataSets/ExampleDataset/ITreeSAAPTabData/AF_seq.fas [Change] [GFN]

● DNA ○ Protein ☑ Add Gap Columns

[Retrieve Alignment Sites] [View]

[View In Full]

[Close]

M.S. Alignment

● Sequence Selection ○ Site Selection

Select	Taxa	# -	# 6	# -	# 15	# -	# 34	# -	# 40	# -	# 43	# -	# 47	# -
☑	aus_sup1	-	V	-	Q	-	R	-	K	-	A	-	S	-
☑	bun_can2	-	V	-	Q	-	Q	-	K	-	A	-	S	-
☑	rhi_nig3	-	V	-	Q	-	Q	-	K	-	A	-	S	-
☑	dem_ves4	-	V	-	Q	-	R	-	K	-	A	-	S	-
☑	dry_cor5	-	V	-	Q	-	R	-	K	-	A	-	S	-
☑	hop_ste6	-	V	-	Q	-	Q	-	K	-	A	-	S	-
☑	lat_sem7	-	V											
☑	naj_atr8	-	V											
☑	naj_kao9	-	S											
☑	not_scu10	-	V											
☑	oph_han11	-	T											
☑	oxy_mic12	-	V											
☑	oxy_scu13	-	V											
☑	pse_aus14	-	V											
☑	pse_por15	-	V											

□ View Only Selected

PhyML

□ Choose Another Location...

New Tree File: /home/labpc2/D

[Close]

○ Site Selection

[Save Alignment...]

[Run] [View Details]

ta/AF_seq.tree [Browse]

MSA Editor / Viewer - IMPACT_S <seq.fas>

File Edit View PhyML

M.S. Alignment

● Sequence Selection ○ Site Selection

# 32	# 33	# 34	# 35	# 36	# 37	# 38	# 39	# 40	# 41	# 42	# 43	# 44	# 45	# 46	# 47	# 48	# 49
GAT	TAC	CGA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AAA
GAA	AAC	CAA	AAG	CAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AGA
GAA	AAC	CAA	AAG	CAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AGA
GAT	TAC	CGA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AAA
GAT	TAC	CGA	AAG	GAG	ATC	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AAA
GAT	TAC	CAA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AAA
GAA	AAC	CAA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AGA
GAA	AAC	CAA	AAG	CAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AGA
AAA	AAA	CAA	AAG	GAG	ATT	GTT	GAC	TTG	CAC	AAT	TCT	TTA	AGG	AGA	AGA	GTG	AGT
GAT	TAC	CAA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AAA
AAA	AAA	CAA	AAG	GAG	ATT	GTT	GAC	TTG	CAC	AAT	TCT	TTA	AGG	AGA	TCA	GTG	AGT
GAT	TAC	CGA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GAT	TTA	AGG	AGA	TCA	GTG	AAA
GAT	TAC	CGA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GAT	TTA	AGG	AGA	TCA	GTG	AAA
AAT	TAC	CAA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AAA
AAT	TAC	CAA	AAG	GAG	ATT	GTT	GAC	AAG	CAC	AAT	GCT	TTA	AGG	AGA	TCA	GTG	AAA

□ View Only Selected [⬇] [⬆] [Reset Alignment] [Save Alignment...]

PhyML

□ Choose Another Location... [Options...] [Run] [View Details]

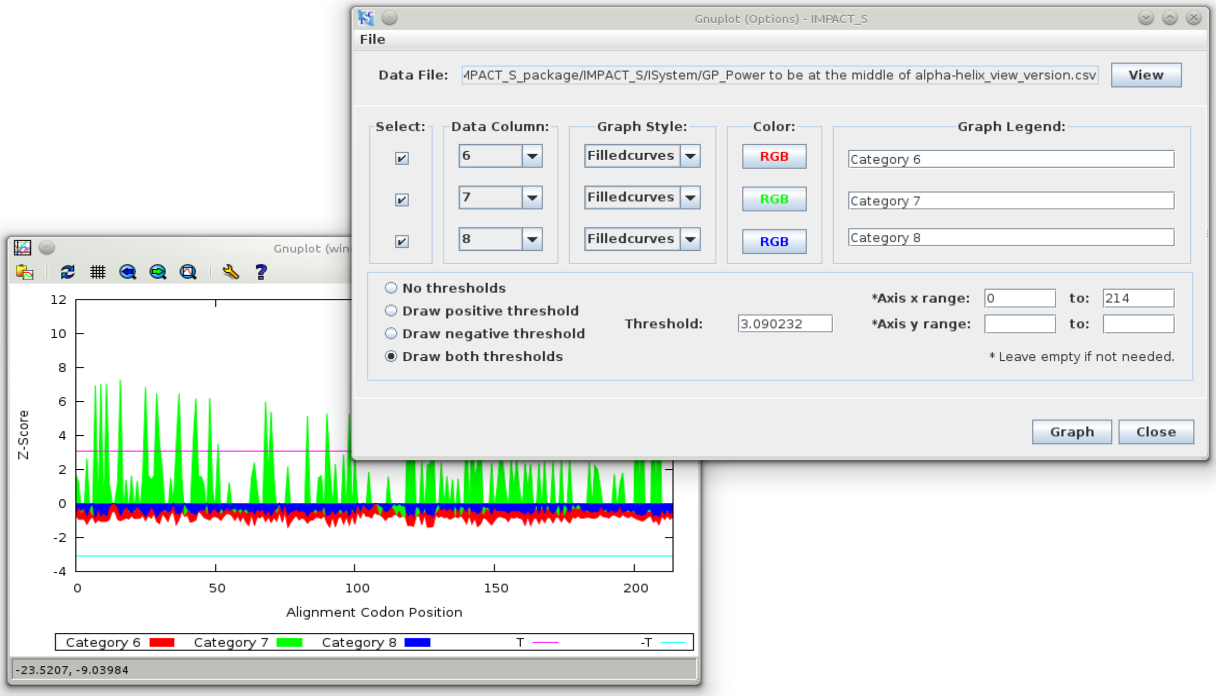
New Tree File: MPACT_S/IDataSets/ExampleDataset/IPAMLTabData/seq.tree [Browse]

[Close]

This tool is also directly available from each tab, **TreeSAAP - *Substs***, **Datamonkey**, **PAML** and **Results & 3D**, through the **“Filter Alignment”** button, automatically using the resulting corresponding file.

3.7. *Gnuplot (Options) window*

This window can be found in the **"Run"** menu from the **main window** or in the **TreeSAAP -*Evpthwy*** tab by clicking the **"View"** button when viewing the selected property content (also **☑"Plot"** must be selected). It allows the user to graph the properties from the *SlidingWindow* directory. The "**File**" menu allows this tool to be run independently, by opening any properties file available.



Gnuplot (Options) - IMPACT_S

Data File: ɪPACT_S_package/IMPACT_S/ISystem/GP_Power to be at the middle of alpha-helix_view_version.csv

Select:	Data Column:	Graph Style:	Color:	Graph Legend:
☑	6	Filledcurves	RGB	Category 6
☑	7	Filledcurves	RGB	Category 7
☑	8	Filledcurves	RGB	Category 8

○ No thresholds
○ Draw positive threshold
○ Draw negative threshold
● Draw both thresholds

Threshold: 3.090232

*Axis x range: 0 to: 214
*Axis y range: to:
* Leave empty if not needed.

-23.5207, -9.03984

1) The user can plot graphs in different styles here. Using the **"Select"** column, the user can draw graphs from one to three data columns at the same time.

2) Choose to graph any of the category **data** columns found in the TreeSAAP property file (from 1 to 8), see the **"Data"** column.

3) Choose the graph **style** from four available options (**"Filledcurves", "Impulses", "Lines" and "Linespoints"**), under the **"Graph Style"** column.

4) Choose the **color** for each selected graph using the colored **"RGB"** buttons in the **"Color"** column.

5) Choose the **legend** in the **“Graph Legend”** column which is automatically generated/updated when the category column data to graph is changed. It can also be chosen by the user.

6) Additionally the user can draw **threshold lines** for either positive (**⦿"Draw positive threshold”**) or negative (**⦿"Draw negative threshold”**) or both (**⦿"Draw both thresholds”**); according to the default value (that can be modified in the **“Threshold”** field) obtained from TreeSAAP itself, which refers to the most significant Z-scores (i.e. 3.090232).

7) Also, the graph can be **adjusted or zoomed** by inserting minimum and maximum values for the *x* and *y* axis, respectively in **“Axis x range”** and **“Axis y range”**. Leave empty, if not necessary.

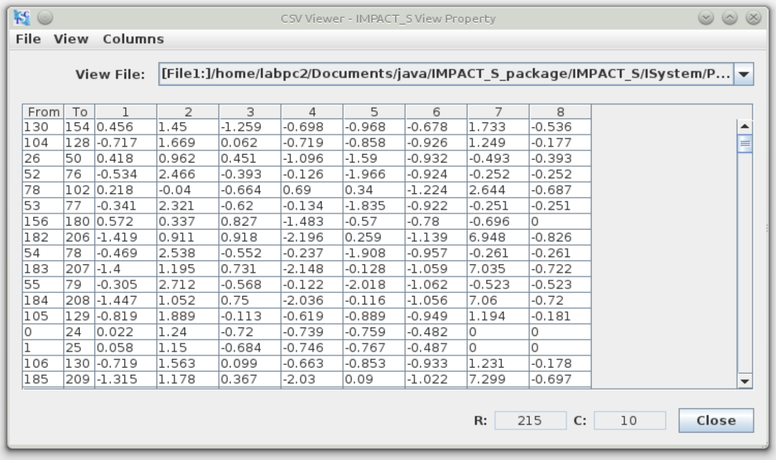
8) Use the **“Graph”** button to build the plot or whenever updates are needed regarding new choices made.

3.8. CSV Viewer window

This function was imported from the IMPACT software (http://impact-gui.sourceforge.net).

1) This window (see figure below) provides the same functionality as in IMPACT. But it was augmented to allow the saving of table files (generated by TreeSAAP, PAML and Datamonkey) currently being viewed.

2) This window allows a user to **visualize two CSV (Comma-Separated Values) files** at the same time. To switch among files (i.e. select) the user selects from the **“View File”** field found above the table. The two files can be shown in separate windows by going to the **“View”** menu and select from the **“New Window...”** sub-menu.



View File: [File1:]/home/labpc2/Documents/java/IMPACT_S_package/IMPACT_S/ISystem/P...

From	To	1	2	3	4	5	6	7	8
130	154	0.456	1.45	-1.259	-0.698	-0.968	-0.678	1.733	-0.536
104	128	-0.717	1.669	0.062	-0.719	-0.858	-0.926	1.249	-0.177
26	50	0.418	0.962	0.451	-1.096	-1.59	-0.932	-0.493	-0.393
52	76	-0.534	2.466	-0.393	-0.126	-1.966	-0.924	-0.252	-0.252
78	102	0.218	-0.04	-0.664	0.69	0.34	-1.224	2.644	-0.687
53	77	-0.341	2.321	-0.62	-0.134	-1.835	-0.922	-0.251	-0.251
156	180	0.572	0.337	0.827	-1.483	-0.57	-0.78	-0.696	0
182	206	-1.419	0.911	0.918	-2.196	0.259	-1.139	6.948	-0.826
54	78	-0.469	2.538	-0.552	-0.237	-1.908	-0.957	-0.261	-0.261
183	207	-1.4	1.195	0.731	-2.148	-0.128	-1.059	7.035	-0.722
55	79	-0.305	2.712	-0.568	-0.122	-2.018	-1.062	-0.523	-0.523
184	208	-1.447	1.052	0.75	-2.036	-0.116	-1.056	7.06	-0.72
105	129	-0.819	1.889	-0.113	-0.619	-0.889	-0.949	1.194	-0.181
0	24	0.022	1.24	-0.72	-0.739	-0.759	-0.482	0	0
1	25	0.058	1.15	-0.684	-0.746	-0.767	-0.487	0	0
106	130	-0.719	1.563	0.099	-0.663	-0.853	-0.933	1.231	-0.178
185	209	-1.315	1.178	0.367	-2.03	0.09	-1.022	7.299	-0.697

R: 215 C: 10

3.9. *Alignment Editor/Viewer (PhyML) window*

This function was imported from IMPACT software (http://impact-gui.sourceforge.net). It can be accessed under the “Tools” menu or in the **PAML tab**, through the **“View”** button associated to the MSA field. Two new menus have been added: **Edit** and **View**.

- **M.S. Alignment section –** from Alignment Editor/Viewer window.

  1) In the **Edit** menu, the user can find functions, such as the **“Find & Replace...”** and **“Rename Taxa...”**.

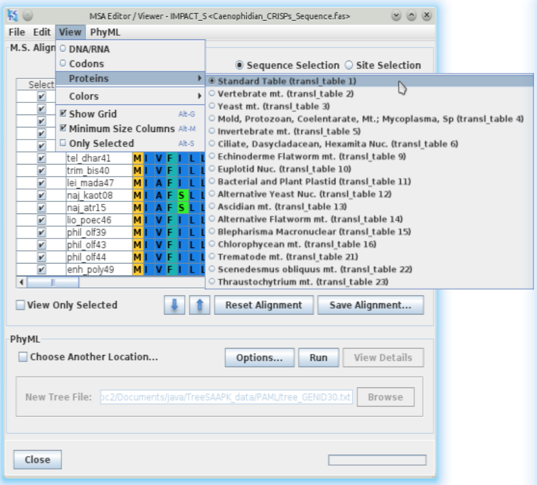
  2) Taxa or sites can be **removed** from the alignment by dragging the mouse across the taxa or sites column (depending of the selected view: either ⊙**“Sequence Selection”** or ⊙ **“Site Selection”** selection). Additional functions can be accessed from the **Edit** menu.

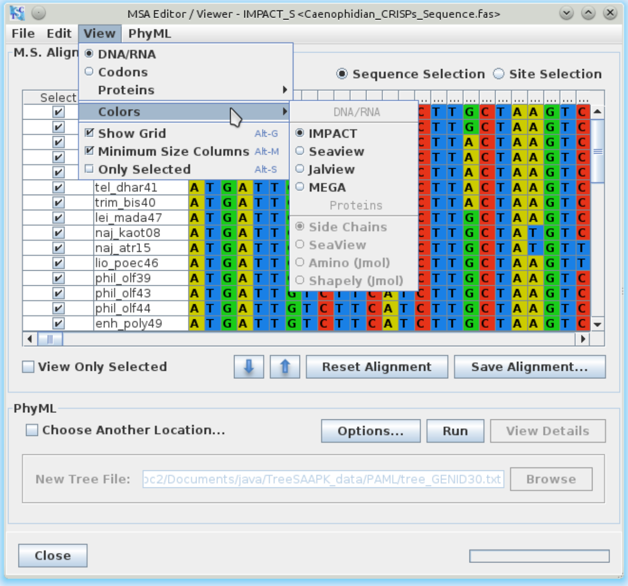
  1) In the **View** menu, the user can select in which mode the alignment data is presented. If it is a nucleotide alignment (the ⊙**“DNA/RNA”** option is selected). The user can choose ⊙**“Codons”** view, which organizes the nucleotides by groups of three (in codons); or, translate the alignment into **protein**, by selecting one of the available protein translation tables from the **“Proteins”** sub-menu (see left figure below).

  2) If it is an amino acid alignment, the ⊙**“DNA/RNA”** and ⊙**“Codons”** options are **deactivated**.

3) Four **color schemes** are available for each case (DNA and Proteins) from the **“Colors”** sub-menu (see right figure below).

4) The option **☑“Show Grid”** shows/hides the table grid.

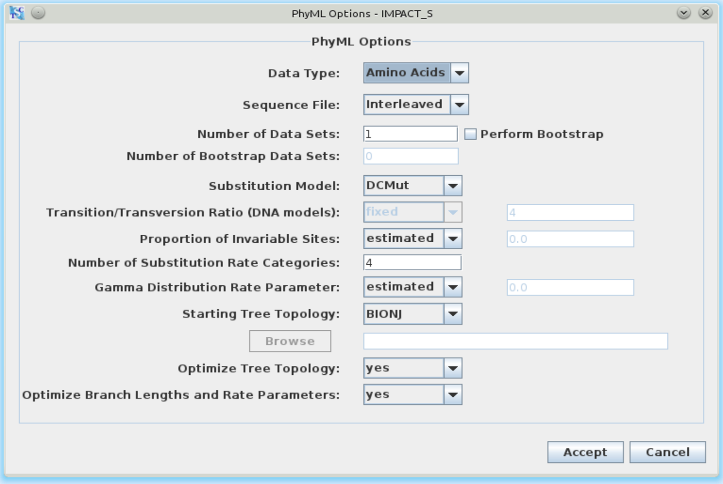

MSA Editor / Viewer - IMPACT_S <Caenophidian_CRISPs_Sequence.fas>
File Edit View PhyML
M.S. Align
○ DNA/RNA
○ Codons
Proteins
Colors
☑ Show Grid Alt-G
☑ Minimum Size Columns Alt-M
□ Only Selected Alt-S
● Sequence Selection ○ Site Selection
● Standard Table (transl_table 1)
○ Vertebrate mt. (transl_table 2)
○ Yeast mt. (transl_table 3)
○ Mold, Protozoan, Coelentarate, Mt.; Mycoplasma, Sp (transl_table 4)
○ Invertebrate mt. (transl_table 5)
○ Ciliate, Dasycladacean, Hexamita Nuc. (transl_table 6)
○ Echinoderme Flatworm mt. (transl_table 9)
○ Euplotid Nuc. (transl_table 10)
○ Bacterial and Plant Plastid (transl_table 11)
○ Alternative Yeast Nuc. (transl_table 12)
○ Ascidian mt. (transl_table 13)
○ Alternative Flatworm mt. (transl_table 14)
○ Blepharisma Macronuclear (transl_table 15)
○ Chlorophycean mt. (transl_table 16)
○ Trematode mt. (transl_table 21)
○ Scenedesmus obliquus mt. (transl_table 22)
○ Thraustochytrium mt. (transl_table 23)
Select
tel_dhar41
trim_bis40
lei_mada47
naj_kaot08
naj_atr15
lio_poec46
phil_olf39
phil_olf43
phil_olf44
enh_poly49
□ View Only Selected
Reset Alignment
Save Alignment...
PhyML
□ Choose Another Location...
Options...
Run
View Details
New Tree File: pc2/Documents/java/TreeSAAPK_data/PAML/tree_GENID30.txt
Browse
Close



5) The **☑“Minimum Size Columns”** expands the column width when deactivated, thus displaying the column headers containing the site positions (**tool-tips** are always available with this information).

6) The order of taxa (sequences) can be changed using the **button arrows** next to the **“Reset alignment”** button. This is only possible in **⊙“Sequence Selection”** mode.

- **PhyML section –** from Alignment Editor/Viewer window.



Data Type: Amino Acids
Sequence File: Interleaved
Number of Data Sets: 1 □ Perform Bootstrap
Number of Bootstrap Data Sets: 0
Substitution Model: DCMut
Transition/Transversion Ratio (DNA models): fixed 4
Proportion of Invariable Sites: estimated 0.0
Number of Substitution Rate Categories: 4
Gamma Distribution Rate Parameter: estimated 0.0
Starting Tree Topology: BIONJ
Optimize Tree Topology: yes
Optimize Branch Lengths and Rate Parameters: yes

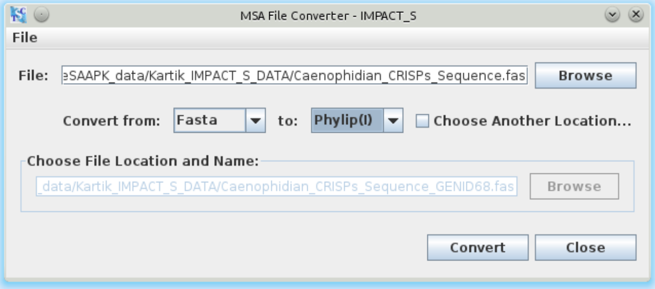
1) The user can run PhyML and build a phylogenetic tree for the loaded alignment. This tree can then be used for running Codeml.

2) The maximum-likelihood tree generated should be in a desired location, and if no location is specified, the resulting files will be located in the same directory as the MSA used.

3) The user can select any **MSA format**. When PhyML is executed, the format is automatically converted to the proper input format accepted by the program (PhyML Options; see figure above).

3.10. *MSA File Format Converter window*

This function was imported from the IMPACT software (http://impact-gui.sourceforge.net). No further modifications were implemented to this window. It allows the user to convert between multiple sequence alignment (MSA) formats: *Mase*, *ClustalW*, *Phylip*(I/S), and *Fasta*. The user selects the MSA file to be converted, the input format of the file, the required format and the output file location.



File: ɘSAAPK_data/Kartik_IMPACT_S_DATA/Caenophidian_CRISPs_Sequence.fas

Convert from: Fasta to: Phylip(I) ☐ Choose Another Location...

Choose File Location and Name: _data/Kartik_IMPACT_S_DATA/Caenophidian_CRISPs_Sequence_GENID68.fas

3.11. *Help and Hints*

Some text fields are designed to make sure that the user has entered all the required information appropriately.

1) Certain fields verify file paths and display hints through **color changes**. If

incorrect path is entered then the field turns **red with white foreground color**. In other fields, if the file already exists, the field turns **white with blue foreground color**.

2) **Tool-tips** are shown in text fields, buttons and check boxes, displaying messages that provide additional explanation.

3) Under **PAML** and **Results & 3D** tabs, **blue squares** with a blue '**i**' at the center can also present useful information.

4) Under the **"Help"** menu the user can find information regarding the **color schemes** used by Jmol (**"Color Schemes")**, along with other useful information about **IMPACT_S** (**"About")**.

Part IV – References

4.1. References

Yang, Z. (2007)

PAML 4: Phylogenetic analysis by maximum likelihood.

Molecular Biology and Evolution 24:1586-1591.

Home Page: http://abacus.gene.ucl.ac.uk/software/paml.html

Yang Z. (1997)

PAML: a program package for phylogenetic analysis by maximum likelihood.

Computational Applied Biosciences 13(5):555-556.

Home Page: http://abacus.gene.ucl.ac.uk/software/paml.html

Wayne D., Poon A.F, Frost S.D.W. and Pond S.L.K. (2010)

Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology.

Bioinformatics 26(19): 2455-2457.

Home Page: http://www.datamonkey.org/

Pond S.L.K. and Frost S.D.W. (2005)

Datamonkey: rapid detection of selective pressure on individual sites of codon alignments.

Bioinformatics 21(10): 2531-2533.

Home Page: http://www.datamonkey.org/

Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. (2003)

TreeSAAP: selection on amino acid properties using phylogenetic trees.

Bioinformatics 19(5):671-672.

Home Page: http://dna.cs.byu.edu/treesaap/

Herraez A. (2006)

Biomolecules in the computer: Jmol to the rescue.

Biochemistry and Molecular Biology Education 4(4):255-261.

Home Page: http://jmol.sourceforge.net/

Maldonado E., Dutheil J.Y., da Fonseca R.R., Vasconcelos V. and Antunes A. (2011)

IMPACT: Integrated Multiprogram Platform for Analyses in ConTest.

Journal of Heredity 102 (3): 366-369.

Home page: http://impact-gui.sourceforge.net/

Arnold K., Bordoli L., Kopp J., and Schwede T. (2006)

The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling.

Bioinformatics 22,195-201.

Home Page: http://swissmodel.expasy.org/

Schwede T., Kopp J., Guex N., and Peitsch M.C. (2003)

SWISS-MODEL: an automated protein homology-modeling server.

Nucleic Acids Research 31: 3381-3385, 2003.

Home Page: http://swissmodel.expasy.org/

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. and Schwede, T. (2009)

Protein structure homology modelling using SWISS-MODEL Workspace.

Nature Protocols, 4,1-13.

Home Page: http://swissmodel.expasy.org/

Guindon S., Gascuel O. (2003)

A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.

Systematic Biology 52(5):696-704.

Home Page: http://www.atgc-montpellier.fr/phyml/

Zmaseck CM, Eddy SR. (2001)

ATV: display and manipulation of annotated phylogenetic trees.

Bioinformatics 17(4):383-384.

Home Page: https://sites.google.com/site/cmzmasek/home/software/archaeopteryx

(Archaeopteryx sucessor of ATV)

Rose, P.W., Beran, B., Bi, C., Bluhm, W.F., Dimitropoulos, D., et al (2011)

The RCSB Protein Data Bank: redesigned web site and web services.

Nucleic Acids Research 39 (suppl 1):D392-D401.

RCSB website http://www.pdb.org/

Rose, P.W., Bluhm, W.F., Beran, B., et al (2011)

The RCSB Protein Data Bank: site functionality and bioinformatics use cases.

NCI - Nature Pathway Interaction Database. doi:10.1038/pid.2011.1

RCSB website http://www.pdb.org/

Gouy M., Guindon S. and Gascuel O. (2010)

SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building.

Molecular Biology and Evolution 27 (2): 221-224.

Home Page: http://pbil.univ-lyon1.fr/software/seaview.html

Gnuplot

Home Page: (http://www.gnuplot.info/)