T-ACE Manual

30.03.2012

IKMB, UK S-H

Lars Kraemer

        - Add GO terms
        - Add InterProScan results
        - Add polyA information
    - Import sequences
        - Add ACE file
        - Add nucleotide file
        - Add protein file
        - Add SAM file
    - New nucleotide sequence
    - Translate contigs to proteins
    - Update
        - Update blast2interpro entries
        - Update coverage
        - Update database info
        - Update domain2interpro entries
        - Update GO-statistic
        - Update KEGG-maps
        - Update protein descriptions
        - Update protein entries
        - Update read distribution
        - Update read status
        - Update sequence descriptions
        - Update sequence entries
- Invert selection
- Remove selected sequences
- Save
- Show

- File menu
    - Delete selected DB
    - Export
    - Import
    - Load config
    - Log in
    - New DB
    - New project
    - Open project
    - Quit
    - Save config
    - Save project
    - Select DB

Why T-ACE

Next generation sequencing (NGS) technologies allow a rapid and cost-effective accumulation of large RNA sequence data sets in model and non-model organisms. However the subsequent handling of sequence information from different NGS platforms is still a significant bottleneck, leading to a delay in data processing. We present here the Transcriptome Analysis and Comparison Explorer (T-ACE); a tool designed for the organization and analysis of large sequence datasets, and especially suited for non-model organisms lacking genome information. T-ACE offers a TCL based database interface, which accesses a PostgreSQL database via a php-script. Within T-ACE, information belonging to a single sequence or contig, such as annotation or read coverage, is linked to the respective sequence and immediately accessible. Sequences and assigned information can be searched via keyword- or BLAST search, and whole transcriptomes analyzed on the level of expression, GO terms, KEGG pathways and protein domains. Results are visualized and can be easily exported for external analysis or further graphical processing. We developed T-ACE for laboratory environments which have only a limited amount of bioinformatic support, and cooperation-projects in which different partners work on the same dataset from different locations or platforms (Windows/Linux/MacOS).

Have also in mind that T-ACE is not designed as a genome browser, but for de-novo transcriptomes. This means, in the current state of T-ACE, contig sequences should not be much longer than 60.000 bp and should not contain over 100.000 reads (at least if those reads shall be displayed).

Installation

Setting up a T-ACE Client

If there is already a running T-ACE database server, there is not much to do to set up a T-ACE client. At first TCL/TK 8.510.1 or higher should be installed. The easiest way to do this is to use an ActiveState package; it contains already most necessary additional TCL packages. The additional packages are listed in the Required Software section below.
When TCL is installed, just download T-ACE from http://www.ikmb.uni-kiel.de/tace/ and extract it to any convenient directory. Then double-click on the T-ACE.tcl file in the T-ACE folder and the client interface should open. Alternatively T-ACE can be started from command-line with either ‘tclsh’ or ‘wish’ (windows), both executables can be found in the /bin folder of the TCL installation. Just execute the following command from the T-ACE directory:

```
tclsh T-ACE.tcl
```

To set up the connection to the T-ACE database server open the 'Database connection' in the config menu.
It is also recommended to install additional software like NCBI-Blast+, Phobos and Primer3; otherwise some T-ACE features will not function. Look in the Required Software section below for specifications.

Setting up a T-ACE database server

To set up a T-ACE database server a PostgreSQL server is necessary. The additional software pgAdmin3 should also be installed for an easier handling of the Postgres database. For the functions used in T-ACE parent database the additional module 'postgresql-pltcl' is needed.
After the installation of the PostgreSQL server some configurations have to be made:

Resetting the password of the postgresuser:

```
sudo su postgres -c psql template1
template1=# ALTER USER postgres WITH PASSWORD 'password';
template1=# \q
```

Then it is necessary to change some lines in the 'postgresql.conf':

```
#listen_addresses = 'localhost'   →   listen_addresses = '*'
#password_encryption = on         →   password_encryption = on
```

The 'pg_hba.conf' file should be edited to use md5 for password checks.
If T-ACEpg is used for non-local access the following line should be added:

```
# TYPE      DATABASE        USER        CIDR-ADDRESS        METHOD
# IPv6 local connections:
Host        tace_parentdb   tace_user   0.0.0.0/0           md5
```

This would allow the user 'tace_user' to access the database 'tace_parentdb' from anywhere in the internet. Change the CIDR-ADDRESS to limit the access further.
The normal T-ACE version accesses the Postgres database via a php script; this means it is not necessary to make the Postgres server accessible for external IP-addresses.

Now the server has to be restarted:

```
sudo /etc/init.d/postgresql restart
```

Then an Apache web server is needed to allow external T-ACE clients the access to the Postgres database. An external T-ACE client will access the Postgres database through a php script which runs on the Apache server, so make sure that php scripts can be executed correctly on the server.

Now TCL/TK and T-ACE can be installed (see above). But there is one difference, the additional TCL package Pgtcl is needed to use the T-ACE_DB_manager.
When T-ACE is extracted, create a directory (e.g. 'tace') in the /htdocs folder of the Apache web server (could be /var/www under linux) and copy the T-ACE.php into it. This folder needs to be set in the server authentification for the 'Database connection' in the T-ACE client.
To set up the T-ACE parent database, functions and login roles follow the instructions in the T-ACE_DB_manager section. When this is done the first T-ACE transcriptome database can be created, as described in 'Creating the first database'.
Additionally wwwblast should be installed on an apache server, but it does not need to run on the same machine as the Postgres server. Then blast databases of the T-ACE databases have to be copied to the /db directory of blast server and added to the blast.rc file, otherwise the blast function of T-ACE will not work.

Different T-ACE versions

T-ACE:
This version accesses the Postgres database through a php script, which has to run on the database server.
Pgtcl is not needed for running the T-ACE client.

T-ACEpg:
This version accesses the Postgres database directly. For this version the Pgtcl package is needed. It should only be used for local database access.

The T-ACE_DB_Manager is the same for both versions it always accesses the database directly and therefore needs the Pgtcl package.
After the required software is installed the T-ACE.tcl and T-ACE_DB_Manager.tcl should be executable. Start with the T-ACE_DB_Manager to create a parentDB.

Required software:

- **TCL/TK 8.5**

  Required packages:
  - Bwidget v1.9.2
  - http
  - Pgtcl v1.7 (needed for T-ACEpg and T-ACE_DB_Manager)
  - Tablelist
  - Thread 2.6.5

  TCL/TK can be downloaded From ActiveState under:
  http://downloads.activestate.com/ActiveTcl/releases/.

  The threaded version is needed. After that only the Pgtcl package needs to be added. It is available at http://sourceforge.net/projects/pgtclng/files/pgtclng/.
  Under Windows the libraries should be copied into the Tcl library directory, like it is described in the README.txt under 'Option 1) Package require'.

- **PostgreSQL 8.4**

  Download at: http://www.postgresql.org/download/

  The package 'postgresql-pltcl' has to be installed; otherwise the T-ACE functions cannot be used.
  Download also pgAdmin3 for management of the postgresql databases.

All following programs are optional, but some features of T-ACE depend on these programs and will not be executable if they are not installed.

- **InterProScan v4.6**

  Download at: ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/index.html

  InterProScan needs be installed with a web interface; otherwise it will not be available through T-ACE. Also a correct setting of the access rights of the interpro folders is needed otherwise T-ACE will not be able to access the search results.

- **NCBI – blast-2.2.25+**

  Download at: ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/

  It should be installed on the client machine and the executables should be added to the system path, otherwise the 'local blast' functions of T-ACE will not work.

- **Phobos v3.3.12**
  Download at: http://www.ruhr-uni-bochum.de/spezzoo/cm/cm_phobos.htm

  Phobos is a highly accurate and fast search tool for DNA-(micro/mini)satellites. To be installed on client machine.
  ATTENTION: make sure executables are added to the system path.

- **Primer3 v1.1.4**
  Download at: http://sourceforge.net/projects/primer3/files/

  PCR Primer design tool. To be installed on client machine.
  ATTENTION: make sure executables are added to the system path.

- **Apache 2.0 HTTP Server**
  Download at: http://httpd.apache.org/download.cgi
  (Necessary for running the T-ACE.php and a Blast server)
  - libapache2-mod-php5
  - apache2-mpm-prefork

- **Blast server**
  wwwblast-2.2.19 from NCBI, at: ftp://ftp.ncbi.nlm.nih.gov/blast/executables/release/LATEST/

  The blast server is necessary to make the blast function of T-ACE available for ever client without the necessity of a local NCBI-blast+ installation and local blast databases of all T-ACE databases.
  Of each T-ACE database two blast databases should be created and copied into the /db folder of the blast server. The databases must be added to the blast.rc file of the blast server otherwise they will not be available.

  The following blast databases of each T-ACE database should be created:

<T-ACE database name>	(containing all contig sequences)
<T-ACE database name>_reads	(containing all read sequences)

  optional:

<T-ACE database name>_prot	(containing all protein sequences)
<T-ACE database name>_singletons	(containing all singleton sequences)

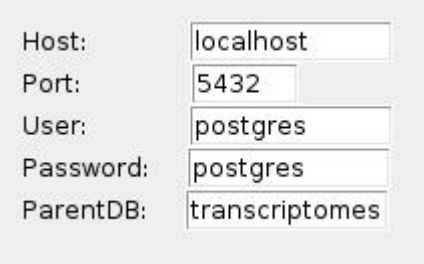
  The blast databases can be created with the 'Create Blast-Database'-function of T-ACE or with the command-line program makeblastdb (in each case NCBI-blast+ has to be installed).
  When using the command-line fasta files of the according sequences are needed, also make sure the option -parse_seqids is used; otherwise the blast hits cannot be reassigned.

T-ACE DB Manager

This script creates the 'parent database', which will contain all databases created with T-ACE as schemata. It also creates the reference schema, which holds most of the annotation information, such as domain descriptions.
Just execute the script to open the GUI or use '-h' for help about the command line usage.

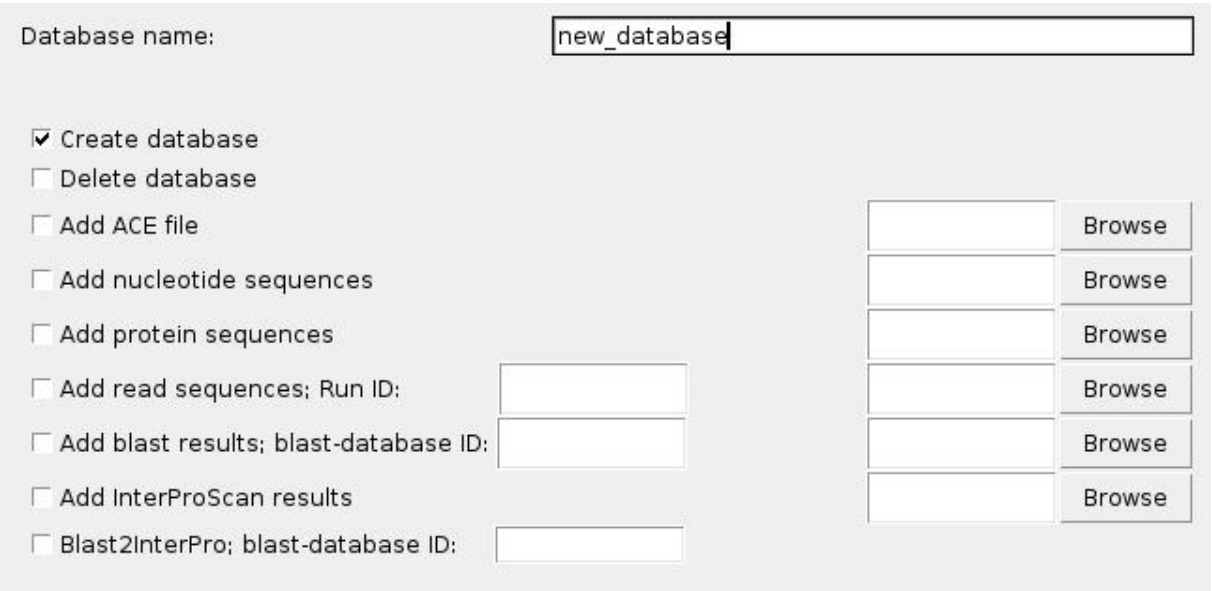
Database connection

The 'Database connection'-tab, sets all needed parameters to establish a connection to the chosen parent database. For all operations in the 'Reference database'-tab or 'User management'-tab it is necessary to login as a superuser. It is not necessary to press the 'Run'-button to set the database connection.



Database options

The 'Database options'-tab allows the creation/deletion of databases. It is also possible to fill databases with sequences and annotations, but it is easier to do all this directly with T-ACE.



Database name

Sets the database schema that shall be modified.

Create database

Creates a new database with the name set in the 'Database name' entry field.

Delete database

Deletes the database with the name set in the 'Database name' entry field.

Add ACE file

Adds the contents (contig and read information) of an ACE file to the selected database. The contig names in the database will consist of the database name and the contig name from the ACE file.

Add nucleotide sequences

Adds the sequences from a fasta file to the selected database as contigs.

Add protein sequences

Adds the sequences from a fasta file to the selected database as proteins. For this an InterProScan protein input file should be used. If a sequence from the file does not have a corresponding contig in the database it will be omitted.

Add read sequences

Adds the sequences from a fasta file to the selected database as reads with the given run id.

Add blast results

Adds the blast hits from a blast output file, in table format, to the selected database. If a blast hit has no corresponding contig in the database it is omitted.

Add InterProScan results

Adds the InterProScan results from an InterProScan output file, in raw format, to the selected database. If a domain hit has no corresponding protein sequence in the database it is omitted.

Blast2InterPro

Searches the 'protein2interpro' reference table for InterPro hits associated with blast hits of the selected database. The InterPro hits, and the corresponding GO terms, are added to the annotations of the selected database.

The 'Blast2InterPro'-function is also available from T-ACE.

Reference database

The 'Reference database'-tab allows the creation of the parent database, functions and reference schema. To fill the different tables of the reference database, various data files are needed. In most cases, the download source of the needed file is shown beneath the corresponding option. Some of the files have to be unpacked before they can be used as input.

It is possible, to start working with T-ACE after the execution of steps 1 and 2. The other steps are more or less optional, depending on the requirements for the sequence annotation.

1. Create parent database

This creates a parent database accordant to the parameters set under 'Database connection'.

2. Create functions

Creates the functions schema of the parent database. The functions are necessary for many T-ACE features. Through the 'User management'-tab the function rights for each user can be set.

If the functions are not created by executing this step make sure 'pltcl' is available in postgres. Then open the parent database via psql and try the following command:

```
psql  -U <user> <parent database>
<parent database>=# CREATE LANGUAGE pltclu
<parent database>=# \q
```

Do not forget to grant function rights to each T-ACE user.

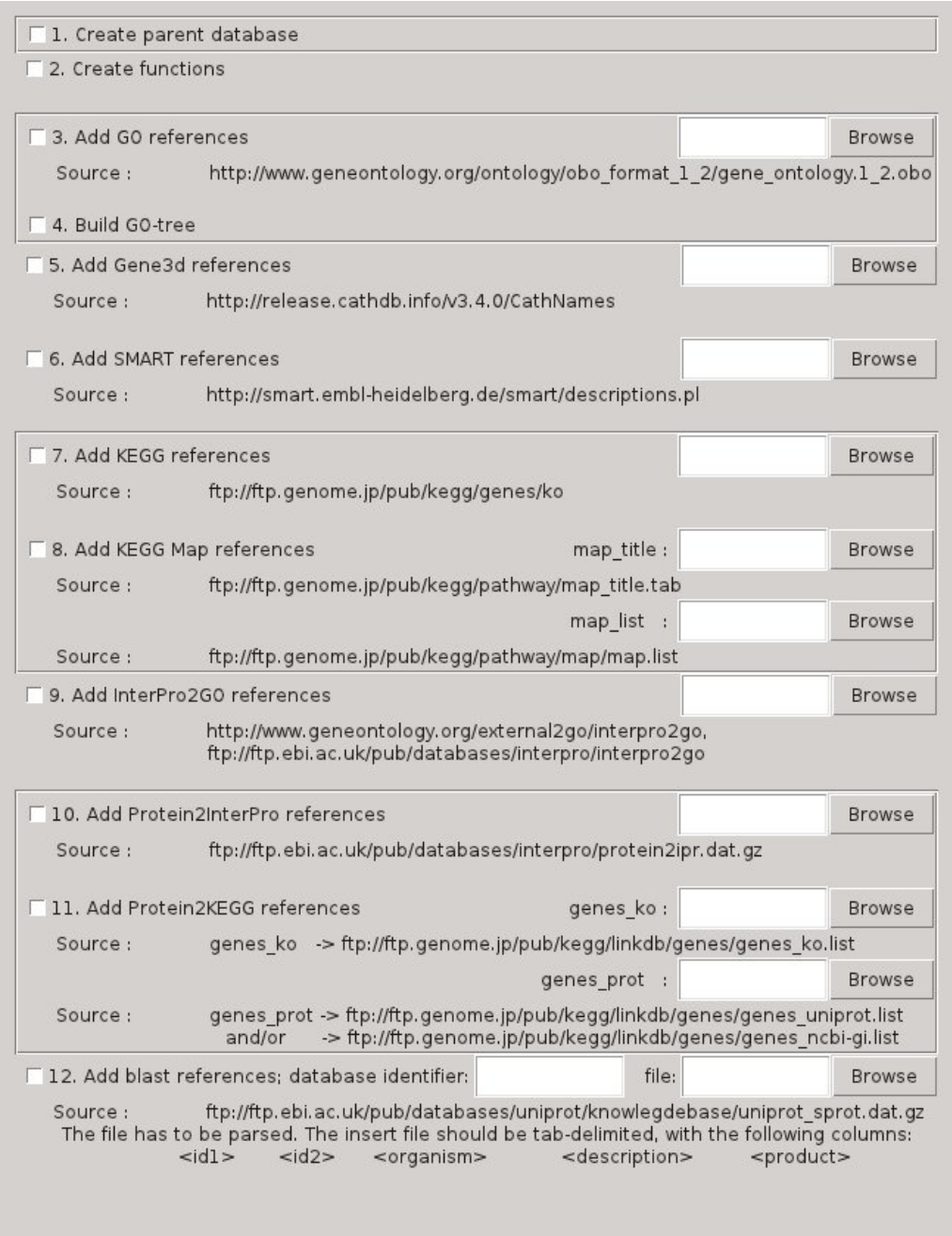
3. Add GO references

Writes the gene ontology terms from an .obo file to the GO reference table. The reference can be downloaded at:

http://www.geneontology.org/GO.downloads.ontology.shtml

4. Build GO-tree

Creates a GO-tree from the GO references. The tree is necessary for the 'GO statistics'-tab of T-ACE.



5. Add Gene3d references

Adds the gene3d domain entries from a file to the gene3d reference table. The reference file can be downloaded at:

http://release.cathdb.info/v3.4.0/CathNames

6. Add SMART references

Adds the SMART domain entries from a file to the SMART reference table. The reference file can be downloaded at:

http://smart.embl-heidelberg.de/smart/descriptions.pl

7. Add KEGG references

Adds the KEGG ontology terms from a file to the KEGG reference table. The reference file can be downloaded at:

ftp://ftp.genome.jp/pub/kegg/genes/ko

8. Add KEGG Map references

Adds the KEGG pathways from two files to the KEGG Maps reference table. The reference files can be downloaded at:

ftp://ftp.genome.jp/pub/kegg/pathway/map_title.tab

ftp://ftp.genome.jp/pub/kegg/pathway/map/map.list

9. Add InterPro2GO references

Adds InterPro2GO links to the interpro2go reference table. The reference file can be downloaded at:

ftp.ebi.ac.uk/pub/databases/interpro/interpro2go

or

http://www.geneontology.org/external2go/interpro2go

10. Add Protein2InterPro references

Adds Protein2InterPro links to the protein2interpro reference table. The reference file can be downloaded at:

ftp://ftp.ebi.ac.uk/pub/databases/interpro/protein2ipr.dat.gz

11. Add Protein2KEGG references

Adds Protein2KEGG links to the protein2interpro reference table. The reference files can be downloaded at:

ftp://ftp.genome.jp/pub/kegg/linkdb/genes/genes_ko.list

and

ftp://ftp.genome.jp/pub/kegg/linkdb/genes/genes_uniprot.list

ftp://ftp.genome.jp/pub/kegg/linkdb/genes/genes_ncbi-gi.list

12. Add blast references

Creates a blast reference table accordant to the given 'database identifier', e.g. nr, nt, uniref100, uniprot_sprot. The references from the file will be written to the created reference table. The reference file needs to be in a special format. It should be tab-delimited and has to contain the following columns:

[id1] [id2] [organism] [description] [product]

A reference file can be downloaded at:

ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowlegdebase/uniprot_sprot.dat.gz

User management

The 'User management'-tab allows the creation/deletion of user accounts for the postgresql databases. Also access rights to functions and individual databases can be given or denied. If a new user is added the password for the account will be the same as the user name. Before a user can be deleted all his databases have to be removed, otherwise the user will only loose his access rights.

Add new user

This function creates a new user for the postgresql parent database. The name of the new user is set in the 'User name' entry field. The checkbutton 'cDB' allows the new user to create schemata in the parent database.

ATTENTION:

The user will get access and insert right to all tables of the reference schema. If tables are added to the 'refdb'-schema after the user creation, the user will not have access right to these tables (e.g. blast reference tables). This could cause errors in T-ACE. To prevent this just create the user anew, it is not necessary to delete the user first.

Grant function rights to user

This function grants execution rights to functions of the function schema to a new user. To which functions rights are granted depends on the radiobutton setting:

access: Rights are granted only to functions, which are necessary for the database access.

all: Rights to all functions are granted. Allows the user to modify or delete contents of the database.

create: Except for delete-function, rights to all functions are granted. The user can add new information to the database (sequences, annotations), but is not allowed to delete database contents.

If the functions are created anew all users will lose their function rights, so it will be necessary to grant function right to each user again.

Grant access rights to user on database

This function grants access rights for the given database to the user. The user may also have write access if the accordant checkbutton is selected. It also adds an entry to 'dbs'-table in the public schema of the T-ACE parent database. This table shows T-ACE which user has access rights to which tables.

Delete user

Deletes the user login role, if no database object depends on it. If the user cannot be deletes the user just looses all access rights to the parent database.

Revoke function rights of user

Revokes the execution rights of all functions from the user.

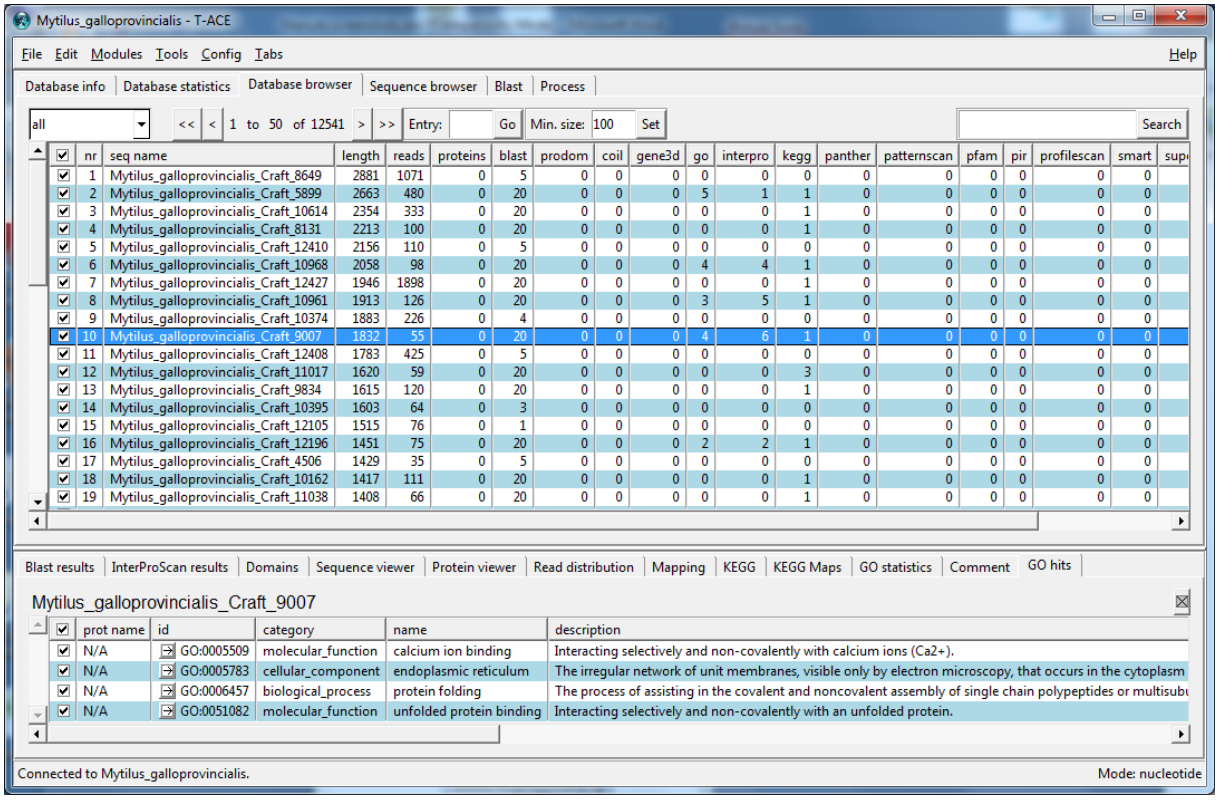
Revoke access rights of user on database

Revokes all access rights to the given database from the user.

T-ACE (Transcriptome annotation and comparison explorer)

Introduction to T-ACE

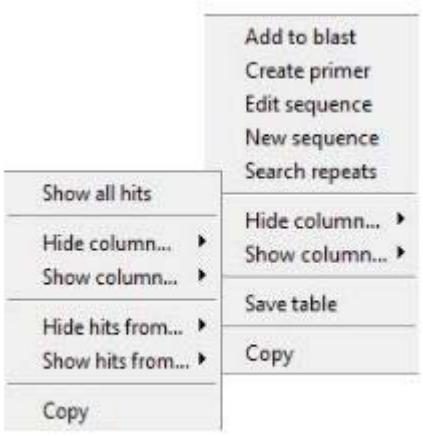
T-ACE mainly consists of two parts. The upper part of the window shows information about the database and its contents, the lower part shows detailed information for a single database entry.



Both parts contain several tabs, which will be described in detail in the following sections. Many of the tabs are optional and can be closed through the 'X'-button in the right corner of the tab.

Right-click menus

Most of the tables in the tabs possess a menu which can be opened through a click on the right mouse button. Those menus allow changes to the table configuration or the execution of tab specific features. Below you see two examples for such menus. Some options refer to the whole table, such as 'Hide/Show column', 'Hide/Show hits from...' or 'Save table', the other options are normally row specific.



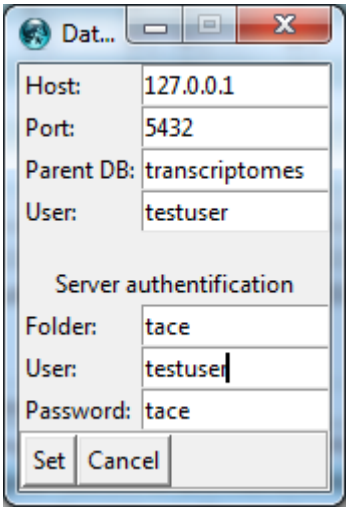
The 'Copy'-option is the most common one and refers always to the selected table cell. It copies the cell content to the clipboard.

Other options, like 'Add to blast' and 'Edit sequence', will be explained in the appropriate chapter.

Configuring T-ACE

Before you start working with T-ACE it is necessary to set some variables. To do this, open the 'Config'-menu of T-ACE.

Database connection

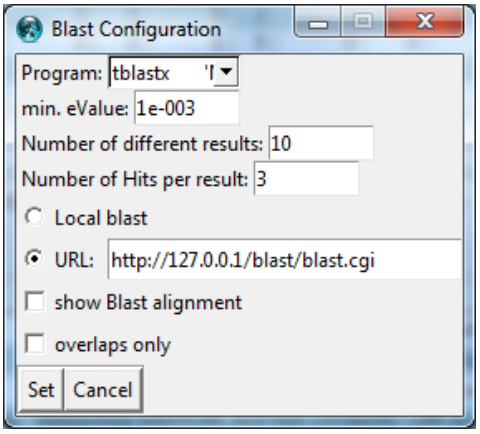


The most important setting is the 'Database connection'. Here the user sets the IP-address and port to his postgresql database and the name of the parent database, which were created with the 'T-ACE_DB_manager.tcl'. The user name is not necessary, but the entry will appear as login name when starting T-ACE.

The 'Server authentification' is needed, when T-ACE is used with a php script. The 'Folder' entry, together with the host IP, shows location of the T-ACE.php. In this case the URL to the php script would look like this:

http://127.0.0.1/tace/T-ACE.php

Blast parameters



Here you set your standard parameters for the 'Blast'-tab. You can choose to blast with a blast server, by setting an URL, or locally. If you choose the 'Local blast'-option you can only blast against blast-databases which are in the /blast_dbs folder of the T-ACE directory. For the local blast *NCBI-blast+* has to be installed.

Browser path and links

To set the 'BrowserPath' select the executable of your favored browser. This is important, because some of the tables in T-ACE contain columns with link buttons.

You can find a list of all links in the T-ACE directory under: /conf/links_tv.conf

There you can change existing links or add new ones.

Home folder

The 'HomeFolder' is standard folder for opening and saving files.

Protein translation table

Sets the path to file with information about protein translation codon usage.

Some such files can be found in the T-ACE directory under: /gcodes

Sequence types

T-ACE contains two types of sequences, nucleotide and protein sequences. The nucleotide sequences consist of three different subtypes: contigs, reads and singletons

Contigs

A contig (from contiguous) is a set of overlapping DNA segments (reads), which can be assembled to represent the original DNA sequence of the source.

Here, the word contig is also used to refer to one segment of the original DNA sequence, not necessarily the complete one.

Reads

In T-ACE all sequences, which were used in the contig assembly, are called reads. Normally they would be ESTs.

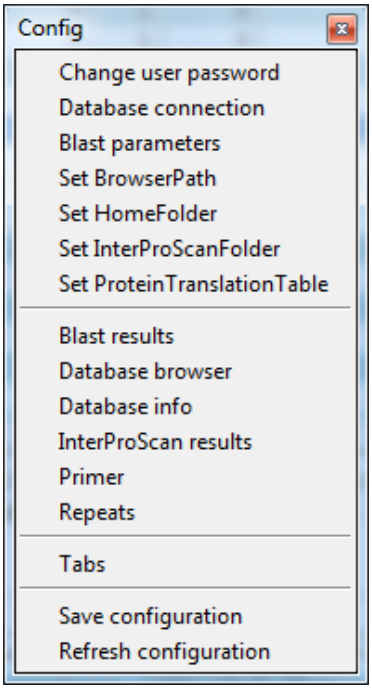
Singletons

A read, which does not cluster with other reads, is called a singleton.

The T-ACE Menus

Config
Edit
File
Help
Modules
Tabs
Tools

The Config menu



The 'Config'-menu allows the setting of the standard parameters for T-ACE. Those parameters are saved in the T-ACE directory under: /conf/T-ACE.conf

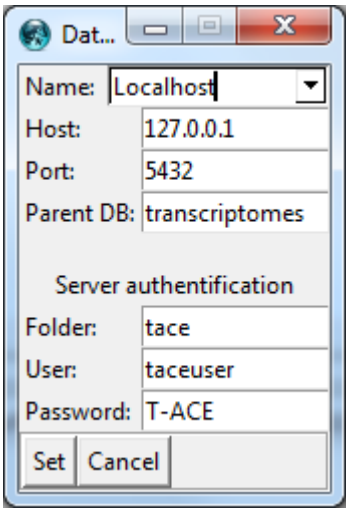
All changes to the configuration made through options of the 'Config'-menu will be written automatically to this file.
It is also possible to save and load a specific configuration to/from an additional file. This can be done in the 'File'-menu, with the options 'Save config' and 'Load config'.

Change user password
Database browser
Database connections
Database info
Blast parameters
Blast results
InterProScan results
Primer3
Refresh configuration
Repeats
Save configuration
Set BrowserPath
Set HomeFolder
Set InterProScanFolder
Set ProteinTranslationTable
Tabs

Change user password

This function allows the current user to change the login password.

Database connections

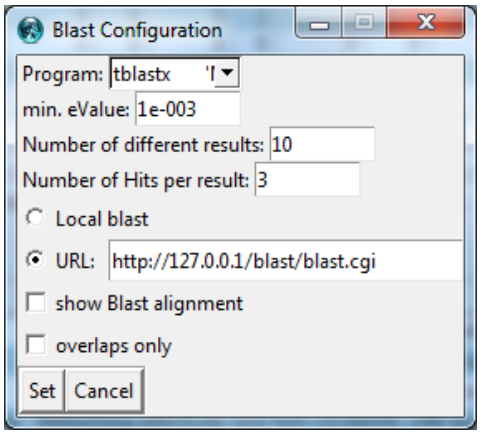


Here the IP-address and port to a postgresql database can be set. Also the name of the parent database, which was created with the 'T-ACE DB_manager.tcl', has to be set; otherwise the connection will not be possible.
It is possible to enter more than one connection, just enter a new connection name and host settings. After pressing the 'Set'-button the new connection will be saved. Existing connections can also be altered that way.

After setting the 'Database connection' the current connection will be closed and a new 'Login'-window will appear.

Blast parameters

Here you set your standard parameters for the 'Blast'-tab. You can choose to blast with a blast server, by setting an URL, or locally. If you choose the 'Local blast'-option you can only blast against blast-databases which are in the */blast_dbs* folder of the T-ACE directory. For the local blast NCBI-blast+ has to be installed.



Set BrowserPath

To set the 'BrowserPath' select the executable of your favored browser. This is important, because some of the tables in T-ACE contain columns with link buttons.

You can find a list of all links in the T-ACE directory under:

/conf/links_tace.conf

There you can change existing links or add new ones.

Set HomeFolder

The 'HomeFolder' is standard folder for opening and saving files.

Set InterProScanFolder

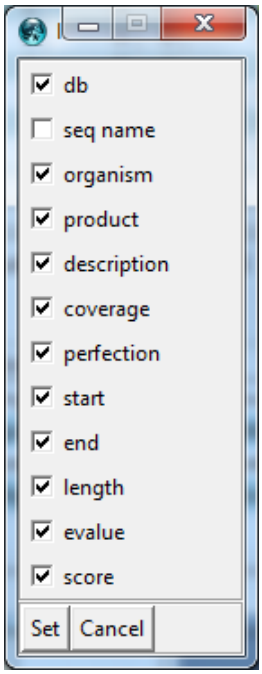
The 'InterProScanFolder' is the folder of a local InterProScan installation.

Set ProteinTranslationTable

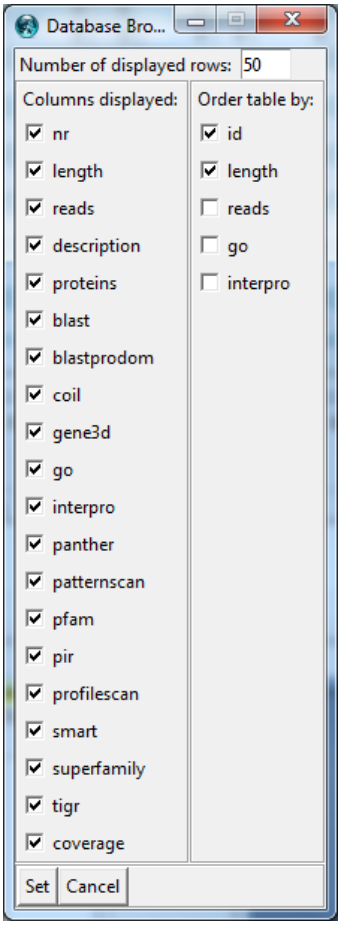
Sets the path to file with information about protein translation codon usage. Some such files can be found in the T-ACE directory under: */gcodes/*

Blast results

This window sets the standard for showing or hiding columns of the 'Blast results'-tab. Only the columns selected in this window will be shown in the table of 'Blast results'-tab, after starting T-ACE.



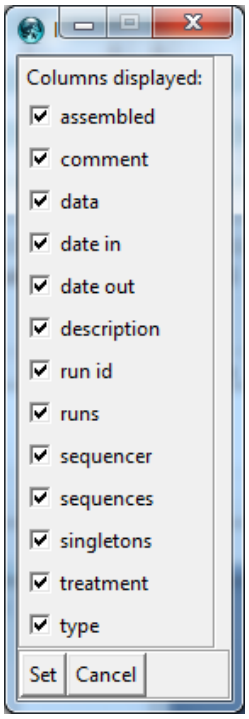
Database browser



The checkbuttons in the left half of the window set the standard for showing or hiding columns of the 'Database browser'-tab. Only the columns selected in this window will be shown in the table of 'Database browser'-tab, after starting T-ACE.

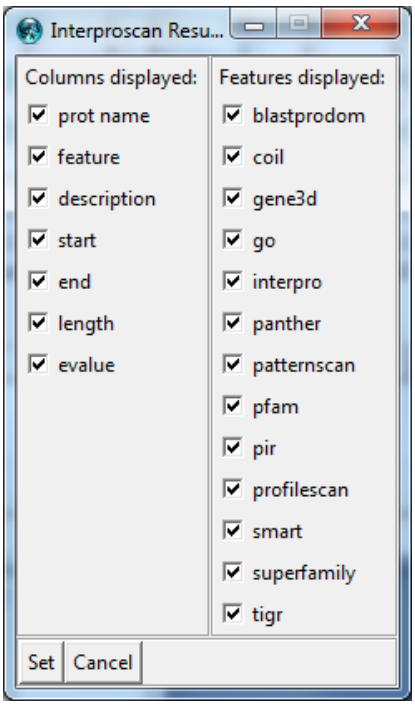
The checkbuttons in the right half set the standard order for listing the entries of the 'Database browser'-table.

The 'Number of displayed rows'-entrybox sets exactly that; the number of displayed rows in the 'Database browser'-table.



Database info

The checkbuttons set the standard for showing or hiding columns of the 'Database info'-tab. Only the columns selected in this window will be shown in the table of 'Database info'-tab, after starting T-ACE.



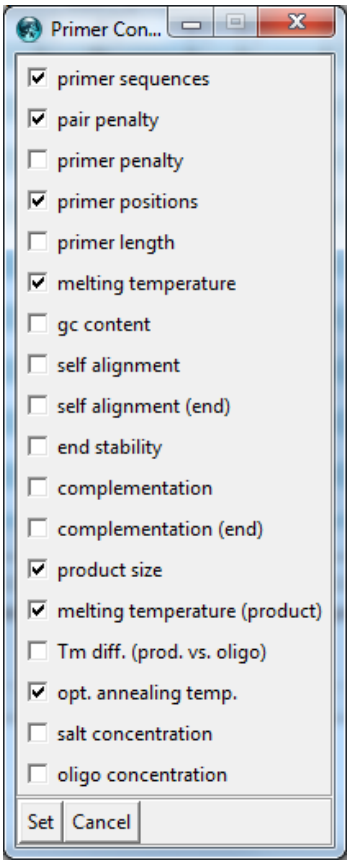
InterProScan results

The checkbuttons in the left half of the window set the standard for showing or hiding columns of the 'InterProScan results'-tab. Only the columns selected in this window will be shown in the table of 'InterProScan results'-tab, after starting T-ACE.

The checkbuttons in the right half set the standard for the listing of the different types of hits, only the hits of selected features will be displayed in the 'InterProScan results'-table.

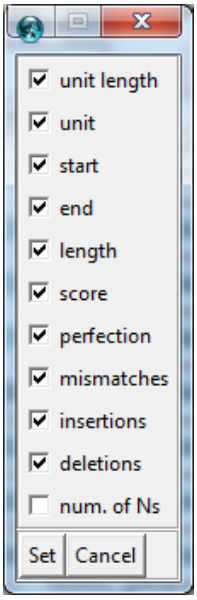
Primer3

The checkbuttons set the standard for showing or hiding columns of the 'Primer'-tab. Only the columns selected in this window will be shown in the table of 'Primer'-tab, after starting T-ACE.



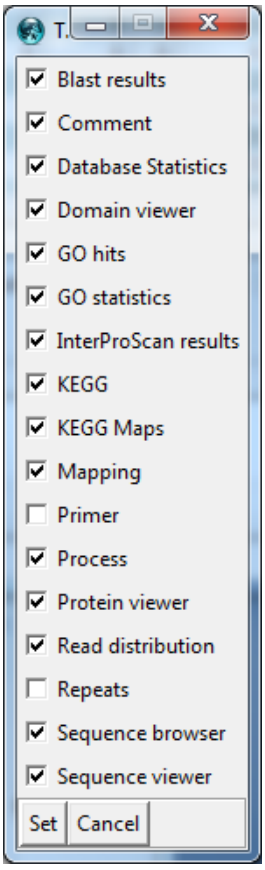
Repeats

The checkbuttons set the standard for showing or hiding columns of the 'Repeats'-tab. Only the columns selected in this window will be shown in the table of 'Repeats'-tab, after starting T-ACE.



Tabs

The checkbuttons set the standard for showing or hiding the different tabs of T-ACE. Only the tabs selected in this window will be opened after starting T-ACE.



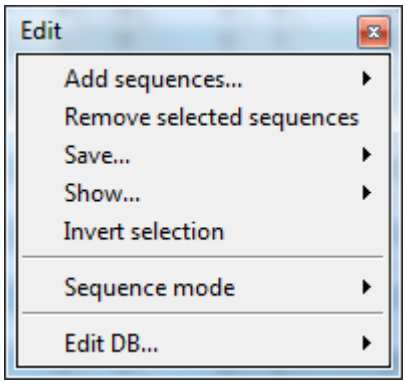
Save Configuration

Saves the current configuration of T-ACE to the 'T-ACE.conf'.

Refresh Configuration

Refreshes the current configuration of T-ACE, by loading the 'T-ACE.conf'.

The Edit menu



Add sequences
Edit DB
Invert selection
Remove selected sequences
Save
Sequence mode
Show

Add sequences

With 'Add sequences' it is possible to add the checked entries from the 'Database browser', 'Sequence browser', 'Blast'-tab or a project tab to a project tab. After clicking on 'Add sequences' a list of all available project tabs will appear.

Edit DB

'Edit DB' contains a lot of options to modify and update a database.

Change comment
Create database copy
Delete selected contigs
Delete selected hits
Import sequences…
- Add ACE file
- Add nucleotide file
- Add protein file
- Add SAM file

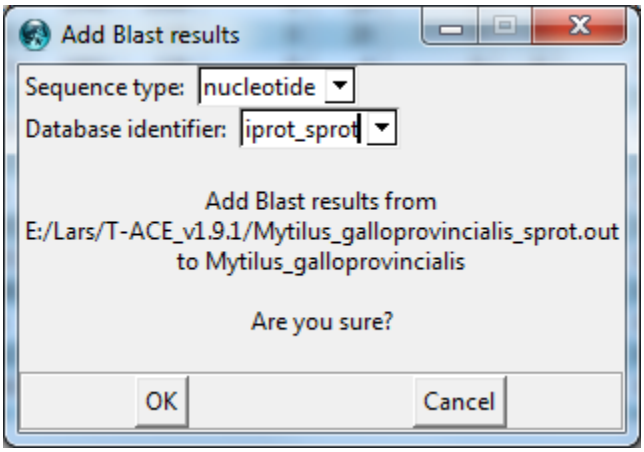
Import annotations…
- Blast results
- GO terms
- InterProScan results
- PolyA information

New nucleotide sequence
Translate contigs to proteins
Update…
- blast2interpro entries
- coverage
- database info
- domain2interpro entries
- GO-Statistic
- KEGG-Maps
- protein descriptions
- protein entries
- read distribution
- read status
- sequence descriptions
- sequence entries

Add ACE file

This option loads the contents of an ACE file into the database. The read mapping information for each contig of the ACE file will be available.

Add blast results



It adds the blast hits from a tabular tab-delimited blast output file (-outfmt 6) to the database. Take care to choose the right blast reference table from the 'Database identifier'-combobox, otherwise the references for the blast hits will not be available.

It is important to choose the correct sequence type (nucleotide or protein) of the annotated sequences; otherwise the annotations cannot be related correctly to the sequences.

After adding the blast hits 'Update sequence entries' and 'Update sequence descriptions' should be performed.

Add GO terms

It adds GO hits to the database, as input file serves an .annot or .txt file containing GO terms. The files have to be in the Blast2GO (http://www.blast2go.com) format.

It is important to choose the correct sequence type (nucleotide or protein) of the annotated sequences; otherwise the annotations cannot be related correctly to the sequences.

After adding the annotations 'Update sequence entries' and 'Update sequence descriptions' should be performed.

Add InterProScan results

It adds the domain, InterPro and GO hits to the database, as input file serves a results file of InterProScan in the .raw format. Alternately .txt files in the Blast2GO (http://www.blast2go.com) format, containing InterProScan domain annotations can be used, but only the InterPro and GO annotations will be written into the database.

It is important to choose the correct sequence type (nucleotide or protein) of the annotated sequences; otherwise the annotations cannot be related correctly to the sequences.

After adding the annotations 'Update sequence entries' and 'Update sequence descriptions' should be performed.

Add nucleotide file

The option adds the sequences from a fasta file as contigs to the database.

Add polyA information

This option allows the user to mark reads contained in a database as ‘poly-A tailed’. For this the polyA boolean value, of the 'sequences_read'-table of the selected database, is set to TRUE for all reads listed in the input text file. The file should contain only one read id per inserted row.

Add protein file

The option adds protein sequences from a fasta file as proteins to the database, as input serves a protein file from InterProScan in the .input format.

Add SAM file

This option loads the contents of a SAM file into the database. The read mapping information for each contig of the SAM file will be available. Before the SAM file is uploaded the reference contigs must be imported into the database, otherwise the reads in the SAM file cannot be assigned.
The upload of SAM file will take a considerable amount of time. Have also in mind that T-ACE is not designed as a genome browser, but for de-novo transcriptomes. This means contig sequences should not be much longer than 60.000 bp and should not contain over 100.000 reads (at least if those reads shall be displayed).

Change comment

This changes the database comment.

Create database copy

Creates a copy of the complete database.

Delete selected contigs

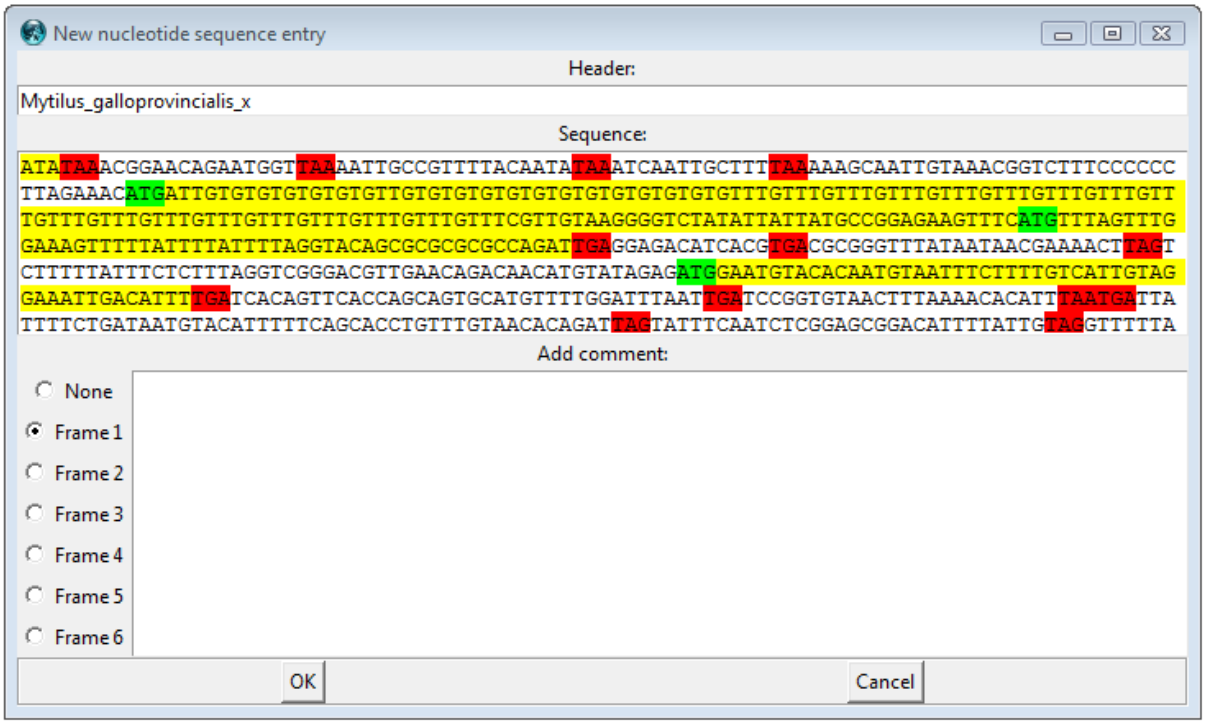
Deletes the checked contigs and all its proteins and hits from the database.

Delete selected hits

Deletes the checked hits of the 'Blast results'-tab, 'InterProScan results'-tab, 'GO hits'-tab or 'KEGG'-tab from the database, depending on which tab is raised.

New nucleotide sequence

With 'New nucleotide sequence' a new contig sequence can be added to the database. The 'Header'-field sets the name for the new contig and the 'Sequence'-field the sequence. With the 'Add comment'-field a comment can be added to the new contig. The radio buttons display the open reading frames (ORFs) of the different frames of the sequence. The ORFs are displayed in yellow, while start codons are green and stop codons are red.

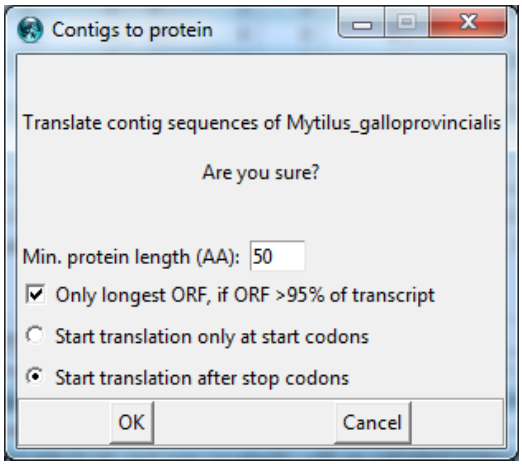


Translate contigs to proteins

This function allows the user to translate all the nucleotide sequences in the database into additional amino acid sequence entries.

All six frames of each nucleotide sequence will be translated. Every open reading frame (ORF), with the given length or longer, is written to the 'sequences_protein'-table of the database.

The user can also choose, if the translation should only begin at start codons, or directly after each stop codon.



Update

Update blast2interpro entries

This option tries to associate the blast hits of a contig with InterPro hits, if there are associated InterPro hits they will be added to the contig annotation. The blast2interpro associations can be found in the 'interpro2protein'-table of the 'refdb'-schema of the parent database.

Update coverage

This function calculates the average contig coverage, it is only necessary after adding contig sequences with read mapping information, for example after adding an ACE file. The coverage is needed for the 'Coverage'-functions of the 'Database statistics'-tab.

Update database info

It updates all the information shown in the 'Database info'-tab.

Update domain2interpro entries

This option tries to associate the domain hits of a contig with InterPro hits and GO terms, if there are associated InterPro hits they will be added to the contig annotation. The domain2interpro associations can be found in the 'interpro2domain'-table of the 'refdb'-schema of the parent database.

Update GO-Statistic

Updates the GO-tree of the 'GO statistics'-tab. It is necessary to update the GO-tree after adding new GO hits to the database.

Update KEGG-Maps

This option tries to associate the blast hits of a contig with KEGG ontologies and adds them to the contig annotation. Then all KEGG hits will be reviewed for updating the 'KEGG maps'-table.

Update protein descriptions

Updates the standard descriptions of all proteins, which are shown in the description column of the 'Database browser'-tab. It necessary to update the descriptions after adding new annotations over 'Add blast results' or 'Add InterProScan results'.

Update protein entries

Updates the annotation status of all proteins. It counts the number of hits for the different annotation categories. It is necessary to update the protein entries after adding new annotations over 'Add blast results' or 'Add InterProScan results'.

Update read distribution

Updates the affiliation of the reads to the different sequencing runs contained in the database, for each contig. This update should be run after adding sequences and/or sequencing runs to the database, otherwise features like 'Expression analysis' or 'Mapping' will not work correctly.

Update read status

This options checks if a read in the database is associated with a contig or if it is a singleton. This is only necessary after adding new read sequences, or contigs with read mapping information to the database.

Update sequence descriptions

Updates the standard descriptions of all contigs, which are shown in the description column of the 'Database browser'-tab. It necessary to update the descriptions after adding new annotations over 'Add blast results' or 'Add InterProScan results'.

Update sequence entries

Updates the annotation status of all contigs. It counts the number of hits for the different annotation categories. It is necessary to update the sequence entries after adding new annotations over 'Add blast results' or 'Add InterProScan results'.

Invert selection

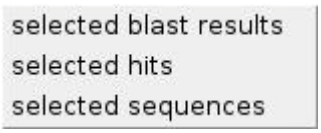
Inverts the checked status of a table in the upper half of T-ACE, depending on which tab is raised.

Remove selected sequences

Removes the checked rows from a project tab table.

Save

It saves the 'selected blast results', 'selected hits' or 'selected sequences', depending on which option is chosen.



selected blast results

To use this option the 'Blast'-tab has to be raised. It writes the information of all checked rows in the results table of the 'Blast'-tab into a tab-delimited text file.

selected hits

This option writes the information of all checked rows in the hit table of the 'Blast results'-tab, 'InterProScan results'-tab or 'Go hits'-tab, depending on which tab is raised, into a tab-delimited text file.

selected sequences

This option writes the contig sequences of all checked rows of the 'Database browser', the 'Blast'-tab or a project tab into a fasta file, depending on which tab is raised.

Sequence mode

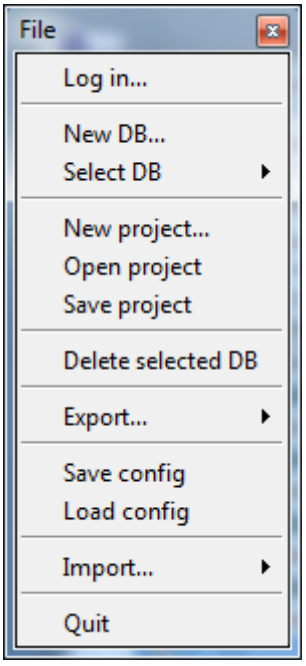
This option allows the user to switch between the nucleotide- (standard) and protein-mode. This mainly affects the entries of the 'Database browser'-tab. In the nucleotide-mode the contig entries of the database will be displayed, the protein-mode shows all protein entries of the database.

Show

The 'Show'-option does the same as the 'Save'-option, with the difference of displaying the information in a window before saving.

The File menu

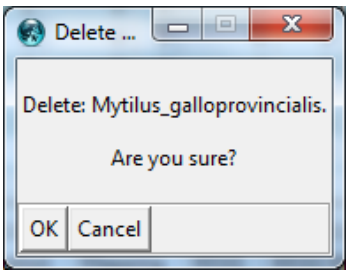
The 'File'-menu contains the options for loading and saving of configurations, databases and projects.



Delete selected DB
Export
Import
Load config
Log in
New DB
New project
Open project
Quit
Save config
Save project
Select DB

Delete selected DB

This function deletes the currently opened database. After clicking the 'Delete selected DB'-button a popup window will opens and asks, if the database really should be deleted.

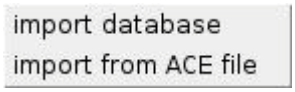


Export

This option exports the complete selected database, including all sequences and annotations. It creates a .db file, from which the database can be reconstructed with the 'Import'-function.

Import

The 'Import'-function can create new databases from a .db or .ace file. A .db file can be created from an existing database, with the 'Export'-function.

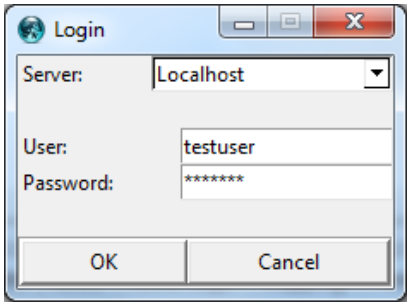


Load config

This function loads a .conf file and replaces the current configuration of T-ACE with the information of the loaded file. With 'Save config' the configuration of T-ACE can be written to a .conf file.
For more information about the T-ACE configuration see the 'Config'-menu.

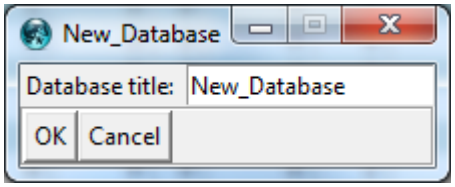
Log in

The Login window opens directly after starting T-ACE. To change the user or the parent database a new login is necessary, to do this, the 'Log in'-function of the 'File'-menu can be used.



New DB

This function creates a new and complete empty database. The newly creating database should now be filled with read and contig sequences, to do this use the options 'Add run' for the reads and 'Add ACE file' or 'Add nucleotide file' for the contigs.



New project

With this function a new project tab can be created.

Open project

Here, a project tab can be saved.

Quit

This button closes T-ACE.

Save config

This function saves the current configuration of T-ACE into a .conf file. With 'Load config' such a .conf file can be used to replace the configuration of T-ACE.
For more information about the T-ACE configuration see the 'Config'-menu.

Save project

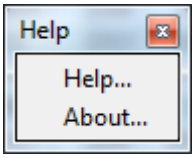
It writes the information of a project tab into a .prj file. With the 'Open project'-option such a .prj file can be reopened.

Select DB

'Select DB' contains a list of all databases in the parent database, to which the logged in user has access. By clicking on a database it can be opened.

The Help menu

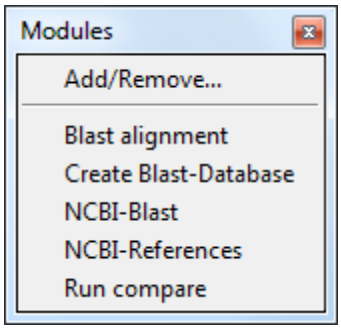
The 'Help...'-button of this menu open the index page of the html manual.



The Modules menu

The 'Modules'-menu gives access to additional modules, which are not directly implemented into T-ACE. There are already several modules integrated into T-ACE. These can serve as examples for writing new modules. How to write such a module is explained under 'Writing custom modules' at the end of this manual.

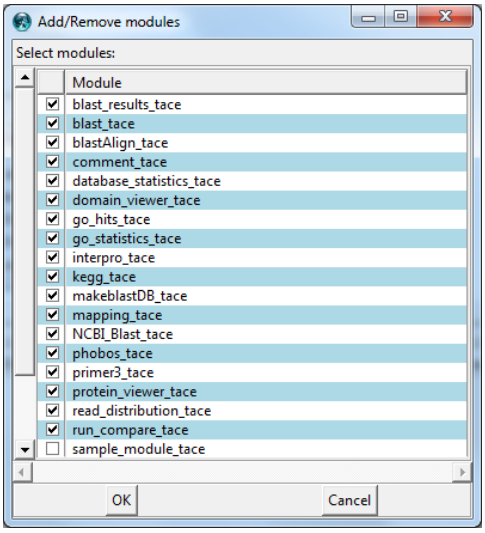
Add/Remove...
Blast alignment
Create Blast-Database
NCBI-Blast
NCBI-References
Run compare



Add/Remove

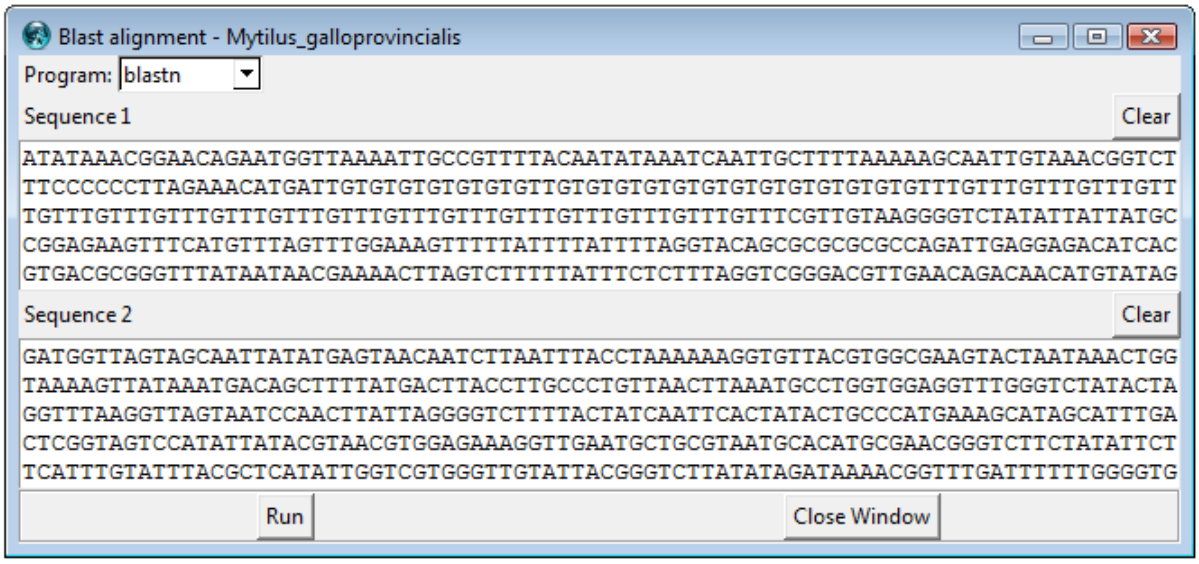
This function allows to add or remove TCL modules, which are located in the /modules folder of the T-ACE directory. Just check or uncheck the desired TCL script and press okay. The new script configuration will be loaded.

See 'Writing a T-ACE module' for more information.

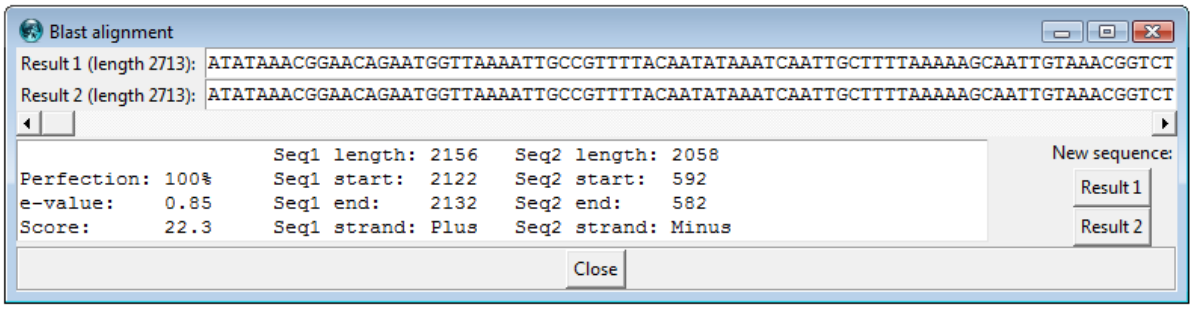


Blast alignment

'Blast alignment' allows to blast two sequences against each other. It is possible to blast nucleotide or protein sequences.

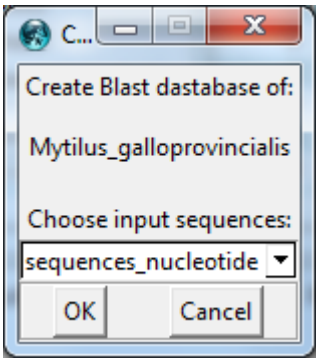


The blast result window shows two result sequences. 'Result 1' contains the query match sequence; 'Result 2' the subject match sequence. Additionally, sequence parts from 'Sequence 1' and 'Sequence 2', which do not appear in the blast hit, are attached to the start or end of the result sequences.
By clicking the 'Result 1'-button or 'Result2'-button the according result sequence will be opened in a 'New sequence'-window and can be added to the T-ACE database.



To use this module *NCBI-blast+* has to be installed.

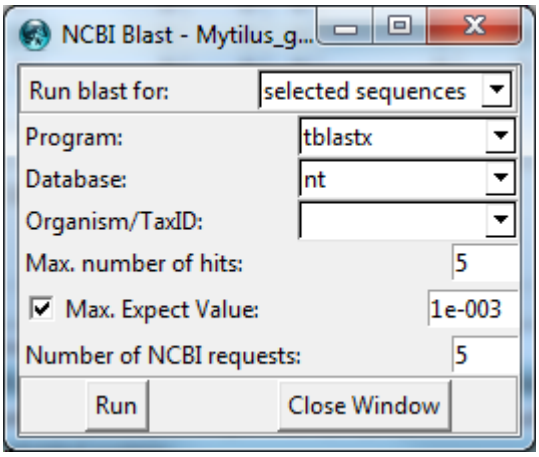
Create Blast-Database



This module can create blast databases from the sequences in a T-ACE database. It is possible to create blast database from the contig, protein or read sequences. The created will be named like the corresponding T-ACE database, if the proteins are used a '_prot' will be appended to the name. If the reads are used a '_reads' is appended to the database name.

To use this module *NCBI-blast+* has to be installed.

NCBI-Blast



This module can blast contig sequences from a T-ACE database against any blast database available at NCBI. It sends the query sequence directly to the NCBI web-blast and writes the results into the database.

The 'Run blast for'-combobox allows to select a subset of sequences from the T-ACE database for blasting at NCBI. The option 'selected sequences' refer to all sequences in the 'Database browser'-table or a project tab table, which have 'checked'-status in the first column. 'Number of NCBI requests' set the number of threads running parallel NCBI requests.

If you blast against a specific NCBI database for the first time a new blast reference table needs to be created. For this the user needs writing rights for the 'refdb'-schema of the T-ACE parent database.

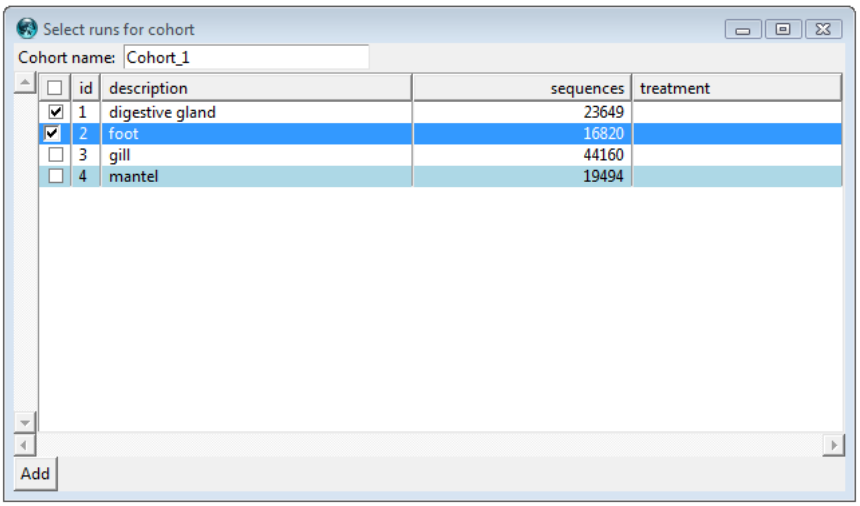
NCBI-References

This module tries to get all missing references of a NCBI database for all blast hits of the selected contig group and writes results into the according reference table.

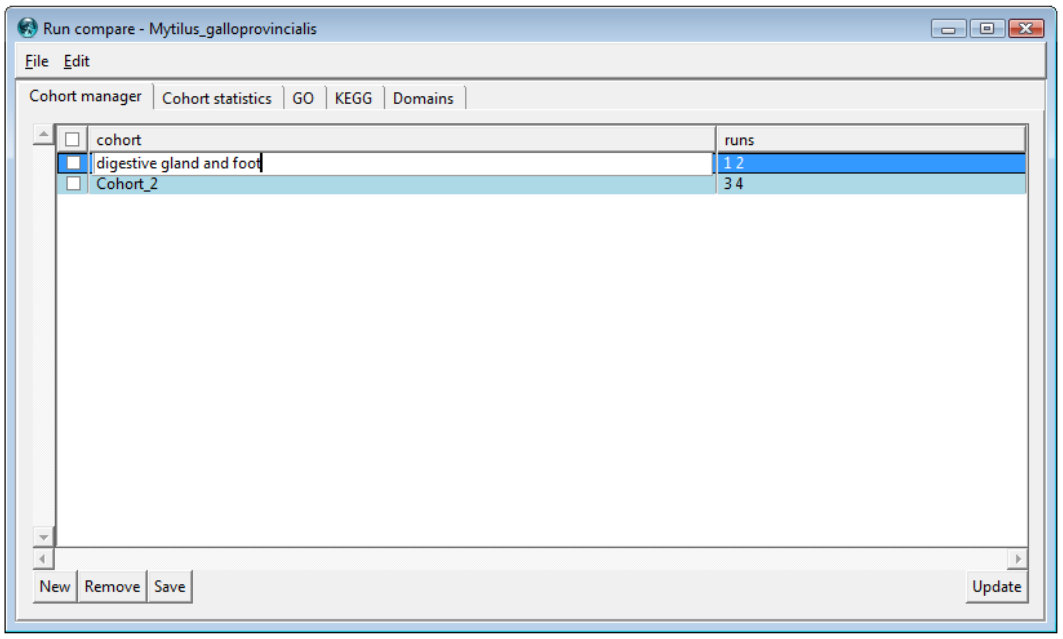
Run compare

The 'Run compare' tool compares multiple subsets of contigs (called cohorts) against the combined contigs of all subsets. This gives a statistical estimation on which GO terms, KEGG pathways or domains are enriched or depleted within the respective transcriptome or treatment group. Calculations can be conducted for the whole set of different GO terms, KEGG pathways or for specific domain databases. P-values are calculated by performing the Fisher's exact test but it has to be noted that values are not corrected for multiple testing

After opening the 'Run compare'-window new 'Run'-cohorts can be added through the 'Cohort manager'-tab. Just click on the 'New'-button and choose the runs or sort them in the different cohorts you want to compare with each other.

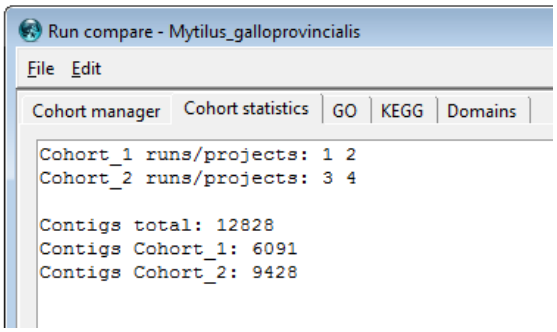


The columns 'cohort' and 'runs' can be modified by clicking on a cell. In this way the cohort name can be changed or runs can be added or removed from a cohort.

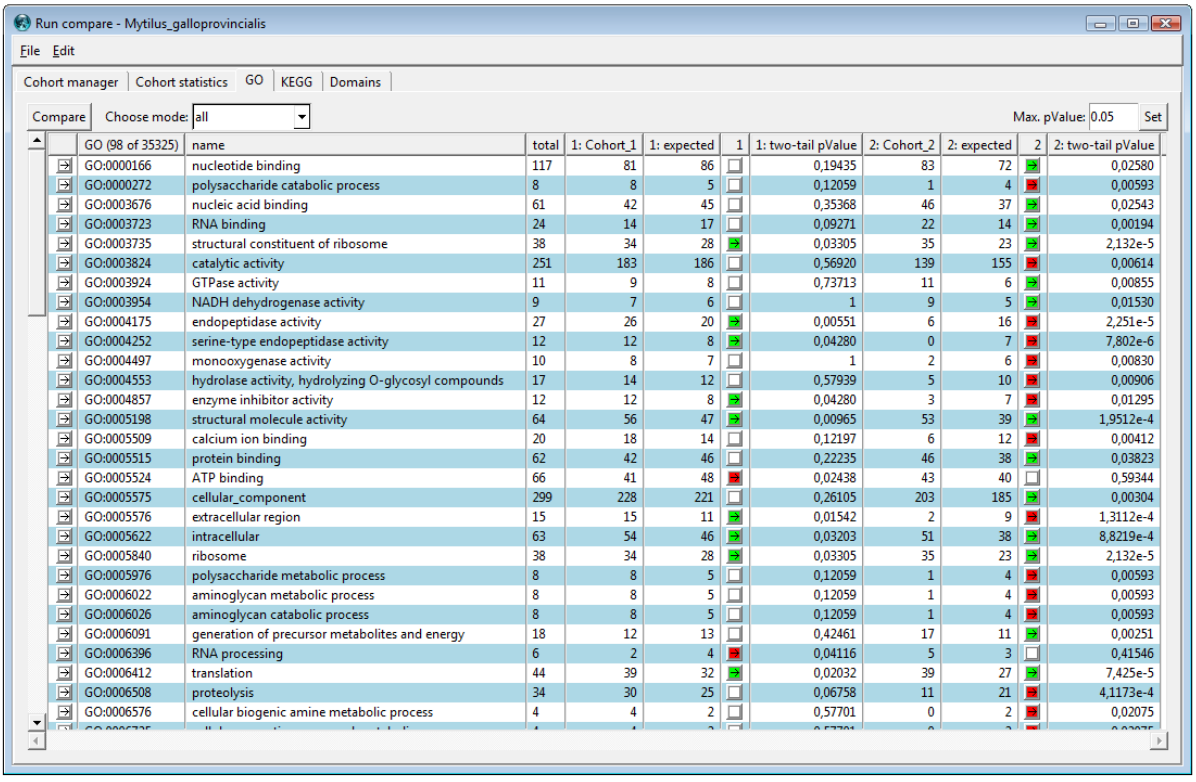


Press the 'Save'-button to adopt those changes. When all cohorts are added, it is necessary to press the 'Update'-button in the lower right corner of the tab. The update will create contig lists for each cohort and the reference list, consisting of all contigs of all cohorts. The results of the update will be visualized in the 'Cohort statistics'-tab. An update is always necessary, when a cohort was added, removed or changed.

After the update is completed the cohorts can be compared against the reference contig list concerning the distribution of GO terms, KEGG terms or protein domains. To do this, select the 'GO'-, 'KEGG'-, or 'Domain'-tab and press the 'Compare'-button in the upper left corner of the tab. The comparing can take several minutes to complete.



The result table of the comparison consists of several columns. The first column shows the GO, KEGG or domain ID, the second column the according name. The third column shows the number of contigs in the reference, which are associated with the ID in the first column.
Then always four columns for each cohort are following. The first cohort column shows the number of contigs in the cohort, which are associated with the ID in the first column. The second column shows the number of contigs that could be expected for the respective cohort and the third and fourth column shows the visualization and value for the two-tailed p-value.



GO (98 of 35325)	name	total	1: Cohort_1	1: expected	1	1: two-tail pValue	2: Cohort_2	2: expected	2	2: two-tail pValue
GO:0000166	nucleotide binding	117	81	86		0,19435	83	72		0,02580
GO:0000272	polysaccharide catabolic process	8	8	5		0,12059	1	4		0,00593
GO:0003676	nucleic acid binding	61	42	45		0,35368	46	37		0,02543
GO:0003723	RNA binding	24	14	17		0,09271	22	14		0,00194
GO:0003735	structural constituent of ribosome	38	34	28		0,03305	35	23		2,132e-5
GO:0003824	catalytic activity	251	183	186		0,56920	139	155		0,00614
GO:0003924	GTPase activity	11	9	8		0,73713	11	6		0,00855
GO:0003954	NADH dehydrogenase activity	9	7	6		1	9	5		0,01530
GO:0004175	endopeptidase activity	27	26	20		0,00551	6	16		2,251e-5
GO:0004252	serine-type endopeptidase activity	12	12	8		0,04280	0	7		7,802e-6
GO:0004497	monooxygenase activity	10	8	7		1	2	6		0,00830
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	17	14	12		0,57939	5	10		0,00906
GO:0004857	enzyme inhibitor activity	12	12	8		0,04280	3	7		0,01295
GO:0005198	structural molecule activity	64	56	47		0,00965	53	39		1,9512e-4
GO:0005509	calcium ion binding	20	18	14		0,12197	6	12		0,00412
GO:0005515	protein binding	62	42	46		0,22235	46	38		0,03823
GO:0005524	ATP binding	66	41	48		0,02438	43	40		0,59344
GO:0005575	cellular_component	299	228	221		0,26105	203	185		0,00304
GO:0005576	extracellular region	15	15	11		0,01542	2	9		1,3112e-4
GO:0005622	intracellular	63	54	46		0,03203	51	38		8,8219e-4
GO:0005840	ribosome	38	34	28		0,03305	35	23		2,132e-5
GO:0005976	polysaccharide metabolic process	8	8	5		0,12059	1	4		0,00593
GO:0006022	aminoglycan metabolic process	8	8	5		0,12059	1	4		0,00593
GO:0006026	aminoglycan catabolic process	8	8	5		0,12059	1	4		0,00593
GO:0006091	generation of precursor metabolites and energy	18	12	13		0,42461	17	11		0,00251
GO:0006396	RNA processing	6	2	4		0,04116	5	3		0,41546
GO:0006412	translation	44	39	32		0,02032	39	27		7,425e-5
GO:0006508	proteolysis	34	30	25		0,06758	11	21		4,1173e-4
GO:0006576	cellular biogenic amine metabolic process	4	4	2		0,57701	0	2		0,02075

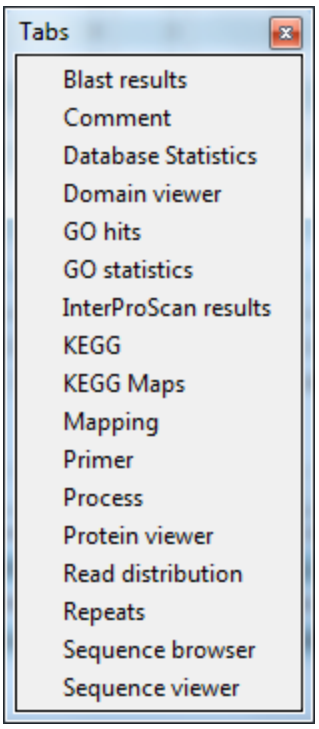
The expected contig number is calculated with the following formula:

$$\frac{(contigs\ in\ reference\ associated\ with\ ID) \times (total\ number\ of contigs\ in\ cohort)}{(total\ number\ of\ contigs\ in\ reference)}$$

The Tabs menu

The buttons of the 'Tabs'-menu will open the accordant T-ACE tab, if it does not already exist. Only the tabs that can be closed are listed in the 'Tabs'-menu.

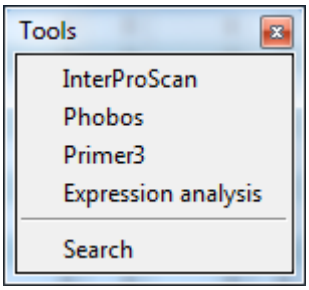
Blast
Blast results
Comment
Database browser
Database info
Database statistics
Domain viewer
GO hits
GO statistics
InterProScan results
KEGG
KEGG maps
Mapping
Primer
Process
Protein viewer
Repeats
Sequence browser
Sequence viewer



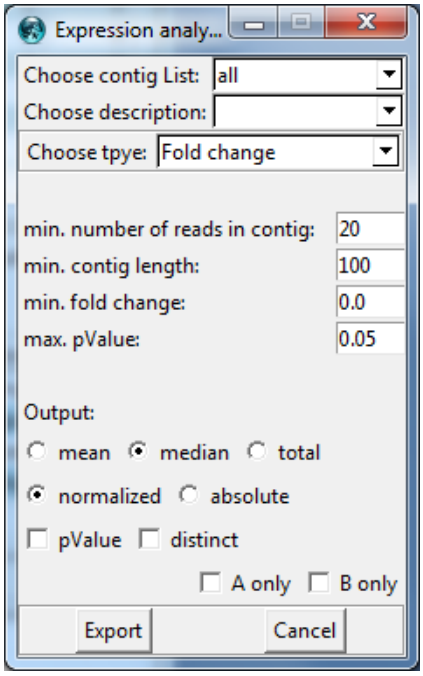
The Tools menu

This menu contains some tools for working with a whole database or just on groups of contigs. The tools 'Primer3' and 'Search repeats' also appear in some right-click menus.

Expression analysis
InterProScan
Primer3
Search
Search repeats

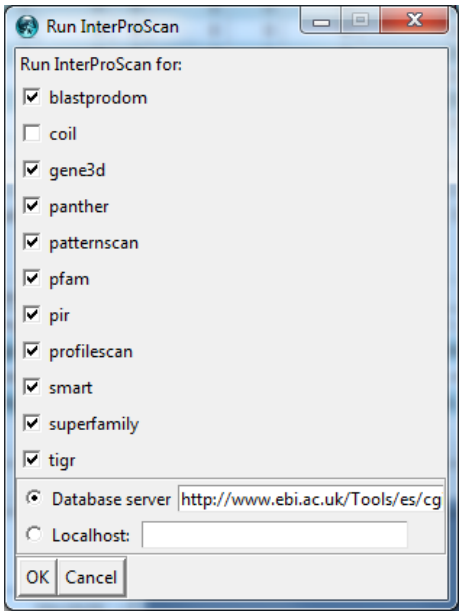


Expression analysis



With this function expression levels between different transcriptomes or groups of transcriptomes can be analyzed. For more details see the description of the 'Database statistics'-tab. Basically this function creates a list with the information of the up- or down-regulation level and/or the numbers of reads per contig in selected *A* and *B* groups. The group classification (*A* or *B*) depends on selection in the run table in the 'Database info'-tab.

The contig list can be exported as a tab-delimited text file with the 'Export'-button.



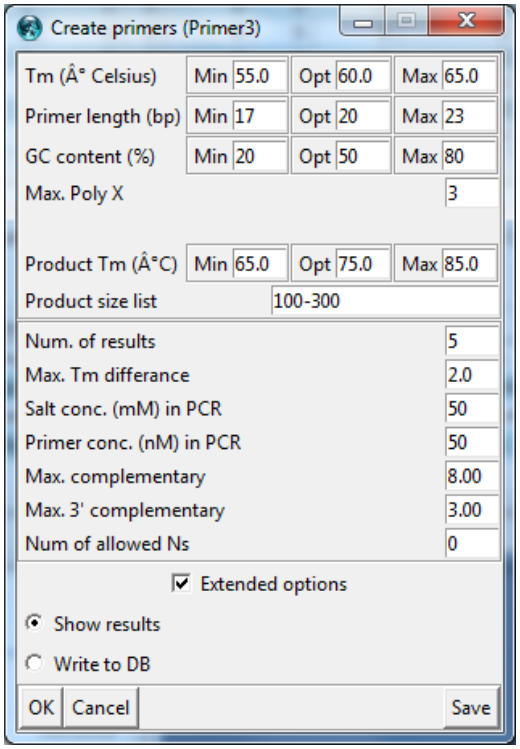
InterProScan

To use this option InterProScan has to be installed with a web interface or on the client computer. The contants of the iprscan/bin folder should be executable for everyone. Also writing and reading rights for the iprscan/tmp folder are necessary.

InterProScan is available at:
ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/index.html

The 'InterProScan'-option will try to annotate every checked table entry of the 'Database browser' or a project tab, depending on which tab is raised. The found domains will be written to the database.

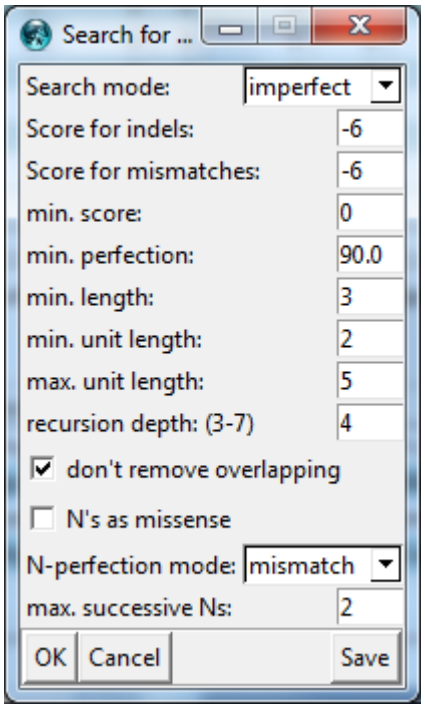
Primer3



Primer3 is a tool for designing PCR primers. To use the 'Primer3'-function or, in the right-click menu, 'Create primer'-function of T-ACE the primer3_core executable has to be in the system path. Primer3 is available at:
http://sourceforge.net/projects/primer3/files/

The 'Primer3'-window allows the setting of most Primer3 parameters. The parameters can be saved as standard by clicking the 'Save'-button.
After clicking the 'OK'-button primers will be generated according to the parameter setting. The created primers will be written directly into the database.
If the 'Primer3'-window was opened through the 'Primer3'-option of the 'Tools'-menu, than it will create primers for each checked contig sequence of the 'Database browser' or a project tab, depending on which tab is raised.
The generated primers will be listed in the 'Primer'-tab.

Phobos



Phobos is a tandem repeat search tool. To use the 'Phobos'-function of T-ACE the phobos_cl executable has to be in the system path.

Phobos is available at:
http://www.ruhr-uni-bochum.de/spezzoo/cm/cm_phobos.htm
The 'Phobos'-window allows the setting of most Phobos parameters. The parameters can be saved as standard by clicking the 'Save'-button.
After clicking the 'OK'-button Phobos will search for repeats according to the parameter setting. The found repeats will be written directly into the database.
If the 'Phobos'-window was opened through the 'Phobos'-option of the 'Tools'-menu, than it will search repeats in each checked contig sequence of the 'Database browser' or a project tab, depending on which tab is raised.
The found repeats will be listed in the 'Repeats'-tab.

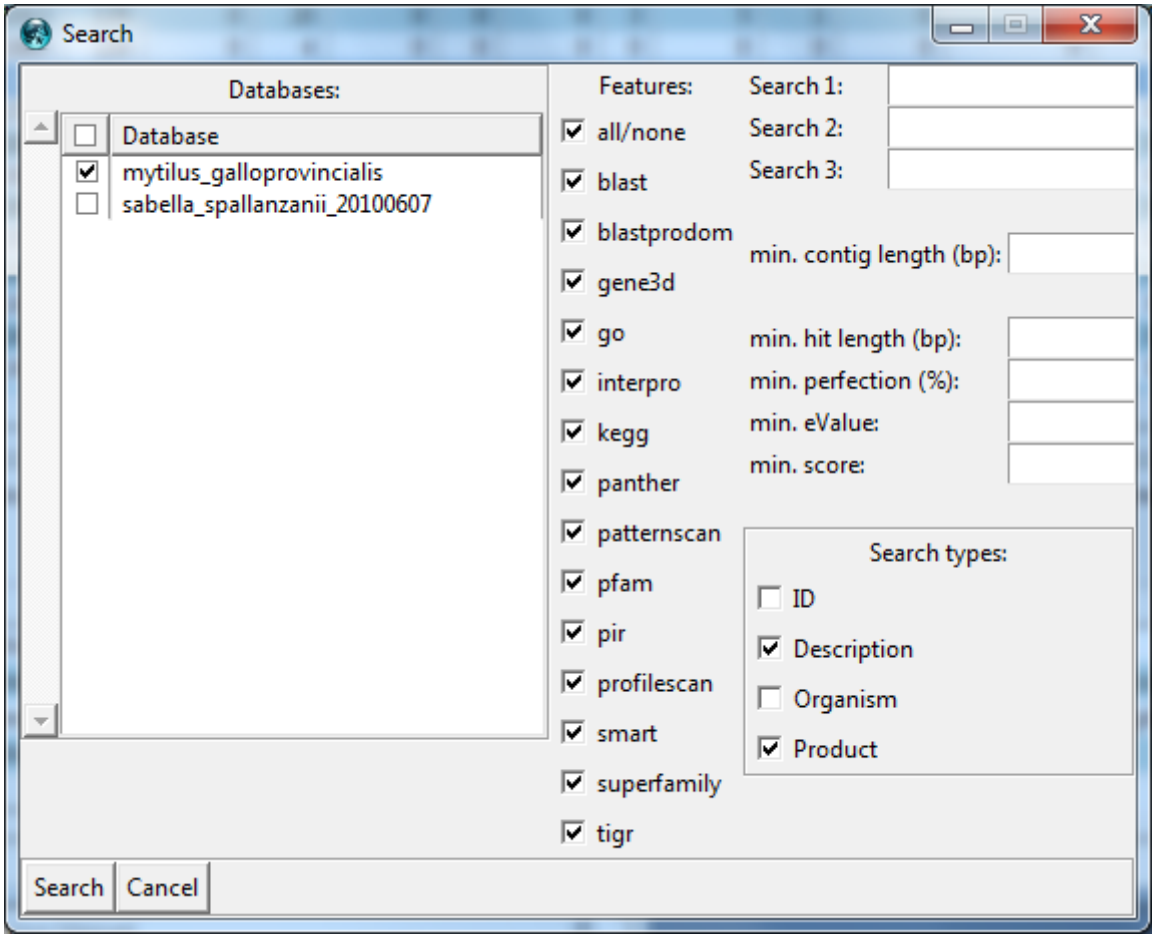
Search

With the 'Search'-window it is easy to search in databases for certain key words. The results will be listed in a project tab.

On the left side of the window the databases are listed, which are available for the search.

Then comes the feature list, were the user can choose for which types of annotation the search shall be performed.

The three 'Search'-fields allow three different searches in the same search run, the results of all three searches will be listed in the same project tab. All words that are typed into the same 'Search'-field have to appear in the annotation of one of the selected 'Search types', otherwise there will be no hit.

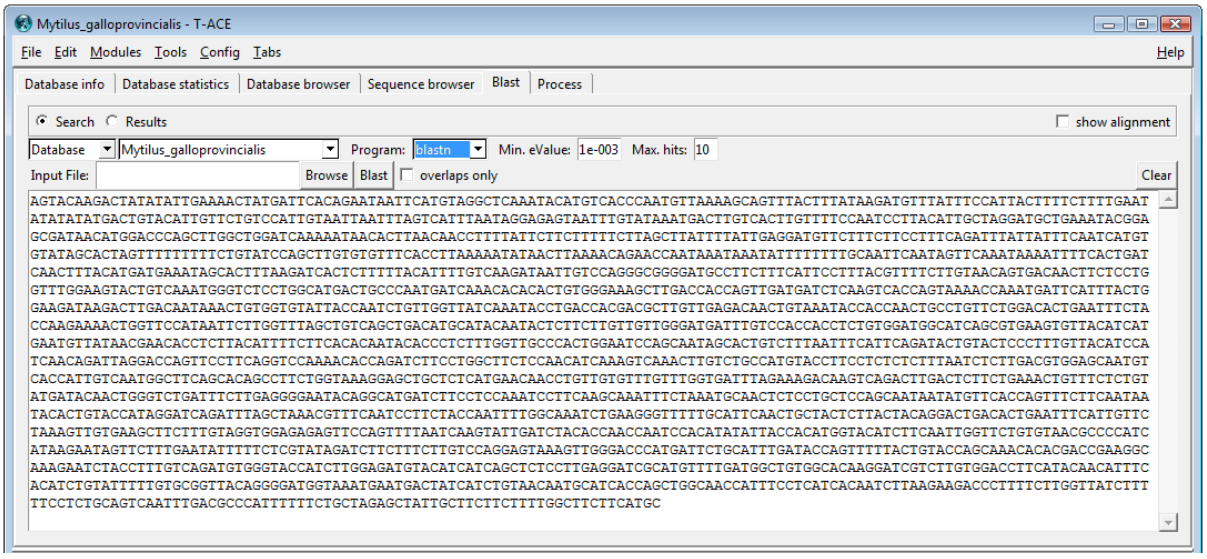


The Tabs of T-ACE

T-ACE mainly consists of two parts. The upper part of the window shows information about the database and its contents, the lower part shows detailed information for a single database entry. Both parts contain several tabs, which will be described in detail in the following sections. Many of the tabs are optional and can be closed through the 'X'-button in the right corner of the tab.

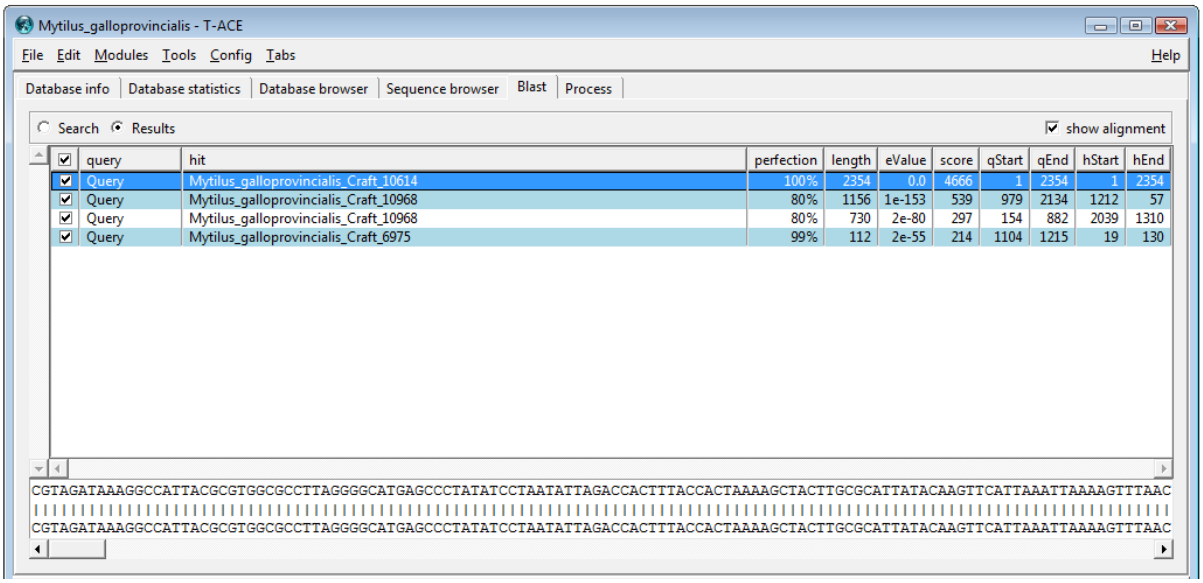
Blast

This tab allows to blast sequences against available databases. The standard blast parameters can be set through the 'Blast parameters'-option in the 'Config'-menu. If a blast server is used for blasting, the 'Database'-combobox will contain a list of every database available to the user, these databases do not necessarily exist on the blast server. When using the 'Local'-option only the databases situated in the /blast_dbs folder in the T-ACE directory are listed in the 'Database'-combobox.



If a mapping is performed from the 'Mapping'-tab while the results table of the 'Blast'-tab is raised, the mapping will show the blast query and its hits.

The 'Add to blast'-option in some right-click menus inserts the according sequence into the sequence panel of the 'Blast'-tab.



Blast results

This tab displays all blast hits in the database associated with the currently selected contig, sorted by evalue. The right-click menu option 'Show all hits' will create a project tab, which lists all contigs with this blast hit.

Comment

The 'Comment'-tab displays the comment text for the selected database entry. It is also possible to write a text in the 'Comment'-panel and save it to the database.

The 'Database browser' includes the option to list only such database entries, which have a comment text.

Database browser

The 'Database Browser' contains a table which shows the different sequence entries of a database. Each row of the table gives an overview of the information available for the according sequence, for example the number of blast or domain hits.

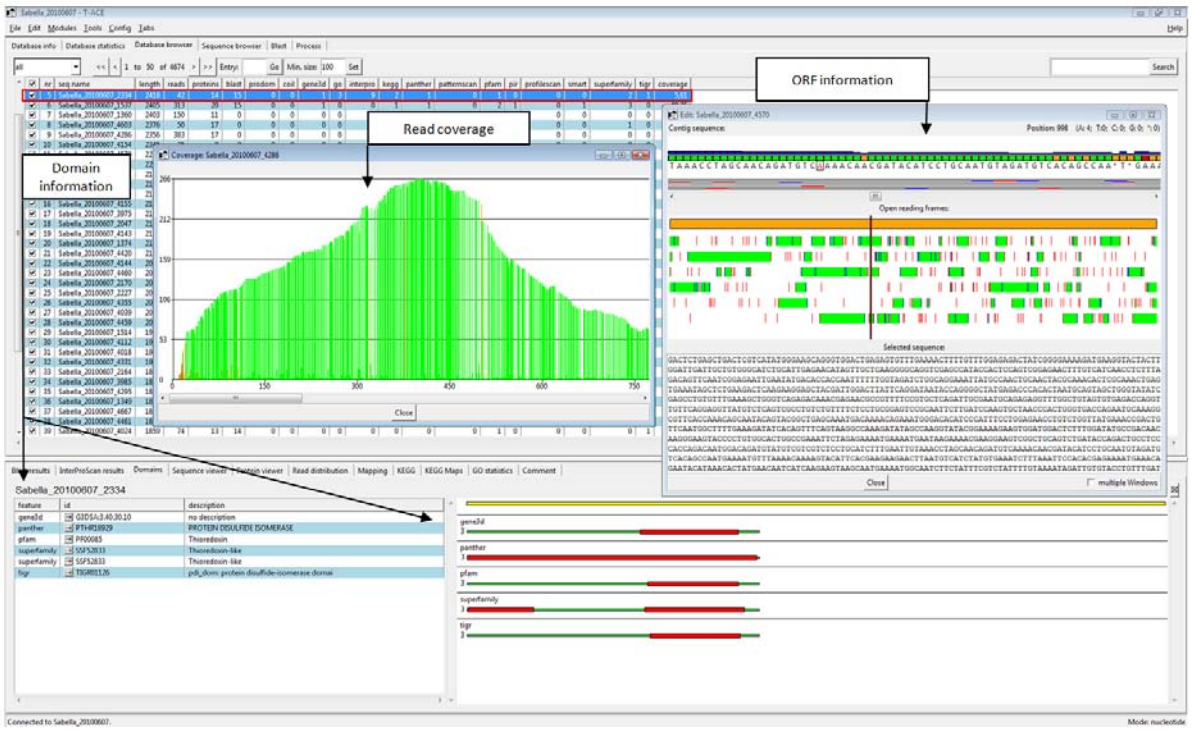
The combobox at the left corner allows you to make a selection of the sequence entries in the table. Such a selection will also add the description column to the table.

The 'Go'-button allows you to jump to a specific entry in the table. To do this you have to enter the number of the sequence entry or its name, like it is written in the 'seq name'-column of the table.

The 'Min. size'-option will remove all table entries with a sequence length smaller than the given size.

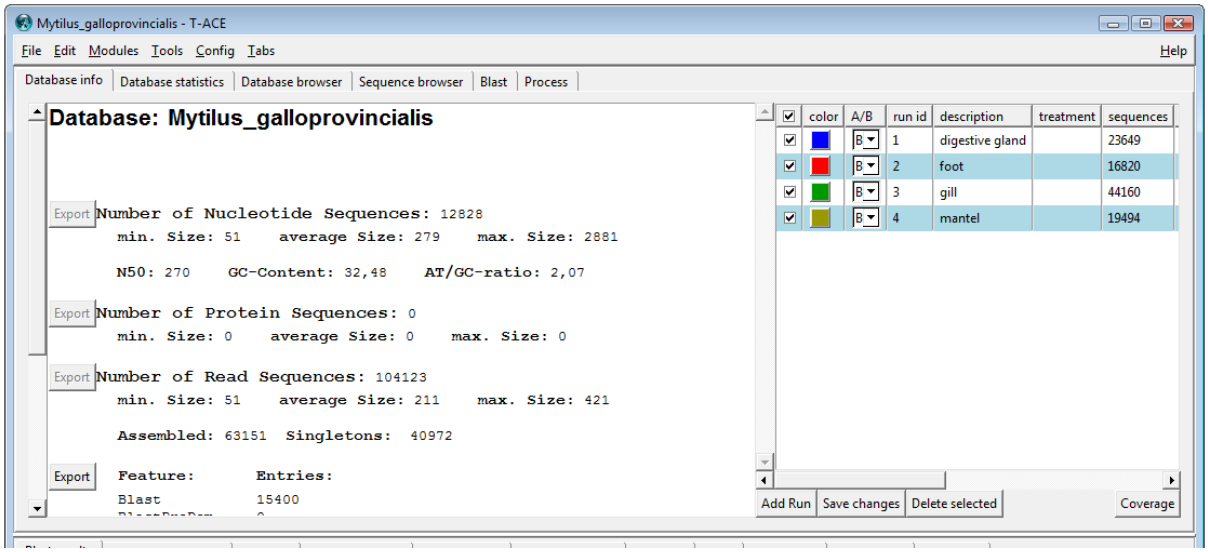
At last there is a search option; it searches only through the description column of the table. To do this, the description column must be shown.

The right-click menu of the 'Database browser' contains a lot of options for working with the sequence entries, for example 'Coverage', 'Edit sequence' and 'New sequence'.



Database info

This tab shows some information about the database content. Additionally it is possible to export most of this content into files. To do this, use the 'Export'-buttons on the left side of the tab.



The right half of the tab contains the list of the different sequencing runs included in the database. Every cell in the table can be edited, the modifications will be saved by pressing the 'Save changes'-button. The first 3 columns of the table are the most important ones, because they have a direct

effect to other functions such as 'Database statistics' or 'Mapping'. With the first column, the runs are selected, that shall be used by other functions. The second column sets the color in which the reads of the accordant run are displayed. The third column shows, if a run is treated as group *A* or *B*, this is important for comparing two groups of runs (Expression analysis).

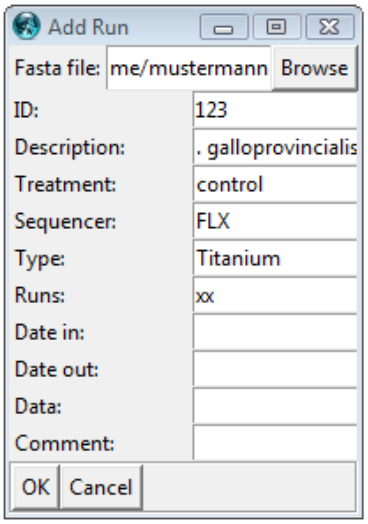
With the 'Add Run'-button the read sequences of a new sequencing run can be added to the database.

The 'Coverage'-button calculates how much of the sequence information in the database covered by the selected runs.

The 'Export read distribution'-option in the right-click menu exports a table, which contains a list of all contigs with its read distribution over all runs in the database.

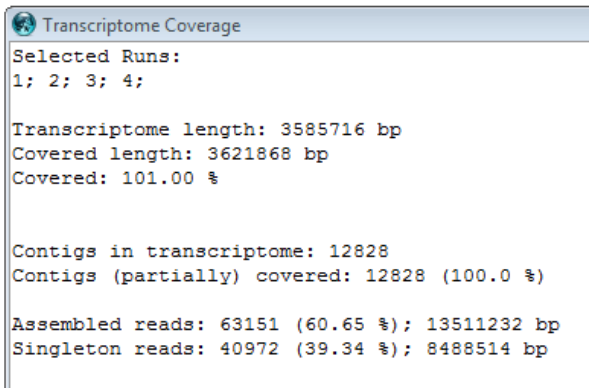
Add run

The 'Add run'-option writes the sequences from a fasta file as reads into the database, if a file is given. Only the 'ID' has to be set, otherwise the run cannot be added. The other fields could be left empty; they can be edited at any time.



Coverage

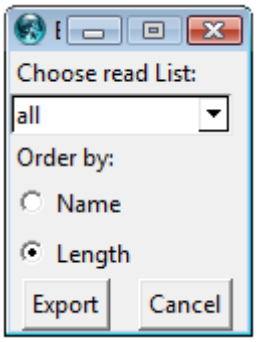
The 'Coverage'-option calculates the number of bases, of the whole transcriptome, covered by the reads of the selected runs and counts all contigs, that are partially covered those reads. The results are not correct to the last base, because of insertions and deletions in the reads.

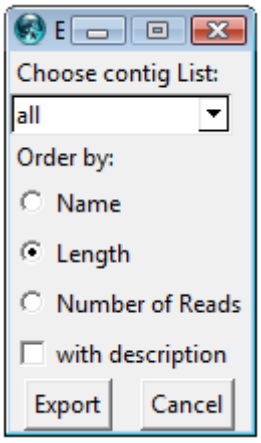


'Export'-buttons

Export sequences

With the first three export buttons the contig, protein and read sequences can be exported to a fasta file. While exporting the contig sequences it is possible to select a subset of contigs, which have blast annotations or annotations of a specific domain database (pfam, smart, etc.). If such a subset is chosen the description of the best annotation hit of the selected category can be attached to the fasta header of each sequence.

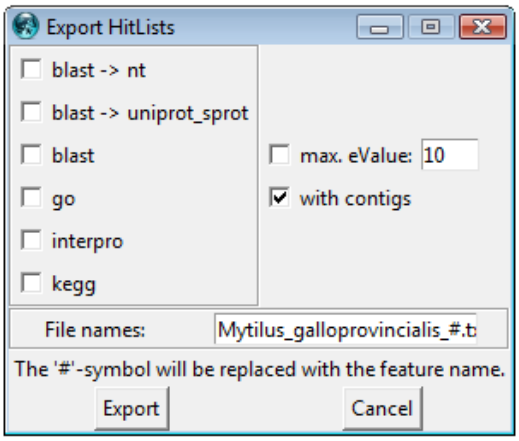




While exporting read sequences it is possible to choose to export just the singleton reads, the assembled reads, the reads of a specific run or all reads.

Export features

With the fourth export button a list of all different hits, including their descriptions, will be exported to a tab-delimited text file, for each selected feature. If the option 'with contigs' is selected each hit entry also lists it associated contigs.



Database statistics

The 'Database statistics'-tab allows different kinds of overviews of the selected database through the 'Coverage' and 'Frequency' options. The 'Expression analysis'-option calculates the fraction of reads, in the *A* and *B* groups, in each contig, depending on the run settings in the 'Database info'-tab.

The '<' (previous) and '>' (next) buttons enable switching between already created 'Database statistics'-lists, with the 'Remove'-button the currently opened list can be deleted. The tab shows normally a diagram, but it can be switched to a tab-delimited list of the results by selecting the 'List'-radiobutton instead of the 'Diagram'-radiobutton.

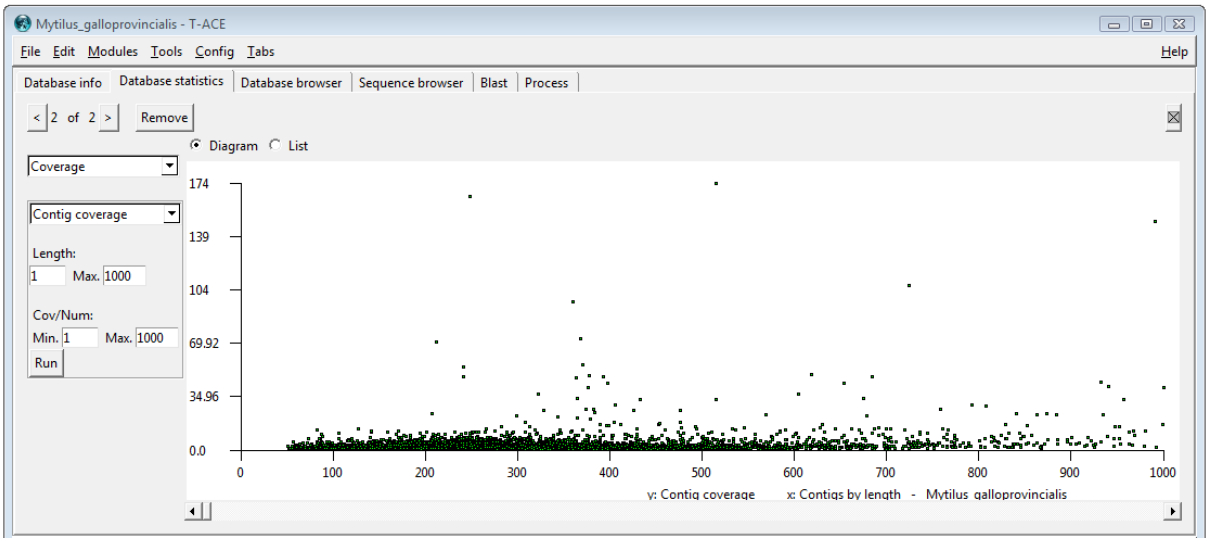
Coverage

Contig coverage

This shows the average coverage for each contig within the given parameters. The diagram shows a dot for each contig; its position depends on the contig length (x-axis) and the contigs average coverage (sum of all bases of all reads in the contig / number of bases in the contig; y-axis).

Coverage num. of reads

This diagram also shows a dot for each contig, but here its position depends on the contig length (x-axis) and the number of reads contained in the contig (y-axis).

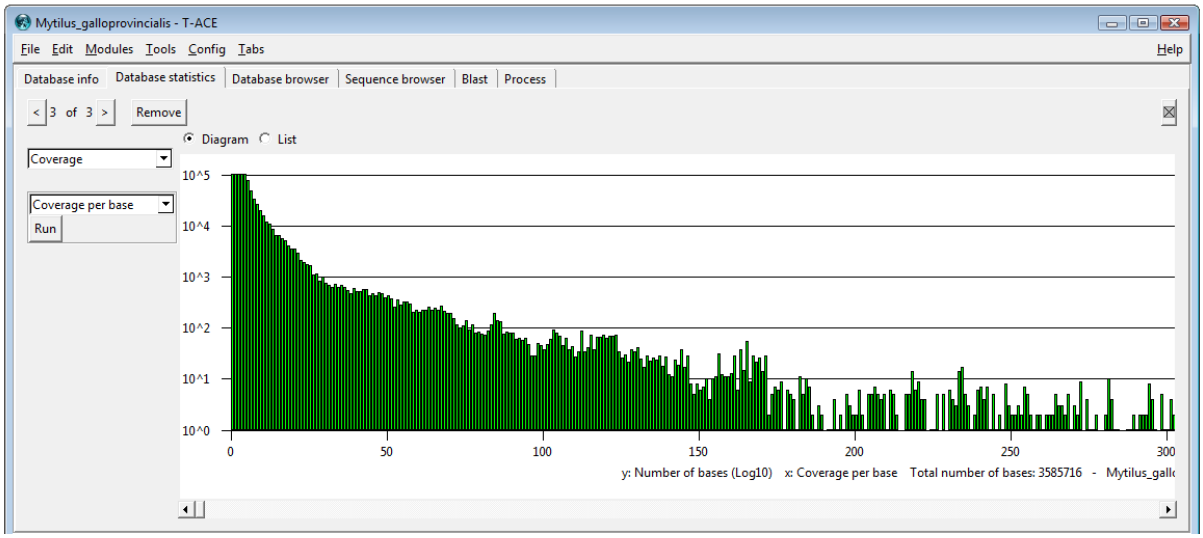


Coverage per base

This function calculates the coverage for each base in each contig of the database. For each different coverage value, the number of bases which have the same coverage value is then calculated and displayed in the graph. The number of bases (y-axis) is displayed in logarithmic values; the maximal displayed value is 100.000. To see absolute values switch to the list.

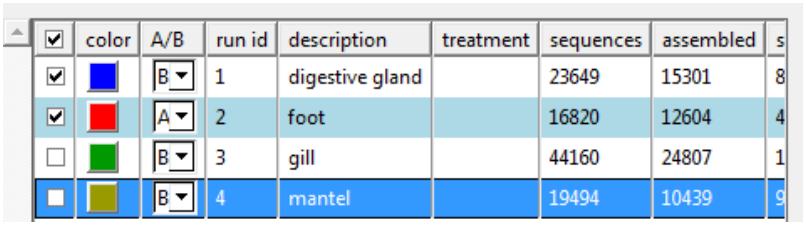
Coverage per base (gaps)

This is the same as 'Coverage per base', with the difference, that also the coverage of the gaps in the contigs is counted.



Expression analysis

'Expression analysis' tries to calculate the number of reads, in the *A* and *B* groups, in a contig. For this the run table in the 'Database info'-tab is very important. Only reads from runs that are selected in the run table are used for this calculation. Also it is important, that some of the selected runs are marked as *A* and others as *B*.

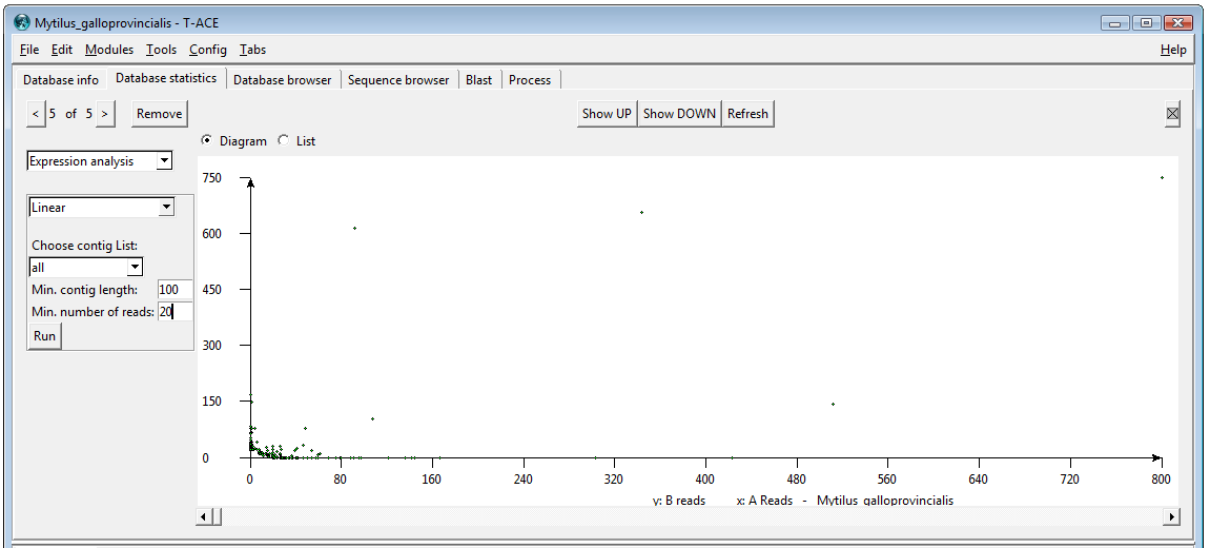


	color	A/B	run id	description	treatment	sequences	assembled	s
☑		B	1	digestive gland		23649	15301	8
☑		A	2	foot		16820	12604	4
☐		B	3	gill		44160	24807	1
☐		B	4	mantel		19494	10439	9

The different options use the following equations for the calculation:

a: reads in group A*; b: reads in group* B

Linear	***Log2***
x-axis: a; y-axis: b	*x-axis: log2(a); y-axis: log2(b)*

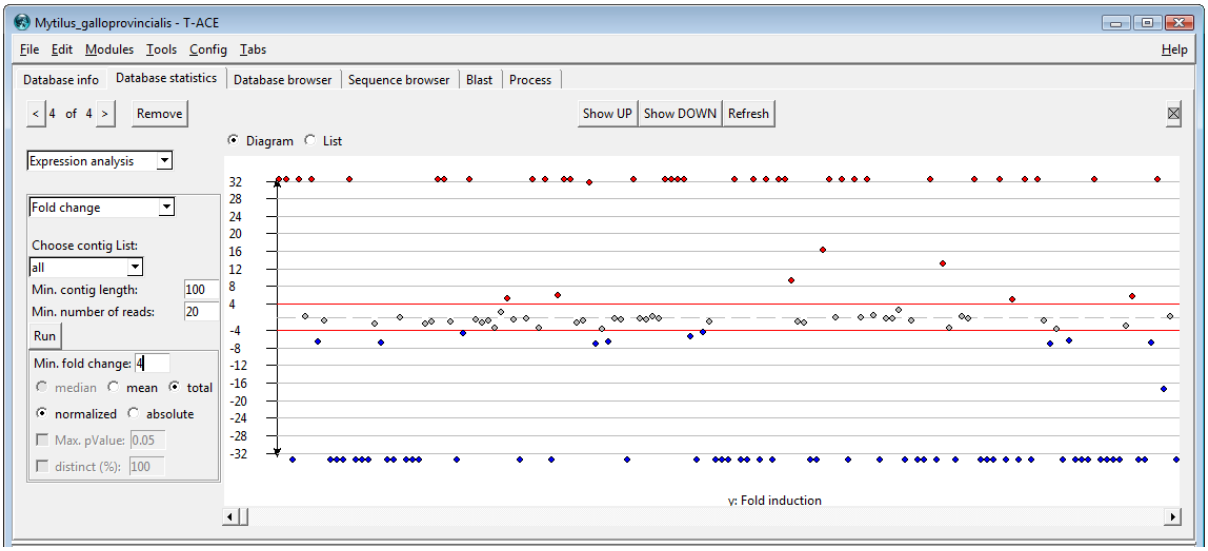


Fold change	***Fold change (Log2)***
if b > a:	*if b > a:*
y-axis: b/a	*y-axis: log2(b/a)*
if b < a:	*if b < c:*
y-axis: -a/b	*y-axis: log2(-a/b)*

When multiple runs are selected for each group (A and B) the user can choose between a calculation with the total, the mean or the median number of reads for each contig. He also can choose if absolute or normalized read numbers should be used.

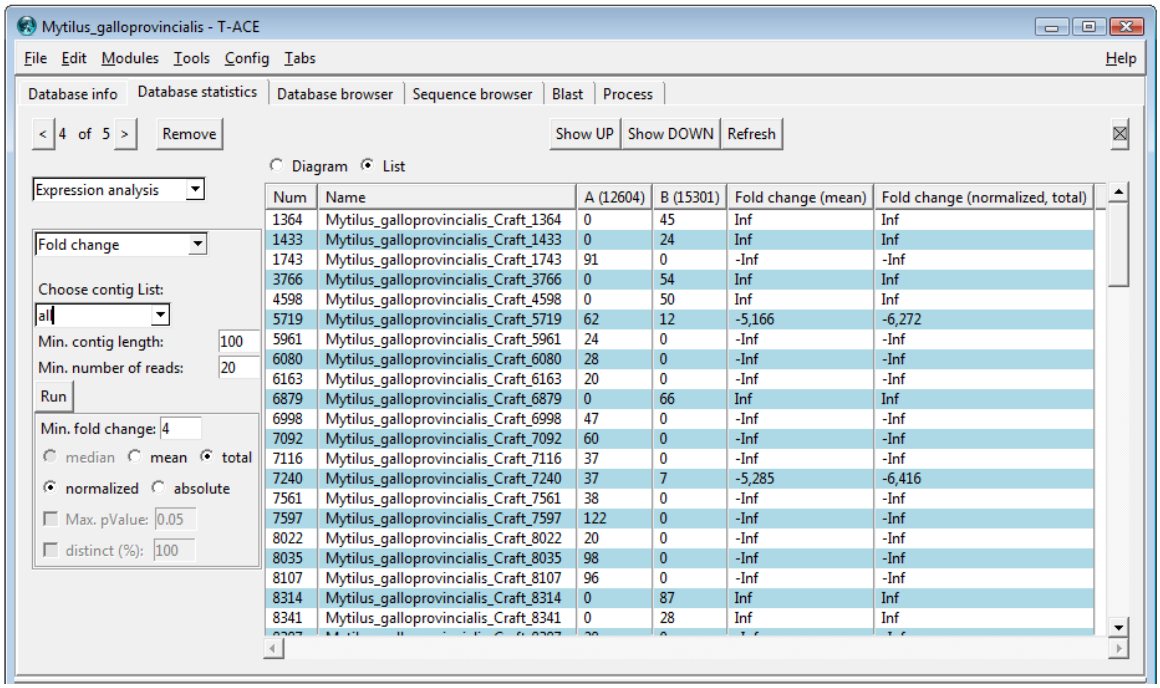
The p-value calculation is based on a Mann-Whitney U test, the p-value corrections is done by a Benjamini-Hochberg Step-Up FDR-controlling Procedure.

Under 'Fold change' the down- and/or up-regulated contigs can be send to project tabs. To do this, just press the 'Show UP'- and/or 'Show DOWN'-button. These projects tabs can then be used for the run-compare tool to investigate whether GO and KEGG terms or domains are enriched or depleted in the up- or down-regulated group of contigs.



The list view shows a table of all contigs which match the parameter settings. The list can be exported through a right-click menu. The menu also allows hiding or unhiding most of the available columns of the list.

When changing parameters it is not necessary to run the calculation again, it suffices to press the 'Refresh'-button. Only after a change in the given dataset (‘min. contig length’ and ‘min. number of reads’) or the percentage of the ‘distinct’-parameter, the calculation has to run again.



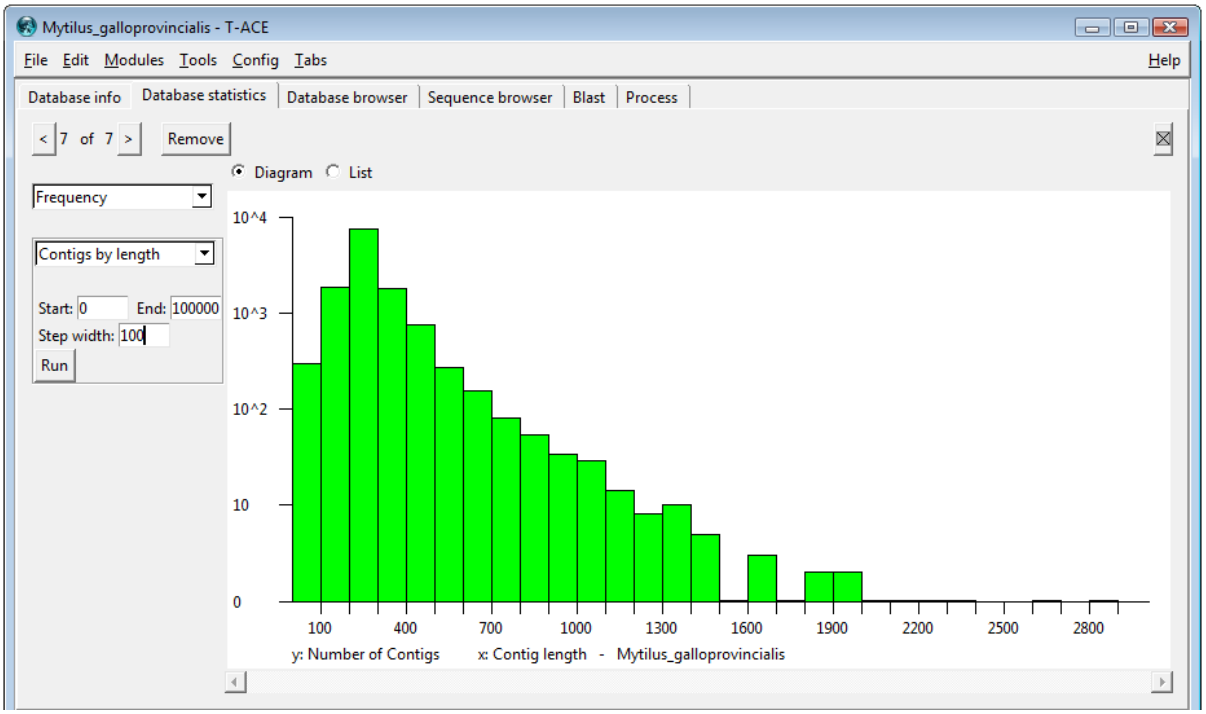
Frequency

Contigs by length

It shows the number of contigs within a certain length range, the range is defined by the 'Step width'-parameter. The diagram shows bars for each contig length range. The height of a bar depends on the number of contigs, which have a length in the range that is associated with the bar. The number of contigs (y-axis) is displayed in logarithmic values. To see absolute values switch to the list.

Reads by length

This is the same as 'Contigs by length', but instead of the contig lengths it displays the read length distribution.

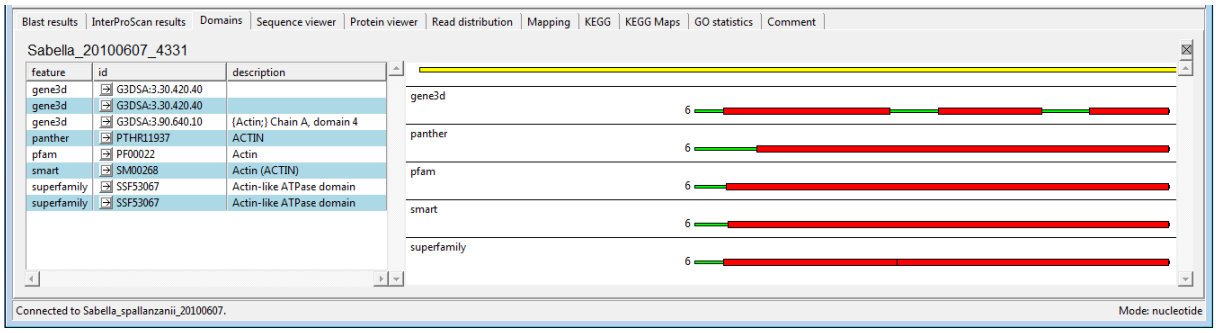


Contigs by reads

This function counts the number of contigs, which consist of the same number of reads. The diagram shows bars for each read number range, the range is defined by the 'Step width'-parameter. The number of contigs (y-axis) is displayed in logarithmic values. To see absolute values switch to the list.

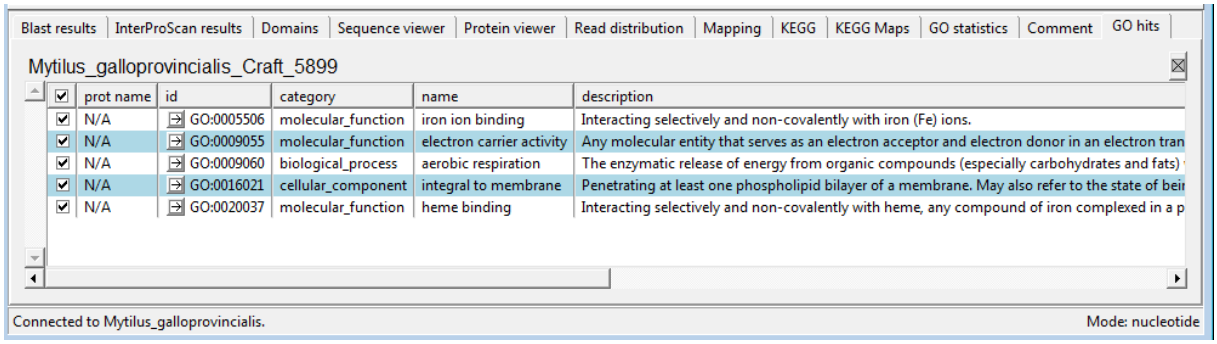
Domains

The 'Domains'-tab visualizes the domain hits of the selected sequence. The green lines display open reading frames (ORFs), the number in front of the lines stands for the frame. The red blocks display the position of domains in the ORFs. By selecting a row of the table the associated domain will be marked in yellow.



GO hits

This tab displays all GO hits in the database associated with the currently selected contig, unlike the 'InterProScan results'-tab, it also shows the category and description of the GO hits.



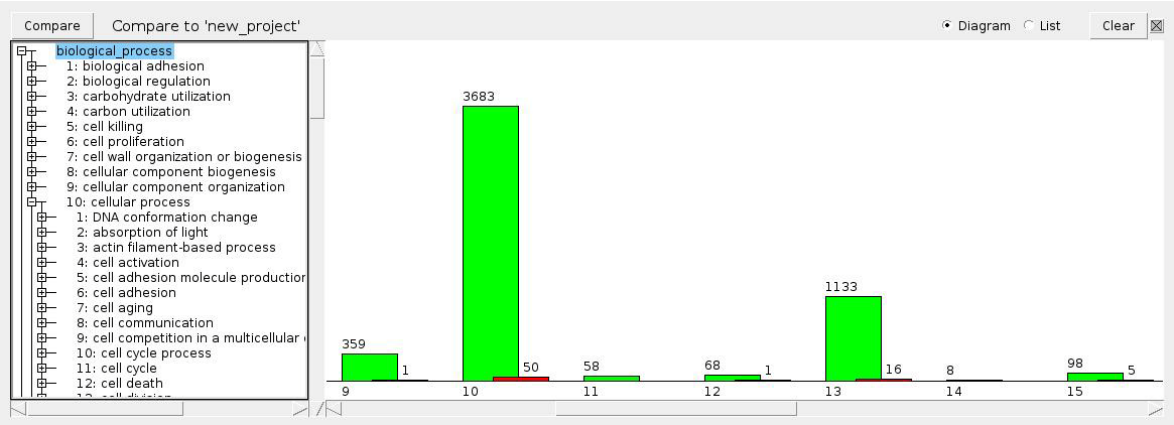
GO statistics

The 'GO statistics'-tab visualizes the numbers of GO ontology entries in the database for all levels of the GO-tree.

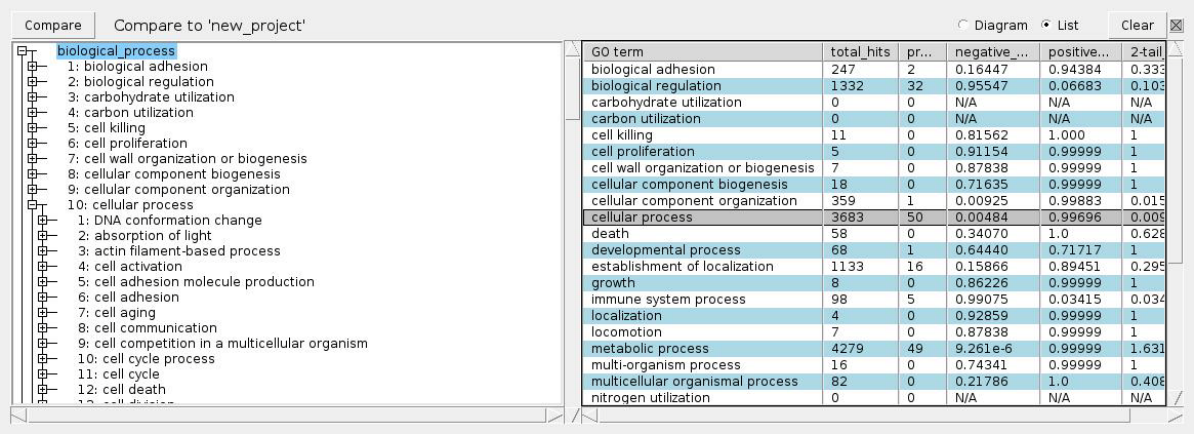
On the left side of the tab is the GO-tree. By clicking on a '+'-button in the tree the associated node will be unfolded. By clicking directly on a node all its children will be displayed as green bars in the 'GO statistics'-canvas. The number above each bar is the number of contigs with GO hits, which are associated with the GO-id represented by the bar, either as direct hit or as a hit to any of the children of this GO-id. The numbers below the bars refer the numbers of the children of the displayed node in the GO-tree.

With the 'Compare'-button the entries of a raised project tab will be added to the GO-tree diagram as red bars.

By right-clicking on a bar the option 'Show all hits' appears, which will open a project tab containing all contigs with a GO-hit belonging to the selected bar.

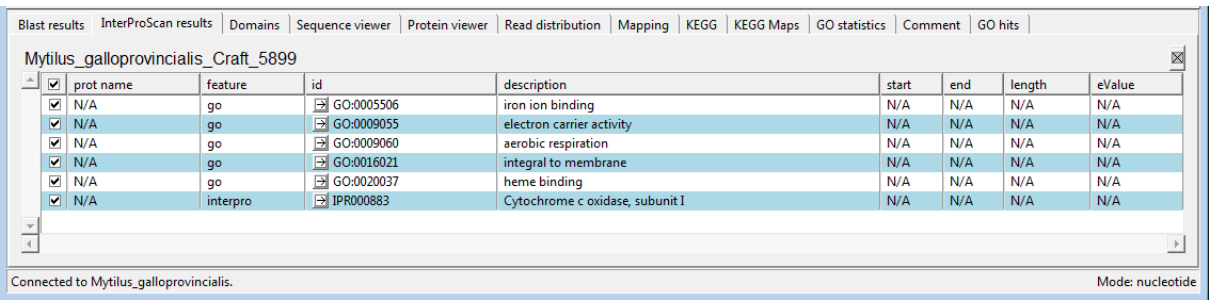


By switching the radiobutton from 'Diagram' to 'List' a table is displayed, which contains all children of the selected node. The table shows the name of each child and the number of contigs, which are associated with it. If 'Compare' was executed with a project tab, it also shows the number of entries from the project tab associated with each child and the corresponding p-values. By right-clicking on the table it is possible to copy the whole table into the clipboard or just the content of the selected cell.



InterProScan results

This tab displays all GO, InterPro and domain hits in the database associated with the currently selected contig, sorted by evalue, if possible. The right-click menu option 'Show all hits' will create a project tab, which lists all contigs with this GO, InterPro or domain hit.



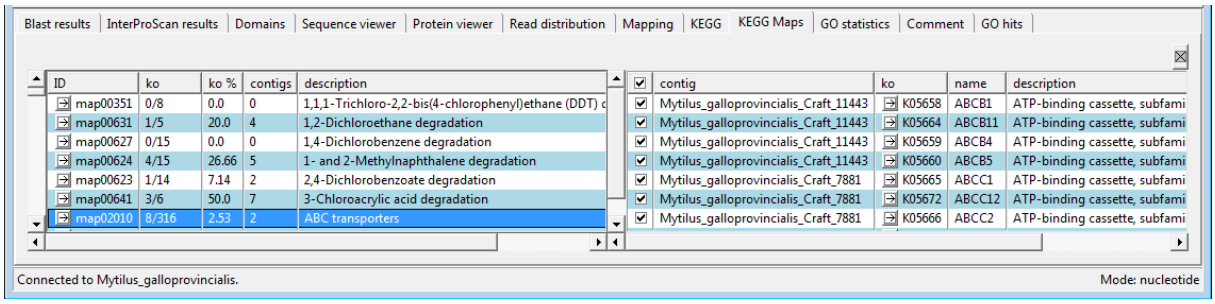
KEGG

The 'KEGG'-tab displays the KEGG orthology (ko) hits associated with the selected sequence.

KEGG Maps

This tab contains two tables. The left table shows a list of all KEGG pathways in the parent database. For each pathway is shown how many contigs are associated with this pathway and how much of the pathway is covered by those contigs.

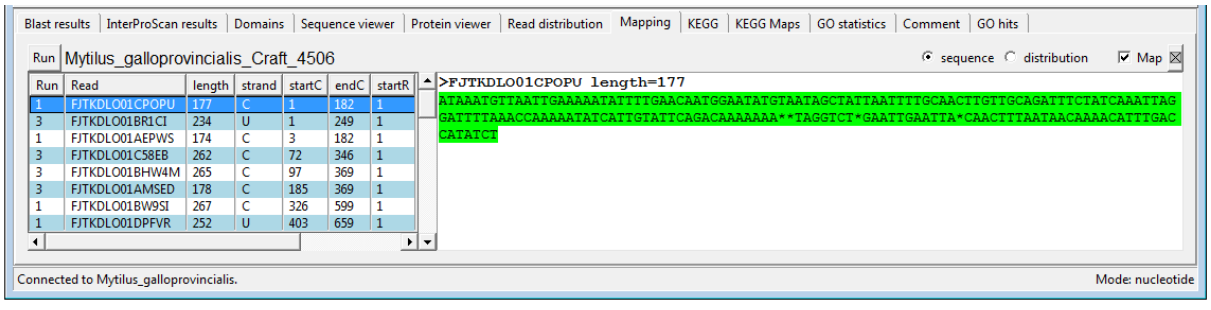
By selecting one pathway the associated contigs, together with their KEGG orthology (ko) hits, will be listed in the right table of the tab.



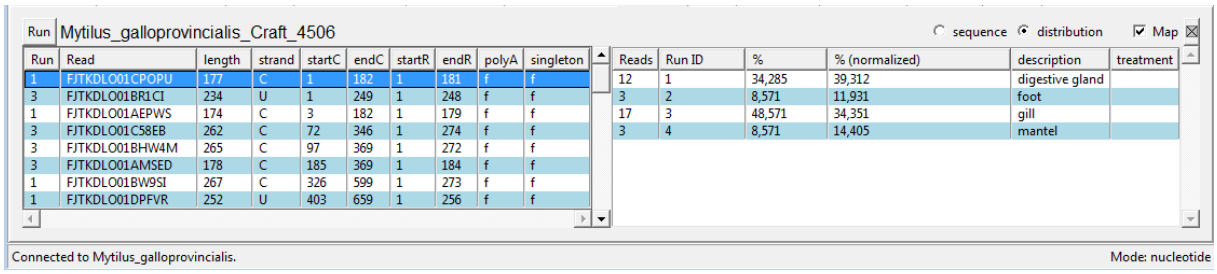
Mapping

This tab allows displaying the mapping of the reads of a contig or the mapping of the blast results of a query sequence. If the 'Blast'-tab is raised the first query sequence and its blast results will be mapped, otherwise the selected contig and its reads will be mapped.

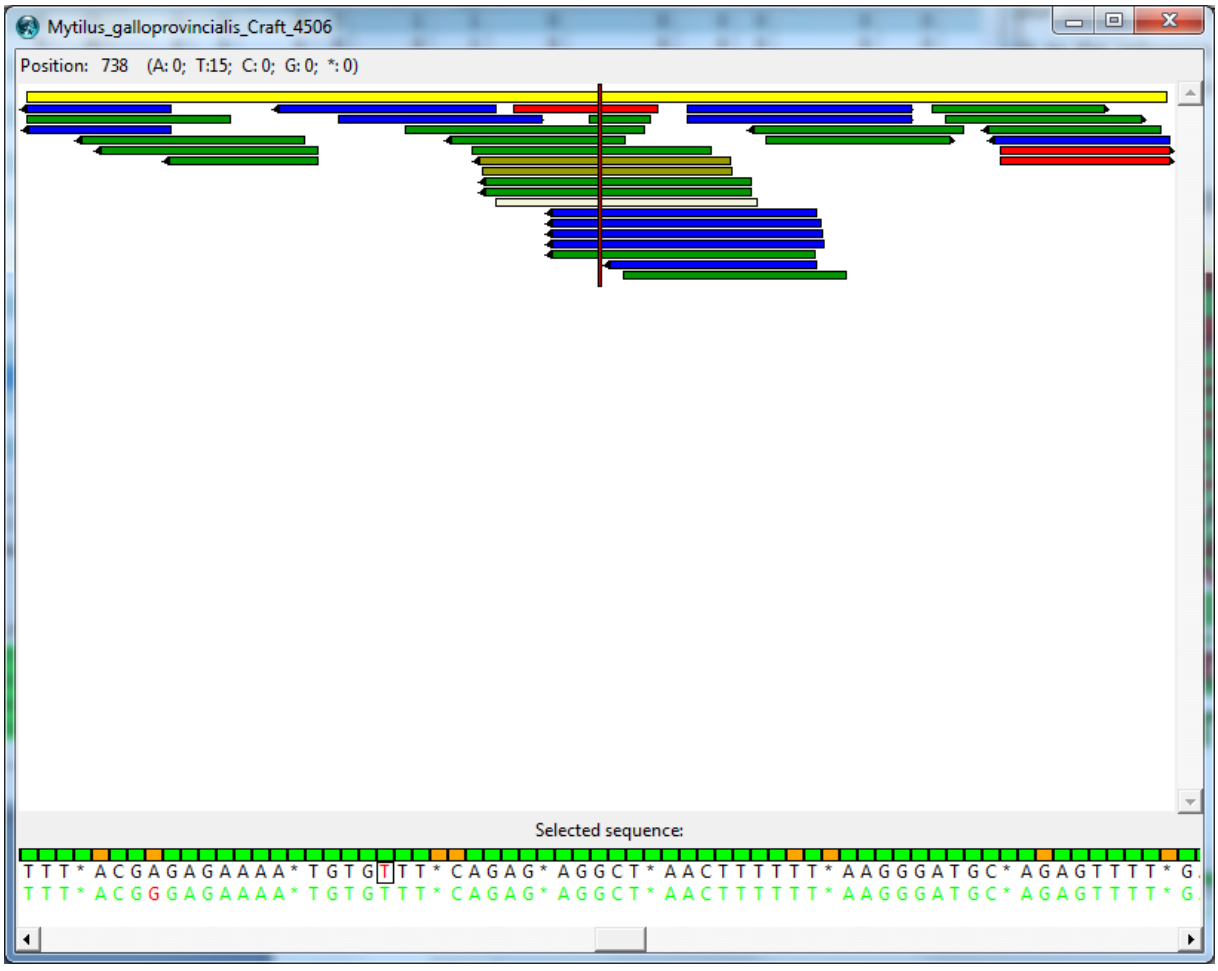
The left half of the 'Mapping'-tab contains a table which lists all reads/blast-results associated with the mapped contig/blast-query. When selecting a row of the list the accordant reads/blast-results sequence will be shown in the right half of the tab, if the 'sequence'-radiobutton is selected. The selection also highlights the reads/blast-results in the mapping window.



When the 'distribution'-radiobutton is selected a table appears in the right half of the tab. This table shows the read distribution of the contig. Through the right-click menu the reads of a single run can be exported.

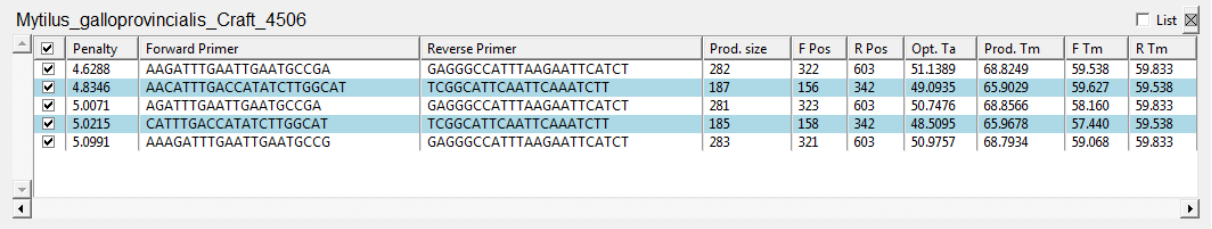


The mapping window only opens if the 'Map'-option of the 'Mapping'-tab is checked. It displays the reference sequence (contig/blast-query) as a yellow bar. The hit sequences (reads/blast-results) are displayed in green or in various colors, depending on the color setting for the runs in the 'Database info'-tab. Regions of the hit sequences, which do not match to the reference, are shown in gray. At the bottom of the mapping window the reference sequence is shown. By clicking on a hit bar the accordant sequence will appear under the reference sequence.



Primer3

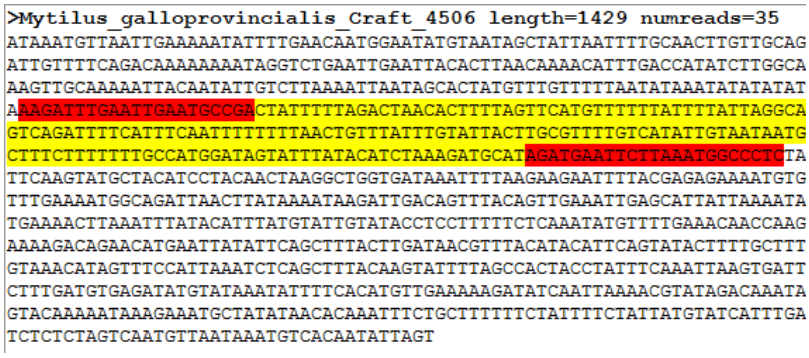
The 'Primer'-tab lists all primer pairs, of the selected contig, which were created by running the 'Primer3'-tool.



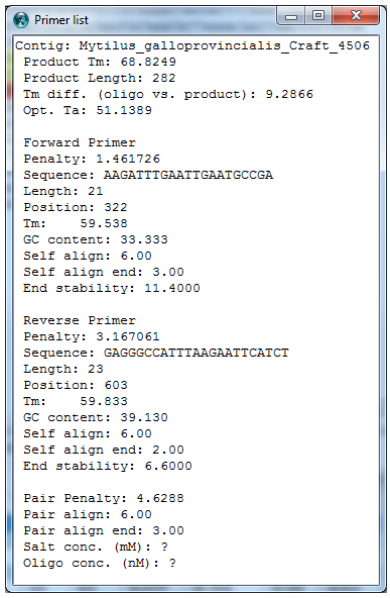
Mytilus_galloprovincialis_Craft_4506 — List

	Penalty	Forward Primer	Reverse Primer	Prod. size	F Pos	R Pos	Opt. Ta	Prod. Tm	F Tm	R Tm
☑	4.6288	AAGATTTGAATTGAATGCCGA	GAGGGCCATTTAAGAATTCATCT	282	322	603	51.1389	68.8249	59.538	59.833
☑	4.8346	AACATTTGACCATATCTTGGCAT	TCGGCATTCAATTCAAATCTT	187	156	342	49.0935	65.9029	59.627	59.538
☑	5.0071	AGATTTGAATTGAATGCCGA	GAGGGCCATTTAAGAATTCATCT	281	323	603	50.7476	68.8566	58.160	59.833
☑	5.0215	CATTTGACCATATCTTGGCAT	TCGGCATTCAATTCAAATCTT	185	158	342	48.5095	65.9678	57.440	59.538
☑	5.0991	AAAGATTTGAATTGAATGCCG	GAGGGCCATTTAAGAATTCATCT	283	321	603	50.9757	68.7934	59.068	59.833

When selecting a row of the 'Primer'-table, the accordant primer pair is highlighted in the 'Sequence Browser'.



```
>Mytilus_galloprovincialis_Craft_4506 length=1429 numreads=35
ATAAATGTTAATTGAAAAATATTTTGAACAATGGAATATGTAATAGCTATTAATTTTGCAACTTGTTGCAG
ATTGTTTTCAGACAAAAAAAATAGGTCTGAATTGAATTACACTTAACAAAACATTTGACCATATCTTGGCA
AAGTTGCAAAAATTACAATATTGTCTTAAAATTAATAGCACTATGTTTGTTTTTAATATAAATATATATAT
AAAGATTTGAATTGAATGCCGACTATTTTTAGACTAACACTTTTAGTTCATGTTTTTTATTTTATTAGGCA
GTCAGATTTTCATTTCAATTTTTTTTAACTGTTTATTTGTATTACTTGCGTTTTGTCATATTGTAATAATG
CTTTCTTTTTTTTGCCATGGATAGTATTTATACATCTAAAGATGCATAGATGAATTCTTAAATGGCCCTCTA
TTCAAGTATGCTACATCCTACAACTAAGGCTGGTGATAAATTTTAAGAAGAATTTTACGAGAGAAAATGTG
TTTGAAAATGGCAGATTAACTTATAAAATAAGATTGACAGTTTACAGTTGAAATTGAGCATTATTAAAATA
TGAAAACTTAAATTTATACATTTATGTATTGTATACCTCCTTTTTCTCAAATATGTTTTGAAACAACCAAG
AAAAGACAGAACATGAATTATATTCAGCTTTACTTGATAACGTTTACATACATTCAGTATACTTTTGCTTT
GTAAACATAGTTTCCATTAAATCTCAGCTTTACAAGTATTTTAGCCACTACCTATTTCAAATTAAGTGATT
CTTTGATGTGAGATATGTATAAATATTTTCACATGTTGAAAAAGATATCAATTAAAACGTATAGACAAATA
GTACAAAAATAAAGAAATGCTATATAACACAAATTTCTGCTTTTTTCTATTTTCTATTATGTATCATTTGA
TCTCTCTAGTCAATGTTAATAAATGTCACAATATTAGT
```



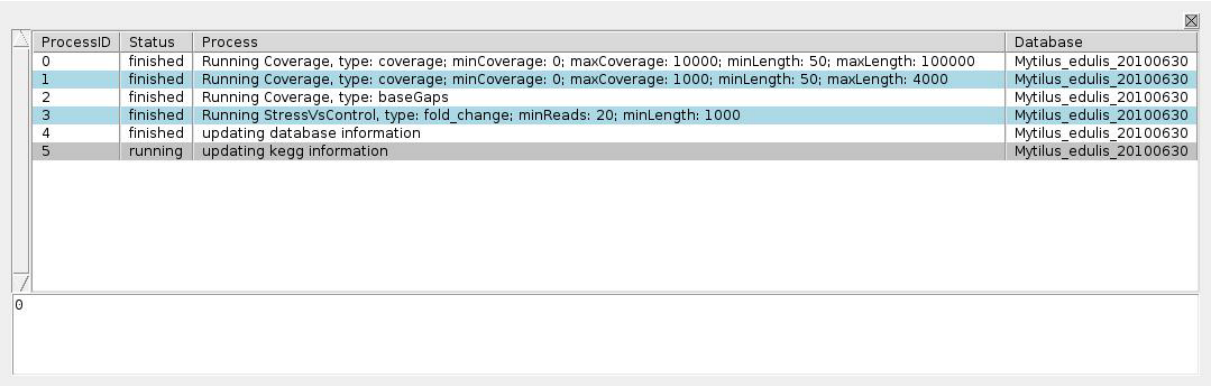
Primer list

```
Contig: Mytilus_galloprovincialis_Craft_4506
 Product Tm: 68.8249
 Product Length: 282
 Tm diff. (oligo vs. product): 9.2866
 Opt. Ta: 51.1389

 Forward Primer
 Penalty: 1.461726
 Sequence: AAGATTTGAATTGAATGCCGA
 Length: 21
 Position: 322
 Tm:    59.538
 GC content: 33.333
 Self align: 6.00
 Self align end: 3.00
 End stability: 11.4000

 Reverse Primer
 Penalty: 3.167061
 Sequence: GAGGGCCATTTAAGAATTCATCT
 Length: 23
 Position: 603
 Tm:    59.833
 GC content: 39.130
 Self align: 6.00
 Self align end: 2.00
 End stability: 6.6000

 Pair Penalty: 4.6288
 Pair align: 6.00
 Pair align end: 3.00
 Salt conc. (mM): ?
 Oligo conc. (nM): ?
```

If the 'List'-field of the 'Primer'-tab is checked the information about the selected primer pair will also be shown in a separate window.

Process

The 'Process'-tab gives an overview of processes executed in the current session. It mainly informs about the state of database updates, searches and blast/InterPro runs.



ProcessID	Status	Process	Database
0	finished	Running Coverage, type: coverage; minCoverage: 0; maxCoverage: 10000; minLength: 50; maxLength: 100000	Mytilus_edulis_20100630
1	finished	Running Coverage, type: coverage; minCoverage: 0; maxCoverage: 1000; minLength: 50; maxLength: 4000	Mytilus_edulis_20100630
2	finished	Running Coverage, type: baseGaps	Mytilus_edulis_20100630
3	finished	Running StressVsControl, type: fold_change; minReads: 20; minLength: 1000	Mytilus_edulis_20100630
4	finished	updating database information	Mytilus_edulis_20100630
5	running	updating kegg information	Mytilus_edulis_20100630

0

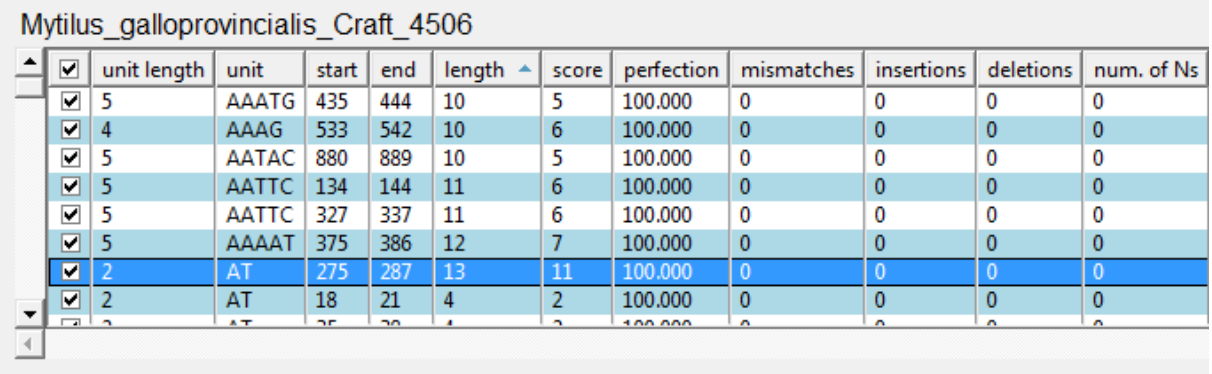
Protein viewer

This tab shows different protein sequences of the selected contig. A green link button in the table means, that the associated protein has domain hits. Red ones have no domain hits.

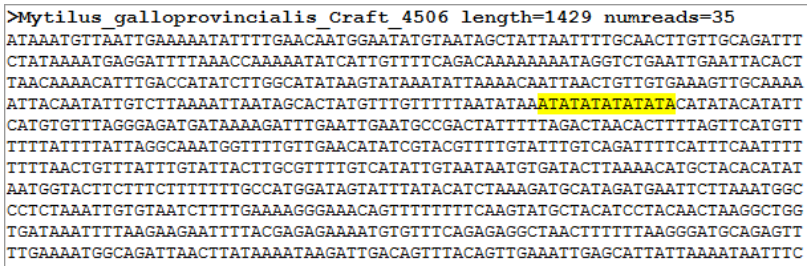
The text panel contains a right-click menu, which allows highlighting different types of amino acids.

Repeats

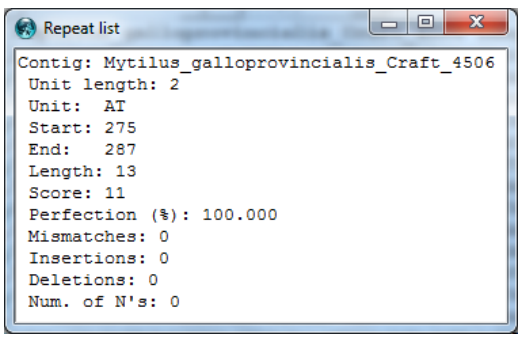
The 'Repeats'-tab lists all repeats, of the selected contig, which were found by running the 'Search repeats'-tool.



When selecting a row of the 'Repeats'-table, the accordant repeat is highlighted in the 'Sequence Browser'.

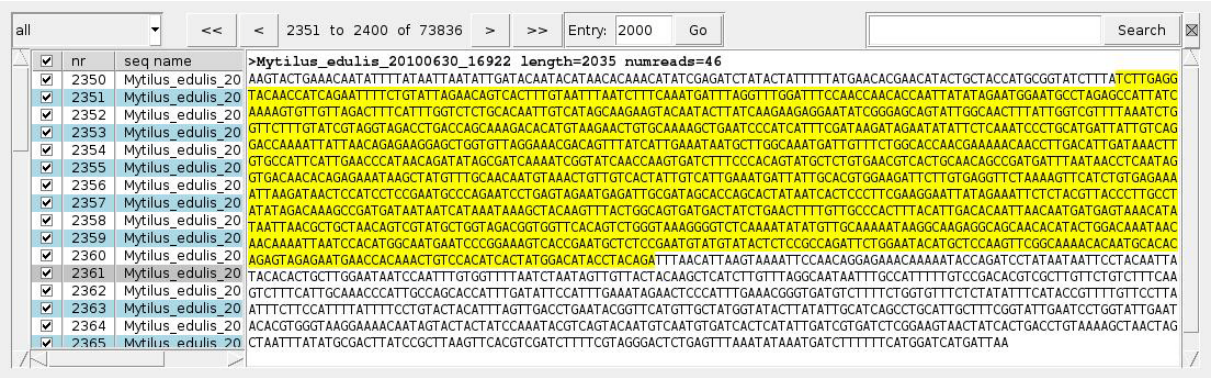


If the 'List'-field of the 'Repeats'-tab is checked the information about the selected repeat will also be shown in a separate window.

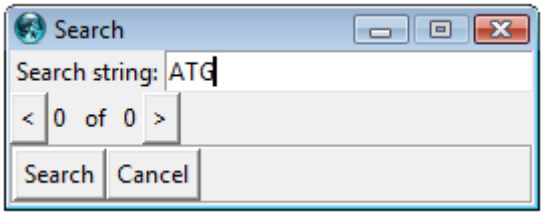


Sequence browser

Most of the 'Database Browser'-options are also available in this tab. It shows the sequence of the selected contig. Additionally, according Blast, InterProScan, primer and repeats hits will be highlighted in the sequence when selected.



It is also possible to make a text search in the panel area of the tab. The search window is available through a right-click menu.



Sequence viewer

This tab just shows the nucleotide sequence of the selected database entry. Like the Sequence Browser it highlights Blast, InterProScan, primer and repeats hits in the sequence when selected. The search option is also available.

Windows

There are some functions of T-ACE, which are only available through right-click menus. These options normally open additional widows.

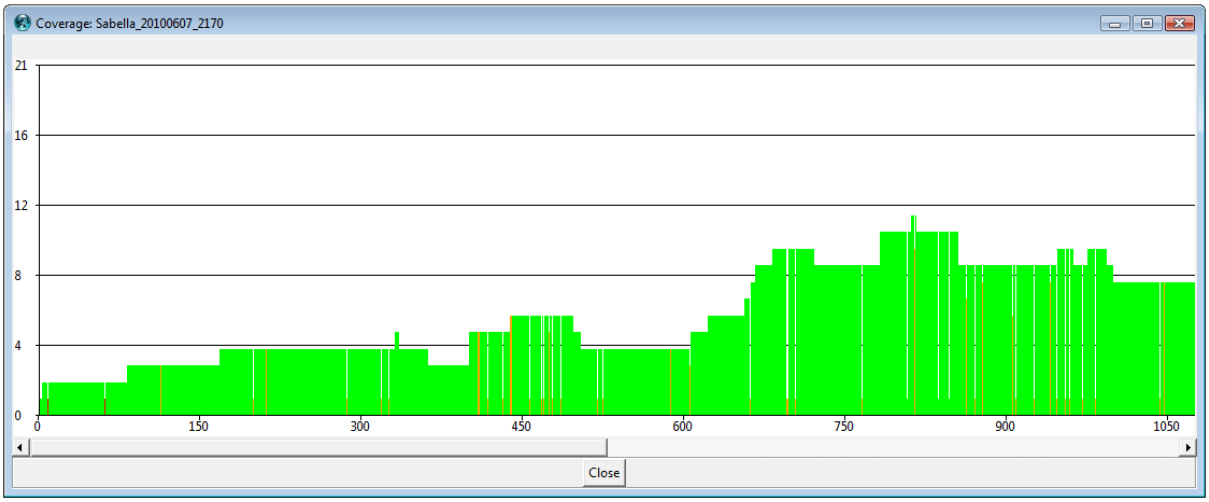
Coverage

Edit sequence

New sequence

Coverage

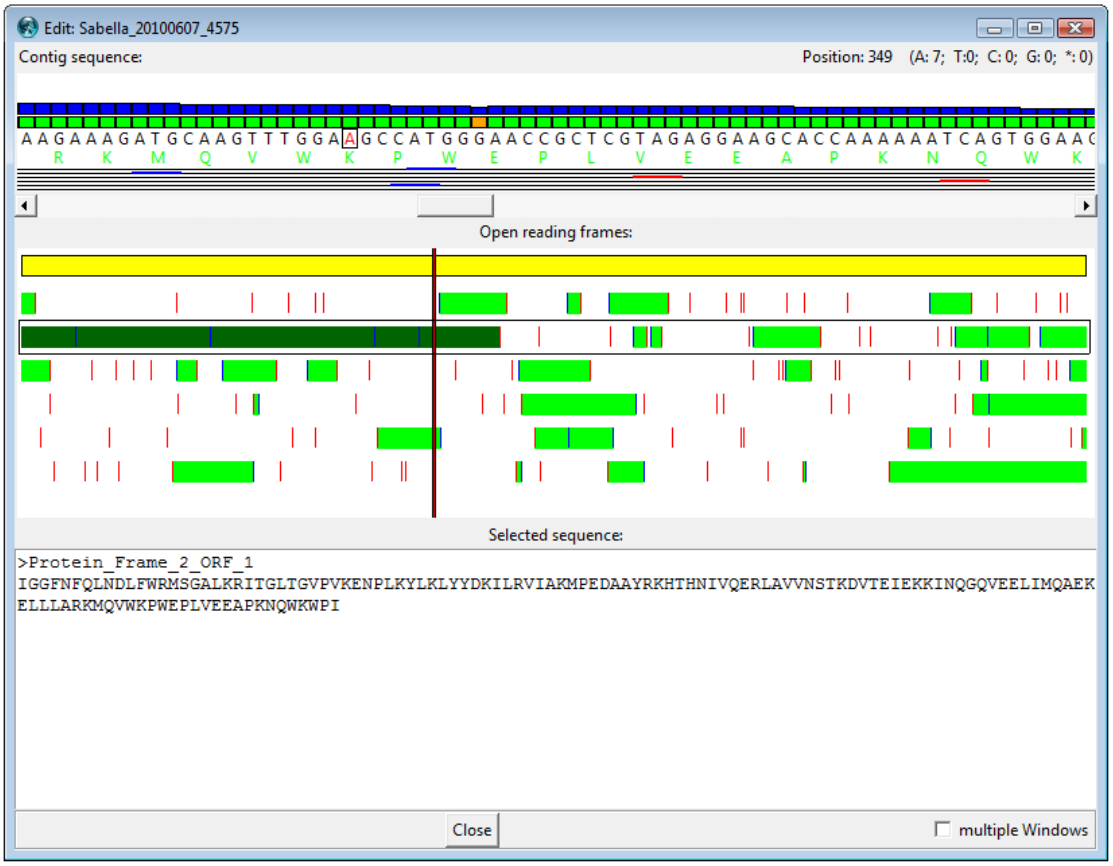
It just displays the coverage for each base of the selected contig.



Edit sequence

'Edit sequence' displays the sequence of the selected contig in detail. It consists of three parts: 'Contig sequence', 'Open reading frames' and 'Selected sequence'

It also contains its own right-click menu.

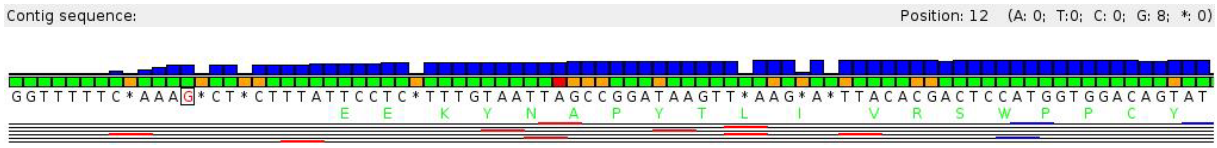


Contig sequence

This part displays the contig sequence. Left-clicking on a base will mark it in red; also a red line appears at the associated position in the 'Open reading frames'-canvas. Right-clicking on a base will open a menu for modifying it.

The blue bars above the bases represent their coverage to a maximum of 30. Then there are green, orange or red squares directly above the bases, these show conformity of the 'read'-bases, which lie at this position. If more than 90% of the 'read'-bases are equal to the 'contig'-base the square is green, it is orange by 60% to 90% and below 60% it is red.

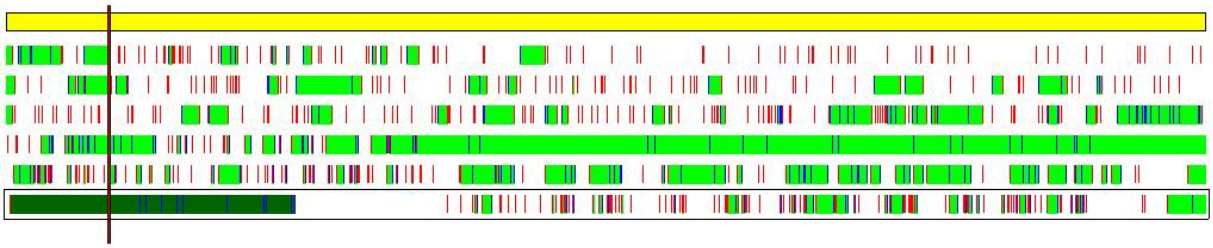
Directly below the bases the amino acids of a selected ORF can be displayed. At last there are six lines, which represent the six frames of the nucleotide sequence. On the lines can appear blue or red regions, these are markers for start (blue) and stop (red) codons in the accordant frame.



Open reading frames

The yellow bar in the canvas represents the contig sequence. There is a right-click menu associated with this bar. Then there are six levels, which represent the different frames of the contig. On each level the ORFs are shown in green, start codons in blue and stop codons in red.

Clicking somewhere in the canvas will move the red line to that position; it will also mark the associated base in the 'Contig sequence'. Clicking on an ORF will mark it in a dark green, the sequence of the ORF will appear below the 'Contig sequence' and in the 'Selected sequence'-panel.



Selected sequence

Displays the selected nucleotide or protein sequence.

Right-click menu

The two right-click menus of 'Edit sequence' contain mainly special function, but also some of the standard right-click menu.

Add to blast
Coverage
Delete
Insert
New sequence
Refresh
Replace with
Search
Show complementary strand
Translate whole frame

Add to blast

See 'Add to blast' in right-click menu.

Coverage

See 'Coverage' in right-click menu.

Delete

Deletes the selected base from the contig sequence. This is no permanent chance to the contig.

Insert

Insert a new base at the selected position of the contig sequence. This is no permanent chance to the contig.

New sequence

See 'New sequence' in right-click menu.

Refresh

Redraws the 'Open reading frames'-canvas, e. g. to fit a new window size.

Replace with

Replaces the selected base in the contig sequence. This is no permanent chance to the contig.

Search

Opens a search dialog, which allows searching for patterns in a reference string. The search results will be highlighted.

This option is available for the 'Contig sequence'-part and the 'Selected sequence'-panel.

Show complementary strand

Displays the complementary strand of the contig sequence in the 'Selected sequence'-panel.

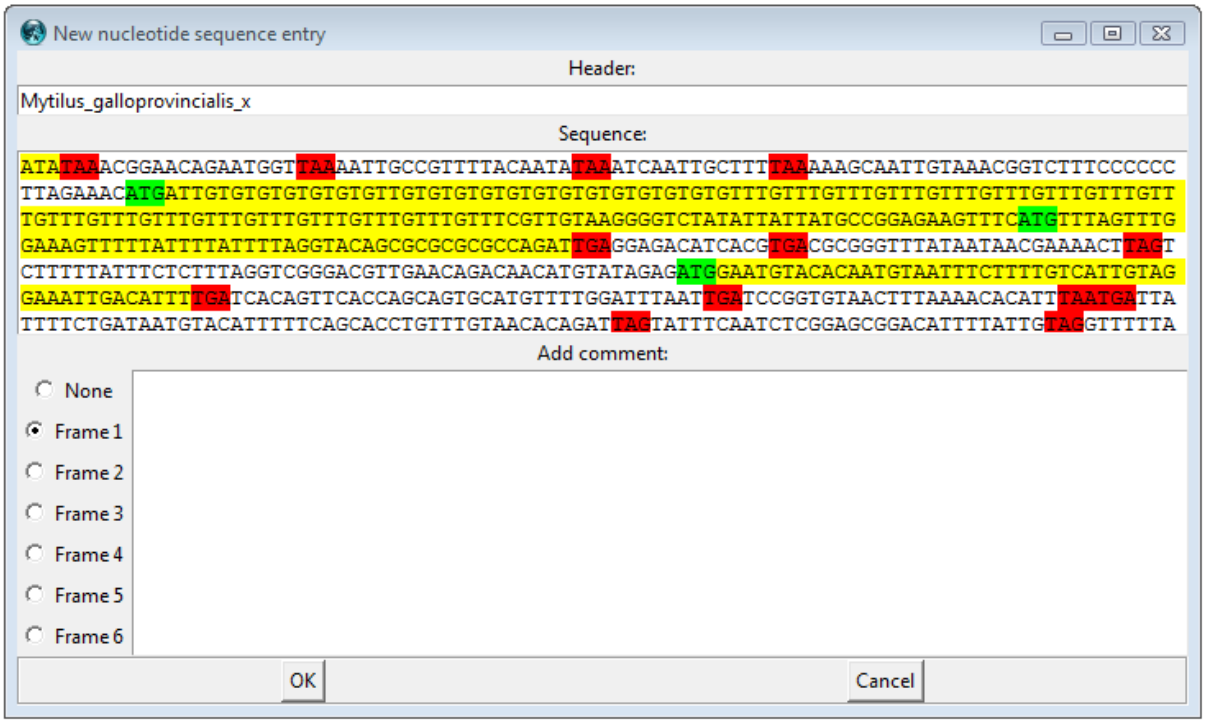
Translate whole frame

Translates the whole frame from nucleotides into amino acids. Stop condons will appear as '*'.

New sequence

With this a new contig sequence can be added to the database. The 'Header' entry field sets the name of the new sequence, the 'Sequence'-panel hold the new sequence and the 'Comment'-panel allows to directly add a comment to the new sequence entry.

The radiobuttons in the lower left corner of the window will highlight the ORFs of the selected frame in yellow, start codons in green and stop codons in red.

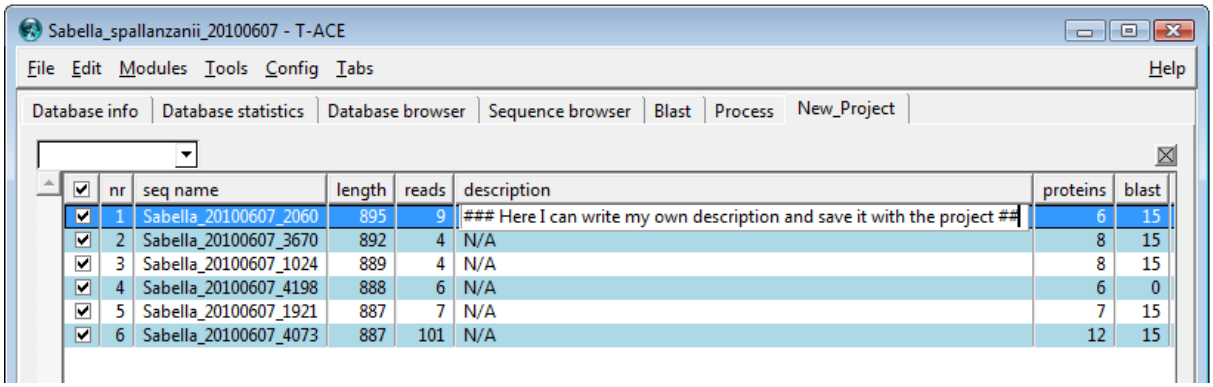


Project tabs

Project tabs consist mainly of a table which look like the 'Database Browser'-table, they share also most of its right-click menu functions. The difference is that the content of a project tab can be altered.

Through the 'File'-menu project tabs can be created, loaded or saved. With the 'Add sequences'-option of the 'Edit'-menu selected sequences from the 'Database Browser', 'Blast'-tab or another project tab can be copied into a project tab.

It is possible to alter the text of the description column of a project tab table just by clicking on a description cell.



Right-click menus

Most of the tables and some canvas objects in T-ACE have an attached right-click menu. Each menu can vary in its contained functions.

Add to blast
Copy cell
Copy column
Copy row
Copy sequence
Copy table
Coverage
Create primer
Edit sequence
Export read distribution
Export reads
Hide/Show column
New sequence
Save config
Search
Search repeats
Show all hits

Add to blast

'Add to blast' adds the sequence of the selected database entry into the query panel of the 'Blast'-tab.

Copy cell

Copies the text of the selected table cell to the clipboard.

Copy column

Copies the text of all rows of the selected column to the clipboard.

Copy row

Copies the text of the selected row to the clipboard.

Copy sequence

This copies the corresponding sequence of the selected contig to the clipboard, in FASTA format.

Copy table

Copies the text of the whole table to the clipboard.

Coverage

Opens the 'Coverage'-window of the selected contig. This option is available from the 'Database browser' and project tabs.

Create primer

It opens the 'Primer3'-window. The primer search will be performed only for the selected sequence. This option is available from the 'Database browser' and project tabs.

Edit sequence

This function opens the 'Edit sequence'-window for the selected contig. It is available from the 'Database browser', project tabs, 'Blast'-tab and contig-objects of a 'Mapping'-window.

Export read distribution

This function is available from the 'Database info'-tab or a project-tab. it exports a table, which contains a contig list, that displays the read distribution, over all runs in the database, for each contig. If 'Export read distribution' is executed from the 'Database info'-tab it creates a list for all contigs of the database. When called from a project tab only the selected contigs will be exported.

Export reads

It exports the reads of the selected contig into a FASTA file. This option is available from the 'Database browser', project tabs and the 'Mapping'-tab.

Hide/Show column

Depending on the current column state it hides the selected column from view or makes the column visible.

New sequence

This function opens the 'New sequence'-window for the selected contig. It is available from the 'Database browser', project tabs, 'Blast'-tab and contig-objects of a 'Mapping'-window.

Save config

Saves the current configuration of the table. Available for the 'Primer'-tab and 'Repeat'-tab.

Search

Opens a search dialog, which allows searching for patterns in a reference string. The search results will be highlighted in green.
This option is available for the 'Sequence browser', 'Sequence viewer' and 'Edit sequence'.

Search repeats

It opens the 'Search repeats'-window. It will search for repeats only in the selected sequence. This option is available from the 'Database browser' and project tabs.

Show all hits

Opens a project tab, which contains all contigs associated with the selected hit.

Creating the first database

When the postgresql server is running, the parent database and user accounts created and you have access to the server, you can start to create your first database.
To do this, just select the option 'New DB…' in the 'File'-menu of T-ACE. A 'New database'-window appears where you can set the name of the new database. Click the 'OK'-button and a new, empty database with the given name will be created.
The best way to fill a database is to use an ACE file. The ACE file contains all information about the contigs and the positioning of their corresponding reads. To add an ACE file to your database open the 'Edit DB…'-cascade of the 'Edit'-menu, then 'Import sequences…' and select the 'Add ACE file'-option. A file dialog opens and you can select your ACE file. Now all contigs and reads in the ACE file will be written into the database. The name of every database contig obtained from the ACE file will be composed of the database name and the contig ID from the ACE file. Loading the ACE file into the database takes a lot of time, to see if the upload has finished you can look at the 'Process'-tab of T-ACE.
A normal fasta file, containing nucleotide sequences, can also serve as input, but the added contigs would not have any read and coverage information available.
Now the run information should be added. This happens through the 'Add run'-button in the 'Database info'-tab. Just set a 'Run ID' and add the original read fasta file, all other fields can be modified later.
At last some updates should be run. For this open the 'Edit DB…'-cascade of the 'Edit'-menu and go to the 'Update'-cascade. From here you should execute 'coverage', 'database info', 'read distribution' and 'read status'. The other update options are only of interest after adding some annotation data.

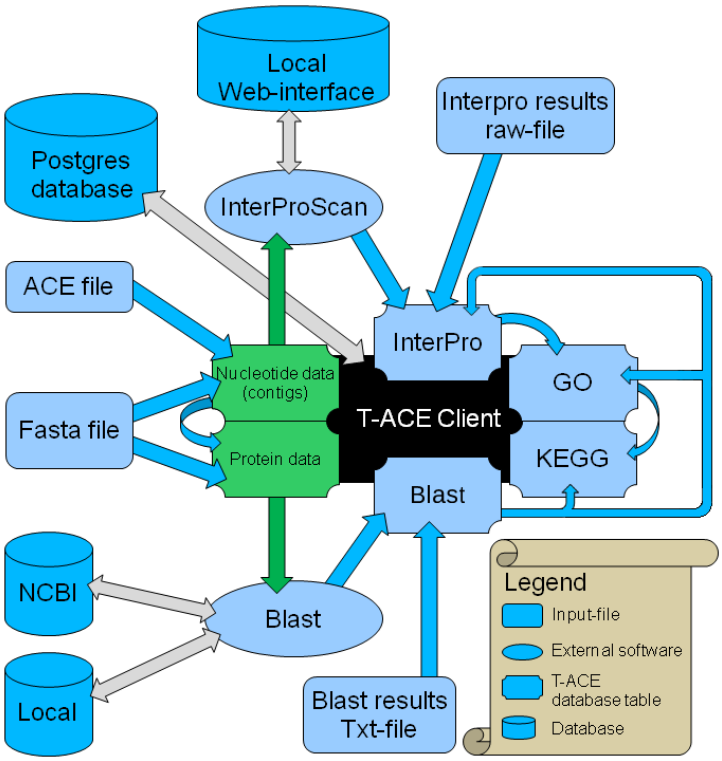
Annotating a database

After the contigs where loaded into the database the annotation can begin.

At first the blast annotation should be performed. There are two possible ways for blast annotation.
The first one is to perform a local blast search. For this NCBI-blast-2.2.25+ has to be installed and the desired blast database has to be downloaded (for example: uniprot_sprot or a NCBI database). The command line blast should be performed with the parameter ***–outfmt 6***, the resulting blast output file can be uploaded into the T-ACE database with the 'blast results'-option of the 'Import annotations'-cascadein the 'Edit'-menu. The contigs or proteins for the blast input can be exported from the database with the export options of the 'Database info'-tab.
The second possibility is to use the 'NCBI-Blast'-module of T-ACE. This module enables the user to blast just specific or all contigs of the T-ACE database against a NCBI database.
The blast annotation can also be used to obtain some GO, InterPro and KEGG associations. If this is possible depends on the references which were inserted into the 'refDB'-schema of the 'parent database'. To upload references to the 'refDB'-schema the T-ACE_DB_manager.tcl can be used, for

more information look here. If there are such references available, the functions 'Update blast2interpro entries' and 'Update KEGG-Maps' in the 'Edit DB'-menu will add the according associations into the hit tables of the T-ACE database.

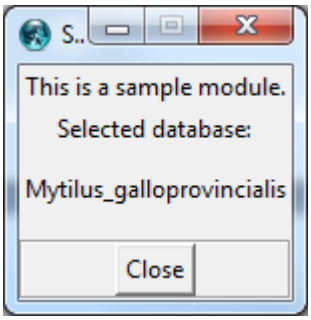
The second annotation step is the InterProScan. For this a local installation of InterProScan has to be available. It can be used via command line, like the NCBI-blast+, with contig or protein sequences as input. The output should be in .raw format and can be uploaded with the 'Add InterProScan results'-option of the 'Edit'-menu. If contig sequences were used as InterProScan input the .input result file of InterProScan should also be uploaded into the database. The .input file contains the contig ORFs which were used for the InterPro scan.



If there is a web-interface available for the local installation of InterProScan the 'InterProScan'-function, in the 'Tools'-menu of T-ACE, can also be used for this annotation.

Writing a T-ACE module

There are already some TCL scripts in the T-ACE modules folder, which provide additional functions. Some of them can serve as examples to create your own T-ACE module. The simplest one is the sample_module_tace.tcl; this module does nothing but opening a dialog which says “This is a sample module” and displaying the name of the selected database.



To add this sample module to T-ACE open the 'Add/remove...'-option in the T-ACE 'Modules'-menu and check the box for the 'sample_module_tace' script, press 'OK' and the module will be loaded. Now the 'Modules'-menu should contain the option 'Sample Module'.

There are some important points for adding a custom made TCL script into T-ACE:

1. The script has to be located in the modules folder of T-ACE (it will automatically appear in the 'Add/remove modules'-window of T-ACE).
2. The script must contain an 'init' and a 'remove' procedure. The procedure names must begin with the script name, for example: The script *myModule.tcl* must contain the procedures *myModule_init* and *myModule_remove*.
3. Menu buttons which link to the new module must be added to the init procedure of the new script.

Database access

Depending on the T-ACE version that is used, the new module has to access the database in different ways.

If T-ACEpg is used the module requires pgtcl to access the database. The normal pgtcl commands are used, like:
pg_connect, pg_exec and pg_select

When using the normal T-ACE version the database is accessed through a php script to do this the module should use the T-ACE methods:
php_checkConnection, php_getSQLresults and php_getDBupload

To get code examples for accessing the database just, look into the files in the modules folder of T-ACE, like:
NCBI_blast_tace.tcl, phobos_tace.tcl or run_compare_tace.tcl

The php methods

php_checkConnection

This method just checks if a connection to the database is possible. It is used for the login procedure.

php_getSQLresults {sql}

This is most used method; it sends an SQL request to the database and returns the results.

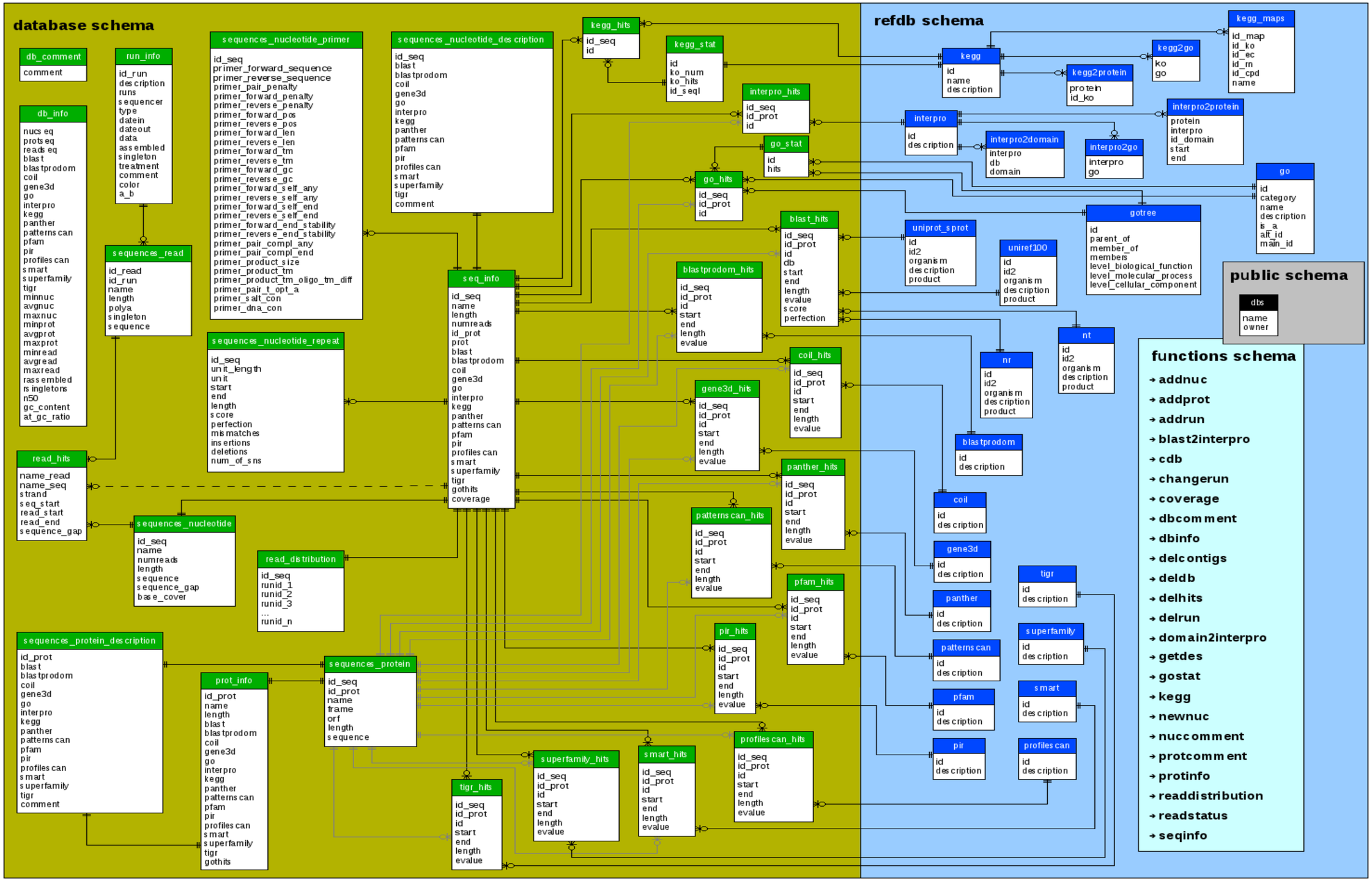
php_getDBupload {table input}

This method is used to write big data sets into the database. It needs as input two parameters:

table: The name of the table of the database into which the data shall be written.

input: The data that shall be written into the table. The data has to be in a tab-delimited, multi row string format.

T-ACE parentDB schema



T-ACE parentDB functions

addnuc

This function writes a nucleotide sequence as a new contig entry into the 'sequences_nucleotide'-table of a T-ACE database and creates the according entry in the 'seq_info'-table.
It is called from 'New nucleotide sequence' and 'Add nucleotide file'.

addprot

This function writes a protein sequence as a new entry into the 'sequences_protein'-table of a T-ACE database and creates the according entry in the 'prot_info'-table.
It is called from 'Translate contigs to proteins' and 'Add protein file'.

addrun

This function adds a new entry into the 'run_info'-table of a T-ACE database. It is called from 'Add Run'.

blast2interpro

This function tries to assign new interpro and GO annotation to contig sequences by comparing existing blast hits from the 'blast_hits'-table of a T-ACE database with the entries of the 'interpro2protein'-table of the refDB schema. The found interpro IDs are checked against the 'interpro2go'-table of the refDB schema and the results (interpro and GO IDs) are written into the 'interpro_hits'- and 'go_hits'-tables of the accordant T-ACE database.
The function is called by the 'blast2interpro entries' update option.

cdb

'cdb' stands for 'create database'. It creates a new empty T-ACE database and is called from 'New DB'.

changerun

This function updates the entries of the 'run_info'-table of a T-ACE database. It is called from the 'Save changes'-button in the 'Database info'-tab.

coverage

This function calculates the coverage for each contig in a T-ACE database and updates the values of the coverage column in the 'seq_info'-table of the T-ACE database. The 'base_cover'-column entry of the 'sequences_nucleotide'-table is used for the calculation. The coverages of base of the contig sequence is added up and divided by the contig length.
The function is called by the 'coverage' update option.

dbcomment

It updates the entry of the 'db_comment'-table in a T-ACE database. It is called from 'Change comment'-option of the 'Edit'-menu.

dbinfo

This function calculates the values seen in the panel of the 'Database info'-tab and updates the entry in the 'db_info'-table of a T-ACE database. It is called by the 'database info' update function, but also after various other processes, e.g. after adding nucleotide and protein sequences or blast and domain entries.

delcontigs

It deletes the given contig entries and all accordant annotations from a T-ACE database. It is called from 'Delete selected contigs' in the 'Edit'-menu.

deldb

It drops a complete T-ACE database schema. It called from 'Delete selected DB' in the 'File'-menu.

delhits

It deletes the given annotation entries from the accordant '*_hits'-tables of a T-ACE database. It is called from 'Delete selected hits' in the 'Edit'-menu.

delrun

This function deletes entries from the 'run_info'-table of a T-ACE database. It is called by the 'Delete selected'-button of the 'Database info'-tab.

domain2interpro

This function tries to assign new interpro and GO annotation to contig sequences by comparing existing domain hits from different '*_hits'-tables of a T-ACE database with the entries of the 'interpro2domain'-table of the refDB schema. The found interpro IDs are checked against the 'interpro2go'-table of the refDB schema and the results (interpro and GO IDs) are written into the 'interpro_hits'- and 'go_hits'-table of the accordant T-ACE database.
The function is called by the 'domain2interpro entries' update option.

getdes

This function updates or generates entries of the 'sequences_nucleotide_description'- or 'sequences_protein_description'-table, depending from where it is called. To do this the function searches for the best (lowest evalue) annotation entry in each '*_hits'-table for each contig and gets the description for each of the entries from the accordant reference table in the refDB schema.
The function can be called from 'sequence descriptions' or 'protein descriptions' in the 'Edit'-menu.

gostat

It deletes all entries of the 'go_stat'-table of a T-ACE database and fills the table anew. To create the new entries, an array of all GO IDs in the 'gotree'-table of the refDB schema is generated. After that each entry of the 'go_hits'-table in the T-ACE database is compared to the 'member_of'-column in the 'gotree'-table of the refDB schema. After removing duplicate hits from the results of all 'go_hits' entries of the same contig (id_seq), the remaining results are used to increase the accordant GO ID values in the array. Then the content of the array is written to the 'go_stat'-table of a T-ACE database.
The function is called by the 'GO-Statistic'-option in the 'Edit'-menu.

kegg

This function deletes all current entries from the 'kegg_hits'-table of a T-ACE database and fills it anew. This is done by comparing existing blast and GO hits from the 'blast_hits'- and 'go_hits'-table of a T-ACE database with the entries of the 'kegg2protein'- and 'kegg2go'-table of the refDB schema. The found KEGG IDs are written into the 'kegg_hits' of the accordant T-ACE database.
After that the 'kegg_stat'-table is updated by comparing the 'kegg_hits'-table with the 'kegg_maps'-table of the refDB schema.
The function is called by the 'KEGG-Maps'-option in the 'Edit'-menu.

newnuc

It adds a new contig sequence entry to the 'sequences_nucleotide'-table of a T-ACE database and creates the according entries in the 'seq_info'- and 'sequences_nucleotide_description'-table.
The function is called by the 'OK'-button of the 'New sequence'-function.

nuccomment

This function updates the 'comment'-column entry of the 'sequences_nucleotide_description'-table for the given 'id_seq'. It is called through the 'Save'-button of the 'Comment'-tab.

protcomment

This function updates the 'comment'-column entry of the 'sequences_protein_description'-table for the given 'id_prot'. It is called through the 'Save'-button of the 'Comment'-tab.

protinfo

This function updates or generates the contents the 'prot_info'-table of a T-ACE database. To do this it counts the number of annotation entries for each 'prot_info'-table entry in the various '*_hits'-tables and updates the appropriate column of the 'prot_info'-table.
It is called by the 'protein entries'-option in the 'Edit'-menu.

readdistribution

At first the function empties the 'read_distribution'-table of a T-ACE database, then gets a list of all run IDs of the 'run_info'-table and adds each run ID as a new column in the 'read_distribution'-table, if the accordant column does not already exist. After that the function gets the number of reads, contained in a contig, for each run for each contig (id_seq). This is done by combining the information of the 'sequences_read'- and 'read_hits'-table. The resulting array is written to the 'read_distribution'-table of the T-ACE database.
The function is called from the 'read distribution'-option in the 'Edit'-menu.

readstatus

It updates the 'singleton'-column status for all entries in the 'sequences_read'-table of a T-ACE database. This happens by looking at the 'read_hits'-table, all reads with an entry in this table are contained in a contig sequence. At first the 'singleton'-column is set to TRUE for all entries in the 'sequences_read'-table. Then each entry in the 'read_hits'-table will cause an entry of the 'sequences_read'-table to switch the status of its 'singleton'-column to FALSE.
After that the number of assembled (singleton status FALSE) and singleton reads for each run in the 'run_info'-table is counted and the 'assembled'- and 'singleton'-columns in the 'run_info'-table are updated.
The function is called by the 'read status'-option in the 'Edit'-menu.

seqinfo

This function updates or generates the contents the 'seq_info'-table of a T-ACE database. To do this it counts the number of annotation entries for each 'seq_info'-table entry in the various '*_hits'-tables and updates the appropriate column of the 'seq_info'-table.
It is called by the 'sequence entries'-option in the 'Edit'-menu.