**CASPER**

**C**RISPR **A**ssociated **S**oftware for Pathway **E**ngineering and **R**esearch

The Manual

Brian Mendoza

Cong T. Trinh, Ph.D.

Department of Chemical and Biomolecular Engineering

University of Tennessee, Knoxville

April 2016

Version 1.0

**Preface.** This program is designed to assist the researcher/scientist/engineer in utilizing the genome editing tools of the CRISPR/Cas system. We stand on the shoulders of giants and hold ethical and moral responsibilities with the power this system provides. Please use for the betterment of the world, not its destruction.

TABLES OF CONTENTS

[I. INSTALLATION 4](#_Toc447983198)

[I.1. System Specifications 4](#_Toc447983199)

[I.2. Download 4](#_Toc447983200)

[I.3. Program Layout 4](#_Toc447983201)

[I.4. Changing Directories 5](#_Toc447983202)

[II. QUICK START 6](#_Toc447983203)

[III. GENERAL FEATURES 7](#_Toc447983204)

[III.1. Gene Search 7](#_Toc447983205)

[III.2. Organism and Endonuclease Support 7](#_Toc447983206)

[III.3. Custom Sequence Search 9](#_Toc447983207)

[III.4. Genome Upload 9](#_Toc447983208)

[III.5. Exporting Target Sequences 10](#_Toc447983209)

[III.6. On-Target Activity Analysis 11](#_Toc447983210)

[III.7. Off-Target Activity Analysis 11](#_Toc447983211)

[APPENDIX 13](#_Toc447983212)

[REFERENCES 14](#_Toc447983213)

# I. INSTALLATION

## I.1. System Specifications

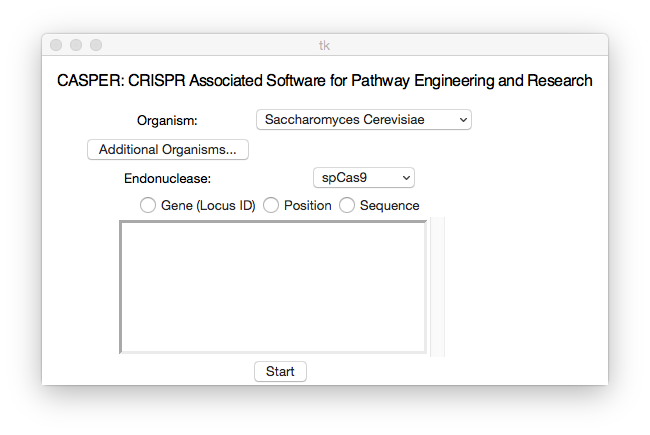
CASPER requires an operating system of Windows XP or later or Mac OS X with a minimum of 1.0 GB of RAM and 255 MB of hard drive space.

## I.2. Download

CASPER is available for download at *http://www.utk.edu/~ctrinh/download*. After download, place the program in your desired directory.

## I.3. Program Layout

When initializing CASPER, the default screen shows the parameter input screen as shown in FIG. 1.

**FIG. 1**: User-interface CASPER

## I.4. Changing Directories

When CASPER is downloaded, it uses the local directory for all files it creates. Some users may find it favorable to create a different directory for files such as the downloaded genomes (should one want to save them) or the indexed files (CASPER database files) created when a genome is run through the “Create New Genome” functionality. To change directories for information, simply go to the menu and select Genome → Change Directory.

# II. QUICK START

Below is a Quick Start description for using CASPER with the user interface.

**Step 1:** Upon starting the program simply select the organism of choice from the dropdown menu along with an endonuclease from the second dropdown menu.

**Step 2:** In the text entry field, enter the desired loci for investigation, either by the identification number or the position, specified by the chromosome number, followed by the nucleotide start and end positions. A sample data input screenshot in FIG.2 shows a standard input for *Y. lipolytica* and two genes of interest using the spCas9 endonuclease.

**Step 3:** Click “Start” and after CASPER completes compiling the data, a new button labeled “Show Results” will appear. Click this button to open a new window with the results of your search.

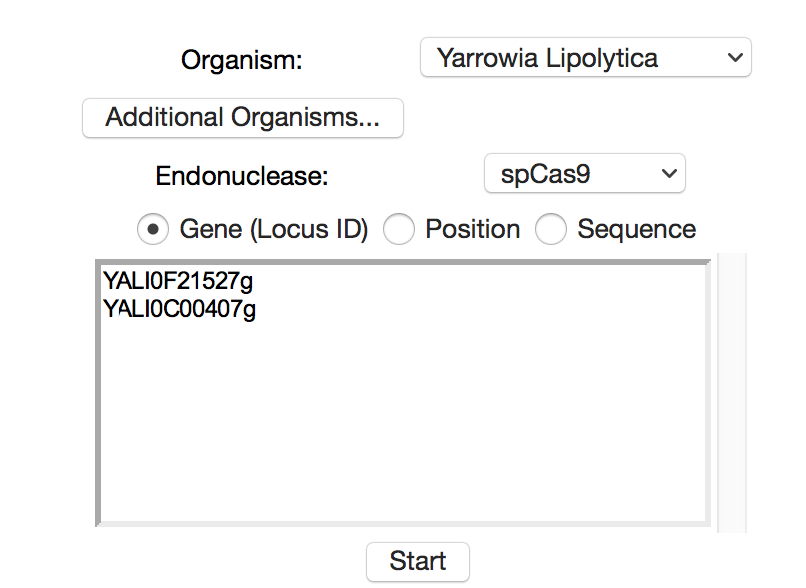


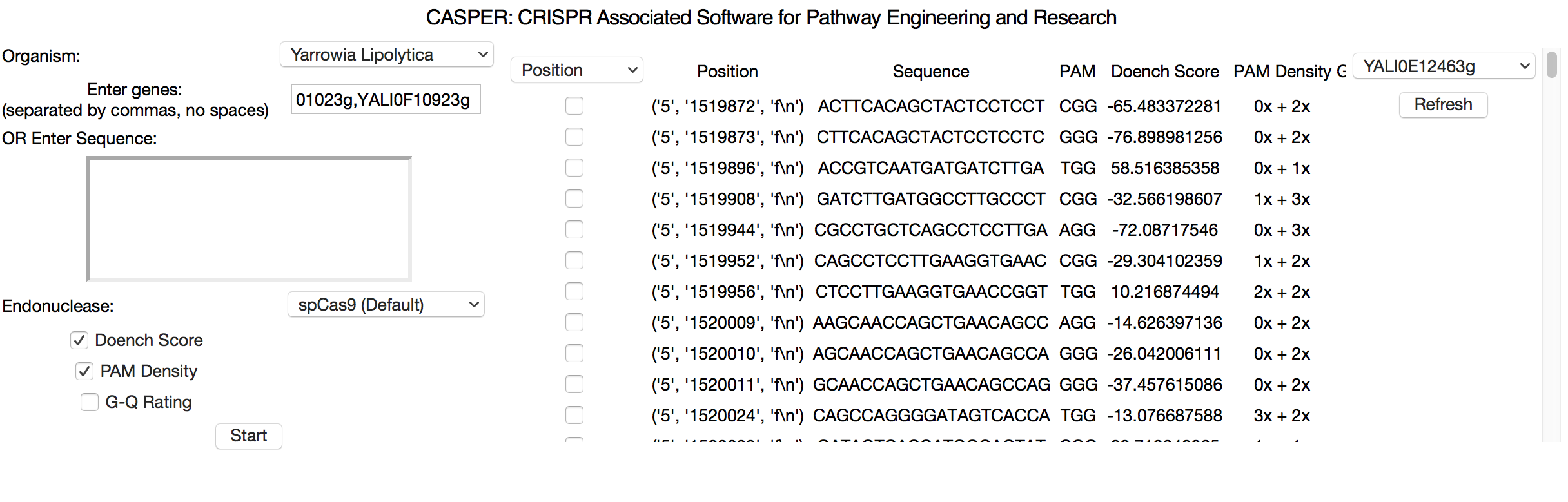
FIG. 2: Sample CASPER data input

# III. GENERAL FEATURES

## III.1. Gene Search

The standard method of analysis uses the user input with a single or list of gene names. The gene names must be given to the program with the correct locus identification number so that it may properly obtain relevant information from KEGG.

**Output:** A sample output screen showing the results of an analysis of three genes from *Y. lipolytica* using the KEGG database and the default spCas9 (FIG. 3). The list of sequences appear on the right side of the program where they can be sorted by any of their features, most notably the score from algorithmic analysis and the density of PAMs within the region.



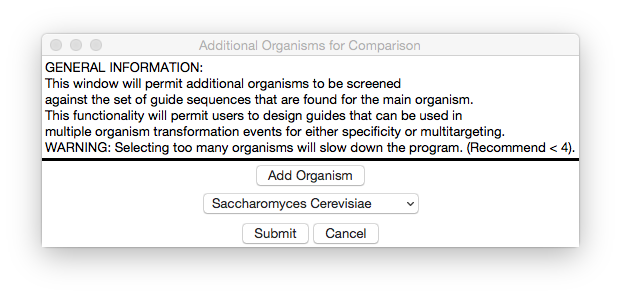
**FIG. 3:** CASPER output

## III.2. Organism and Endonuclease Support

**Organism Support.** The organism that you are searching through must be selected with the dropdown menu, else an error will occur when beginning the search process. The organisms on the dropdown menu are derived from the pre-curated genomes available. For more information on how to add a genome to this list, see the Genome Upload section.

To perform the analysis for more than one organism, a button labeled “Additional Organisms” appears below the Organism dropdown menu (FIG. 1). Selected organisms’ genomes will be included in the process of off-target analysis, and any targets with an exact match on one of the additional organisms will be highlighted in the results section.

To select an additional organism, click the button and a new window will appear with directions and warnings shown below (FIG. 4). For every additional organism you would like to check across click the “Add Organism” button and select the organism from the dropdown menu. When you are finished click “Submit”. The organisms you selected will appear next to the “Additional Organisms button on the start screen.



**FIG. 4:** Instruction for adding multiple organisms for analysis in CASPER

**Endonuclease.** One of CASPER’s unique features is the ability to search for target sequences of non-traditional Cas-endonucleases. The preloaded options are:

* spCas9 (PAM: NGG)
* spCas9-VRER variant (PAM: NGCG)
* spCas9-VQR variant (PAM: NGAG)
* stCas9 (from *Streptococcus thermophilus,* PAM: NNAGAAW)
* nmCas9 (from *N. meningititis,* PAM: NNNNGATT)
* FnCpf1 (PAM: TTTN, reversed binding)
* LbCpf1 (PAM: TTTN, reversed binding)

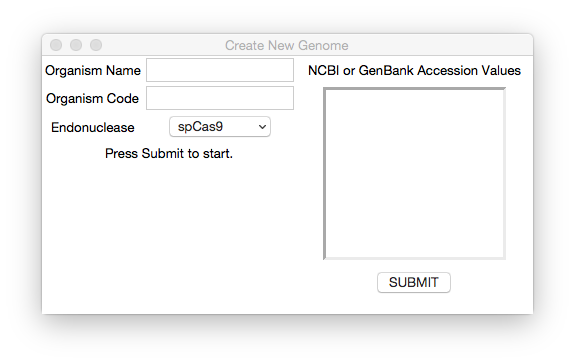
To upload another endonuclease, users can go to the menu bar, select Genome🡪Upload New Endonuclease and fill out the required information, including the name, PAM identification sequence, off-target PAMs and their relative binding to the standard PAM, standard seed sequence length, cut location, and direction of activity (forward or reverse strand).

## III.3. Custom Sequence Search

In addition to obtaining target sites from the input of a locus ID scanning KEGG, the user may also select the options of either entering the position of choice on the chromosome of choice, or a sequence of one’s choice. These custom search methods may be inputted in the same entry field as the locus IDs with the additional requirement of selecting the appropriate radiobutton above the entry field. An example of the appropriate entry style for the custom methods will appear above the entry field when a radiobutton is selected to ensure proper information entry to avoid errors.

## III.4. Genome Upload

If the user desires to add an organism to the available organisms, they must first create a CASPER database file, which obtains indices for occurrences of the desired PAM, and checks for repeat sequences across the genome. The window below (FIG. 5) appears upon accessing the menu and going to Genome → Upload New Genome.

**FIG. 5:** Instruction for creating new genomes in CASPER

Restrictions are as follows: the organism code must be the code used by the KEGG database to identify the genome in question. If this is not entered properly here, the program will be unable to locate the genes for the organism in question. Additionally, the NCBI/GenBank Accession Values must come from the nucleotide database. To be safe, copy and paste the identification codes listed in the table when accessing the genome from NCBI (They will likely start with “NC\_” or “CR” and will appear in columns labeled either “RefSeq” or “INSDC”).

## III.5. Exporting Target Sequences

Target sequences can be simply and quickly exported by selecting the desired sequences with the check button next to the sequence desired. Upon selection the information is stored temporarily so all selections can be made across the desired genes. When the user has finished selecting desired sequences, they may go to the menu and select Tools → Export to → Excel (.csv) for a list of the information as presented in the results window in comma separated value form.

## III.6. On-Target Activity Analysis

In order to rate the potential target, scores are assigned to each target based on the nucleotide pattern of the sequence and whether the corresponding sgRNA would be highly active(Moreno-Mateos, et al., 2015). In addition, this “on-target” score also takes into account the density of PAMs located inside of it and the percentage of guanine, cytosine, and cytosine, all shown to affect Cas binding(Malina, et al., 2015; Wang, et al., 2014). The score is on a scale from 0 to 100 with an increasingly higher score attributed to better sgRNA activity.

## III.7. Off-Target Activity Analysis

The potential for sgRNA to target more than one sequence is possible for a number of guides depending on sequence similarity across the genome. CASPER filters all identical sequences out on the initial pass of identifying possible target locations. In addition, users may choose to run further off-target analysis of potential sites identified by CASPER by clicking on the “Find Off-Targets” button in the column next to the sequence of interest.

Off target analysis is performed by assessing the probability of the guide RNA designed for the target of interest, binding to another site on the genome. This is done by comparing the guide RNA sequence to all other sites via a matrix that assesses a penalty for each mismatched base pairing (Hsu, et al., 2013). The combined penalties are then combined into a probability score. The scores range from 0 to 1, with 0 meaning virtually no chance of activity to 1 meaning inevitable activity. In addition to the matrix, the ratio of the activity (on-target) scores of both the target site and the off target sequence is taken into account. A high on to off ratio will reduce the likelihood of binding, whereas a high off to on ratio suggests increased binding affinity.

When the analysis is finished, the “Find Off-Targets” button label changes to “View Off-Targets”. When clicking on this button, the user is directed to a pop-up window displaying the top 50 off-target sites and their scores.

# APPENDIX

*KEGG* (genome.jp/kegg): KEGG is a powerful database with a variety of organisms and subsequent annotation on their genes and related pathways. CASPER utilizes this collection of information to obtain position and sequence information on genes requested by the user. For more information on using the KEGG database go to the website listed above.

*NCBI* (ncbi.com/genome): The NCBI genome and nucleotide databases are used by CASPER to download genomes of interest. Thus, the genomes available for CRISPR target analysis by this software is limited only by the number of genomes listed on the NCBI website.

# REFERENCES

Hsu, P.D.*, et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* 2013;31:827-832.

Malina, A.*, et al.* PAM multiplicity marks genomic target sites as inhibitory to CRISPR-Cas9 editing. *Nature Communications* 2015;6:10124.

Moreno-Mateos, M.A.*, et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Meth* 2015;12(10):982-988.

Wang, T.*, et al.* Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* 2014;343(6166):80.