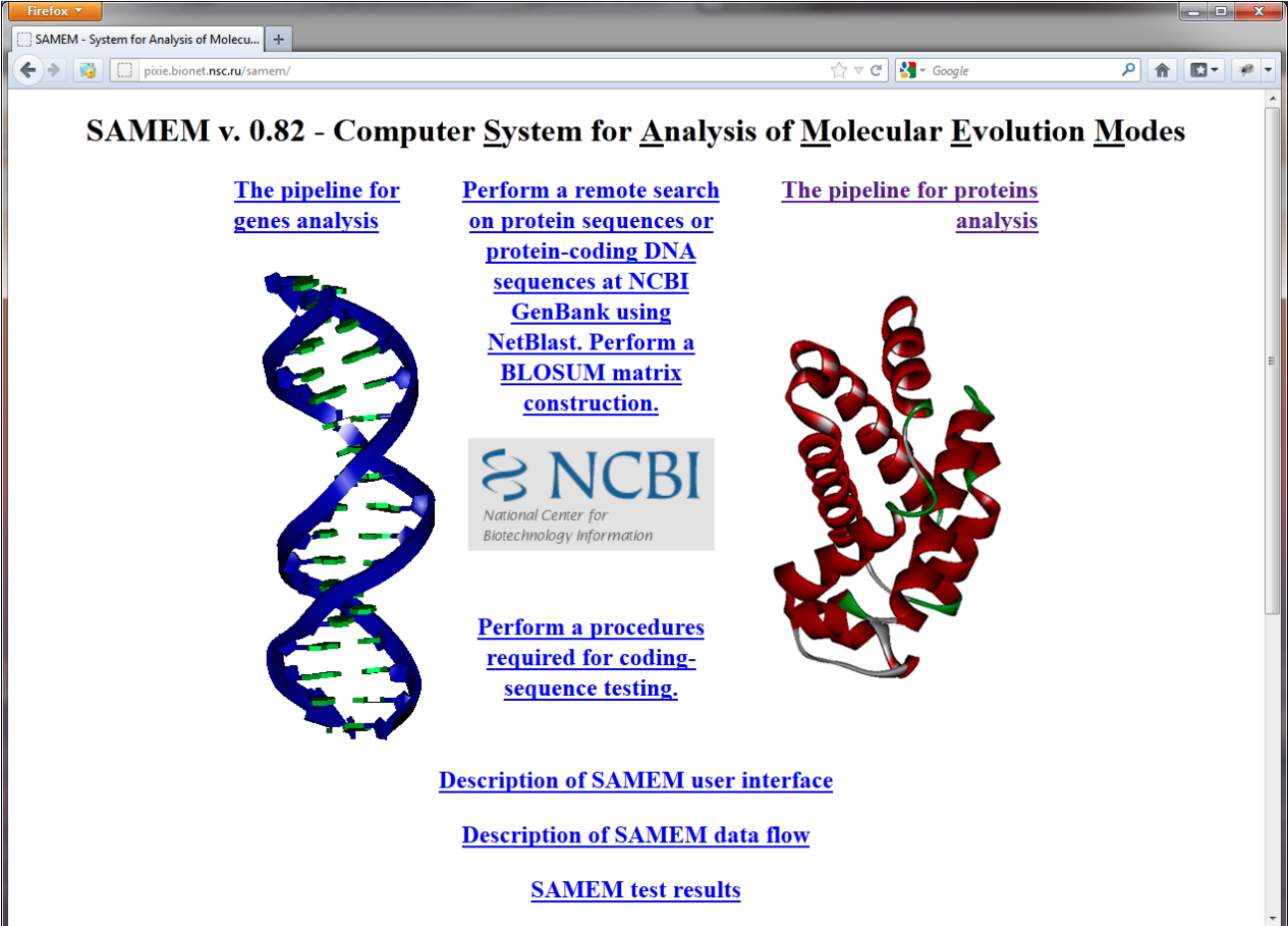


SAMEM user interface

The SAMEM system consists of two main pipelines (gene evolution analysis and protein evolution analysis), and two supplementary pipelines (sample making and coding-sequences testing) (Fig. 1). The common feature of both main pipelines is the possibility of user-guided combination of various methods of common stages of evolutionary analysis such as multiple sequence alignment making, phylogenetic tree construction and ancestral sequences reconstruction.





SAMEM v. 0.82 - Computer System for Analysis of Molecular Evolution Modes

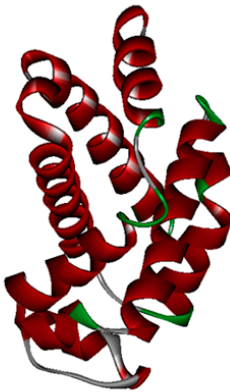
[The pipeline for genes analysis](#)

[Perform a remote search on protein sequences or protein-coding DNA sequences at NCBI GenBank using NetBlast. Perform a BLOSUM matrix construction.](#)

[The pipeline for proteins analysis](#)







[Perform a procedures required for coding-sequence testing.](#)

[Description of SAMEM user interface](#)

[Description of SAMEM data flow](#)

[SAMEM test results](#)

Fig. 1. The SAMEM web-page.

The SAMEM user interface is based on unified structure. The interactive graphical scheme of the pipeline is given on the right side of the screen, as shown in Fig. 2. Additionally, a quick guide explaining how to work with the pipeline is located to the right of the scheme. It includes a description of the input files and parameter settings. This information displayed dynamically depending on the user choice of the pipeline topology.

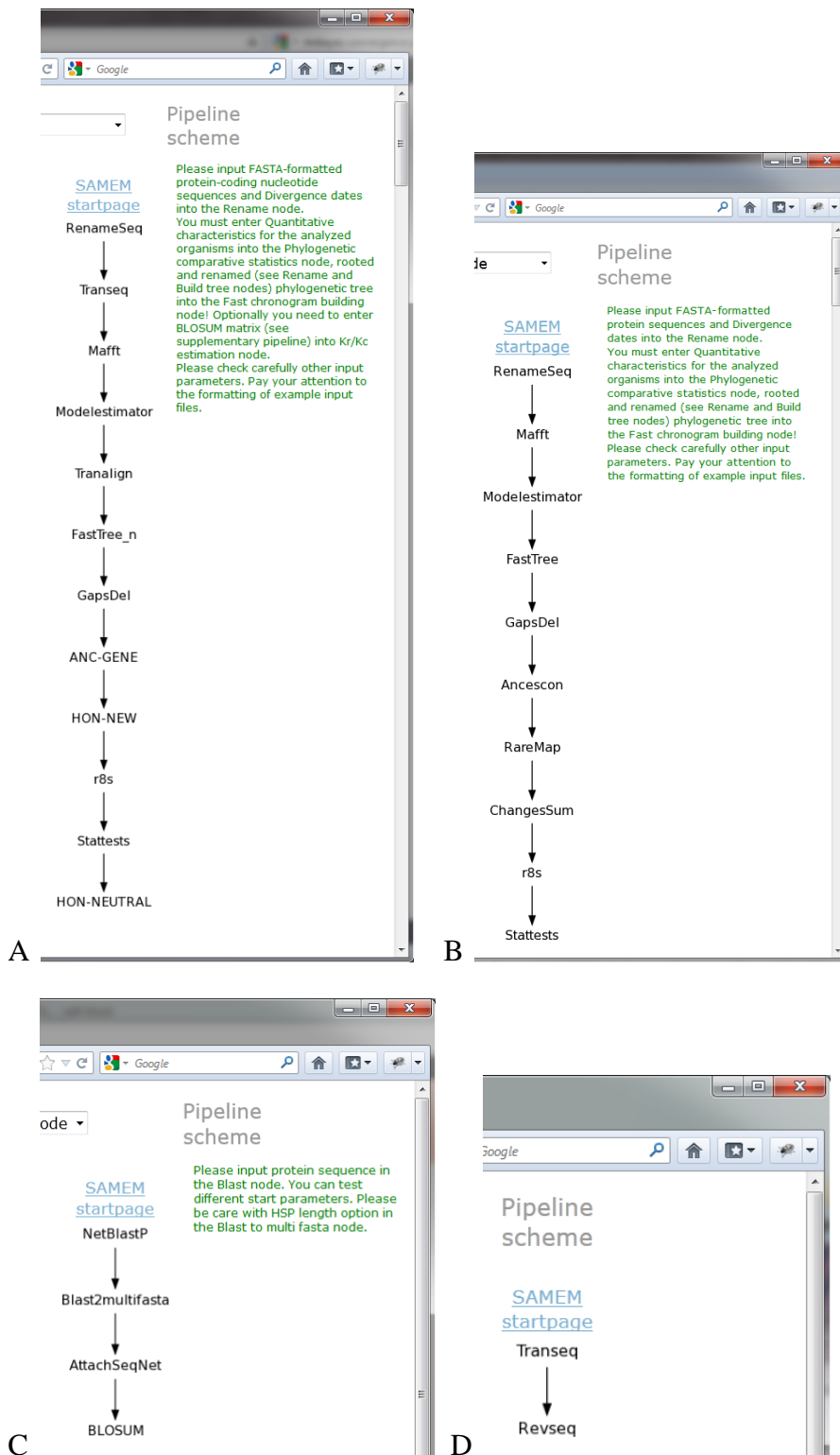


Fig. 2. The interactive graphic scheme of SAMEM pipelines and corresponding quick guide information on the right side displayed in green font. A) Pipeline for analysis of protein-coding genes evolution; B) Pipeline for analysis of protein evolution; C) Supplementary pipeline for sample preparation and BLOSUM matrix generation; D) Supplementary pipeline for checking of protein-coding sequences.

By clicking the name of the program in the graphic scheme of pipeline the user enters its parameter setup panel. User can choose the algorithm of the computational module, change its parameters and specify additional input data if needed. The user can also set their own parameters and intermediate files by clicking on the "Show / Hide input files" link on the node parameter setup panel. (Fig. 3).

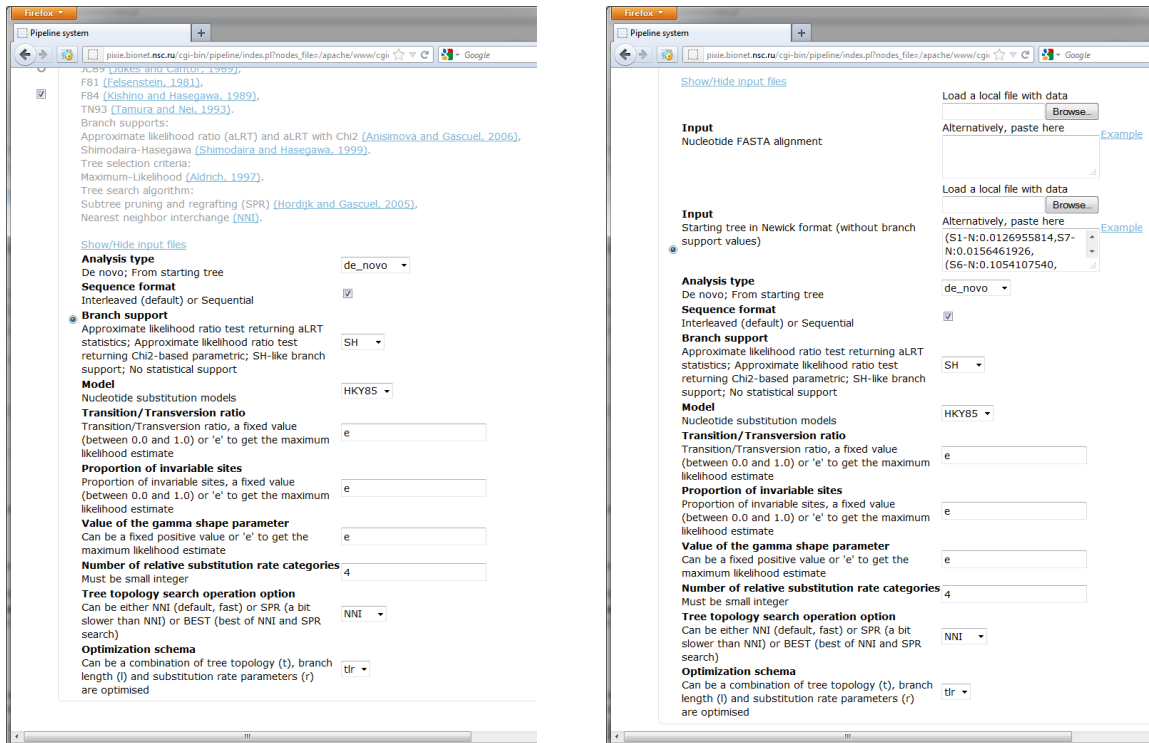


Fig. 3. Examples of the computational module setup panels: setting of parameters by clicking on "Show / Hide input files" link.

The user can set up the start and the stop calculation nodes of pipeline. The start and the stop calculation nodes can be selected by the drop-down menu located at the top of the screen (Fig. 4), or by scrolling through all the calculation nodes of pipeline and changing of start and stop control points (at the left side of the screen).

Firefox Pipeline system

Start: Rename node Stop: Neutral Kr/Kc estimation node

Pipeline scheme

[SAMESM startpage](#)

RenameSeq
↓
Transeq
↓
Mafft
↓
ModelEstimator
↓
Tranalign

Please input FASTA-formatted protein-coding nucleotide sequences and Divergence dates into the Rename node. You must enter Quantitative characteristics for the analyzed organisms into the Phylogenetic comparative statistics node, rooted and renamed (see Rename and Build tree nodes) phylogenetic tree into the Fast chronogram building node! Optionally you need to enter BLOSUM matrix (see supplementary pipeline) into Kr/Kc estimation node. Please check carefully other input parameters. Pay your attention to the formatting of example input files.

1
start
Input
Legenda file

Firefox Pipeline system

Start: Codons to Amino acids Translation node Stop: Neutral Kr/Kc estimation node

Pipeline scheme

[SAMESM startpage](#)

RenameSeq
↓
Transeq
↓
Mafft
↓
ModelEstimator
↓
Tranalign
↓
FastTree_n

Please input FASTA-formatted and renamed (see Rename node) protein-coding nucleotide sequences into the Codons to Amino acids Translation node. You must enter Quantitative characteristics for the analyzed organisms and Legend (see Rename node) into the Phylogenetic comparative statistics node: rooted and renamed (see Rename and Build tree nodes) phylogenetic tree and renamed Divergence dates (see Rename node) into the Fast chronogram building node! Optionally you need to enter BLOSUM matrix (see supplementary pipeline) into Kr/Kc estimation node. Please check carefully other input parameters. Pay your attention to the formatting of example input files.

1
RUN PIPELINE

Rename node

[RenameSeq](#) - Rename sequence names in FASTA file and save original names into separate file

[Show/Hide input files](#)

Rename
Direct or reverse rename: direct

Codons to Amino acids Translation node

[Transeq](#) - Translate nucleic acid sequences into proteins (Rice et al., 2000)

[Show/Hide input files](#)

Firefox Pipeline system

Start: Codons to Amino acids Translation node Stop: Kr/Kc estimation node

Pipeline scheme

[SAMESM startpage](#)

Transeq
↓
Mafft
↓
ModelEstimator
↓
Tranalign
↓
FastTree_n

Please input FASTA-formatted and renamed (see Rename node) protein-coding nucleotide sequences into the Codons to Amino acids Translation node. Optionally you need to enter BLOSUM matrix (see supplementary pipeline) into Kr/Kc estimation node. Please check carefully other input parameters. Pay your attention to the formatting of example input files.

1
RUN PIPELINE

Rename node

[RenameSeq](#) - Rename sequence names in FASTA file and save original names into separate file

[Show/Hide input files](#)

Rename
Direct or reverse rename: direct

Codons to Amino acids Translation node

[Transeq](#) - Translate nucleic acid sequences into proteins (Rice et al., 2000)

[Show/Hide input files](#)

Fig. 4. Setting the start and the stop calculation nodes of pipeline.

The parameter setup panel for each calculation node contains the input fields for input files and the fields for input parameters as shown in Fig. 5. For some calculation nodes the user can change computational algorithm. The choice of the algorithm can be set by the internal control points as shown in Fig. 6.

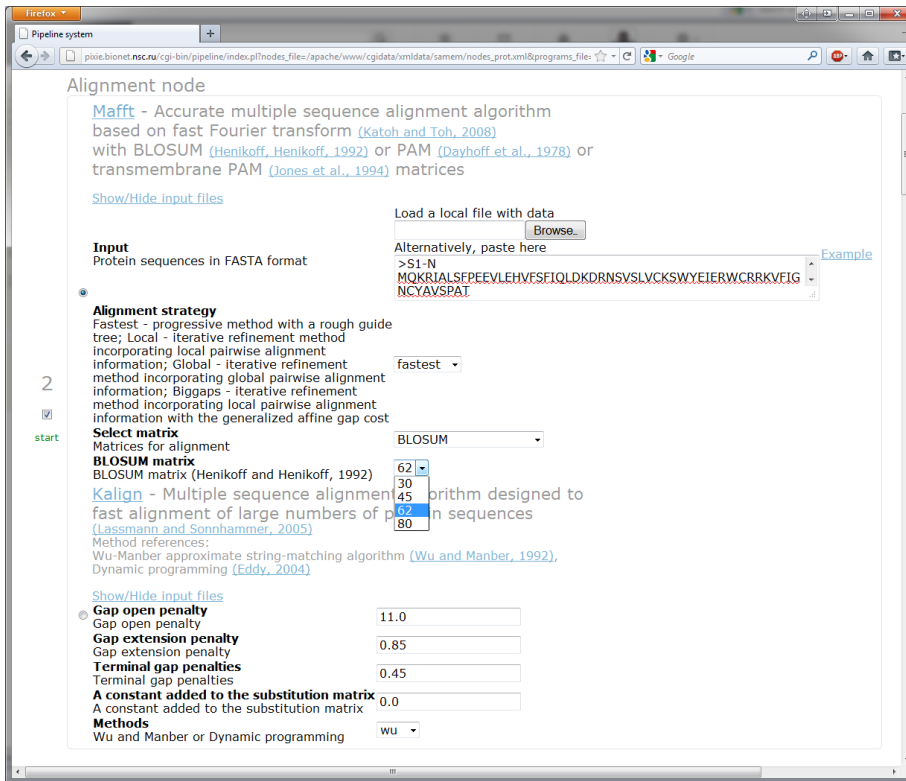


Fig. 5. Examples of the computational module setup panels: setting up parameters using text-boxes and drop-down menus.

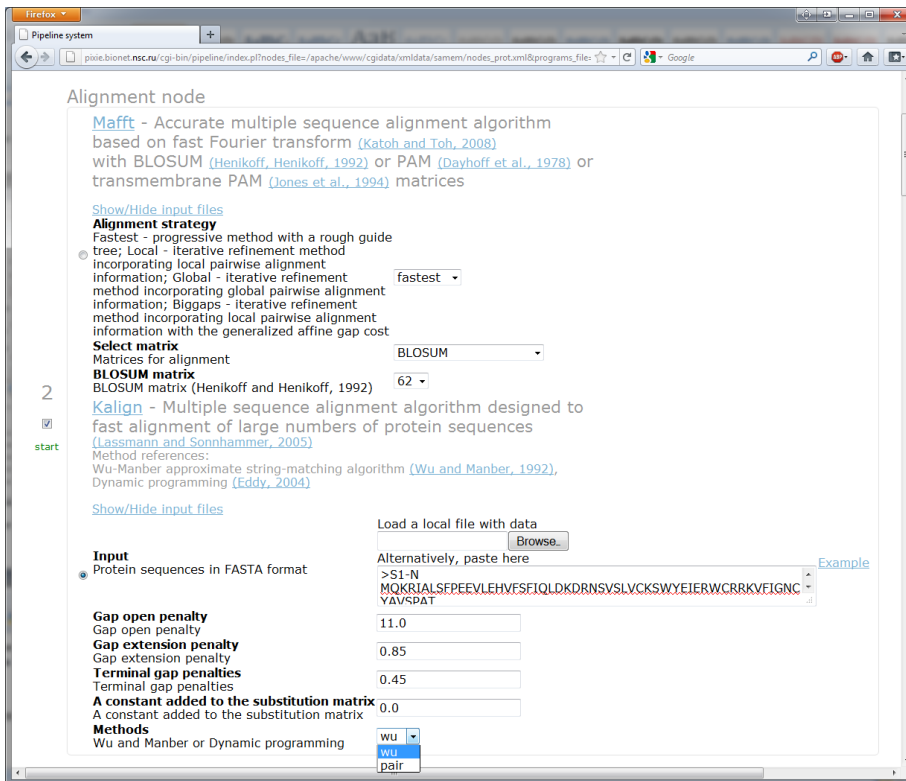


Fig. 6. Changing the computational algorithm into the calculation node (compare with Fig. 5).

Running the pipeline generates a unique task number (Fig. 7) and the corresponding web-page with the information about the progress of task (Fig. 8). This page shows the input and output data and the task execution status (Started, Ended, Failed) for each node of the pipeline.

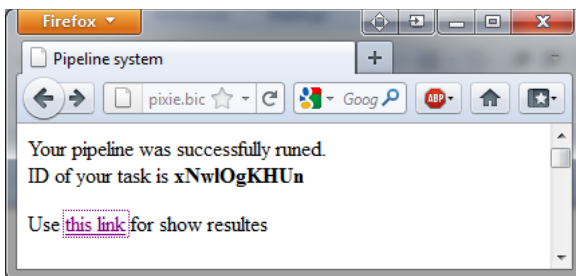


Fig. 7. Unique task number.

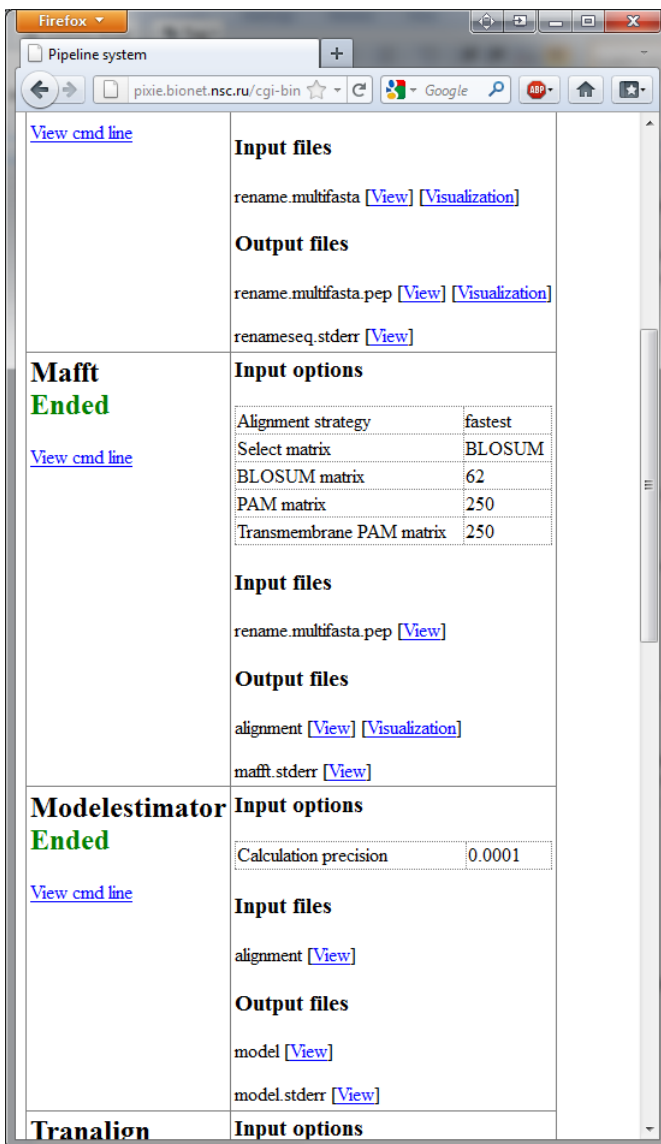


Fig. 8. Web-page with the information about the task execution status.

The web-page containing the information about task execution status is linked with the JalviewLite graphical visualizer of multiple sequence alignment (Fig. 9), the Archaeopteryx graphical visualizer of phylogenetic trees (Fig. 10) and the R (heatmap function) graphical visualizer of numerical tables (Fig. 11).

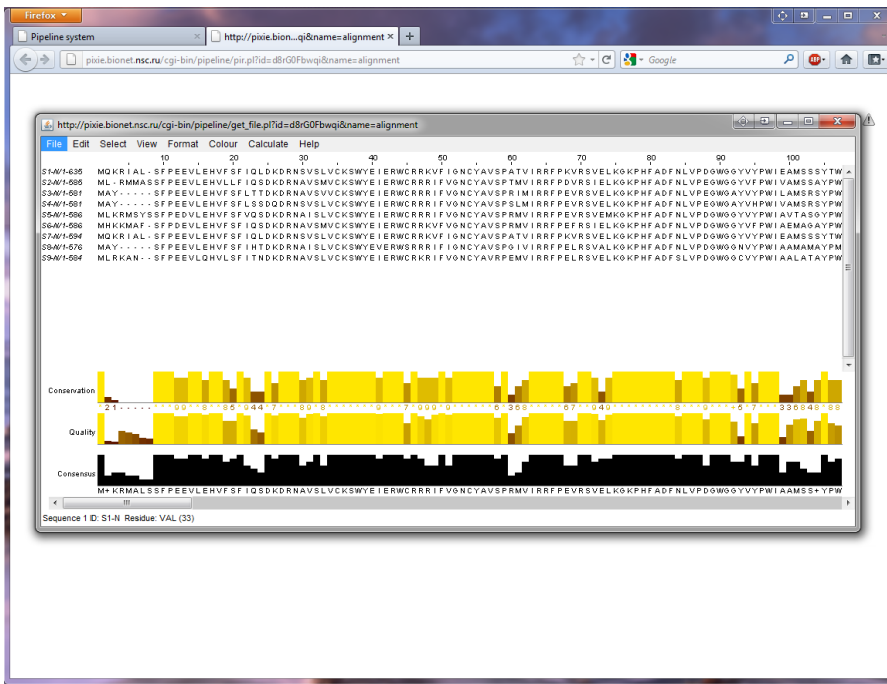


Fig. 9. Visualization of multiple sequence alignment by JalviewLite. It is important to note that the easy-to-use species names were shown. To convert these species names to native ones the first node of the main pipelines must be used.

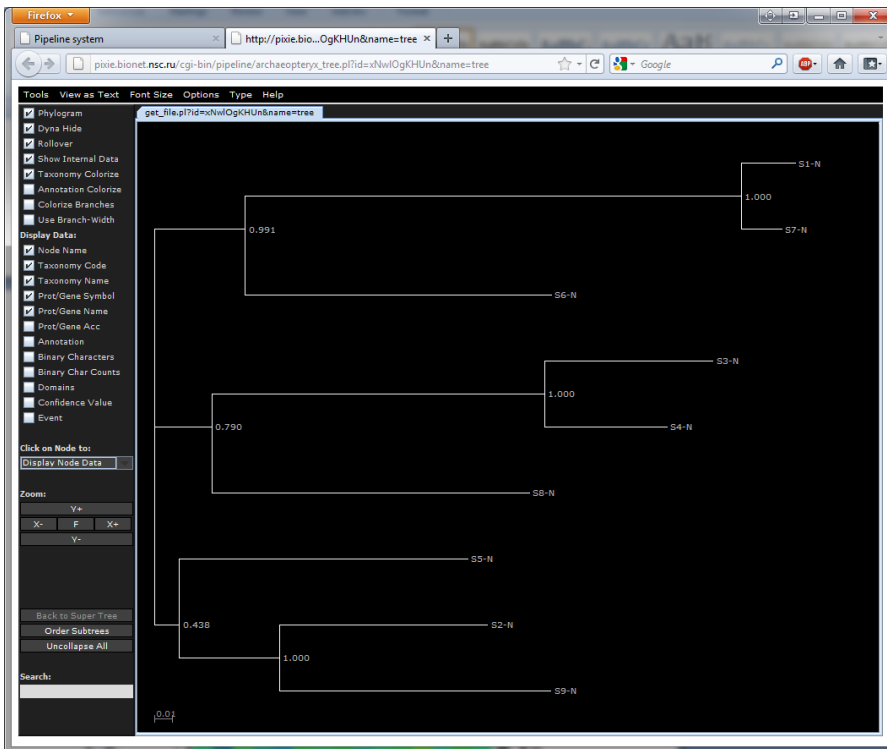


Fig. 10. Visualization of phylogenetic tree by Archaeopteryx. It is also important to note that the easy-to-use species names were shown. To convert these species names to native ones the first node of the main pipelines must be used.

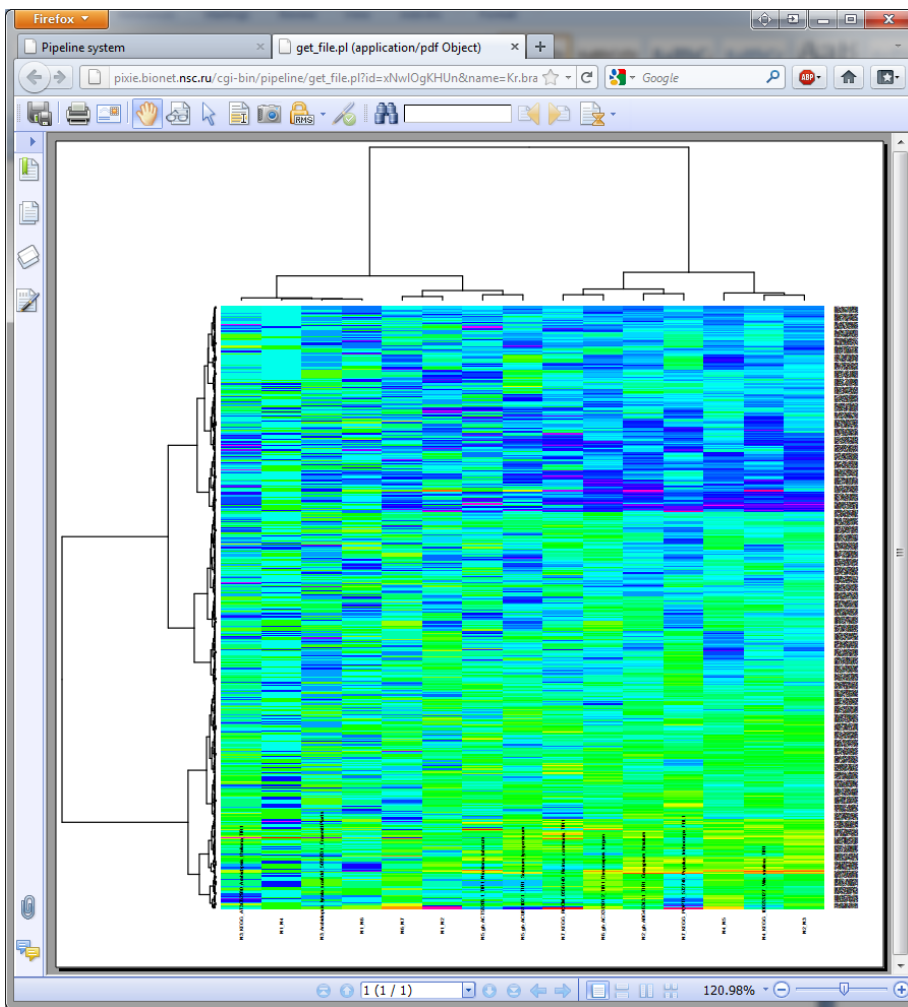


Fig. 11. Heatmap (R) graphical visualizer of numerical tables.

In addition to the visual data representation SAMEM provides a text representation of data (Fig. 12). These text files, information is presented in a simple table form (can be easily exported to MS Excel), in which the rows and the columns represents different data types. For example, the data showed on the Fig. 12 represents species in rows and the physicochemical property changes (from the common ancestor of the analyzed species) in the columns.

	Hydrostatic pressure asymmetry index, FAI Di Giulio, 2005	AA composition of CYT of single-spanning proteins Nakashima-Nishikawa, 1992	Composition of amino acids in nuclear proteins P...
Arabidopsis lyrata scaffold1_303580_1 EnsemblePlanta	0.08879	0.07694 0.08536 0.04586 0.08105 0.05121 0.08006 0.07871 0.04175 0.07369 0.07356 0.07788 0.07353 0.0703 0.07369 0.08339 0.06933 0.0886	
KEGG_RCOM_0556140_Ricinus communis_TIR1	0.05196 0.05095 0.0489 0.03248 0.04785 0.0479 0.05093 0.04477 0.05141 0.04585 0.03847 0.04275 0.03658 0.03762 0.04585 0.04792 0.04167 0.03861 0.04483 0.04153		
gb_ACUS1102_1_TIR1_Solanum lycopersicum	0.0954 0.09753 0.09212 0.04377 0.09112 0.05316 0.08874 0.07333 0.03863 0.0868 0.06778 0.0764 0.07624 0.06771 0.0868 0.09108 0.07727 0.0871 0.08247 0.07002		
gb_ACT53268_1_TIR1_Nicotiana tabacum	0.0851 0.08721 0.07667 0.04783 0.07979 0.04406 0.07429 0.0805 0.04269 0.07959 0.06064 0.07844 0.07419 0.06567 0.07959 0.08489 0.07112 0.08095 0.07526 0.06593		
gb_RCX31301_2_TIR1_Dimocarpus longan	0.05048 0.0569 0.06005 0.03051 0.05468 0.01311 0.05898 0.04522 0.02947 0.04111 0.04406 0.04422 0.03584 0.03166 0.04111 0.05167 0.04201 0.04794 0.03571 0.03885		
gb_AB946343_1_TIR1_Gossypium hirsutum	0.06999 0.06624 0.06719 0.03757 0.06833 0.04489 0.06181 0.06684 0.03344 0.06611 0.06058 0.06928 0.06282 0.05957 0.06611 0.07348 0.05759 0.06915 0.06373 0.05344		
KEGG_AT3G62980_Arabidopsis thaliana_TIR1	0.09281 0.08096 0.08436 0.04586 0.08206 0.05121 0.08207 0.08272 0.04175 0.07871 0.07757 0.08381 0.07956 0.07632 0.07871 0.09142 0.07234 0.09565 0.08806		
KEGG_100233127_Vitis vinifera_TIR1	0.06899 0.06354 0.0768 0.04105 0.07569 0.04216 0.07338 0.0668 0.03689 0.07228 0.05497 0.0668 0.06245 0.06028 0.07228 0.07669 0.06899 0.06471 0.06462 0.05712		
KEGG_POPTR_572746_Populus trichocarpa_FBL1	0.07109 0.06365 0.06373 0.04291 0.06695 0.04477 0.05831 0.06061 0.04181 0.06277 0.06368 0.06502 0.06408 0.06301 0.06277 0.0692 0.0511 0.06615 0.05964		

Fig. 12. Text data representation.

In addition to pipeline data processing, SAMEM has the ability to analyze data by single computational node / program. The user can easily transform the pipeline to the set of individual programs by clicking on the name of the program in the pipeline, as shown in Fig. 13. Note, that when you doing so, the pipeline programs are grouped on the basis of their function (Fig. 13). This SAMEM feature greatly simplifies the task to reanalyzing data using separate pipeline computational nodes or programs.

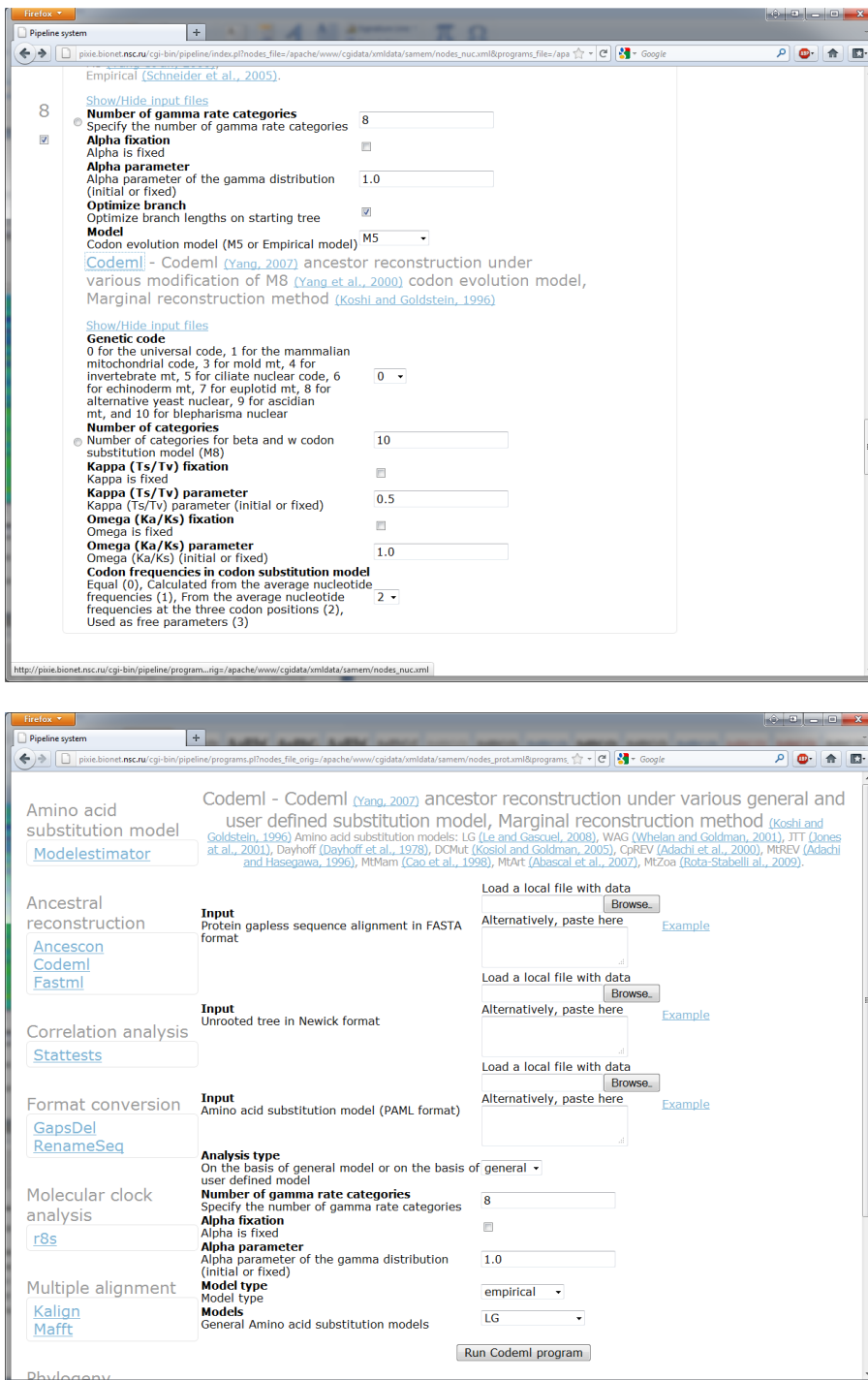


Fig. 13. Analyzing data by single programs.