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.

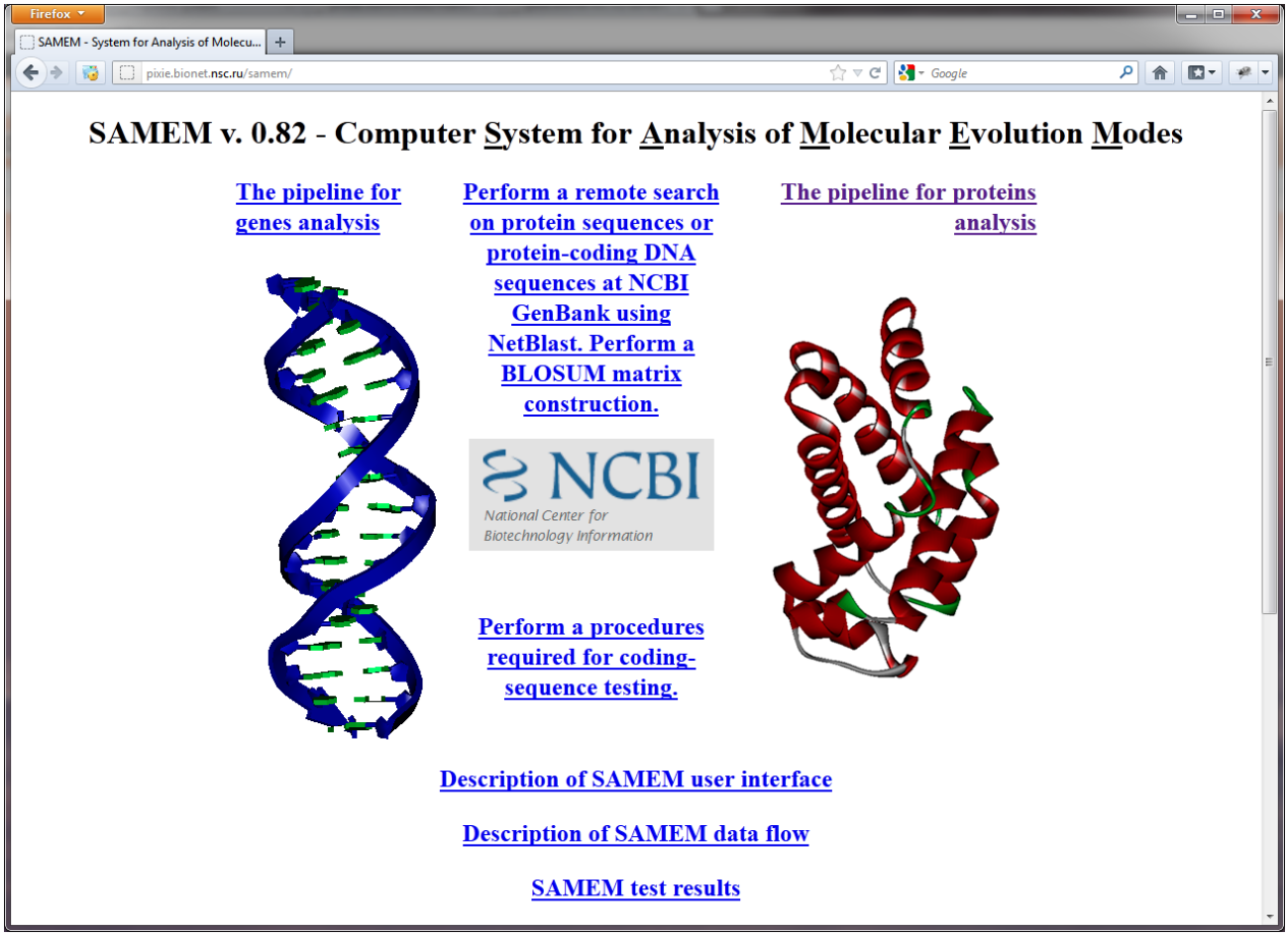

Fig. 1. The SAMEM web-page.

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.

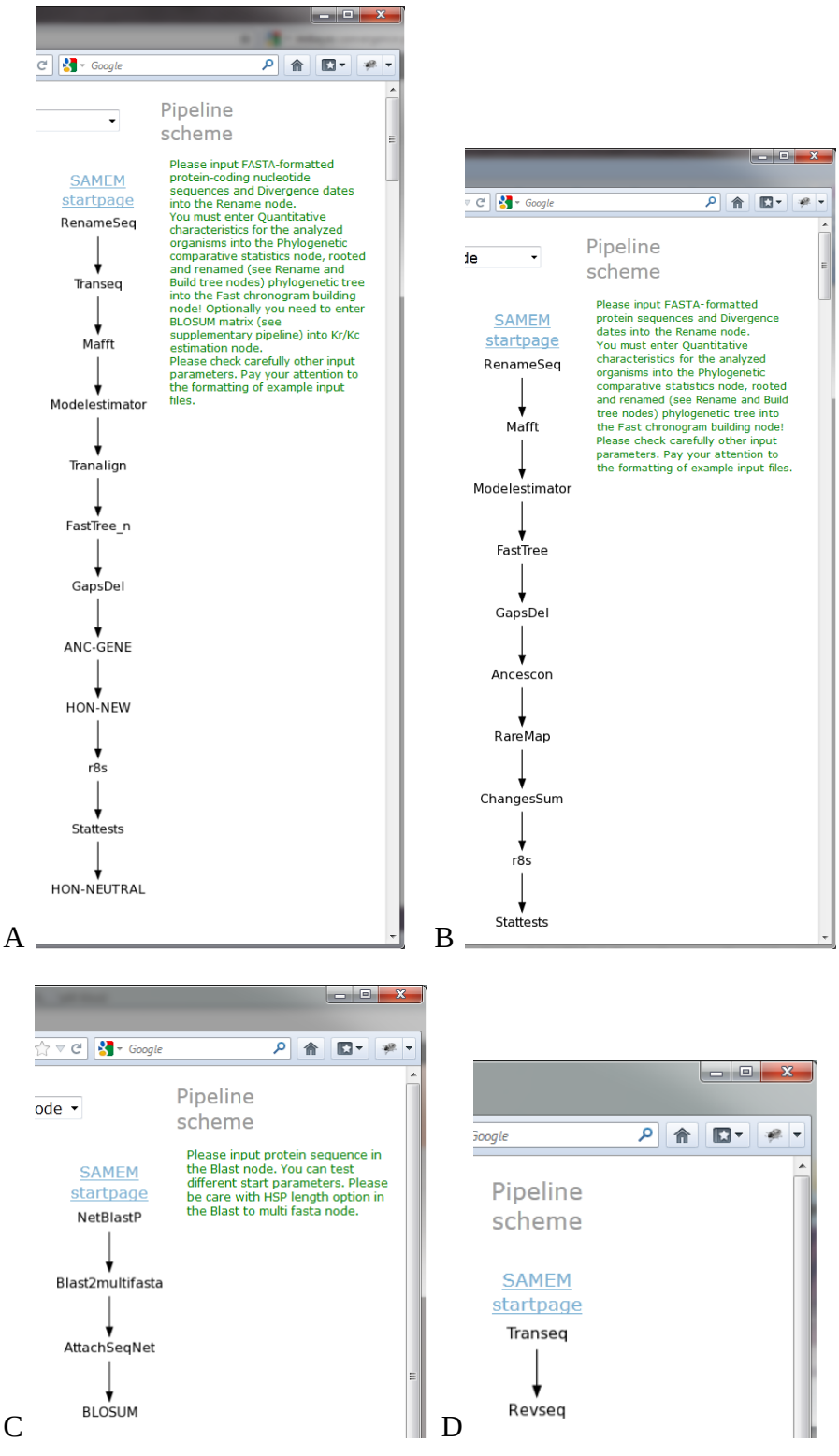

Fig. 2. The interactive graphic scheme of SAMEM pipelines and corresponding quick guide information on the right side displayed in green font. A) Pipeline for analysis of protein-coding genes evolution; B) Pipeline for analysis of protein evolution; C) Supplementary pipeline for sample preparation and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

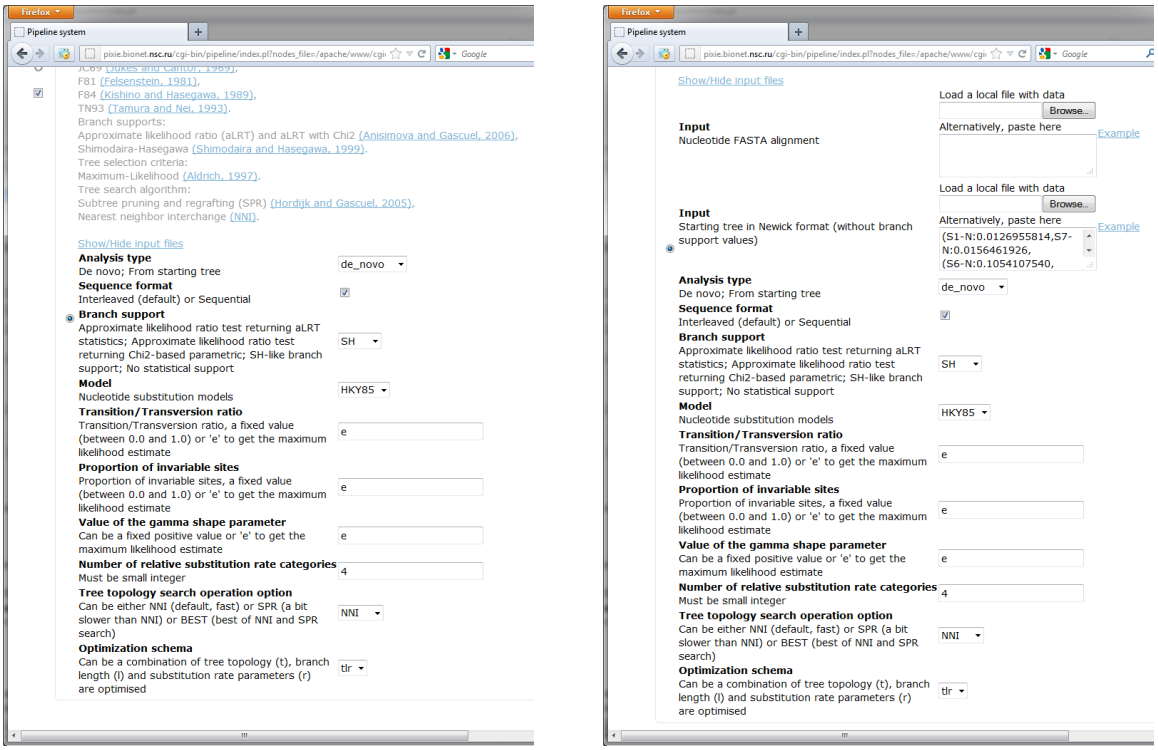

Fig. 3. Examples of the computational module setup panels: setting of parameters by clicking on "Show / Hide input files" link.

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).

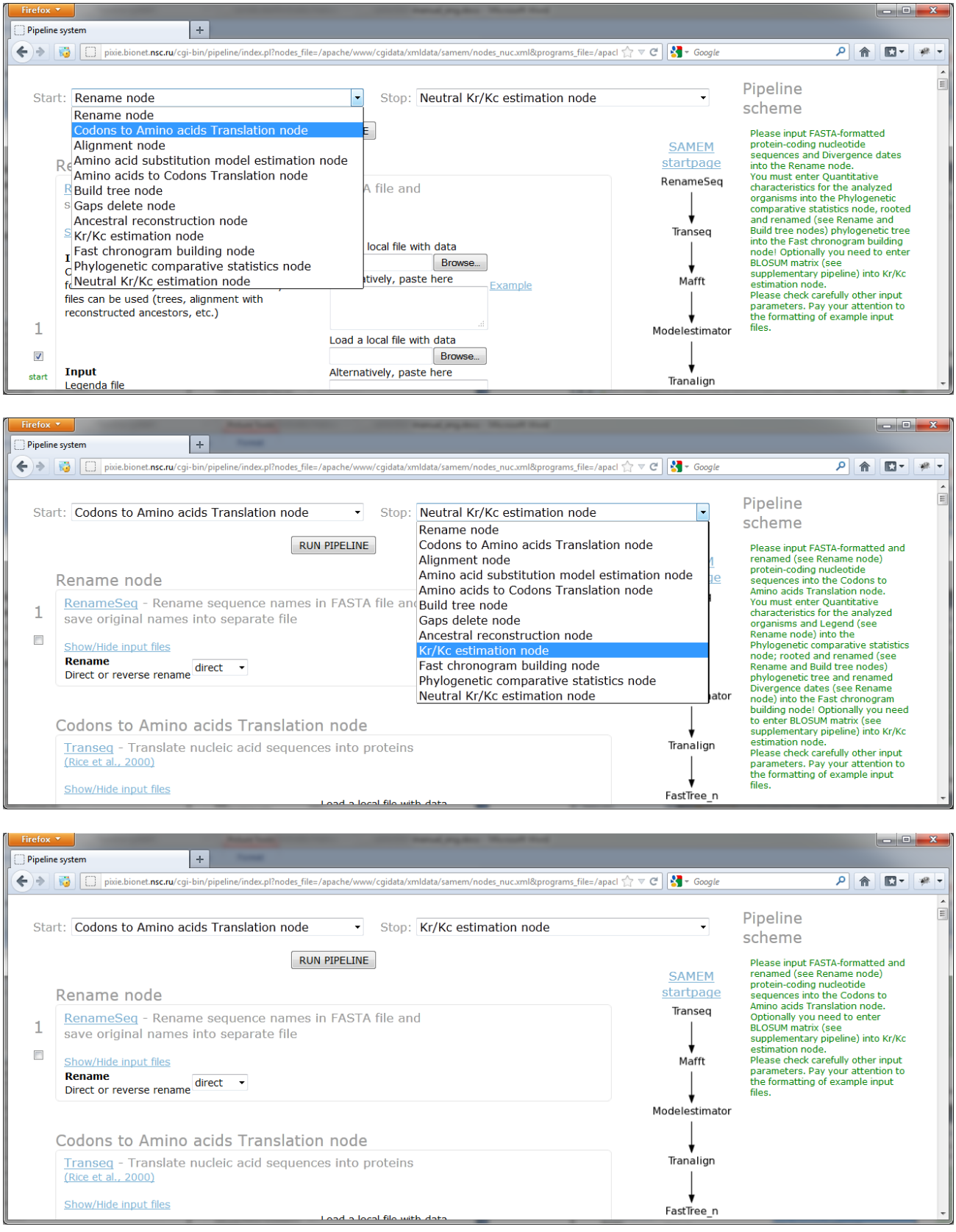

Fig. 4. Setting the start and the stop calculation nodes of pipeline.

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.

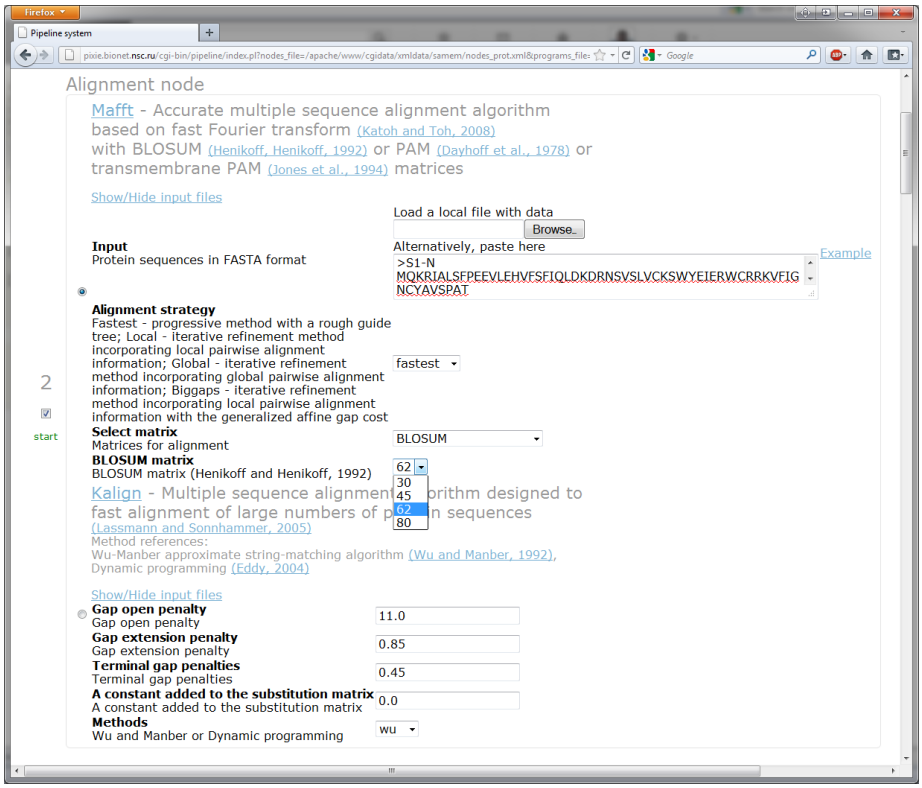

Fig. 5. Examples of the computational module setup panels: setting up parameters using text-boxes and drop-down menus.

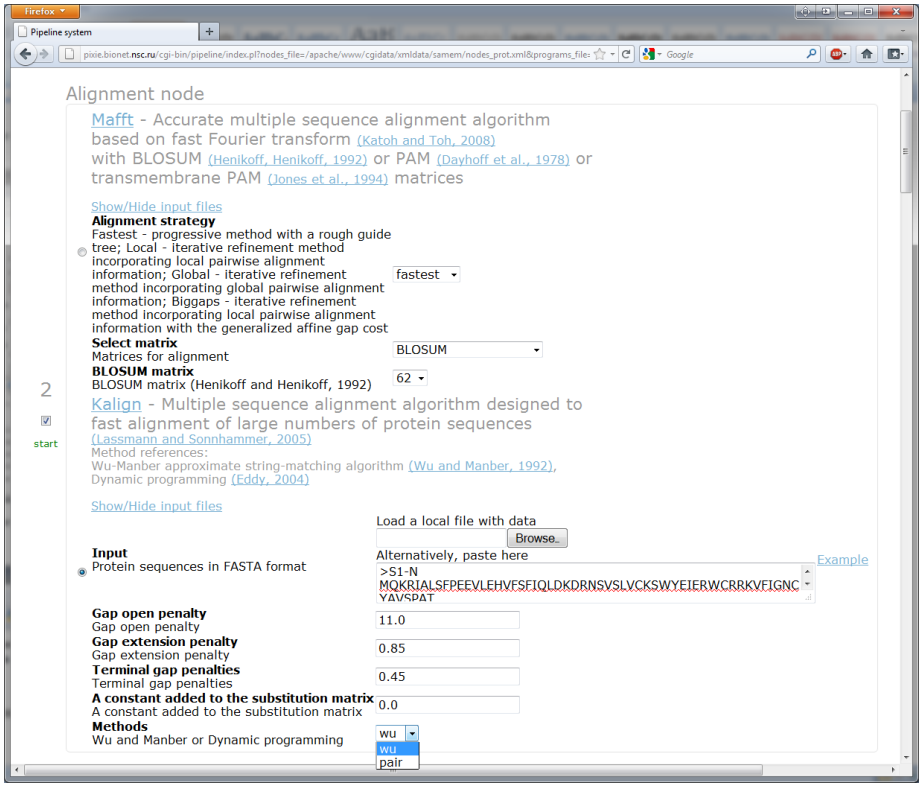

Fig. 6. Changing the computational algorithm into the calculation node (compare with Fig. 5).

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.

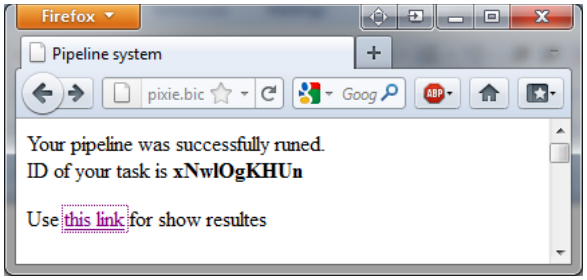

Fig. 7. Unique task number.

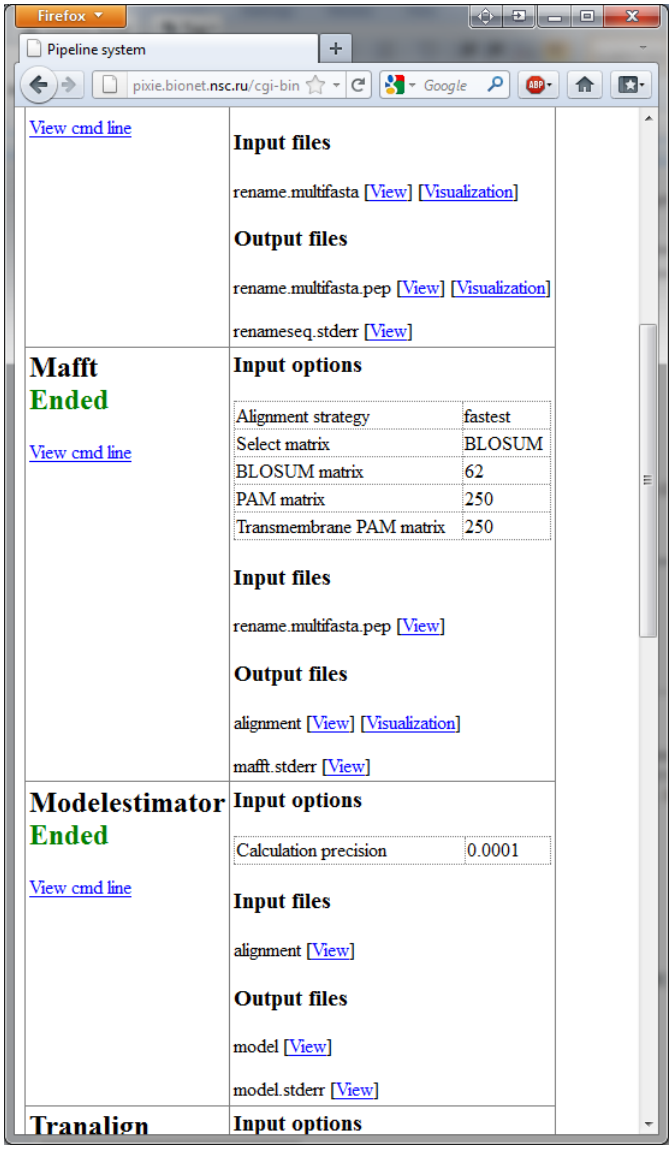

Fig. 8. Web-page with the information about the task execution status.

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).

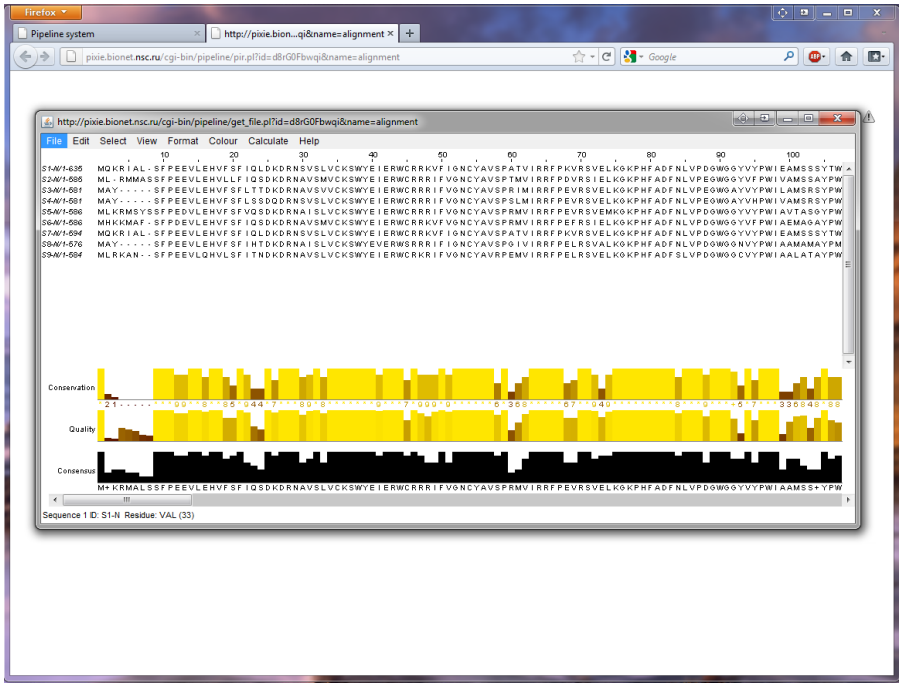

Fig. 9. Visualization of multiple sequence alignment by JalviewLite. It is important to note that the easy-to-use species names were shown. To convert these species names to native ones the first node of the main pipelines must be used.

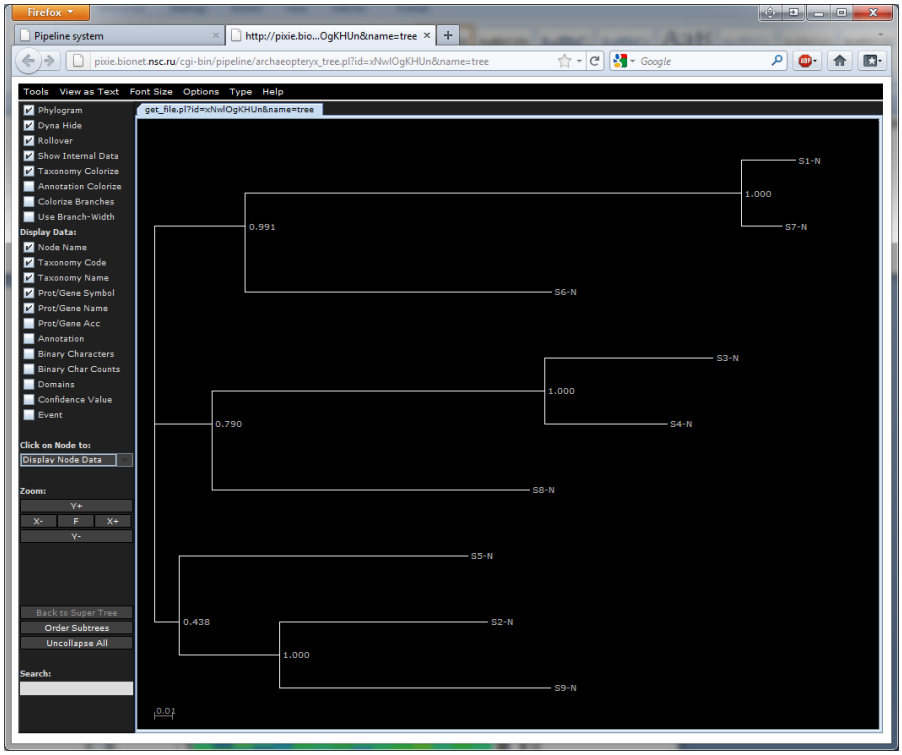

Fig. 10. Visualization of phylogenetic tree by Archaeopteryx. It is also important to note that the easy-to-use species names were shown. To convert these species names to native ones the first node of the main pipelines must be used.

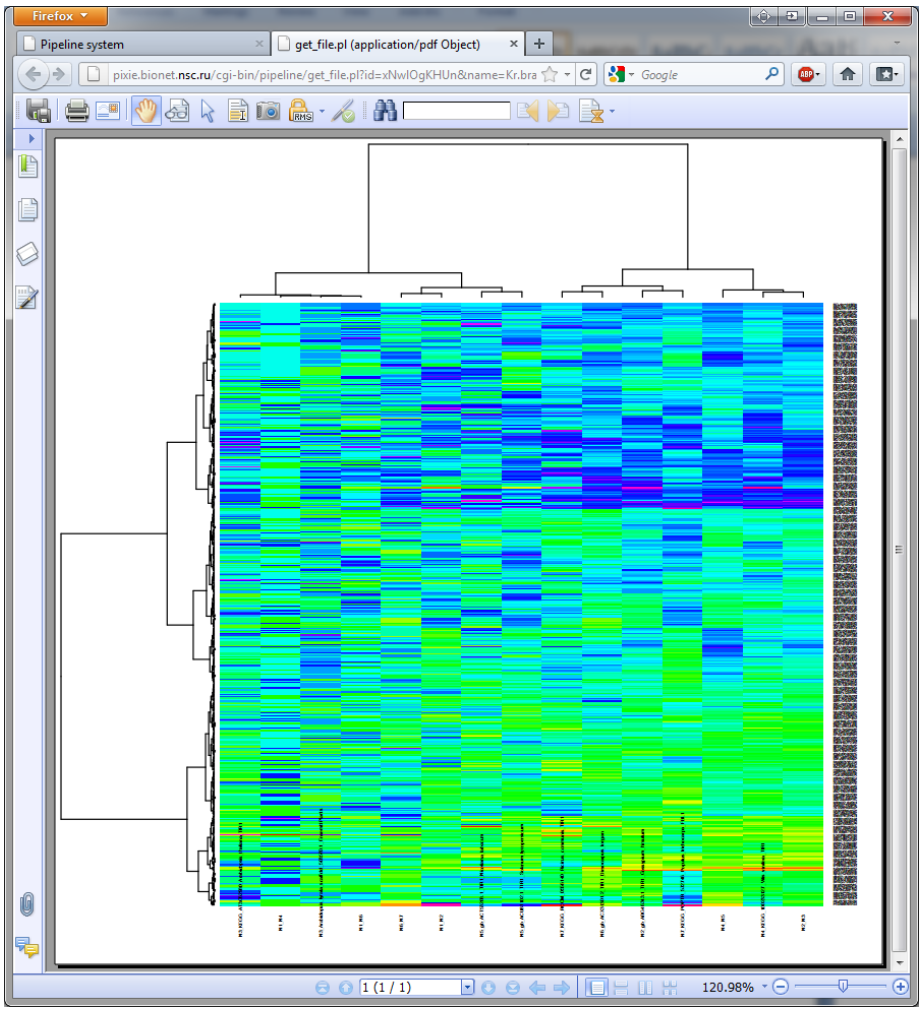

Fig. 11. Heatmap (R) graphical visualizer of numerical tables.

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.

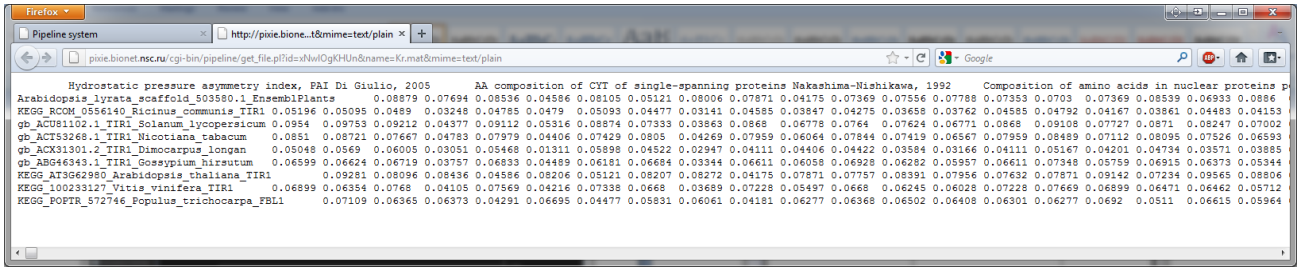

Fig. 12. Text data representation.

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.

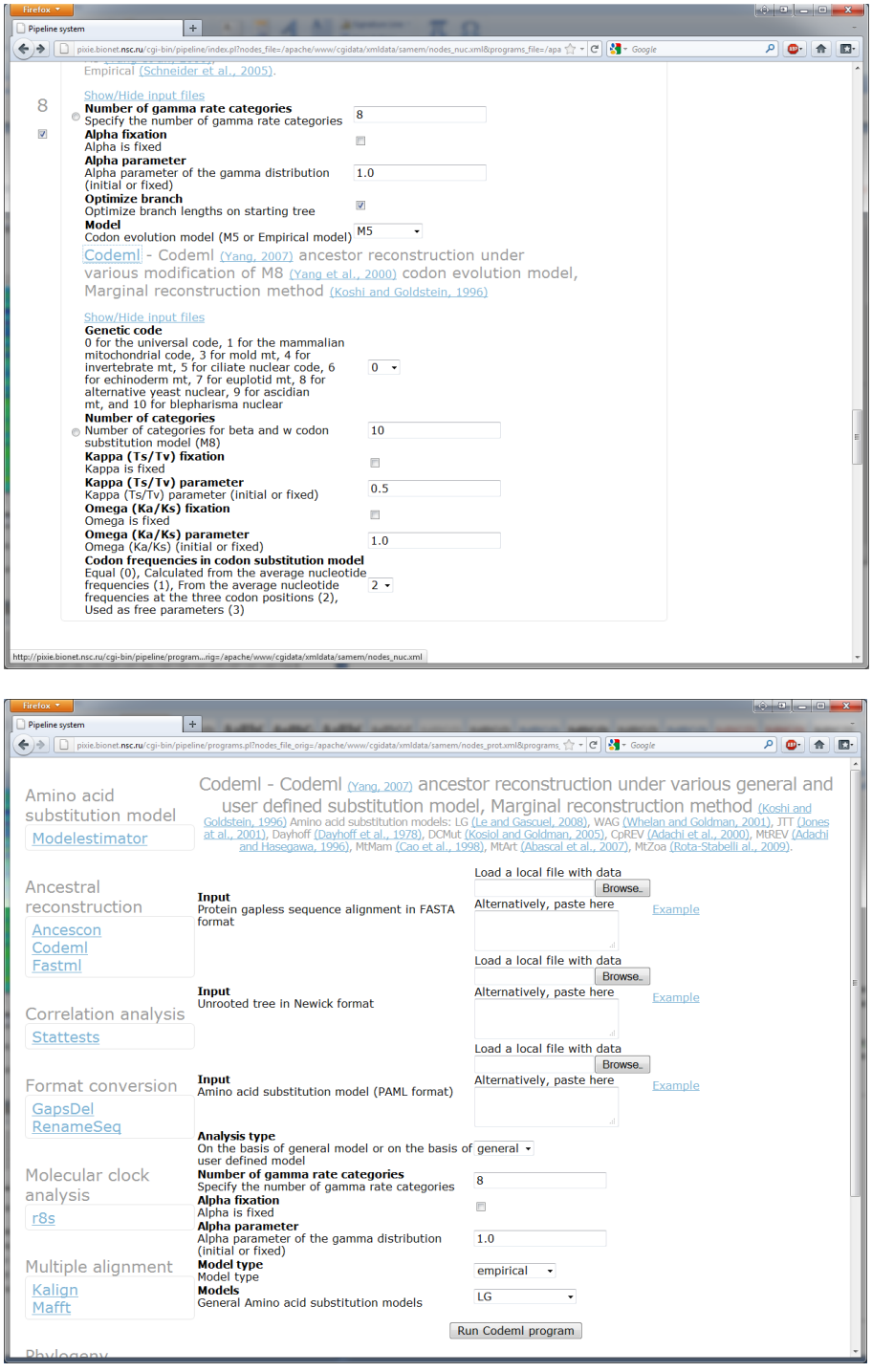

Fig. 13. Analyzing data by single programs.