STAMP – Manual

The Contents:

Phobos is a program for finding and marking microsatellites in DNA-sequences. It is possible to search for different kinds and different qualities of a “Simple Sequence Repeat” (SSR).
With **Primer3** you can search for possible primer pairs, which flank the marked repeats or other kinds of targets you give to Primer3. It’s possible to set many parameters in Primer3 to specify the qualities of the received primer pairs.
These modules give you a fast and easy way to find SSRs in DNA-sequences and isolate them through creation of flanking primer pairs.
After the creation of possible primer pairs through Primer3 it is possible to align these primers against each other and the Gap4 database (for is BLAST is used) and select primers for a multiplex PCR experiment.
All the repeat, primer and multiplex information created can be stored in a **SQLite** database.

The Modules allow a direct access to Phobos and Primer3 through a graphical interface in Gap4.
If you need some information about Gap4, you should use the following webpage:
http://staden.sourceforge.net/manual/gap4_unix_toc.html

Installation:

To install the project you only have to execute the file „STAMP.zip“. You have to choose your „Staden Package“-folder for the installation, otherwise the installation won’t work.

CAUTION: In the process of installation you will be ask to overwrite the “gaprc”- and “GTAGDB”-file. If you use an already modified version of Gap4, it is possible that the modifications would not work after this installation. To avoid this problem you should NOT overwrite the “gaprc”-file and append the following line to the end of the file by yourself:

```
load_package %L/STAMP        {}        STAMP 0
```

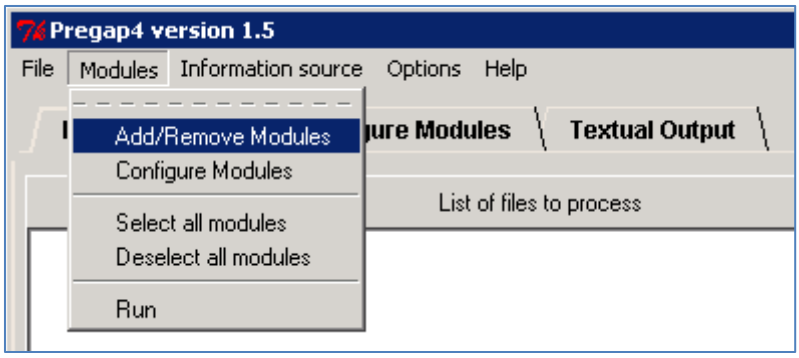
The “GTAGDB”-file includes tag-definitions, the file will be overwritten to create parameters for the tags “REP1” to “REP9”, “PRI1” to “PRI9” and “MPL1” to “MPL9”. If you created some tags for yourself, they will be removed after the file replacement.
“gaprc” and “GTAGDB” can be found in the “tables”-folder of the Staden Package.

1. Pregap4

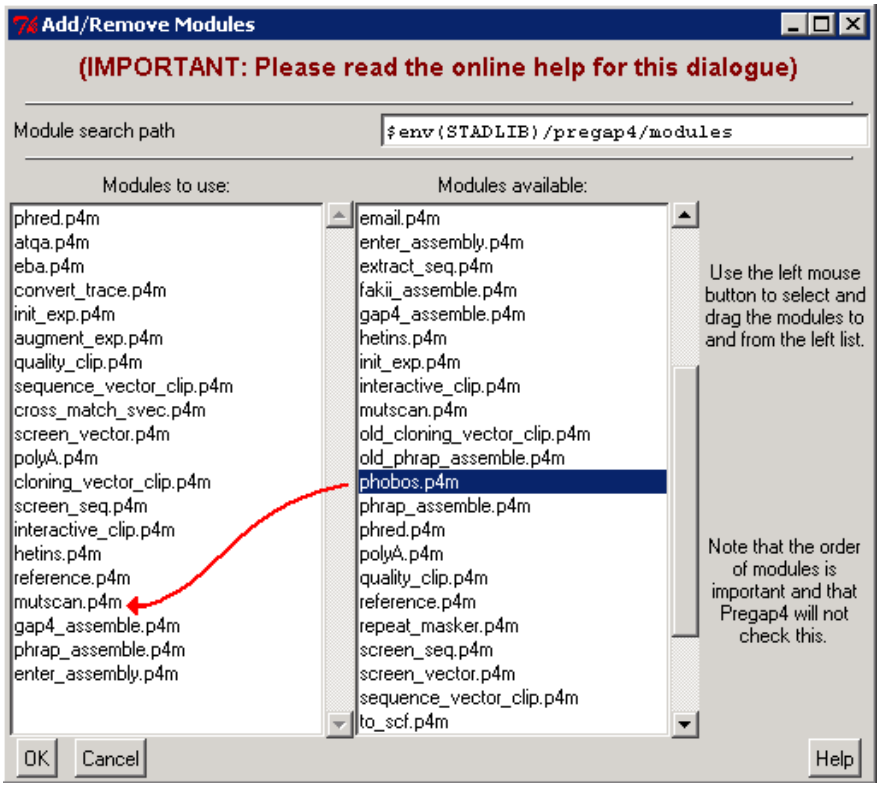
1.1. The Phobos Module:

1.1.1. Adding the module:

The Phobos module can be added to the module list of Pregap4 through the “Modules”-menu and the option „Add/Remove Modules“.



The Phobos-Module should be placed above any assembly tool. To do so it has to be dragged from the right list to its designated place in the left list.



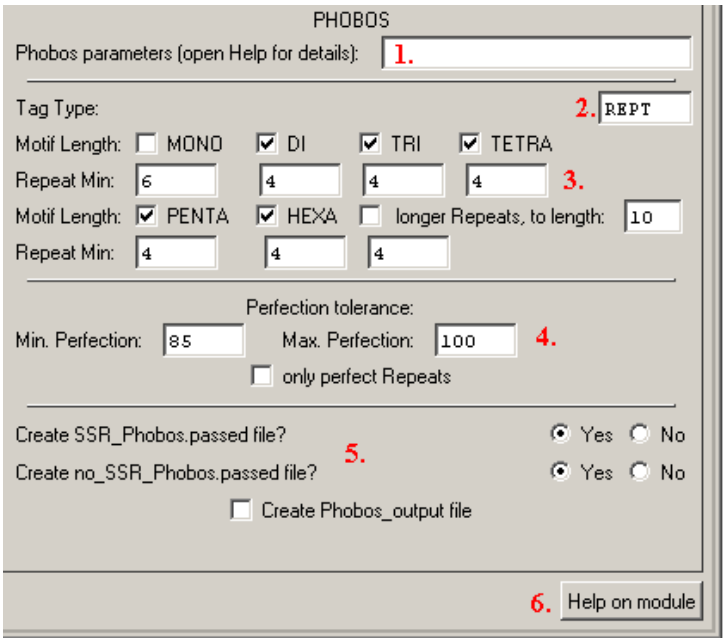
1.1.2. Using the module:

Phobos needs experiment files to work on, these files will be created trough other Pregap4 modules, or they can be loaded directly into the „Files to process“-list of Pregap4.

To use the module it only has to be activated, further there are no parameters which have to be set.

1.1.3. The Interface:

The image below shows the interface of the Phobos Module in Pregap4. The different components of the interface are marked with red numbers in this image. The function of these components will be explained in the following.



1. Phobos Parameters

The text field “Phobos parameters” allows the setting of parameters for the repeat search, a list of the available parameters and their function can be found under “3.3. Phobos Parameters” or the Phobos-Help page (button “Help on Module”).

2. Tag Type

The text field “Tag Type” sets the tag name for the annotation in the experiment files.

3. Parameters for Table generation

With the checkboxes “MONO” to “HEXA” and “longer Repeats” the motif types of SSRs, which should be searched for, can be selected. The text fields under the checkboxes allow a restriction of the minimal length of a SSR which are shown in the list. Also the minimal and maximal perfection for repeats in the list can be set.

4. Perfection tolerance

Here the minimal and maximal allowed perfection of a SSR can be set.

5. Phobos output and passed files

The „SSR_Phobos.passed“-file contains a list of contigs which include repeats. The „no_SSR_Phobos.passed“-file contains a list of contigs which does not include repeats. By loading the „no_SSR_Phobos.passed“-file in Gap4 a more specific SSR search could be run, which will be concentrated on contigs without obvious microsatellites.

If the checkbox “Create Phobos_output file” is selected a text file with the Phobos results will be created.

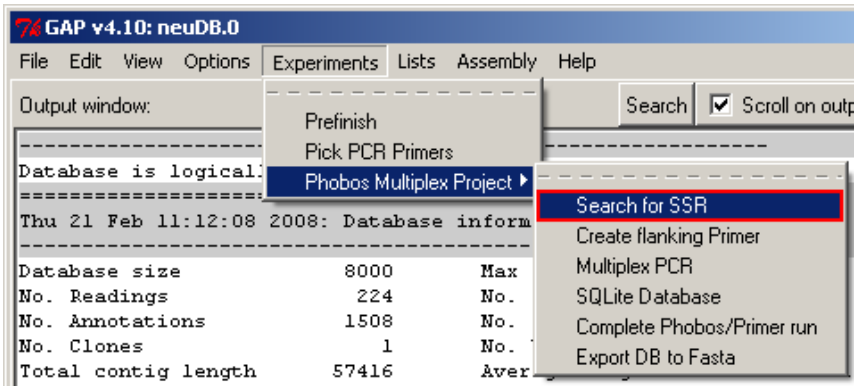
6. Help on Module

The Button “Help on Module” opens a HTML-page with helpful information for the usage of this module.

2. Gap4

2.1. The Phobos Module:

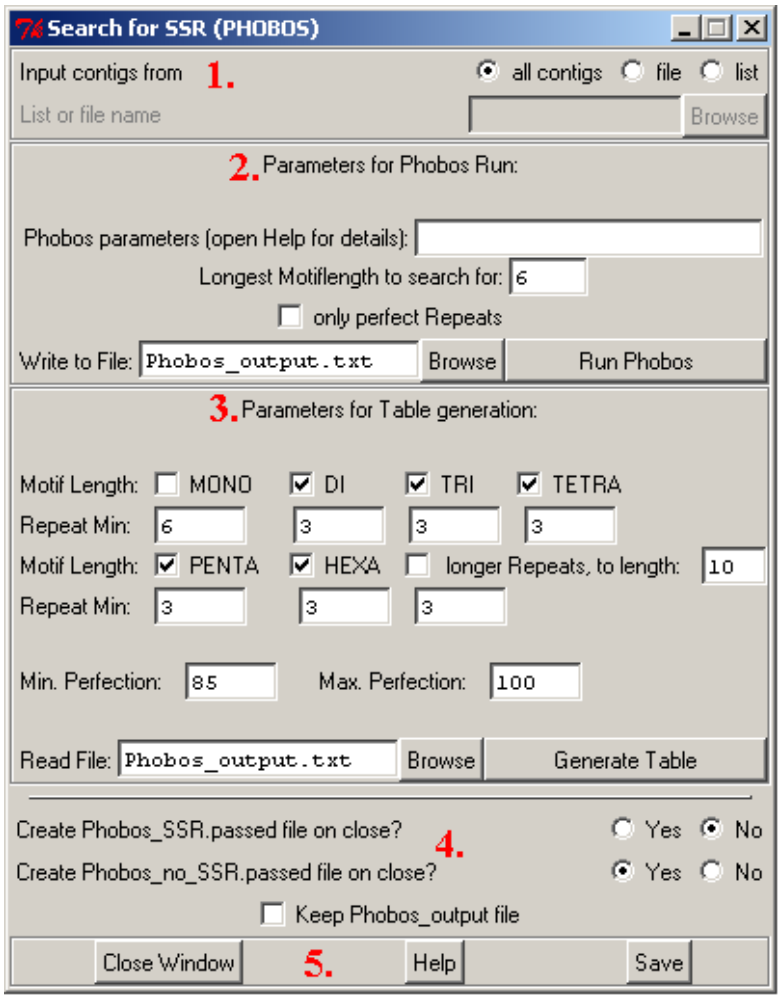
The Phobos Module searches for simple sequence repeats (SSR) in the sequences of the Gap4 database. These repeats could be annotated regions in the Gap4 database or stored in a table of a SQLite database (see 4.)



The Phobos Module can be found under “STAMP”-option of the “Experiments”-menu. It is titled “Search for SSR”.

2.1.1. The Interface:

The image below shows the interface of the Phobos Module. The different components of the interface are marked with red numbers in this image. The function of these components will be explained in the following.



1. The contig input

The sequences from the Gap4 database for the repeat search can here be selected. The standard selection is “all contigs”, this means the whole database.

Under “file” a file could be loaded which has to contain a list of contigs from the database.

After clicking the “Browse”-button when the option “list” is selected a window with a list of different options will open. It can be chosen between “all contigs”, “all readings”, “contigs” and “readings”.

2. Parameters for Phobos Run

The text field “Phobos parameters” allows the setting of parameters for the repeat search, a list of the available parameters and their function can be found under “3.3. Phobos Parameters” or the Phobos-Help page (button “Help”).

“Longest Motiflength to search for” refers to the maximal length of a repeat motif for which Phobos should search. With increasing motif length the time for the repeat search will also increase.

The text field “Write to File” sets the filename for the Phobos output. Press “Run Phobos” to start the microsatellite search.

3. Parameters for Table generation

The button “Generate Table” will create a list of all repeat which are in the Phobos output file and match the parameters set for the list content. With the checkboxes “MONO” to “HEXA” and “longer Repeats” the motif types of SSRs, which should be shown in the table, can be selected. The option “longer Repeats” only makes sense if the option “Longest Motiflength to search for” under *Point 2* was set higher then 6.

The text fields under the checkboxes allow a restriction of the minimal length of a SSR which are shown in the list. Also the minimal and maximal perfection for repeats in the list can be set.

Through the “Generate Table”-button any number of tables with different content can be created.

4. Phobos output and passed files

The „Phobos_SSR.passed“-file contains a list of contigs which include repeats. The „Phobos_no_SSR.passed“-file contains a list of contigs which does not include repeats. By loading the „Phobos_no_SSR.passed“-file (see *Point 1*) a more specific SSR search could be run, which will be concentrated on contigs without obvious microsatellites.

5. The Buttons

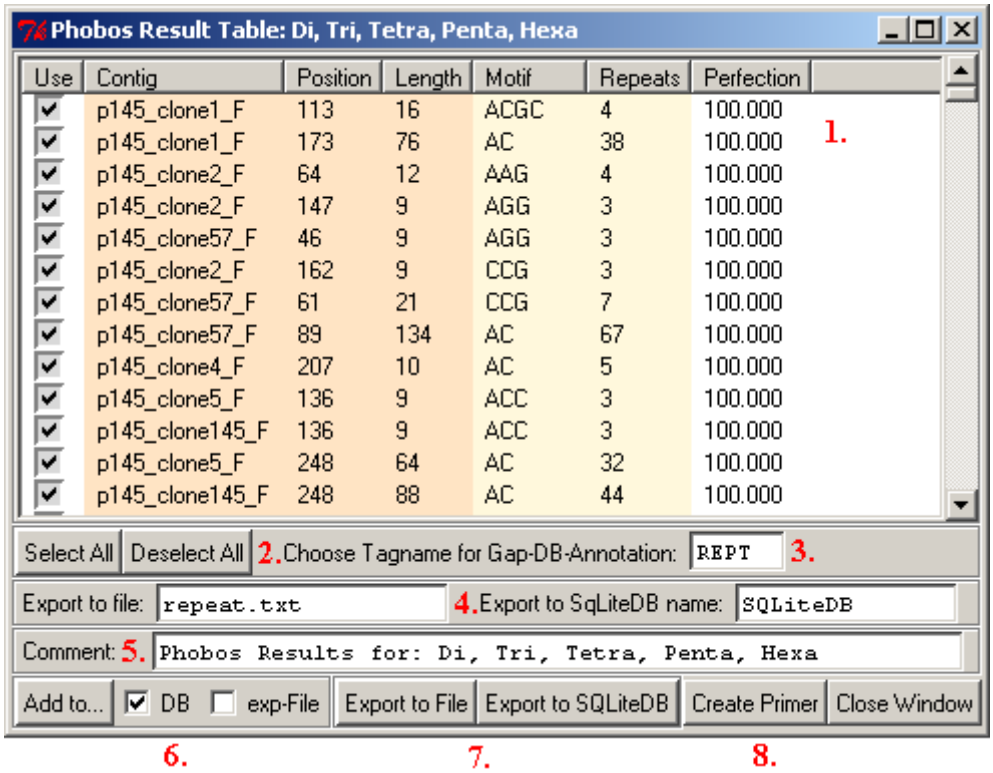
The “Close Window”-button will close the window and delete the Phobos_output file unless the “Keep Phobos_output file”-option under point 4 is selected.

The “Help”-button will open a HTML-page with information for handling the interface.

The “Safe”-button stores the interface parameters in the file “phobos_par.txt”, this file can be found in “{STADENROOT}\lib\stamp\”.

2.1.2. The Repeat Table:

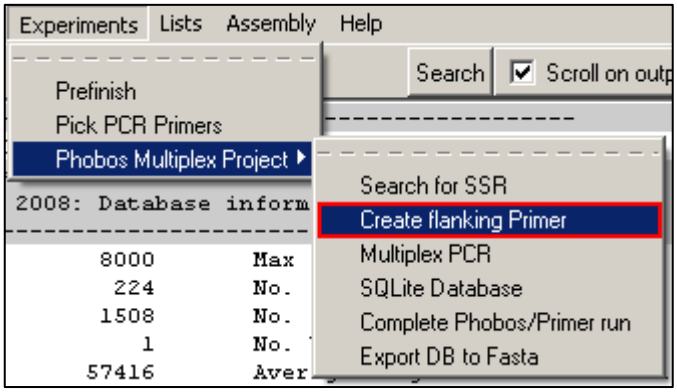
The Repeat Table allows a manual selection of repeats for export to a tab-delimited text file or to the Gap4/SQLite database. The head of the table always shows which kinds of SSRs are contained in the list.



1. A table of the repeats, extracted from the Phobos output file. A double click on a row opens a contig editor (See 3.1.The Contig Editor).
2. The buttons to select or deselect all repeats in the list.
3. The tag-name for the Gap4 database can be set here.
4. Here the names of the export file, SQLite database are set.
5. The comment set in this text field will be used for identification in the SQLite database.
6. The “Add to…”-button write all selected rows to either the Gap4 database or the experiment files, this depends on status of the checkboxes “DB” and “exp-File”.
7. Exports all selected rows to a text file (“Export to File”) or to a SQLite database (“Export to SQLite Database”).
8. The button “Create Primertable” opens the interface of the Primer3 Module and uses all selected rows in the repeat table as targets. After a run the Primer3 interface will not close. It always keeps the same target information. After changing the repeat selection it is necessary to open a new Primer3 interface by clicking the “Create Primertable”-button again.

2.2. The Primer3 Module:

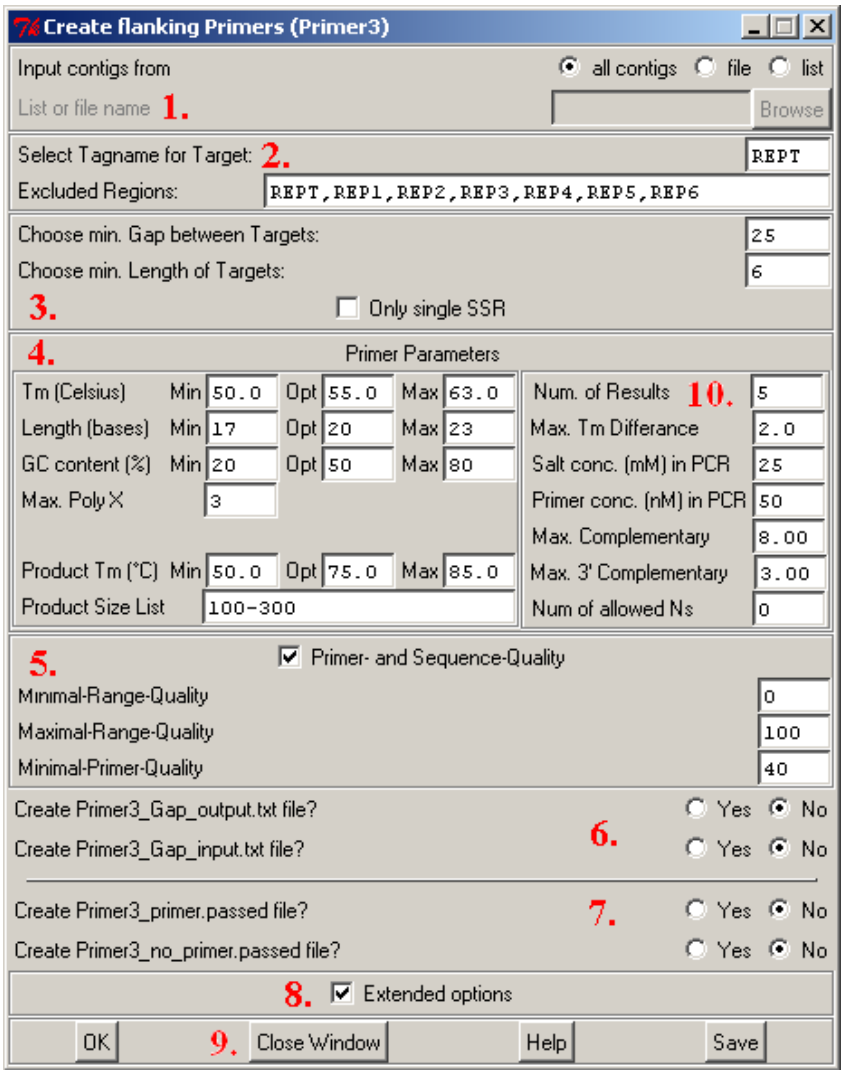
The Primer3 Module allows the selection of primer candidates for targets sequences, for example microsatellites. These targets could be annotated regions in the Gap4 database or stored in a table of a SQLite database (see 4.)



Like the other modules, the Primer3 Module can be found under “STAMP” in the “Experiments”-menu. It is titled “Create flanking Primer”.

2.2.1. The Interface:

The image below shows the interface of the Primer3 Module. The different components of the interface are marked with red numbers in this image. The function of these components will be explained in the following.



1. The contig input

The sequences from the Gap4 database for the primer search can here be selected. The standard selection is “all contigs”, this means the whole database.

Under “file” a file could be loaded which has to contain a list of contigs from the database.

After clicking the “Browse”-button when the option “list” is selected a window with a list of different options will open. It can be chosen between “all contigs”, “all readings”, “contigs” and “readings”.

2. Target- and excluded-regions

“Select Tagname for Target” sets a tag name which refers to the targets regions. The “Excluded Regions” text field should contain all tag names of marked regions were no primers should be placed, for example all repeat tag names (these are already set there as standard).

3. Setting of target parameters

The option “Choose min. Gap between Targets” has a default value of 25, this stands for the distance in bases which shall be between two targets. If two targets have less then this 25 bases between each other they will be counted as one. The default minimum length of a target is 6, only targets as long as the min. length or longer will be used for primer search.

The “Checkbox” with the option “Only single SSR” allow you only to search for targets of a single SSR-motif. (For example: if an AC-SSR lies next to AT-SSR it will be handled like one SSR and one target, if the “Only single SSR”-option is not selected. If the option is selected these SSRs won’t be counted as targets, because there is no room for a primer to be placed between them and the option will not allow two repeats in one primer product)

4. The primer parameters

Here some parameters can be set: the length of the primer in bases, the GC content, the melting temperature of the primer (“TM”) and the melting temperature of the product (“Product TM”). The field “Max. Poly X” sets the number of times the same base can appear in a row in a primer. “Product Size List” allows to different size ranges for the products, example: “100-150 200-300”.

5. Primer- and Sequence-Quality

If the “Primer- and Sequence-Quality”-checkbox is selected only sequences with quality values in between the “Minimal-Range-Quality” and the “Maximal-Range-Quality” are used for the primer search. Also every base of a primer location needs a minimal quality value like the one set at “Minimal-Primer-Quality”, otherwise the primer will not be created at this sequence position.

6. Primer3 input and output files

It can be chosen if the primer3 input and output text files should be kept or not, the default option is “No”.

7. Primer passed files

The „Primer3_primer.passed“-file contains a list of contigs which include primer for targets. The „Primer3_no_primer.passed“-file contains a list of contigs which does not include primers. By loading the „Primer3_no_primer.passed“-file (see *Point 1*) a more specific primer search could be run, which will be concentrated on contigs without useful primer locations.

8. Extended Options

The checkbox displays the second parameters box under point 10.

9. The Buttons

By clicking the “OK”-button Primer3 will be executed and a window opens which contains a table with possible primer candidates.

The “Help”-button opens a HTML-page with the information needed to handle the Primer3 interface.

The “Safe”-button stores the interface parameters in the file “primer3_par.txt”, it can be found in “{STADENROOT}\lib\stamp\”.

10. Extended primer parameters

"Num. of Results" set the number of returned primer pairs. "Max. Tm Difference" is the maximal allowed difference of the forward and reverse primer Tm.

"Salt conc. (mM) in PCR" refers to the salt concentration (milli molar) of the PCR-mix and influences the calculation of the melting temperature.

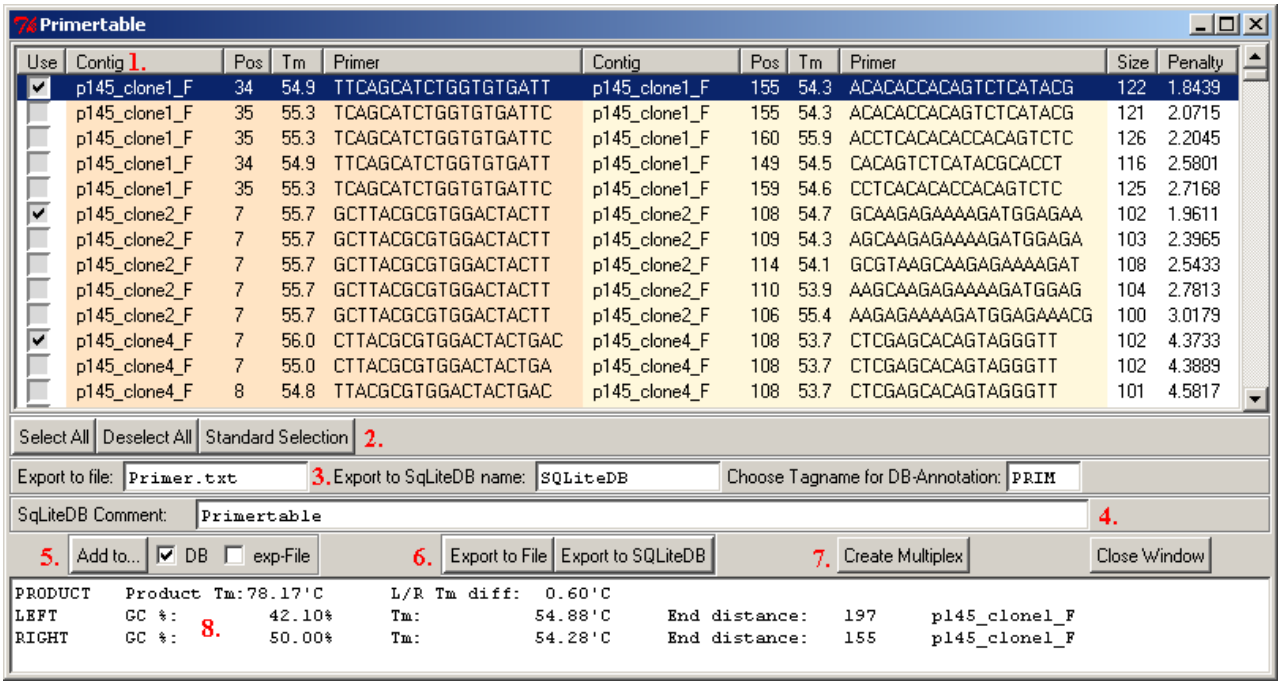
"Primer conc. (nM) in PCR" refers to the concentration of primers (nano molar) in the PCR-mix.

"Max. Complementary" sets the maximal allowed alignment score for self alignment. "Max. 3' Complementary" the same for 3'end.

"Num of allowed Ns" stands for the number of unrecognized bases in a primer, like "N", "-" or "*".

2.2.2. The Primer table:

The Primer Table allows a manual selection of primers for export to a tab-delimited text file or to the Gap4/SQLite database.

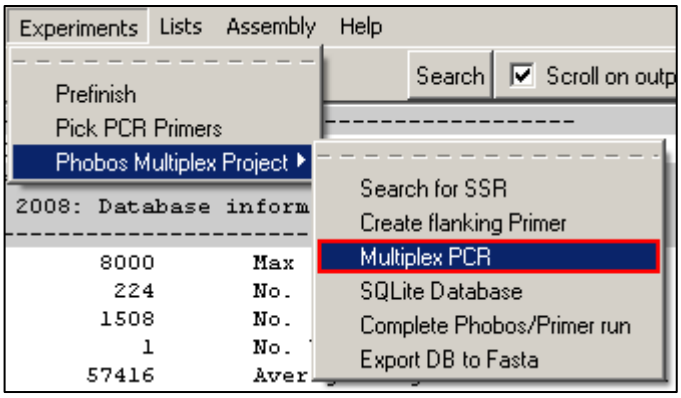


1. A table of the suggested primers, a double click on a row opens a contig editor (See 3.2.The Contig Editor).
2. The buttons to select or deselect all primers in the list, with the button “Standard Selection” the originally selection will be recreated.
3. Here the names of the export file, SQLite database and Gap4 database annotation are set.
4. The comment set in this text field will be used for identification in the SQLite database.
5. The “Add to…”-button write all selected rows to either the Gap4 database or the experiment files, this depends on status of the checkboxes “DB” and “exp-File”.
6. Exports all selected rows to a text file (“Export to File”) or to a SQLite database (“Export to SQLite Database”).
7. The button “Create Multiplex” opens the interface of the Multiplex Module and uses all selected rows in the primer table for the creation of the multiplex. A second use of the “Create Multiplex”-option will destroy the multiplex table window and create a new one, so the multiplex information from the first run should be stored before starting a second run.
8. This text field shows additional primer information like the GC-content of the primer pair and the melting temperature of the primer product.

2.3. The Multiplex Module:

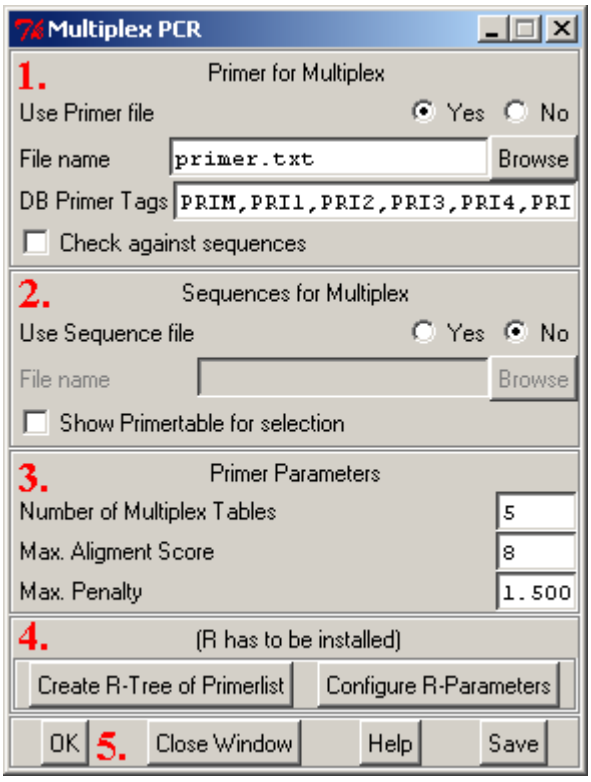
The Multiplex Module allows the selection of primer candidates, which will not interact with each other and so could be used in a Multiplex PCR.

Like the other modules, the Multiplex Module can be found under “STAMP” in the “Experiments”-menu. It is titled “Multiplex PCR”.



2.3.1. The Interface:

The image below shows the interface of the Multiplex Module. The different components of the interface are marked with red numbers in this image. The function of these components will be explained in the following.



1. Primer for Multiplex

In this component of the interface, the source of the primer information for the multiplex has to be set. If the option “Use Primer file” is set to “No” only the primers already written to the Gap4 database, with the tags set in “Primer Tags from DB”, will be used. Otherwise a file with primer information will be loaded. This should be a file created with the Primer3 Module (the “Experiments”-option: “Create flanking Primer”). Next there is a “Check against sequences”-checkbox. Only if this box is selected the interface part “Sequences for Multiplex” will be used.

2. Sequences for Multiplex

If the checkbox “Sequences for Multiplex” in Part 1 was selected the primers for the multiplex will be blasted against the sequences in the Gap4 database. As an alternative it is possible to load a fasta file of sequences to blast against.

If the checkbox “Show Primertable for selection” is checked and the “OK”-Button in Part 5 is clicked a table will open, which allows a manual selection of the primer by means of the blast results for the creation of the multiplex (see also “3.2. The Blast-result table”).

3. Primer Parameters

Here will be set some parameters for the multiplex creation. The “Number of Multiplex Tables” sets the number of different multiplexes which will be created, with an increasing number the time for the calculation will also increase.

The “Max. Alignment Score” sets the maximal allowed number of matches between two primers to be included in one multiplex.

The “Max. Penalty” relates to the penalty value in the primer information, if a primer penalty higher than the set value the primer will not be used for the creation of a multiplex.

4. R-Tree

This component is only of interest, if the program “R” (http://www.r-project.org) is installed. The button “Configure R-Parameters” opens a window with the R-Parameters (see 2.5). After clicking the button “Create R-Tree of Primerlist” a PNG image of a likelihood tree of the primer for the multiplex creation will be created an opened. The PNG will be created in the same folder as the currently opened Gap4 database.

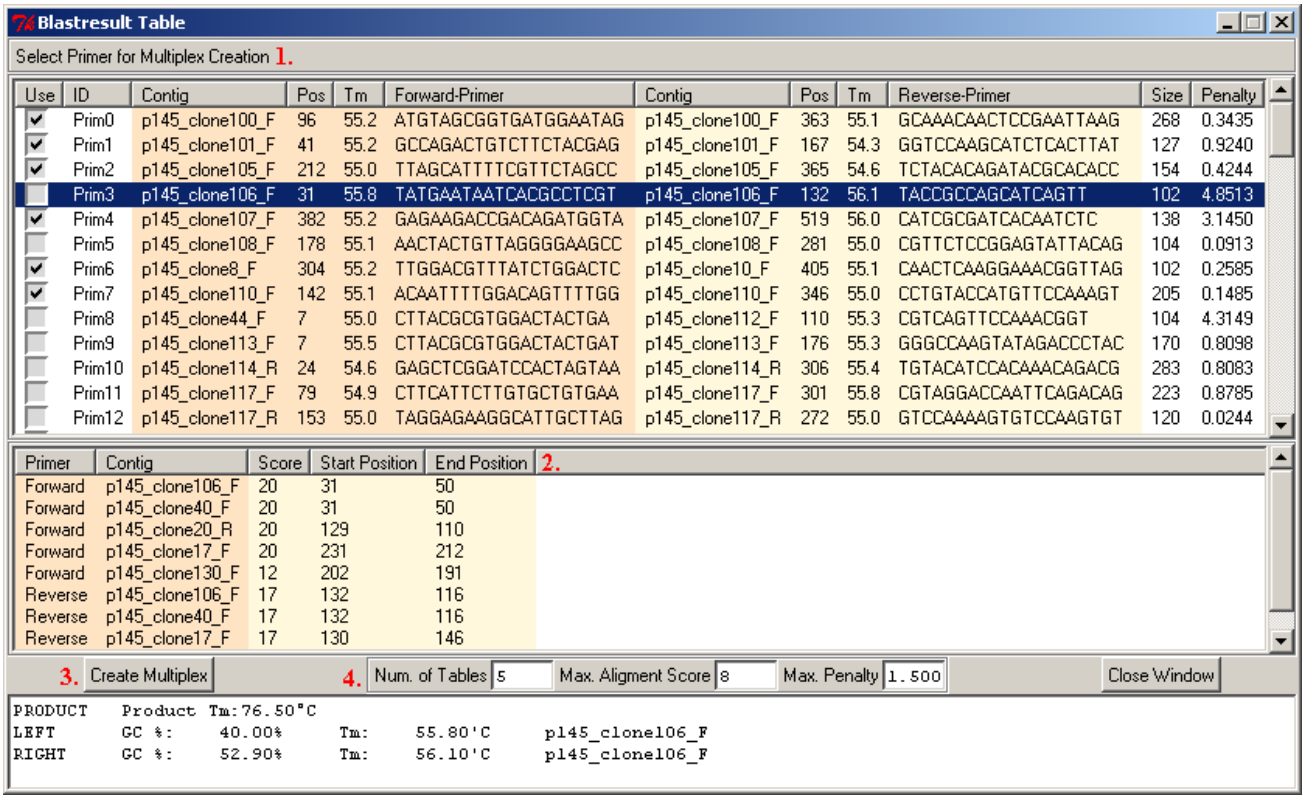
5. The Buttons

“OK” will open a blast-result table or the multiplex table, depending on the selection of the checkbox “Show Primertable for selection” in Part 2. “Help” opens a HTML-page with all information to the Multiplex Module.

The “Safe”-button stores the interface parameters in the file “multiplex_par.txt”, it can be found in “{STADENROOT}\lib\stamp\”.

2.3.2. The Blast-result table:

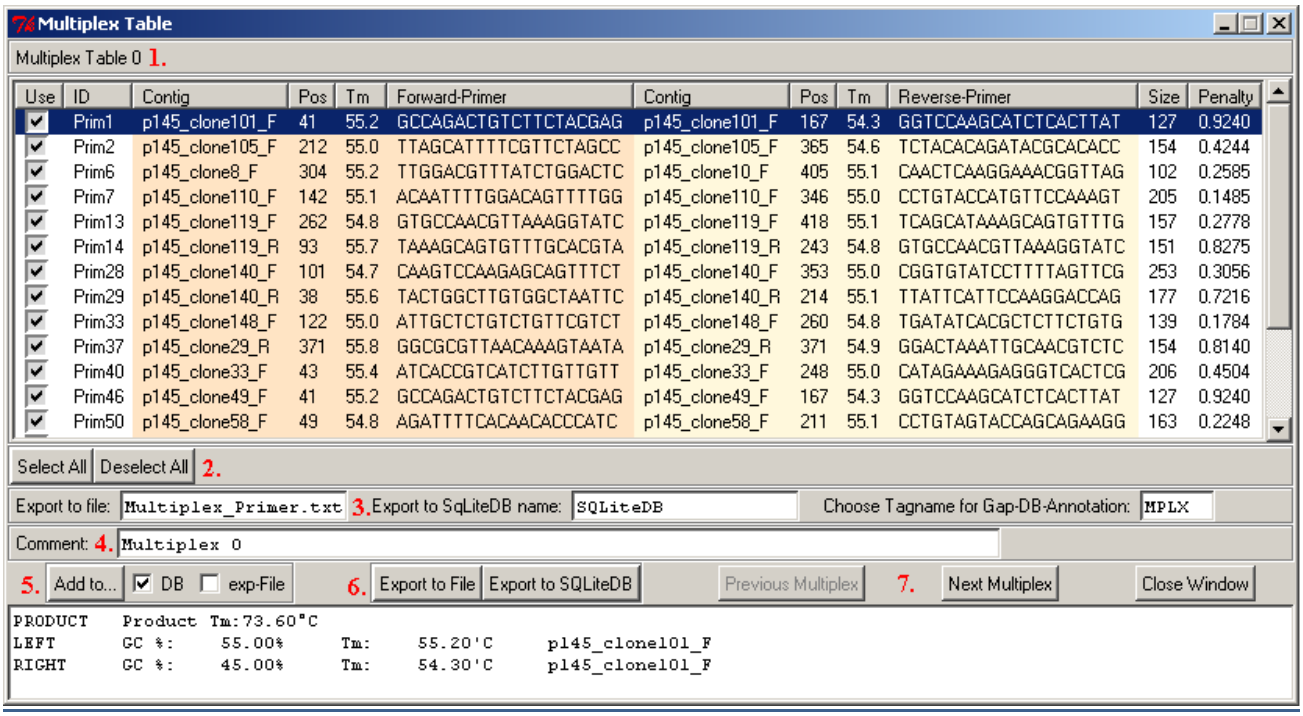
The Blast-result Table allows a manual selection of primers as basis for the multiplex creation process.



1. A table with all primers from the primer file or the Gap4 database. Some primers are already selected on basis of the blast results. Whether a primer is selected or not depends only on the number of matches with the sequences of the Gap4 database or the loaded sequence file. Often the assembly of the Gap4 database is insufficient and the same sequence appears as a number of different contigs. In this case a primer could produce a number of hits, but actually has only a single binding place. Because of this, a manual selection or review is advisable.
   By a double click on a data row a contig editor will open, in which the primer product will be highlighted in the consensus.
2. Displays a table of all binding places of the selected primer pair. The first column of the table holds the information which of the two primers in a pair has a blast hit at the specific position; “Forward” refers to the forward primer of the selected pair and “Reverse” to the corresponding reverse primer. The column “Score” shows the number of alignment matches for the blast hit. This means a score of 20 equals 20 bases of the primer which could anneal at this position of the sequence.
   With a double click on a row in the table a contig editor will open, with a highlighted part in the consensus sequence.
3. The button “Create Multiplex”, which does exactly that, will open a new window with a table of proposed primer pairs to be used in a multiplex experimentally.
4. Shows the primer parameter earlier already set, here it is possible to change them again.

2.3.3. The Multiplex Table:

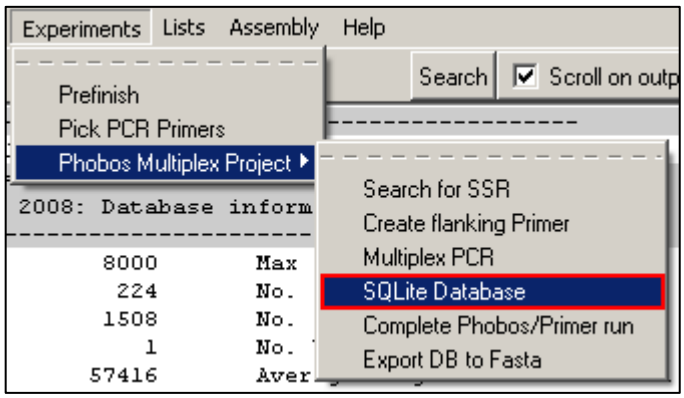
The Multiplex Table displays all primer pairs useable in a single multiplex PCR experiment.

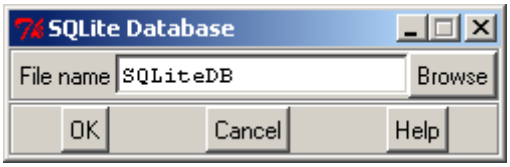


1. A primer table like the one in the blast-result table, but here only the compatible primer pairs are displayed.
2. The buttons under here enable to select/deselect all primers in the list.
3. Here the names of the export file, SQLite database and Gap4 database annotation are set.
4. The text field where a comment can be written, this comment allows for an easy identification of this table when written to the SQLite database and later reopened.
5. Here the names of the export file, SQLite database and Gap4 database annotation are set.
6. With the button “Safe Change” the comment for the table can be changed. The actual comment will be used if the table is again exported to a SQLite database.
7. These two buttons, “Previous Multiplex” and “Next Multiplex”, enable to go through the different multiplex tables which where created. The number of available tables depends of the number set at the parameter “Number of Tables”.

2.4. The SQLite Module:

The SQLite Module allows the usage of a SQL database. With this database it is possible to safe all created repeat, primer and multiplex tables and use these tables again for later research.
Like the other modules, the SQLite Module can be found under “STAMP” in the “Experiments”-menu. It is titled “SQLite Database”.

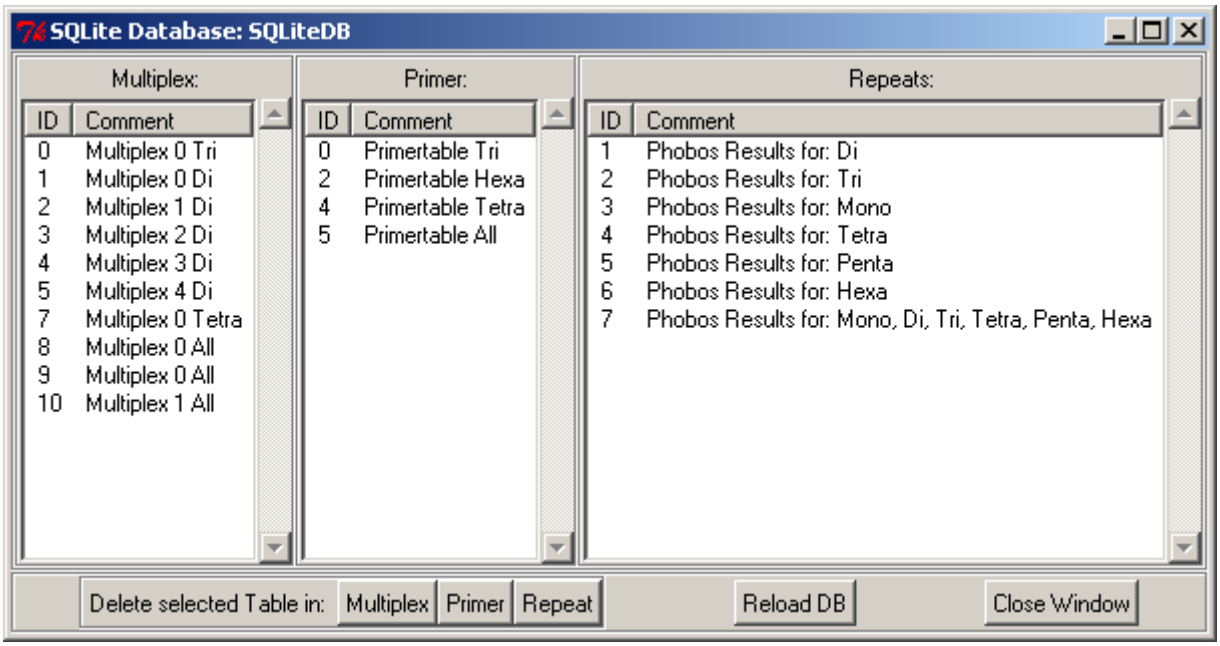




After clicking the “SQLite Database” menu option a little window will open for selection of the desired database.

2.4.1. The SQLite Database

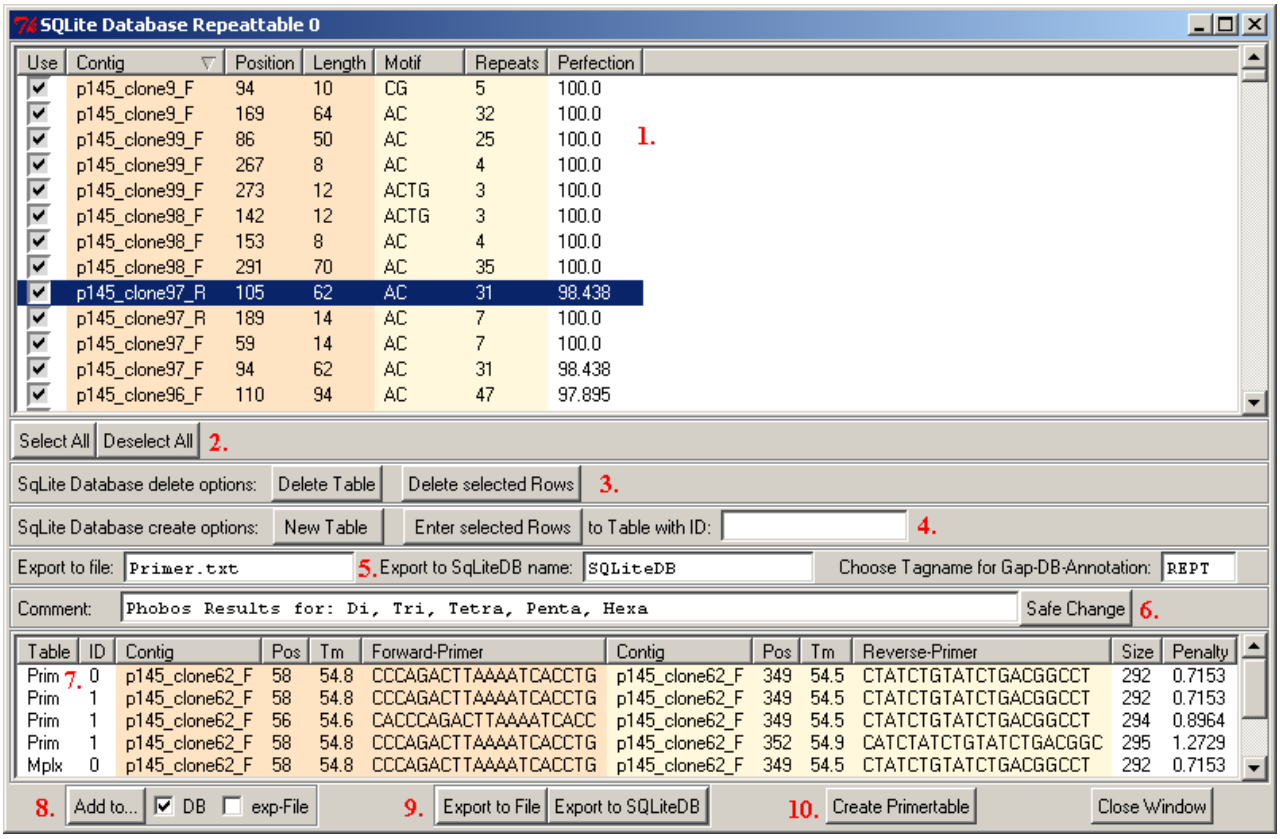
When the database is opened a window like the one in the picture below will be created. This window consists of three lists, a multiplex-, a primer- and a repeat-list. Every list contains tables of according type. Every table in a list has a unique ID. If a primer table is written to the database it automatically will appear in the right list by the next opening of the database. With a double click on a row of a list the corresponding table will open.



Through the buttons “Multiplex”, “Primer” or “Repeat” it is possible to delete a selected table in the related list.
The button “Reload DB” allows to reload the SQLite database, this is necessary when a table is added through one of the other modules.

2.4.2. The SQLite Tables

There are three kinds of tables in the SQLite database; at first the repeat table will be explained:

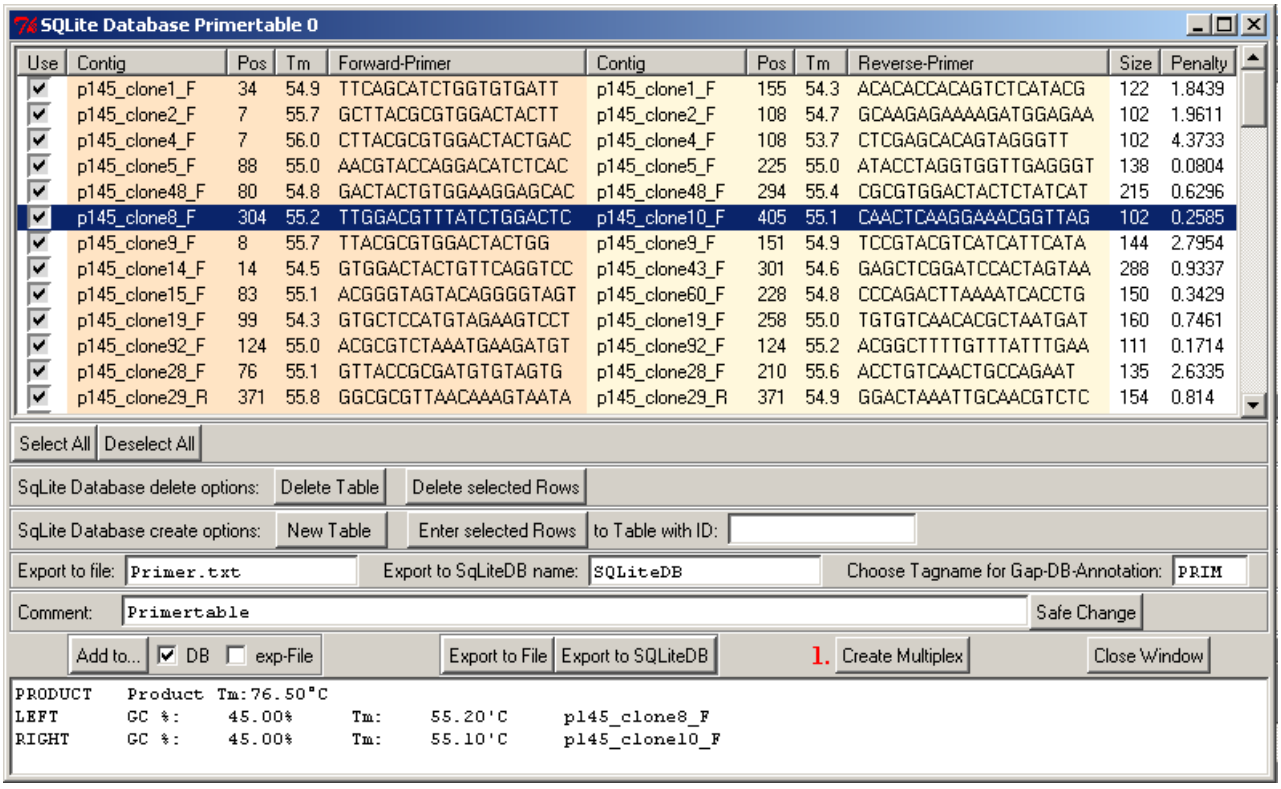


1. A table of the repeats, a double click on a row opens a contig editor.
2. The buttons to select or deselect all repeats in the list.
3. The button “Delete Table” will delete this table from the database when clicked. The button “Delete selected Rows” deletes all rows from this table which are checked in the “Use”-column.
4. The button “New Table” creates an empty table; it appears in the appropriate list of the SQLite database. With the button “Enter selected Rows” all rows which are checked in the “Use”-column can be written to the table with the ID set in the text field.
5. Here the names of the export file, SQLite database and Gap4 database annotation are set.
6. With the button “Safe Change” the comment for the table can be changed. The actual comment will be used if the table is again exported to a SQLite database.
7. Here a table shows all primers in the database which refer to the selected repeat. The columns “Table” and “ID” show the location of the specific primer in the database. Here again a double click will open a contig editor.
8. The “Add to…”-button write all selected rows to either the Gap4 database or the experiment files, this depends on status of the checkboxes “DB” and “exp-File”.

9. Exports all selected rows to a text file (“Export to File”) or to a SQLite database (“Export to SQLite Database”).
10. The button “Create Primertable” opens the interface of the Primer3 Module and uses all selected rows in the repeat table as targets. After a run the Primer3 interface will not close. It always keeps the same target information. After changing the repeat selection it is necessary to open a new Primer3 interface by clicking the “Create Primertable”-button again.

The primer table nearly equals the repeat table with two differences:

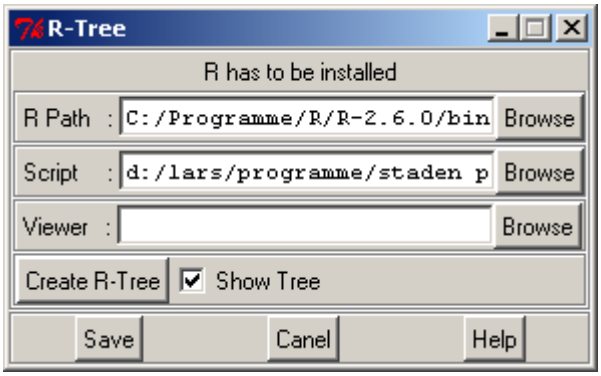
1. There is no “Create Primertable”-button but a “Create Multiplex”-button. This will open the Multiplex Module interface; otherwise the handling of this function is the same.
2. There is no second table list included like for *Point 7* in the repeat table but a text field with some additional primer information.



The multiplex table is of the same kind as the primer table, with one difference; there is no “Create Multiplex”-button.

2.5. R-Tree:

This component is only of interest, if the program “R” (http://www.r-project.org) is installed. The interface seen below can be reached through the Multiplex Module and the button "Configure R-Parameters". The R Module allows the creation of a tree, which shows the compatibility of the different primer pairs.

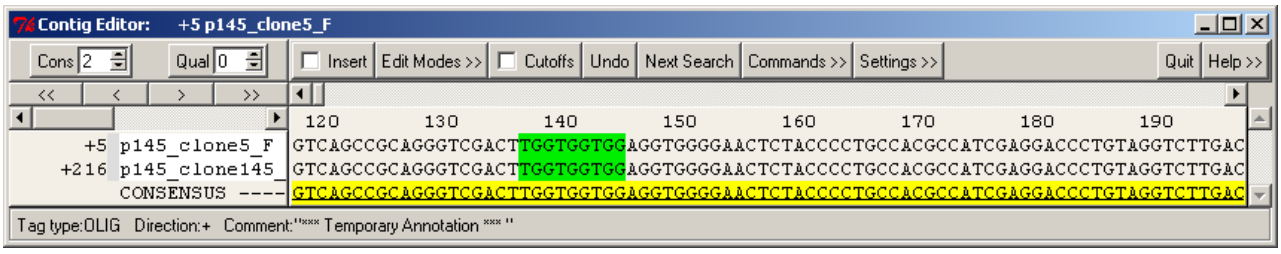


Under “R Path” the path to the “R” program has to be set (see image above for example). The “Script” should already be set correctly to the R-script included in the STAMP folder. For usage of custom scripts should be known, that the script receives a matrix of scores for every primer pair against all others. "Viewer" allows to set the path to the program which will display the tree. If no path is set a standard program will be used, in windows for example it will be the explorer. The button “Create R-Tree” will create a PNG image of a likelihood tree of the current primer pair input of the Multiplex Module. The PNG will be created in the same folder as the currently opened Gap4 database. The button "Save" stores the set paths. It is only necessary to set the paths one time.

3. Additional Information

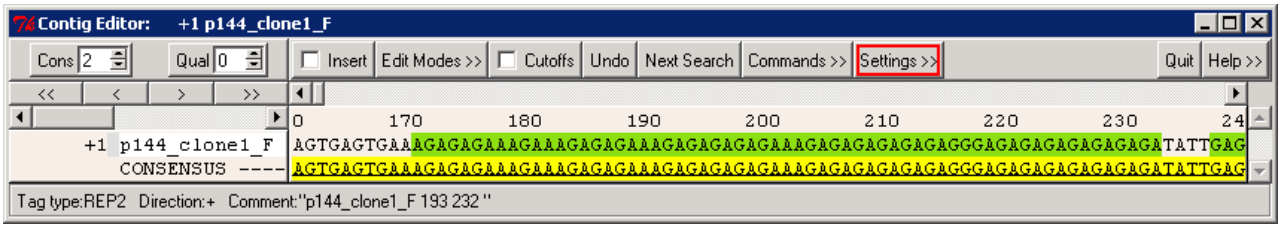
3.1. The Contig Editor:

The contig editor shows the sequence of the selected contig. If it was opened from a table of one of the described modules it normally highlights in yellow a specific part of the sequence in the consensus line of the contig editor. In this example the product of a primer pair is highlighted, the microsatellite which is flanked by the primer pair can be seen highlighted in green in the contig sequences.



3.2. Useful information:

Alternative to the standard tag-names “REPT” and “PRIM”, “REP1” to “REP9” and “PRI1” to “PRI9” can be used. Under the “Settings”-menu in the contig-editor the option “Set Active Tags” can be chosen. Here the tags which shall be marked in the contig can be selected; through this it is possible to get a clear view, if the contig contains many overlapping tags.



Sometimes the repeat markings can be slightly displaced, like on the picture below. This offset is caused through gaps in the consensus sequence provoked by “N”s in the contig sequence.

The repeat position is generated on basis of the consensus sequence. In this sequence are no gaps. The gaps were created through the contig-editor by comparison with the original contig sequence. Therefore the consensus sequence seems longer in the contig-editor than it actually is, when used for repeat marking. Every gap in the sequence, which appears before a repeat tag, moves this tag one base to the left.



3.3. Phobos Parameters

These terms can be used to customize the Phobos output:

-N , --succN
(value required) The maximum number of successive N's allowed in a satellite. Default: 2.

--searchMode
(value required) Phobos probes three different search modes. exact: Search for exact repeats, only. extendExact: Searches for exact repeats and extends them by inserting mismatches and indelds. imperfect: Searches directly for imperfect repeats.

--indelScore
(value required) Score for indels - must be negative. Default: -6. Match score is fixed to one.

--mismatchScore

(value required) Score for mismatch - must be negative. Default: -6. Match score is fixed to one.

--minScore
(value required) The minimum score of a repeat is computed according to maximum(minScore, minScore_a + minScore_b*(unit-length)). Default value of minScore: 0

--minScore_a
(value required) The minimum score of a repeat is computed according to maximum(minScore, minScore_a + minScore_b*(unit-length)). Default value of minScore_a: 0

--minScore_b
(value required) The minimum score of a repeat is computed according to maximum(minScore, minScore_a + minScore_b*(unit-length)). Default value of minScore_b: 0

--minLength
(value required) The minimum length of a repeat is computed according to maximum(minLength, minLength_a + minLength_b*(unit-length)). Default value of minLength: 1

--minLength_a
(value required) The minimum length of a repeat is computed according to maximum(minLength, minLength_a + minLength_b*(unit-length)). Default value of minLength_a: 0

--minLength_b
(value required) The minimum length of a repeat is computed according to maximum(minLength, minLength_a + minLength_b*(unit-length)). Default value of minLength_b: 0

-U , --maxUnitLen
(value required) Maximum unit length. Default: 5

-u , --minUnitLen
(value required) Minimum unit length. Default: 2

-r , --recursion
(value required) The recursion depth used in the search. Values in the range 3-7 are recommended. A value of 0 implies a search for perfect repeats only.

--lastSeq
(value required) Number of last sequence to be processed in this run.

--firstSeq
(value required) Number of first sequence to be processed in this run.

--dontRemoveHidden
Phobos sometimes finds repeats that partially or completely overlap with other repeats, e.g. if alignments with alternative repeat patterns exist at the same locus in a sequence. The default is to remove these mostly hidden repeats in favour of the repeat with the highest score. With this option, Phobos reports also hidden repeats.

-P , --minPerfection
(value required) Minimum perfection of satellites. Satellites below this value are broken up or reduced in length if this yields a satellite above the minimum perfection. If N's are treated as mismatch

when computing the perfection, this value also has an influence when searching for exact repeats only.

-f , --flanking
(value required) If the satellite sequence is printed, this is the number of flanking nucleotides to be printed to the left and right of it.

--maskX
Writes the sequences to a file with the extension ".masked" in which repeats are masked.

--printFormat
(value required) Phobos provides different output formats for printing the repeat information. 0: phobos output format, 1: extended phobos output format, 2: sputnik compatible output format.

--printRepeatSeqMode
(value required) Phobos provides different modes to print the repeat sequence along with its information. 0: dont't print sequence, 1: print sequence, 2: print alignment.

--NPerfectionMode
(value required) Phobos provides different modes to treat N's when computing the perfection of a repeat. 0: asMismatch, 1: asNeutral, 2: asMatch. Default: 0

--reportUnit
(value required) Repeat units can be reported in three different modes. 0: asIs, 1: Alphabetical normal form, 2: Alphabetical normal form also considering the reverse complement. Default: 2

--, --ignore_rest
Ignores the rest of the labeled arguments following this flag.

-v, --version
Displays version information and exits.

-h, --help
Displays usage information and exits.

(required) (value required) The FASTA file containing the input sequences.