TaqSim -
TaqMan PCR Simulator

Matthew D. Dyer\textsuperscript{1,2}, Shea N. Gardner\textsuperscript{3}, Jason R. Smith\textsuperscript{3}, Beth Vitalis\textsuperscript{3}, Clinton Torres\textsuperscript{3}, Thomas A. Kuczmarski\textsuperscript{3}, and Thomas Slezak\textsuperscript{3}

\textsuperscript{1}Genetics, Bioinformatics, and Computational Biology Program, \textsuperscript{2}Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA, and \textsuperscript{3}Pathogen Informatics Group, Lawrence Livermore National Laboratory, Livermore, CA, 94550
The GNU General Public License (GPL)
Version 2, June 1991

The following is the GPL open source license template to be used in each of the
source files of a program where we assert LLNS copyright.
Attach the following notices to each source file. In this notice there should be a
pointer to where the Full notice is to be found.

Copyright (c) 2008, Lawrence Livermore National Security, LLC. Produced at
the Lawrence Livermore National Laboratory. Written by Matt Dyer (dyermd@vt.edu).

CODE-402944 All rights reserved.

This file is part of TaqSim.

This program is free software; you can redistribute it and/or modify it
under the terms of the GNU General Public License (as published by the

This program is distributed in the hope that it will be useful, but WITHOUT
ANY WARRANTY; without even the IMPLIED WARRANTY OF
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
terms and conditions of the GNU General Public License for more details.

You should have received a copy of the GNU General Public License
along with this program; if not, write to the Free Software Foundation,
Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

OUR NOTICE AND TERMS AND CONDITIONS OF THE GNU
GENERAL PUBLIC LICENSE

Our Preamble Notice

A. This notice is required to be provided under our contract with the U.S.
Department of Energy (DOE). This work was produced at the Lawrence
Livermore National Laboratory under Contract No. DE-AC52-07NA27344
with the DOE.

B. Neither the United States Government nor Lawrence Livermore National Security,
LLC nor any of their employees, makes any warranty, express or implied, or assumes
any liability or responsibility for the accuracy, completeness, or usefulness of any
information, apparatus, product, or process disclosed, or represents that its use would
not infringe privately-owned rights.

C. Also, reference herein to any specific commercial products, process, or services by
trade name, trademark, manufacturer or otherwise does not necessarily constitute or
The precise terms and conditions for copying, distribution and modification follows.

GNU Terms and Conditions for Copying, Distribution, and Modification

1. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The “Program,” below, refers to any such program or work, and a “work based on the Program” means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term “modification”.) Each licensee is addressed as “you.”

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

2. You may copy and distribute verbatim copies of the Program’s source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

3. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

   A) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

   B) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of
this License.

C) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate work. But when you distribute the same section as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program. In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

   A) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

   B) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

   C) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b
above.)

The source code for a work means the preferred form the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable. If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the

A) Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

B) You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

C) Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients’ exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

D) If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this license, they do not excuse you from the conditions of this License. If you can not distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free
redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose to this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

A) If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

A) Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and “any later version,” you have the option of following the terms and conditions either of that version of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

B) If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision to grant permission will be guided by the two goals of preserving the free status of all derivatives of our free software and or promoting the sharing and reuse of software generally.
NO WARRANTY

BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF

A) THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

B) IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS
Contents

Description 9
Overview 10
Running TaqSim 11
  Setting Up TaqSim 11
  Options 12
  Running TaqSim 18
  TaqSim Output 20
  TaqSim Examples 21
Help 22
Description

Polymerase chain reaction (PCR) has become a key molecular biology tool with many applications. Although PCR techniques are simple and well defined, they can be time consuming as well as unpredictable in the case of degenerate primers or multiplex primer mixes where multiple reactions are tested simultaneously to reduce expenses. The unpredictability arises from the fact that the components of the individual primer sets may cross-react with each other, yielding unexpected and confusing results. In an effort to eliminate these problems and we have developed TaqSim, a perl program that simulates various types of PCR reactions.

The advantage of TaqSim is that it is run from a command-line and can generate output files in a number of formats allowing it to serve as a front-end or back-end for other software. Users can select between a number of PCR applications including the traditional forward/reverse primer pair (simple), inclusion of an internal oligo (triplet), and multiplex reactions. TaqSim allows users to define the search set so you can easily search a single genome or all known sequences. TaqSim allows users to adjust runtime settings like the number of allowed mismatches, maximum amplicon length, primer hybridization temperatures, and the allowance of degenerate primers which allows the user to model their specific experimental needs. If specified, TaqSim partitions input data into subsets and runs them each in parallel allowing users to make use of multi-processor machines and reduce running time.
Overview

To run TaqSim, a user provides an input set of primers (may be multiple) of primers that contain both a forward and reverse component. If desired the user may also include the optional internal oligo component. The user then defines a set databases, against which a BLAST analysis is performed (Altschul et al. 1997). The resulting high-scoring pairs are filtered based on mismatches and primer melting temperature and those that pass the user defined parameters are mined for possible amplicons given the reaction type. A reaction type can be one of three options; simple, triplet, or multiplex. A simple reaction only requires that a forward and reverse primer set hybridize within a given stretch of DNA where a triplet requires the additional presence of an internal oligo between the two primers. A multiplex reaction can either be simple or triplet and does not require that the two primers belong to the same set. For each of the resulting amplicons TaqSim gathers the amplicon sequence information, and formats the data into a user defined output style. Currently output can be reported in Excel, XML, flat, and tab-delimited formats.
Running TaqSim

Setting Up TaqSim

TaqSim is designed for Linux/Unix based machines and has also been tested on Mac OSX. For Windows users we suggest running TaqSim through Cygwin which is available here [http://www.cygwin.com/](http://www.cygwin.com/). Windows users will also need to ensure that perl is installed. If not then you may install ActivePerl by following the directions located here [http://www.activestate.com/Products/ActivePerl/](http://www.activestate.com/Products/ActivePerl/).

TaqSim requires the installation of blast. Blast executables may be obtained at the following site, [http://www.ncbi.nlm.nih.gov/BLAST/download.shtml](http://www.ncbi.nlm.nih.gov/BLAST/download.shtml). Download and install the version corresponding to your system architecture. Note installation path as it will be required for running TaqSim.

Additionally, TaqSim comes prepackaged with the required modules, which includes the following along with any required by those listed below.

Getopt::Long;
HTTP::Lite
Bio::SeqIO;
Bio::SearchIO;
Data::Dumper;
File::Basename;
Spreadsheet::WriteExcel;

These modules along with those specifically developed for TaqSim can be found in the Modules folder. Information regarding the non-TaqSim modules can be found at [http://search.cpan.org/](http://search.cpan.org/). If you have already installed these modules on your computer then you may remove all of the folders in the Modules folders except TaqSim.
Options

• Option File
  Description: Set the path to the xml format option file (See running TaqSim for example)
  Command Line: -option_file=VALUE
  Option File: NOT VALID
  Default: NULL

• Signature File
  Description: Set the path to the FASTA formatted input file containing primer information (See running TaqSim for example).
  Command Line: -signature_file=VALUE
  Option File: <signature_file>VALUE</signature_file>
  Default: REQUIRED

• Max Threads
  Description: Set the maximum number of threads allowed during the TaqSim analysis.
  Command Line: -max_threads=VALUE
  Option File: <max_threads>VALUE</max_threads>
  Default: 1

• Output Type
  Description: Set the format of the output results. Must be one of four types: excel, tab, flat, or xml.
  Command Line: -output_type=VALUE
  Option File: <output_type>VALUE</output_type>
  Default: excel
  Allowed: excel, tab, flat, xml
• **Reaction Type**

  Description: Set the type of reaction. Must be one of three types: simple, taqman, or mux.

  Command Line: -reaction_type=VALUE

  Option File: <reaction_type>VALUE</reaction_type>

  Default: REQUIRED

  Allowed: simple, taqman, mux

• **Output Directory**

  Description: Set the directory for runtime and output files.

  Command Line: -output_dir=VALUE

  Option File: <output_dir>VALUE</output_dir>

  Default: Same directory as the application ("./")

• **Run Type**

  Description: Set the type of analysis run. Must be one of four types: complete, complete-debug, rerun, rerun-debug (See running TaqSim for more information).

  Command Line: -run_type=VALUE

  Option File: <run_type>VALUE</run_type>

  Default: complete

  Allowed: complete, complete-debug, rerun, rerun-debug

• **Result Folder**

  Description: Pointer to the runtime_files directory for a previous run of TaqSim if you wish to redo an analysis with out redoing the blast analysis.

  Command Line: -result_folder=VALUE

  Option File: <result_folder>VALUE</result_folder>

  Default: REQUIRED if run type is rerun or rerun-debug
• Amplicon Extension

Description: When reporting amplicons you can extend the sequence reported on each side.

Command Line: -amplicon_extension=VALUE

Option File: <amplicon_extension>VALUE</amplicon_extension>

Default: 0

• Blast Database

Description: Set the path to a blast formatted database to be searched.

Command Line: -blast_database=VALUE

Option File: <blast_database>VALUE</blast_database>

Default: NULL

• Fasta Database

Description: Set the path to a fasta file to be searched. Fasta files will be formated into a blast searchable database during execution.

Command Line: -fasta_database=VALUE

Option File: <fasta_database>VALUE</fasta_database>

Default: NULL

• FORMATDB Path

Description: Set the path to the formatdb executable (See Setting Up TaqSim for more information).

Command Line: -formatdb_path=VALUE

Option File: <formatdb_path>VALUE</formatdb_path>

Default: Default installation path (/usr/local/bin/formatdb)
• BLASTALL Path

   Description: Set the path to the blastall executable (See Setting Up TaqSim for more information).

   Command Line: -blastall_path=VALUE

   Option File: <blastall_path>VALUE</blastall_path>

   Default: Default installation path (/usr/local/bin/blastall)

• Max Signatures

   Description: Set the maximum number of primers to be included in each blast search.

   Command Line: -max_signatures=VALUE

   Option File: <max_signatures>VALUE</max_signatures>

   Default: 1

• Processors

   Description: Set the number of processors to be used in a blast search. Equivalent to the blast -a option.

   Command Line: -processors=VALUE

   Option File: <processors>VALUE</processors>

   Default: 1

• Word Size

   Description: Set the word size to be used in a blast search. Equivalent to the blast -W option.

   Command Line: -word_size=VALUE

   Option File: <word_size>VALUE</word_size>

   Default: 7
• **E-value**
  
  Description: Set the e-value cutoff to be used in a blast search. Equivalent to the blast -e option.

  Command Line: -evalue=VALUE

  Option File: <evalue>VALUE</evalue>

  Default: 7

• **Mismatches Allowed**
  
  Description: Set the number of mismatches allowed in each primer.

  Command Line: -mismatches_allowed=VALUE

  Option File: <mismatches_allowed>VALUE</mismatches_allowed>

  Default: 2

• **Amplicon Length**
  
  Description: Set the maximum amplicon size allowed.

  Command Line: -amplicon_length=VALUE

  Option File: <amplicon_length>VALUE</amplicon_length>

  Default: 500

• **Primer Tm Cutoff**
  
  Description: Set the Tm cutoff for the primers.

  Command Line: -primer_tm_cutoff=VALUE

  Option File: <primer_tm_cutoff>VALUE</primer_tm_cutoff>

  Default: 50
• Probe Tm Cutoff
  Description: Set the Tm cutoff for the probe.
  Command Line: -probe_tm_cutoff=VALUE
  Option File: <probe_tm_cutoff>VALUE</probe_tm_cutoff>
  Default: 50

• Nucleotide Concentration
  Description: Set the nucleotide concentration for calculating the Tm.
  Command Line: -nucleotide_conc=VALUE
  Option File: <nucleotide_conc>VALUE</nucleotide_conc>
  Default: 0.0004

• Sodium Concentration
  Description: Set the sodium concentration for calculating the Tm.
  Command Line: -sodium_conc=VALUE
  Option File: <sodium_conc>VALUE</sodium_conc>
  Default: 1.0

• Magnesium Concentration
  Description: Set the magnesium concentration for calculating the Tm.
  Command Line: -magnesium_conc=VALUE
  Option File: <magnesium_conc>VALUE</magnesium_conc>
  Default: 0.0

• Annealing Temperature
  Description: Set the annealing temperature for calculating the Tm.
  Command Line: -annealing_temp=VALUE
  Option File: <annealing_temp>VALUE</annealing_temp>
  Default: 1.0
• Sleep Time

Description: When running TaqSim in parallel this sets the number of seconds TaqSim will sleep when there are still files to process, but no open threads.

Command Line: -sleep_time=VALUE

Option File: <sleep_time>VALUE</sleep_time>

Default: 30

Running TaqSim

There are two main methods for adjusting the runtime parameters for TaqSim, command-line and option-file. If you are constantly running the same sort of analyses it may be easier to write an option file instead of reentering option values each time. An example of an option file can be seen below with its equivalent command-line call.

Command-line: ./TaqSim -option_file=options.xml

Options file:

<option_set>
  <run_type>complete</run_type>
  <signature_file>Path to primer file</signature_file>
  <reaction_type>mux</reaction_type>
  <output_dir>Path to output directory</output_dir>
  <output_type>excel</output_type>
  <fasta_database>Path to FASTA file 1 to search against</fasta_database>
  <fasta_database>Path to FASTA file 2 to search against</fasta_database>
  <primer_tm_cutoff>55</primer_tm_cutoff>
  <probe_tm_cutoff>-60</probe_tm_cutoff>
  <nucleotide_conc>0.0004</nucleotide_conc>
  <sodium_conc>1.0</sodium_conc>
  <magnesium_conc>0.0</magnesium_conc>
  <hybridization_temp>60</hybridization_temp>
  <amplicon_length>2500</amplicon_length>
  <max_threads>1</max_threads>
  <max_signatures>1</max_signatures>
  <processors>1</processors>
  <mismatches_allowed>3</mismatches_allowed>
This option file would run a complete analysis creating an excel output-file allowing up to 3 mismatches and maximum amplicon size of 2500 along with some other settings. Keep in mind that you may specify multiple fasta and blast database for any given analysis. If you supply a fasta_database then TaqSim will first use formatdb to format it into a standard blast database. When supplying a blast database path please member to exclude the file suffixes (.phr, .pin, .psq). For example if I had a precompiled blast database called “dna” then in the option file I would include

```xml
<blast_database>/home/path/to/database/dna/dna</blast_database>
```

An example of the primer file input can be seen here

```xml
>my_primer_1|F
AAAAAAAAAAAAAAAAAAA
>my_primer_1|IO
AAAAAAAAAAAAAAAAAAA
>my_primer_1|R
AAAAAAAAAAAAAAAAAAA
>my_primer_2|F
AAAAAAAAAAAAAAAAAAA
>my_primer_2|R
AAAAAAAAAAAAAAAAAAA
```

The input is a standard FASTA format with a few extra restrictions. Primer title may not contain any spaces, please use the underscore character (‘_’) instead of spaces. At the end of each primer title there should appear the bar ‘|’ character (This is simply shift + ‘\’) and either F, IO, or R (use capital letters) to define whether the primer in this set is a forward or reverse primer or an internal oligo.

There are a couple of different types of TaqSim runs. The above option file does a complete run. This is the standard and typical type of analysis and will not produce any output to the screen. If you wish have TaqSim report as it is processing data then for the runtime value insert complete-debug. TaqSim also allows you to rerun a previous complete analysis in the case that you simply wish to change filter settings or the output style so that you do not have to redo the blast analysis. In this case you would set the runtime value at redo (redo-debug for reporting) and you must include the option

```xml
<result_folder>Path to previous runtime_files folder including “runtime_files” </result_folder>
```
which is simply a pointer to the output directory from a previous analysis of TaqSim.

**TaqSim Output**

TaqSim generates a number of files and organizes them before completion. In the defined output directory you will find the folder runtime_files and if you selected to search against a fasta file you will also find the folder personal_db. In the personal_db folder you will find a blast formatted version of the selected fasta file. Within the runtime_files folder you will find a number of files with the prefix blastFile. These files are generated as a result of number of primers sets allowed per blast run. Each file will then be blasted against each of the defined search databases and be labeled with the extension “.bout”. Finally, the resulting hits that make it through the filter are labeled “.filtered_hits”. Directly in the output directory you will find the results file “results.” with the extension xml, txt or xls depending on the output format.

Below is an example of the Excel output. The Excel output contains three sheets. The first is a summary of the primers that were used in the analysis with the corresponding sequence and type.

![Excel Sheet Example](image)

The second sheet gives you a summary of what type of reaction these primers were involved in. The analysis was done using a multiplex reaction. As you can see, both primer A and B took part in a simple reaction and two mixed reactions.

![Excel Sheet Example](image)

Finally the last sheet is where you get the specifics of each resulting amplicon. Amplicons are grouped together by the set of primers that created them. Within each primer
set you will find the sequence name for which the amplicon was found, the start and stop locations, the amplicon size and sequence, and the orientation of the primers.

**TaqSim Examples**

Included in the TaqSim distribution are two examples, one of a complete analysis and the other of a redo analysis. Information for running these examples can be found in the README files of individual folders “Example_Complete” and “Example_Redo”.
Help

For questions regarding problems or suggestions with TaqSim that are not addressed below please send an email to the authors.

- I know these primers should be producing an amplicon, but I don’t get anything in the results file.

One possible reason is that the T_M did not make the cutoff. For calculating the T_M we use the equation \( T_M = 81.5 + 16.6(\log_{10}([Na^+])) + 0.41(\%GC) - 600/\text{length} \). Future work will include a more sophisticated method for T_M calculation. Try setting the Tm cutoffs extremely low (-300) and rerun the analysis.

- My amplicon size is 100bp, but the reported amplicon sequence is 120bp.

The option amplicon_extension allows you to pad your amplicon sequence with a defined number of bases. If you wish to only include the exact amplicon sequence then set this option at 0 (See TaqSim options).

- There is no sequence in the amplicon sequence column.

TaqSim gathers amplicon sequence information directly from NCBI during the analysis. In order to identify which organism was hit and find the sequence information it assumes you are using standard NCBI headers in the search file (blast or fasta). These headers take many forms, but include “|gi|num|” somewhere in the header. TaqSim identifies the “|gi|” tag and takes everything to the next “|” character as the identifier and uses this for grabbing sequence information from NCBI. Additionally, there seems to be a problem with one of the modules when it is running under windows and we are currently looking into this problem.

- TaqSim seems to be hanging with files to be finished.

If for any reason you kill a TaqSim run please make sure you delete the output directory before restarting. If you do not remove the folder then there remains a control_file that will cause TaqSim to believe there is some thread still running.