OAT

Orthologous Average Nucleotide Identity Tool

a similarity measurement tool for genomes

User Manual

version 0.90

July 10 2015

Laboratory of Evolutionary Bioinformatics
Seoul National University

CONTENTS

1. General Information

1.1 Purpose of the Application

OAT (**O**rthologous **A**verage Nucleotide Identity **T**ool) uses OrthoANI to measure overall similarity between two genome sequences. Unlike the original ANI algorithm, it produces almost identical reciprocal similarities. It has been shown by a large comparison study, values generated by the original ANI and OrthoANI are comparable. The proposed cut-off for species demarcation is 95~96% for both OrthoANI and the original ANI. The detailed algorithm is given in Lee *et al*.

OAT has an easy to follow Graphical User Interface that allow researchers to calculate OrthoANI values between genomes of their interest without the need of dealing with an unfamiliar Command Line Environment.

1.2 Developers

OAT was developed by members of the Laboratory of Evolutionary Bioinformatics at Seoul National University.

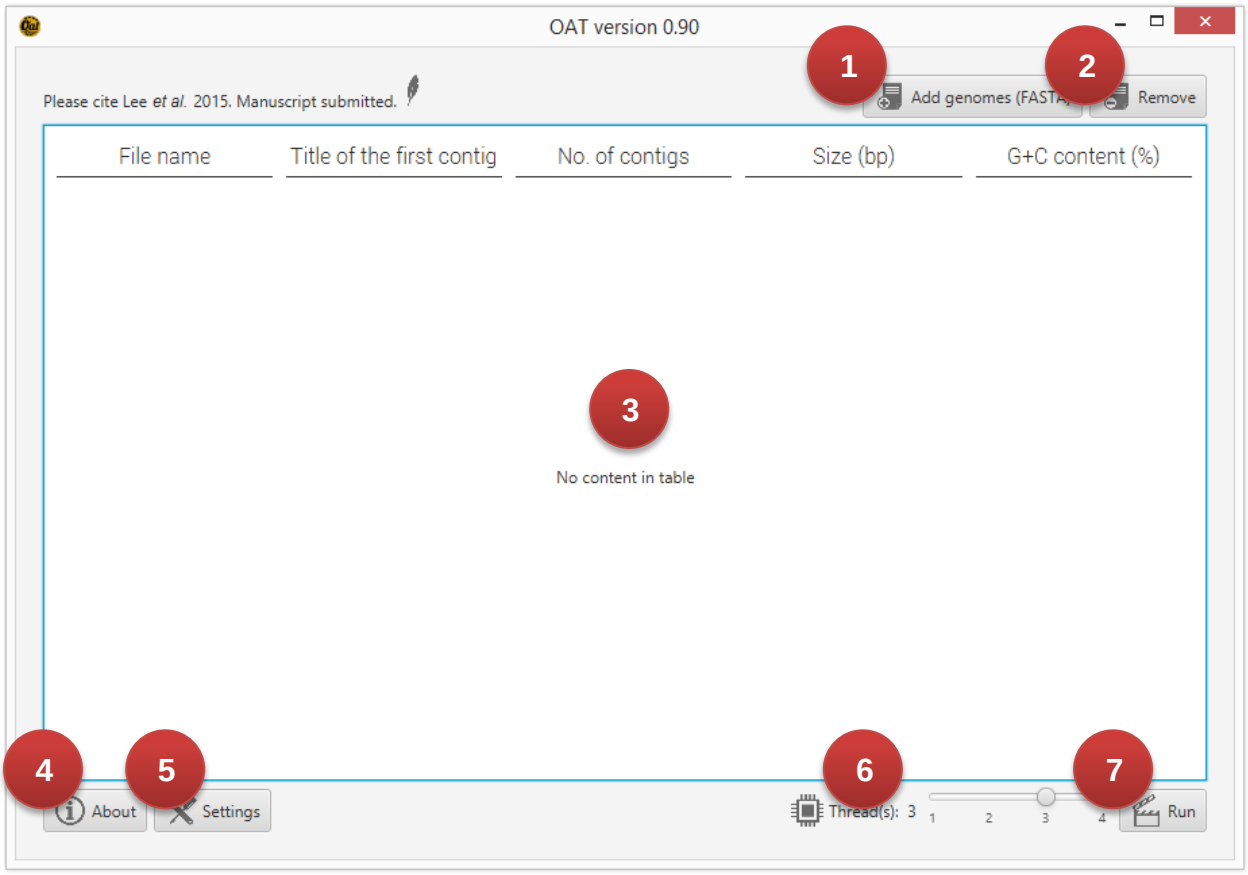
Yeong Ouk Kim, Seoul National Univ.

Imchang Lee, Seoul National Univ.

Sang-Cheol Park, Seoul National Univ.

Jongsik Chun, Seoul National Univ. & ChunLab, Inc.

1.3 Overview

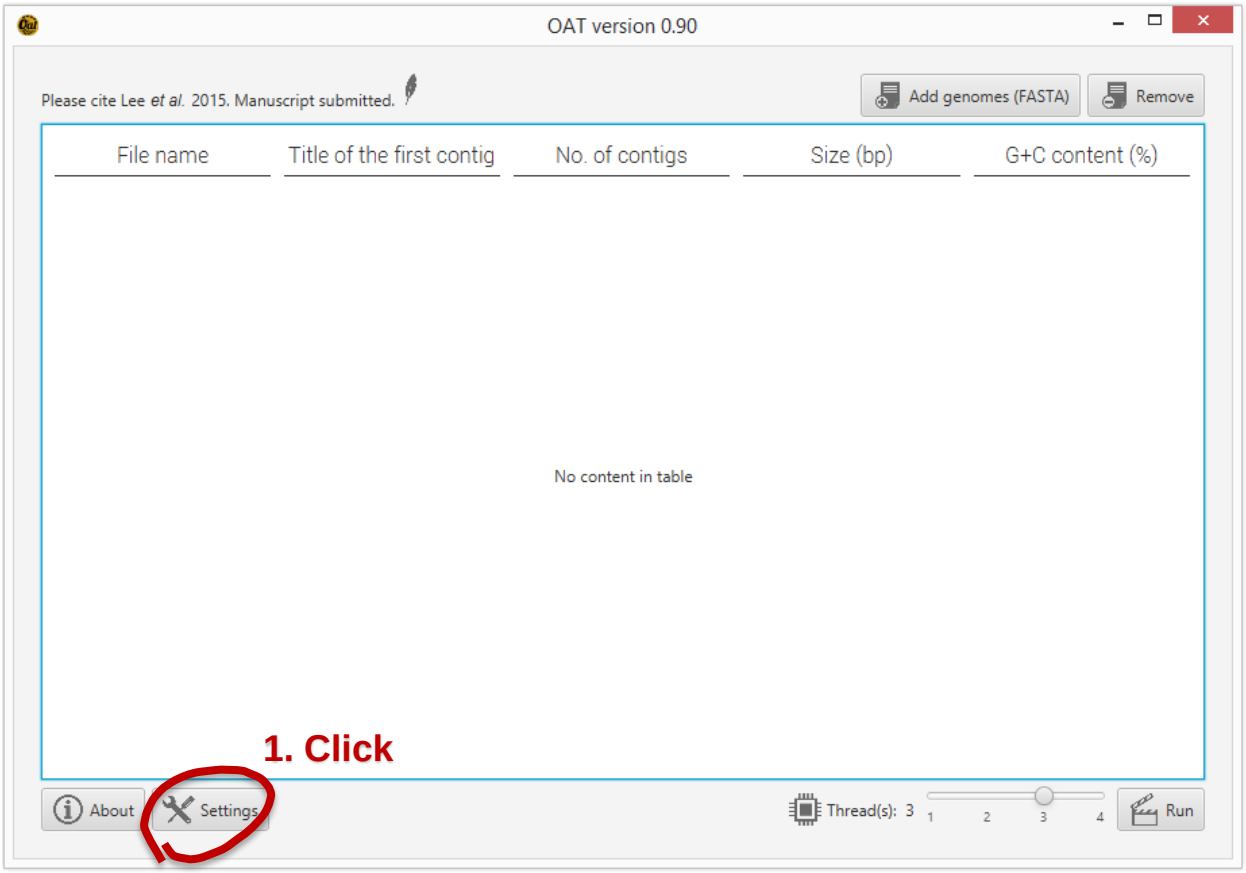


1. Add genome FASTA files as input
2. Remove genome FASTA files
3. Table that keeps your input files
4. Show information about the application
5. Settings for the Blast program
6. Change multithreading options
7. Run the program

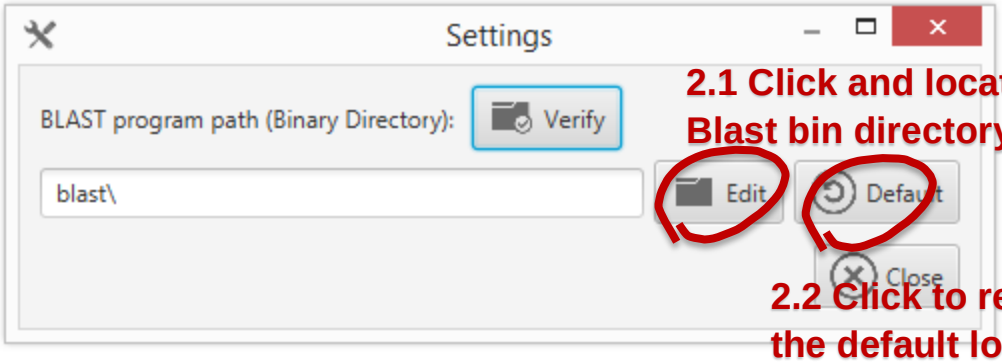
2. Configuration

2.1 Switching the Blast Program

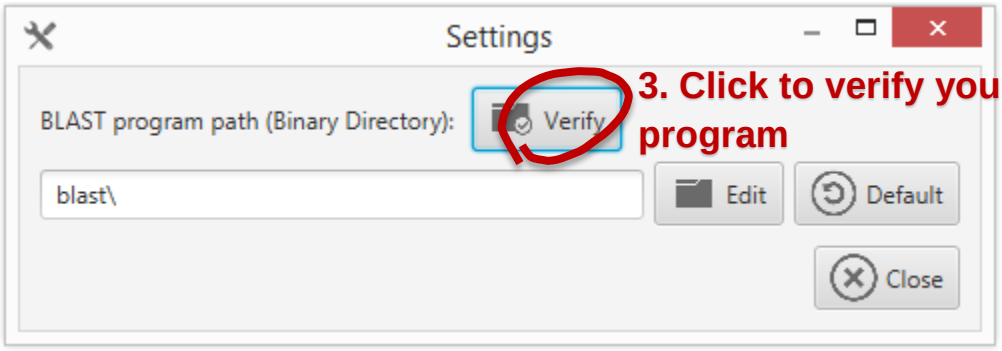
If you wish to use a different Blast version or if you downloaded the Runnable Jar version of the OAT application (which does not include a Blast program), follow the instructions provided below. We recommend users to stick with the provided version of Blast (ncbi-blast-2.2.30+) or a higher version since our application was tested with that version and versions lower than it may potentially cause malfunctions. If you are using the Runnable Jar version of the OAT application you can find the appropriate Blast program for your OS from their website (link).



1. Click the “Settings” button



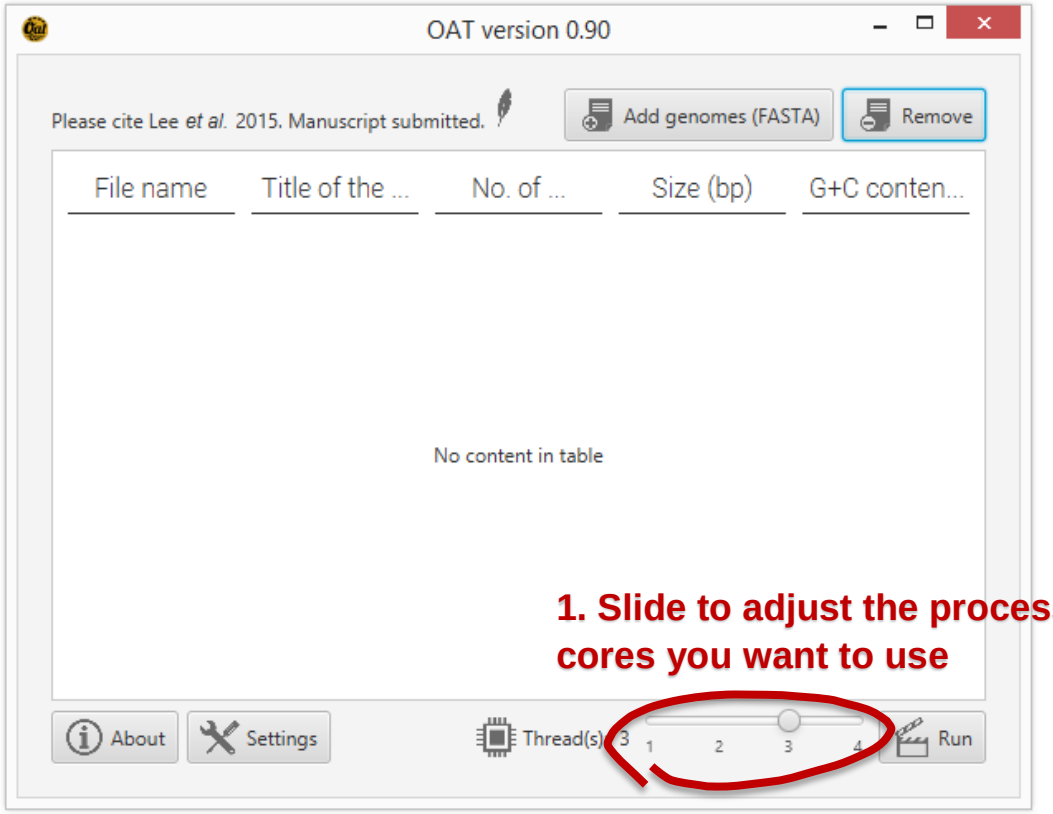
2. Set the Blast program bin directory by…
   2.1 Click the “Edit” button to locate your Blast program’s bin directory
   2.2 Click the “Default” button to set the Blast program’s path to the default location



3. Verify your blast program by clicking the “Verify” button

2.2 Multithreading

If you more than 1 processing cores on your computer you may increase the number of threads the application uses to speed up the calculation process. If you are running other programs on your computer you should avoid setting the threads equal to the number of available processing cores since it will hinder the performance of the other programs.

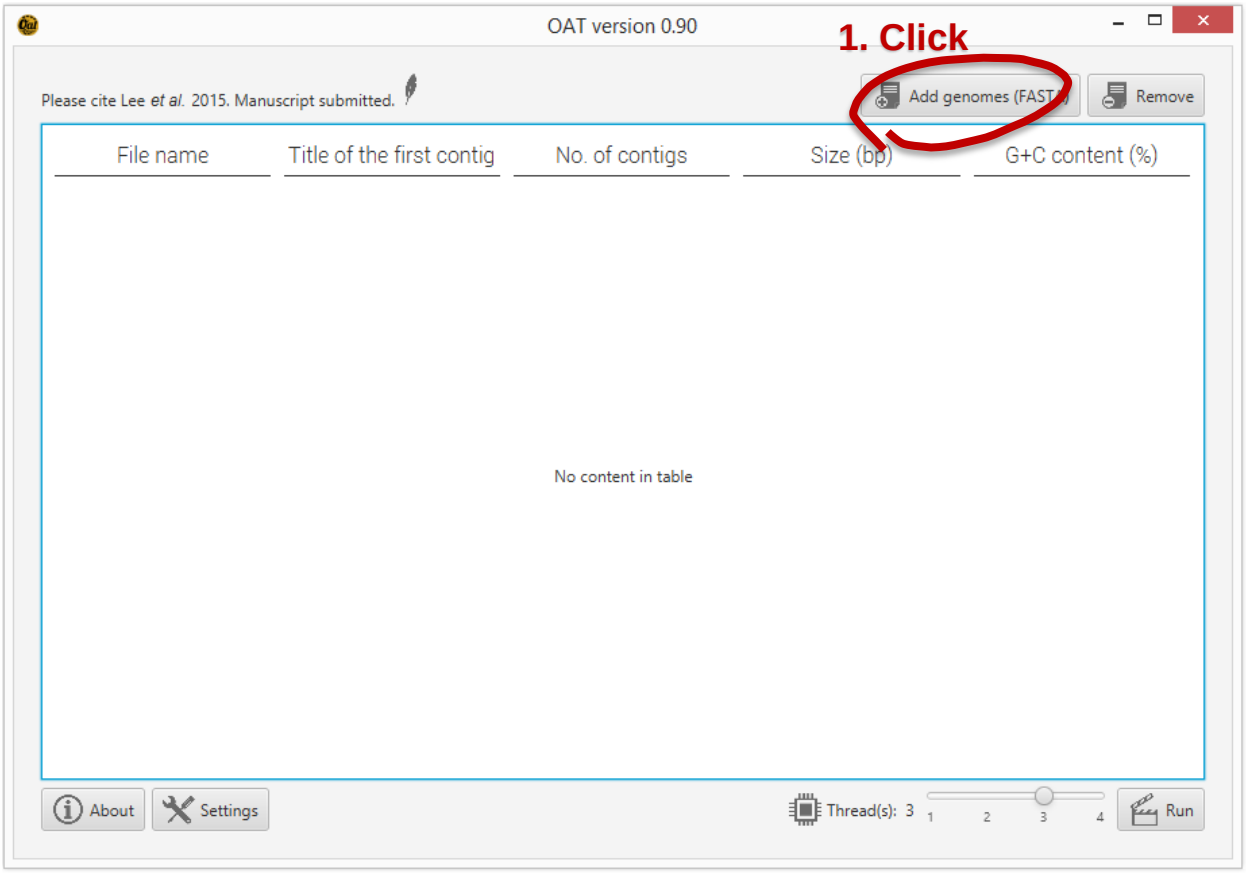


1. Use the slider to adjust the number of threads you wish to use for the calculation process. The maximum value is automatically set to the number of processing cores the computer has.

3. Preparing the Calculation Step

3.1 Adding Genome Data

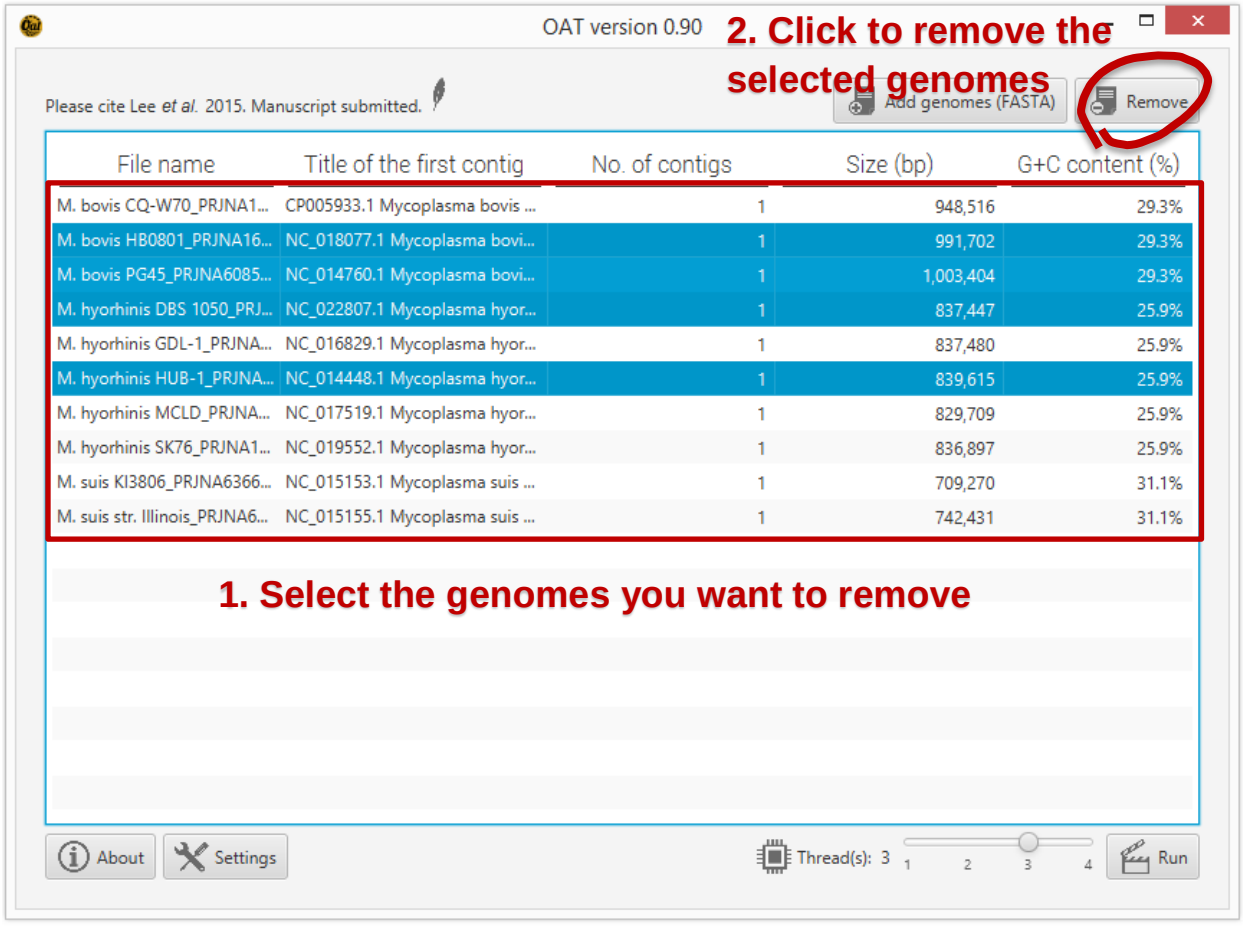
Before starting the calculation, you need to provide genomes that you wish to compare to the Application. The application accepts FASTA files as genome data. Adding genomes to the application can be done easily by locating your FASTA files in your system. You can add up to 10 genomes to the program.



1. Click on the “Add genomes (FASTA)” button to easily add genome data you wish to compare

3.2 Removing Genome Data

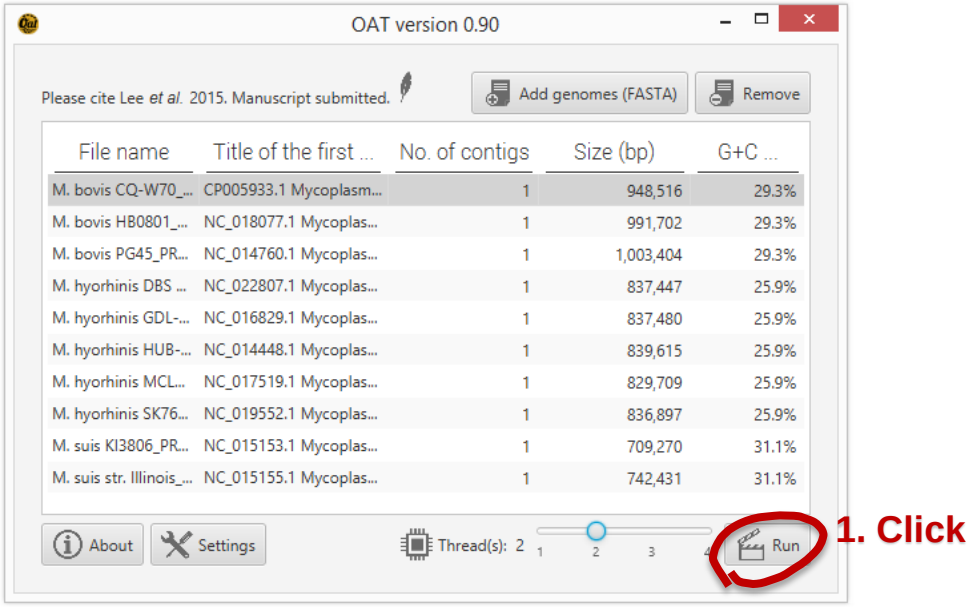
Removing genomes from the application can be done easily by selecting the FASTA files and clicking the “Remove” button



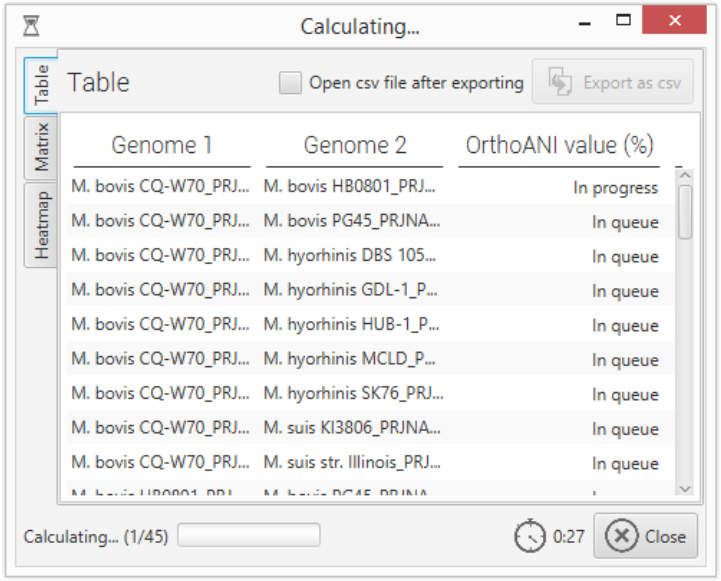
1. Click on the genomes you want remove from the application (click while pressing the shift or control key to select multiple genomes)
2. Click on the “Remove” button to remove the genomes selected from the application

3.3 Run

After managing your input data (genome files) you are now ready to run the calculation step. Simply press the “Run” button and a new window will popup showing the progress and eventually the results of the comparisons. You may start multiple runs at the same time but you should be careful not to exceed the number of processing cores available. Because you are most likely to see no performance (speed wise) gain while overwhelming your system.



1. Click the “Run” button



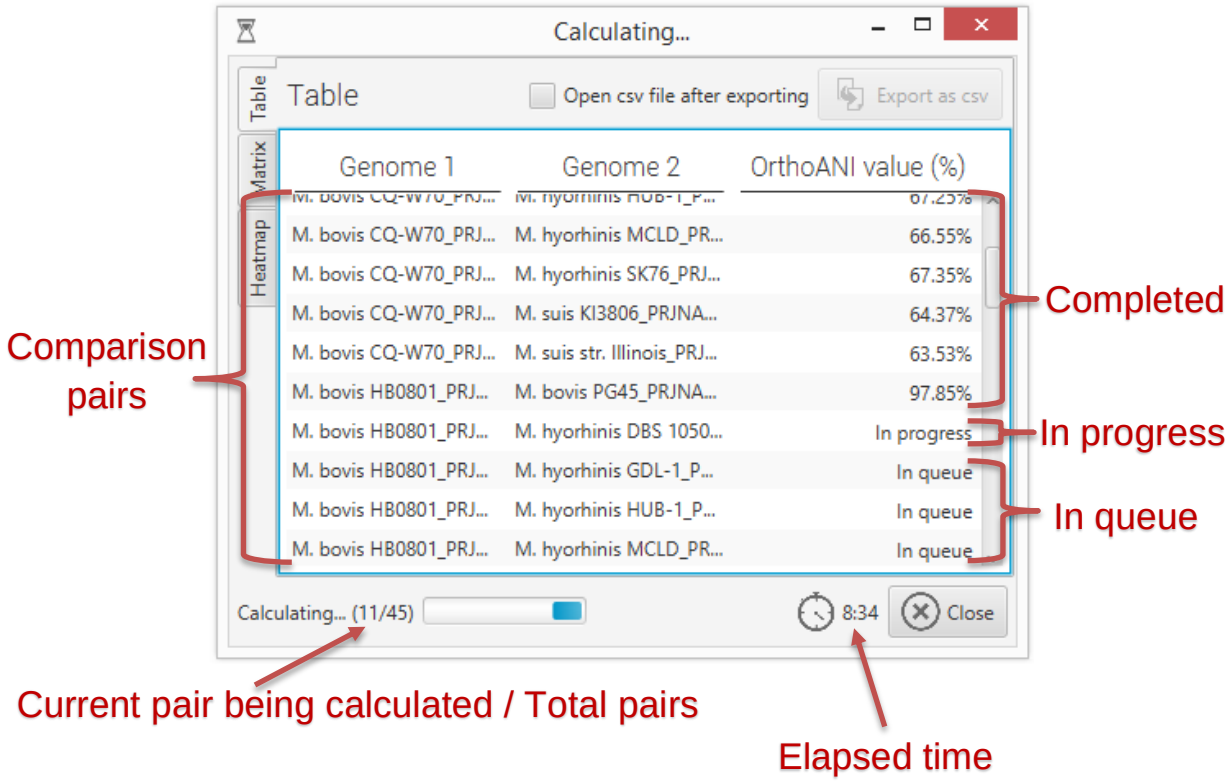
A new window will popup displaying the progress of the run

4. Calculation Step

4.1 Calculation Step

The calculation step will most likely take a while so it is possible to view the progress of the run and also view partial results throughout the calculation process. There will be a total of N*(N-1)/2 comparisons (Where N is the number of genomes).

i.e.) N = 10 → 45 comparisons

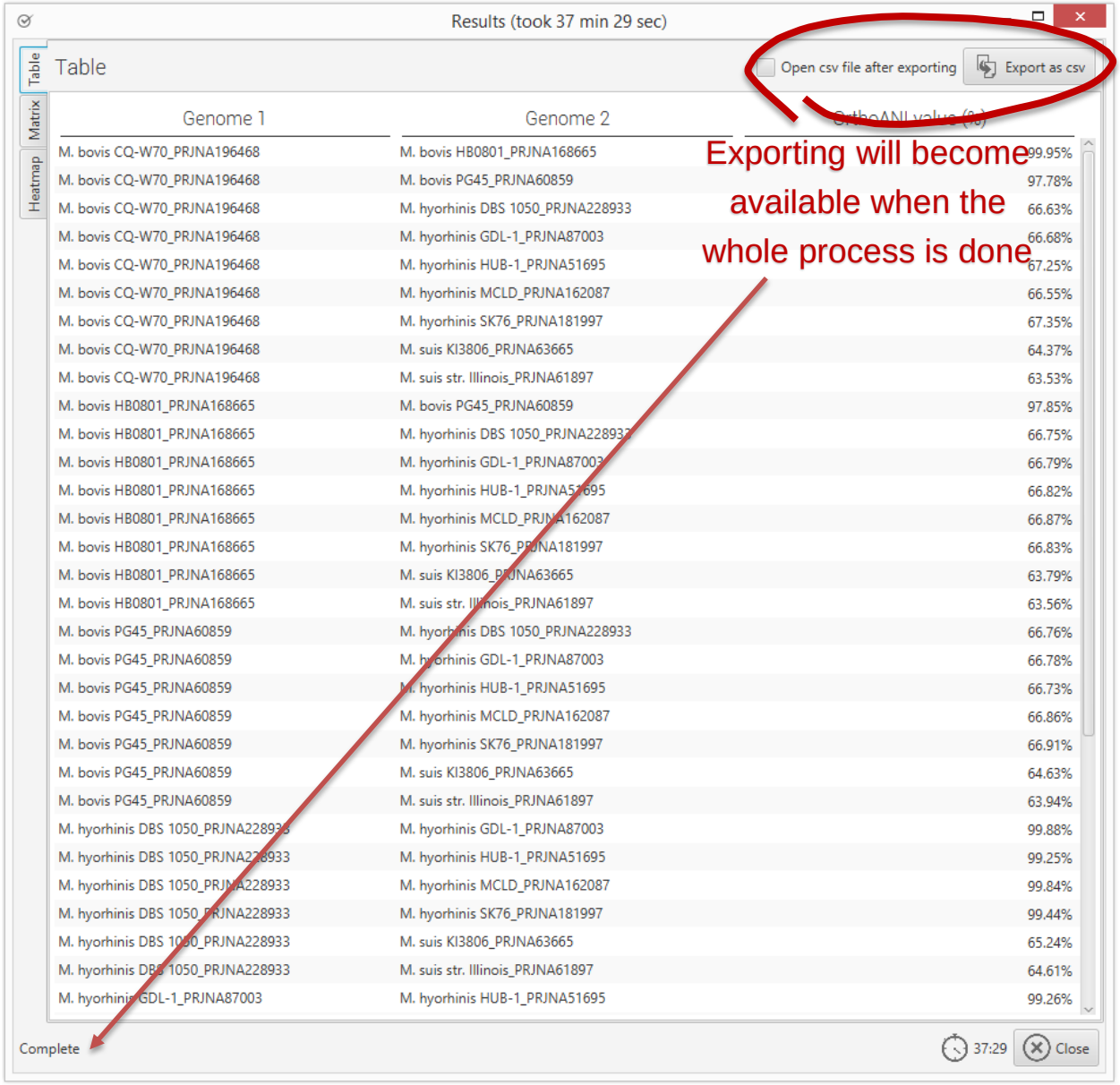


4.2 Result Screen

There are 3 result formats (all populated with the same data) which could be navigated through the tabs. All of them will dynamically update themselves as the OrthoANI values are being computed.

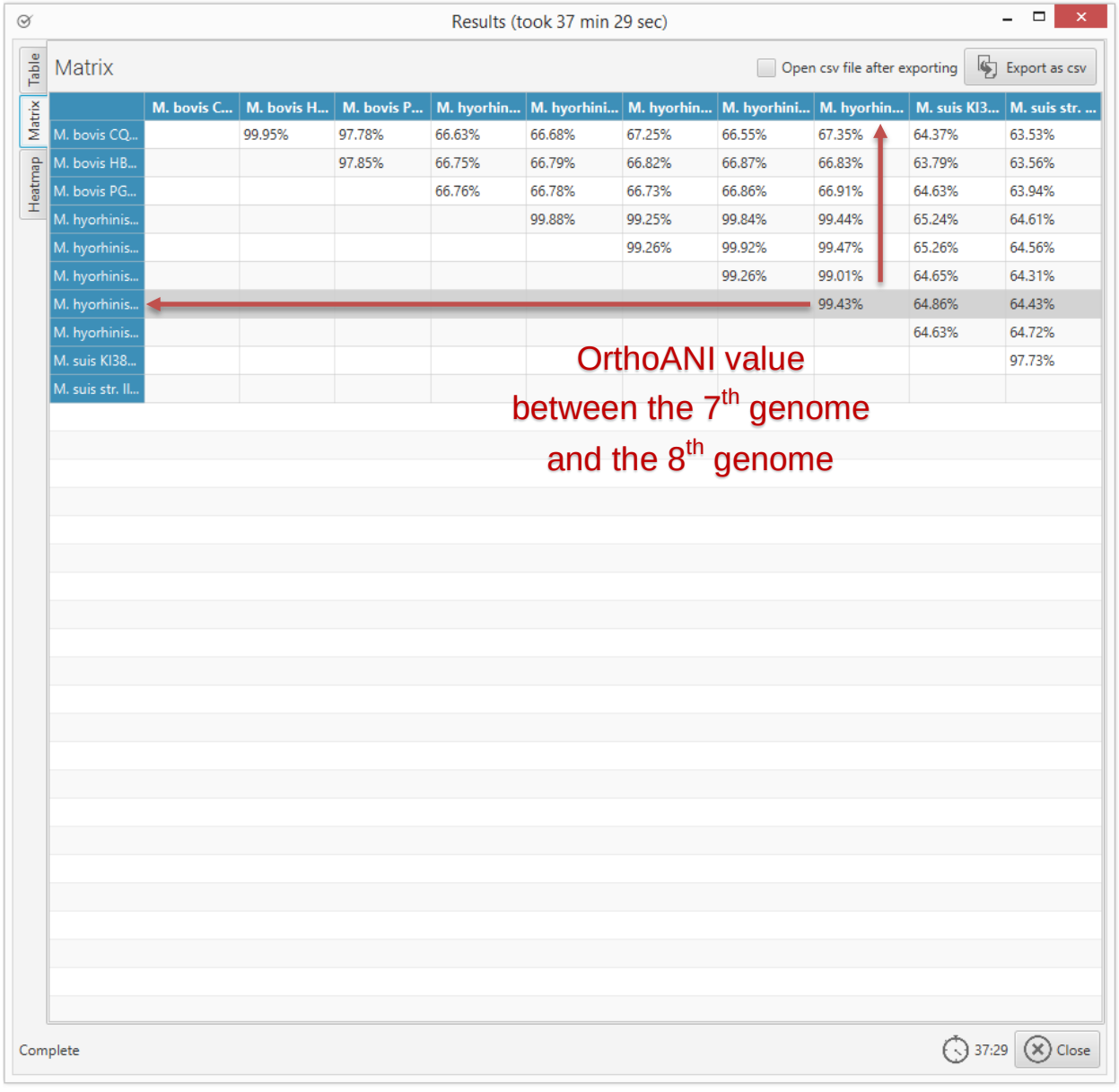
4.2.1 Table View

The table view has a basic format with 3 columns; name of the two genomes being compared and their OrthoANI values. There will be N*(N-1)/2 rows (N being the number of genomes) with a single header row. You can export the results as a CSV file once the whole calculations process is complete.



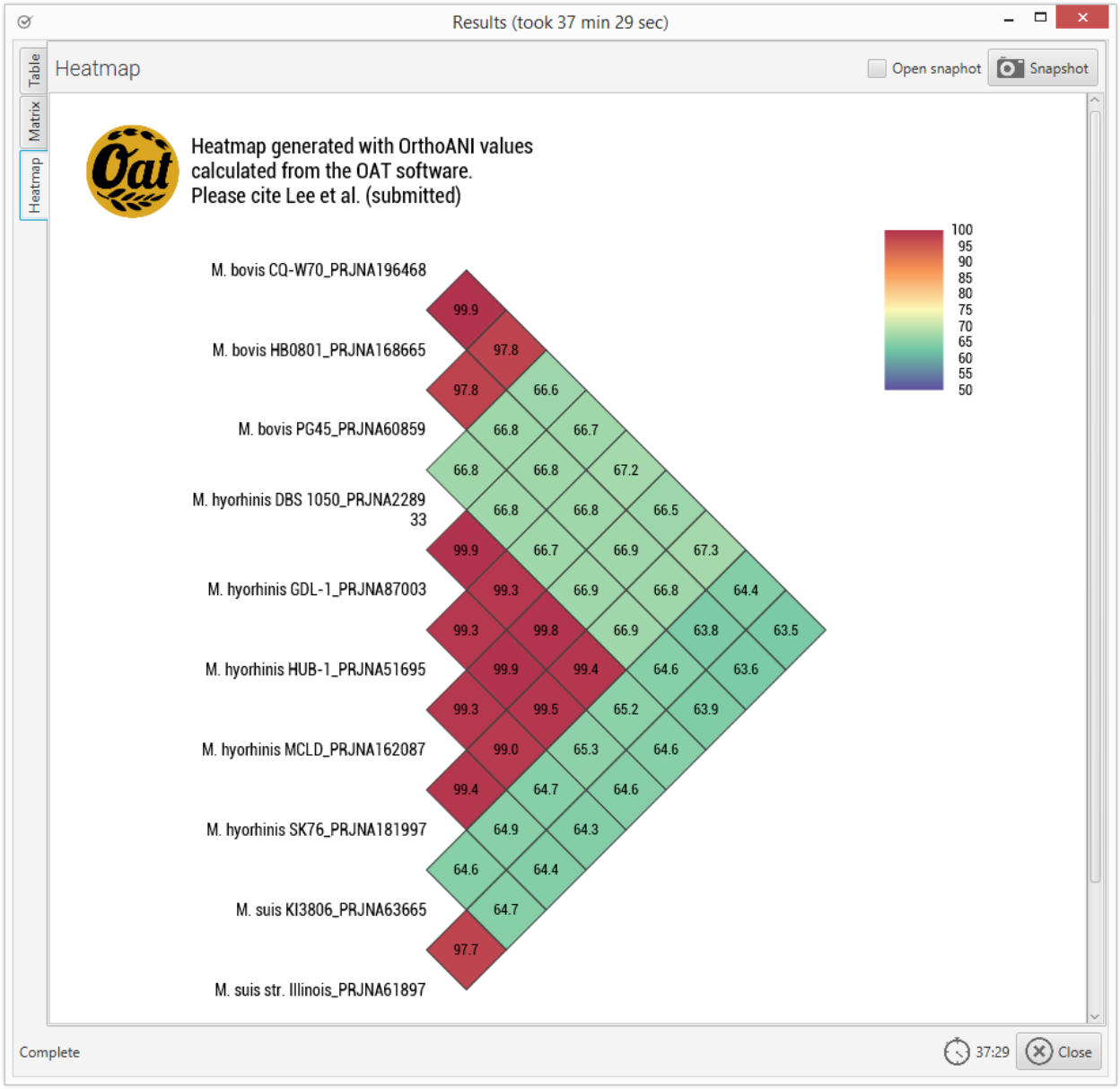
4.2.2 Matrix View

The matrix view has columns and rows for each genomes making it a N by N table (N being the number genomes). Each cell will represent the OrthoANI value for its column (Genome # 2) and row (Genome # 1). You can export the results as a CSV file once the whole calculations process is complete.



4.2.3 Heatmap

The Heatmap is very similar to the matrix view but uses colors to indicate the OrthoANI values. This results in a more intuitive view. You can take a snapshot (PNG file) of the heatmap at any point of the process.



4.3 Support

For any difficulties or questions that you have for the OAT application, please reach out to us via lebmaster.snu+oat.help@gmail.com