TRAPID documentation

General documentation

The TRAPID platform is available at

http://bioinformatics.psb.ugent.be/webtools/trapid

1. User Authentication

Data security is a necessary concern when dealing with online platforms and services. Through the use of user authentication no user has access to the data of any other user. User authentication is performed through username/password combination.

To acquire a username/password combination for the platform, select the “Register” option when visiting the TRAPID website. After supplying a valid email-address an associated password will be send to you. Using the email-address/password combination the user gains access to a – to the user – restricted area within the TRAPID platform.

2. Creating TRAPID Experiments

The transcriptome data should be uploaded to the TRAPID platform. Before doing this, it is important to note that, after authentication, the user has the ability to create different experiments for different transcriptome data sets, with a maximum of five experiments per user. Analyzing different transcriptome data sets at the same time thus becomes feasible. The most important choice to be made here is what kind of reference database the user would like to use. The PLAZA reference database should be very good for transcriptome data sets from plant species, while the OrthoMCLdb reference database should be used for any other species.

After the creation of a TRAPID experiment, the user should upload his transcriptome data to the platform. The transcriptome data should be made available as a multi-fasta file before upload to the server. In order to accommodate for the rather large file-size associated with plain-text multi-fasta files, the uploaded file can also be compressed using zip or gzip. If the transcriptome data is split over several multi-fasta files, the user has the ability to continue uploading data into his transcriptome data set before starting the processing phase.

The processing phase of the TRAPID platform is the next step, and necessary before any of the user custom analyses can be performed. This phase is initiated by selecting the “Perform Initial Processing” link on an experiment page. During this step, the user should consider the options carefully, as they may considerably alter the results. First and foremost is the choice of whether either a single species, a

phylogenetic clade or the gene family representatives will be used for the similarity search. A single species is a good choice if in the reference database a close relative of the transcriptome species is present. Selection of an encompassing or closely related phylogenetic clade is a good choice if the user requires more protein coverage and/or no very close relative is available. Using the gene family representatives is a good choice if the user requires a good sample distribution of the gene content within each reference database.

After this step, the experiment will become available while the server performs the initial processing of the data. Note that this step, depending on the size of the dataset and the option, might take several hours. An e-mail will be sent to you when this processing is completed.

3. Available Analyses

After the initial processing of the data has been performed, several new data types are available for a TRAPID experiment, such as gene families and functional annotation (GO categories or protein domains). Using these extra data types offers exciting new analyses to the user.

An extra item which needs to be given extra attention is the **toolbox**. On mist pages (experiment/transcript/gene family/GO/protein domain) a toolbox is available which contains the most common analyses to be performed on the given data object.

Subsets and labels

If the data set is comprised of transcriptome data from different sources (with sources indicating different tissues, developmental types or stress conditions), then the user has the ability to assign labels to the subsets. This is done through the “import transcript labels” link on the experiment page. By doing so, several new analyses become available, such as comparison of functional annotation between different subsets, or computation of the enrichment against the complete transcriptome.

Searching for data

The user has the ability to search for a number of possible data types within his selected experiment, through the search interface. Functional annotation can be searched for both through direct term identifiers (e.g. GO:0005509) or through the descriptions (e.g. Calcium ion binding).

Exporting data

The TRAPID platform allows the export of both the original data and the annotated and processed data of a user experiment. This data access is available under “import/export data” on the experiment page.

Frameshift corrections

A known flaw in NGS technologies is the tendency to produce more errors in the reads than Sanger sequencing. While the assembly procedure can fix most of these errors in case a high enough coverage is possible, there is still a relative high number of transcripts which can have one or more errors, going from substitutions to indels (inserts/deletions). Within the TRAPID framework users have the ability to

automatically correct transcripts with one or more indels, using the FrameDP software. To do this, the user can go to either a transcript or gene family page, and select the “Correct frameshifts with FrameDP” option within the toolbox on those pages.

Creating multiple sequence alignments and phylogenetic trees

A common task is the creation of a custom multiple sequence alignment (MSA) or phylogenetic tree, using selected transcripts and a set of selected genes. Within the TRAPID platform, this functionality is present and can be accessed from the toolbox on a gene family page.

During the creation of these MSA’s or trees the user has the ability to select, for the reference gene family, which species he would like to have within the MSA or tree. As such, it is easy to create, for example, monocot or dicot specific results. The default setting is to have all species from within the gene family present in the result.

The user can at the same time also select which transcripts he would like to have within the MSA or tree. The default setting allows for all transcripts to be present in the result.

All phylogenetic trees are build using PhyML, which starts from a multiple sequence alignment. However, in order to make a better alignment, and improve the bootstrapping, an editing procedure is applied to the default MSA . Thus, when building a phylogenetic tree, two different extra options are available: the MSA editing (stringent/relaxed) and the required bootstrap support (a higher bootstrap implies longer processing time but also more confidence in the result).

4. Logging system

The TRAPID platform keeps track of all previous actions associated with a TRAPID experiment. This allows the user to review past actions. For each experiment, the log is available from the experiment page.