

CASPER v2.0 Documentation

(<https://github.com/TrinhLab/CASPERapp>)



Trinh Lab
Department of Chemical and Biomolecular Engineering
University of Tennessee Knoxville
Last Updated: April 12, 2022

Contents

1	Introduction	1
1.1	Acknowledgements:	1
1.2	Getting Help:	1
1.3	Background:	1
1.4	CASPER Architecture:	1
2	Installation	3
2.1	Installing from Github Releases:	3
2.2	Installing from Source Code:	4
3	Navigating the Application	5
3.1	New Genome:	6
3.2	Main Window:	8
3.2.1	New Endonuclease:	10
3.2.2	NCBI Genome Browser:	11
3.2.3	NCBI Tool:	12
3.3	View Targets:	13
3.3.1	Gene Viewer:	14
3.3.2	Co-Targeting:	14
3.3.3	Export to CSV:	15
3.3.4	Off-Target Analysis:	15
3.4	Generate Library:	17
3.5	Multitargeting:	18
3.6	Population Analysis:	20
4	YouTube Tutorials	22
4.1	Workflow 1: Installing CASPER	22
4.2	Workflow 2: Analyze a New Genome	22
4.3	Workflow 3: Perform Single Gene/Locus gRNA Generation	22
4.4	Workflow 4: Define New Endonuclease	22
4.5	Workflow 5: Perform Co-Targeting Analysis	22
4.6	Workflow 6: Perform gRNA Library Generation	22
4.7	Workflow 7: Perform Multitargeting/Population Analyses	22

INTRODUCTION

Acknowledgements:

We thank the members of Trinh Lab for their suggestions, corrections, and testing of CASPER.

This research is funded by a DARPA YFA award (#D17AP00023) and the JDRD program at The University of Tennessee, Knoxville. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the funding agencies.

Getting Help:

Questions about CASPER may be directed to David Dooley at ddooley2@vols.utk.edu or posted on the [Issues](#) tab of the Github repository.

Background:

CRISPR-Cas systems have revolutionized biological research and industry alike. While these powerful systems have a broad range of applications, they are best known for their use as genome editors and nucleic acid detectors. Much of CRISPR-Cas' strength stems from its ability to specifically home in on a target sequence using a 20-40 base pair RNA sequence matched to a site on the genome/transcriptome. While there are certain restrictions on which sequences can be targeted, the design space for CRISPR-Cas guide RNAs (gRNAs) is magnificently immense. Therefore, development of rational sequence design tools has been necessary to expedite and improve experimental design.

Enter CASPER (**C**RISPR **A**ssociated **S**oftware for **P**athway **E**ngineering and **R**esearch), a tool which provides a flexible, robust platform for the analysis of non-model organisms and development of CRISPR-Cas tools capable of targeting a broad range of pathogens. CASPER is available for download at <https://github.com/TrinhLab/CASPERapp>.

CASPER Architecture:

CASPER is modular in nature and broken down into two distinct parts: back-end algorithms written in C++ and a graphical user interface (GUI) front-end written in Python (Figure 1).

The following algorithms are written in C++:

- **SeqFinder Algorithm:** Builds CSPr database, which includes on-target scoring. Source code available on [Github](#).
- **Off-Target Algorithm:** Performs off-target scoring of gRNAs. Source code available on [Github](#).

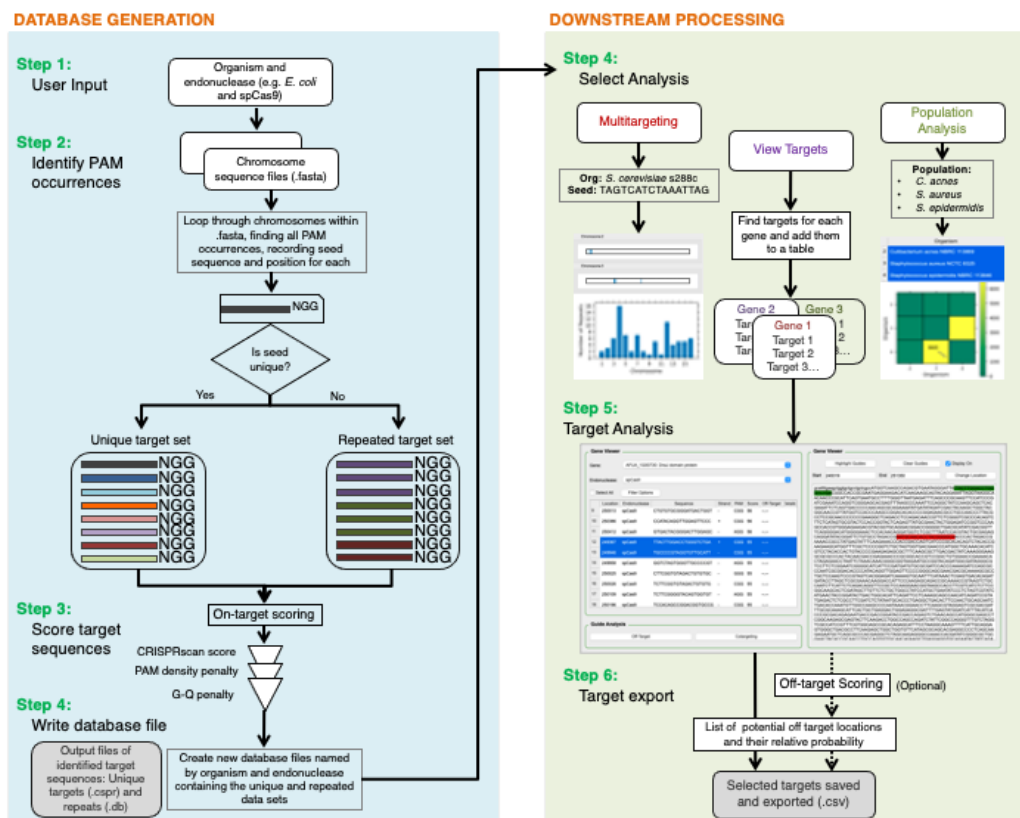


Figure 1: Overview of CASPER's architecture.

For more detailed information on how these algorithms were formulated, please refer to our previous publication [here](#).

The GUI consists of four main modules, written in Python:

1. **New Genome:** Upload a genomic DNA file (FASTA/FNA) to generate a custom CSPR database for gRNA generation and analysis.
2. **Main Window:** Locate and evaluate gRNAs for a given gene(s) or location(s) of interest within a CSPR file (requires cognate annotation file).
3. **Multitargeting:** Analyze a CSPR file for repeated gRNA motifs within an organism.
4. **Population Analysis:** Analyze multiple CSPR files for conserved gRNA motifs across a population of organisms.

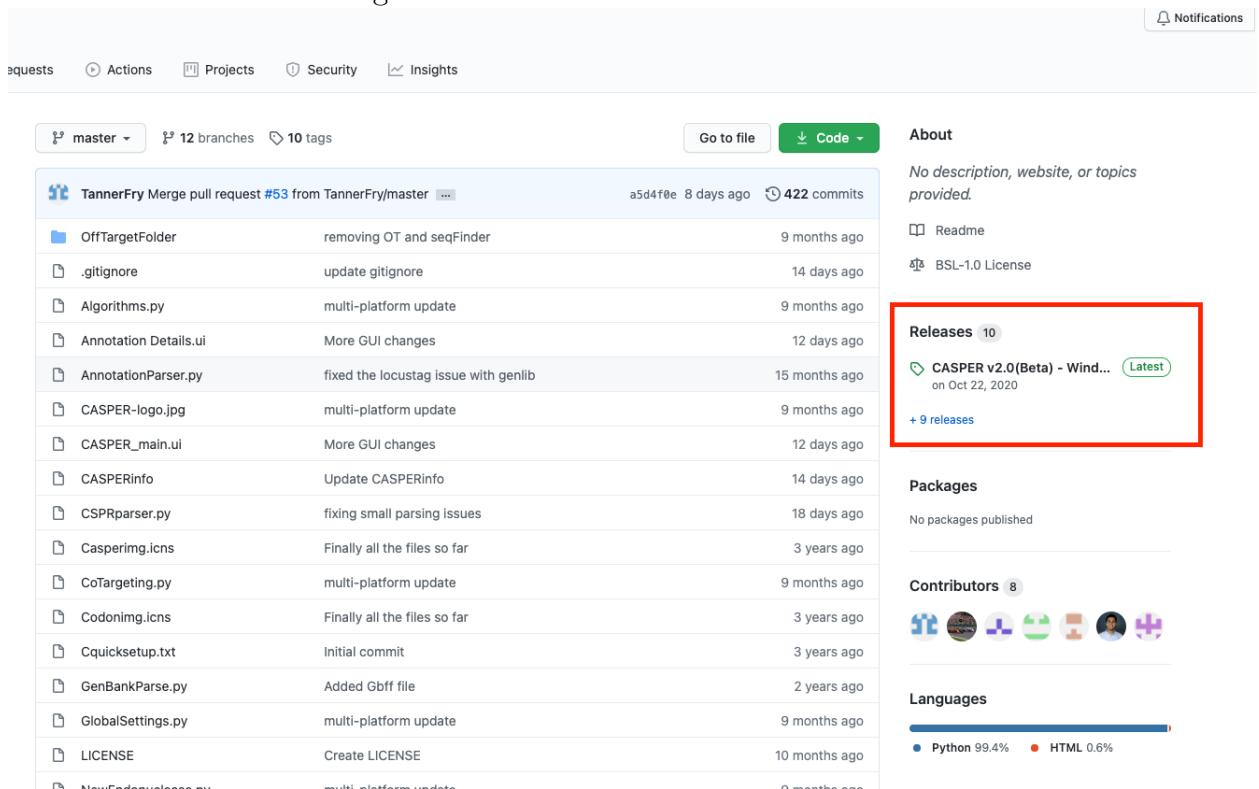
For detailed instructions on how to use CASPER, please refer to the **Navigating the Application** or **YouTube Tutorials** sections.

INSTALLATION

Installing from Github Releases:

To install CASPER from Github releases:

1. Navigate to the CASPER repository: <https://github.com/TrinhLab/CASPERapp>
2. Click on Releases on the right-hand side of the screen:



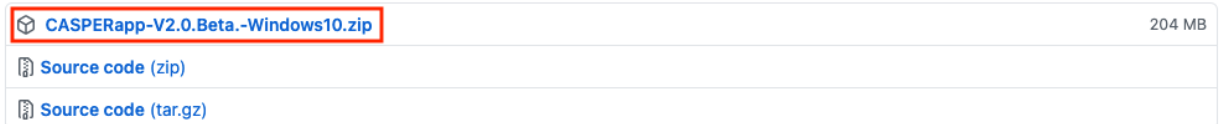
The screenshot shows the GitHub repository page for CASPERapp. The 'Releases' section on the right is highlighted with a red box, showing 'CASPER v2.0(Beta) - Wind...' as the latest release. The left sidebar shows a list of files and their commit history.

File	Commit Message	Time Ago
OffTargetFolder	removing OT and seqFinder	9 months ago
.gitignore	update gitignore	14 days ago
Algorithms.py	multi-platform update	9 months ago
AnnotationDetails.ui	More GUI changes	12 days ago
AnnotationParser.py	fixed the locustag issue with genlib	15 months ago
CASPER-logo.jpg	multi-platform update	9 months ago
CASPER_main.ui	More GUI changes	12 days ago
CASPERinfo	Update CASPERinfo	14 days ago
CSPRparser.py	fixing small parsing issues	18 days ago
Casperimg.icns	Finally all the files so far	3 years ago
CoTargeting.py	multi-platform update	9 months ago
Codonimg.icns	Finally all the files so far	3 years ago
Cquicksetup.txt	Initial commit	3 years ago
GenBankParse.py	Added Gbff file	2 years ago
GlobalSettings.py	multi-platform update	9 months ago
LICENSE	Create LICENSE	10 months ago
NewEndonuclease.py	multi-platform update	9 months ago

3. Scroll to the appropriate release for your operating system, click Assets, and download the installation package:

- Windows

- ▼ Assets 3



The screenshot shows the 'Assets' section for the CASPER v2.0(Beta) release. The 'CASPERapp-V2.0.Beta.-Windows10.zip' file is highlighted with a red box.

Asset	Size
CASPERapp-V2.0.Beta.-Windows10.zip	204 MB
Source code (zip)	
Source code (tar.gz)	

4. Follow the on-screen instructions for installing CASPER to your system. If on Mac or Linux, you must install the Source code file and follow the steps below, replacing Step 3 with the uncompressed Source code download.

Installing from Source Code:

To install CASPER from source code:

Note: this process is covered in detail in **Workflow 1**.

1. Ensure you have the following software packages installed and added to your path:

- Python 3.8: Download [here](#) or install via Homebrew (*recommended*)
- Git: Download [here](#)
- Virtualenv: Download with pip:

```
pip3 install virtualenv
```

- For Windows Users, Visual Studio Build Tools: Download [here](#)

2. Open a terminal/command prompt window and `cd` to the directory you want to download the CASPER files into
3. Clone the repository by executing the command:

```
git clone https://github.com/TrinhLab/CASPERapp.git
```

4. Create a virtualenv in the CASPERapp directory, specifying a 3.8 version of Python:

```
virtualenv <name of environment> -p  
<path to python 3.8 executable>
```

5. Initialize the newly created virtualenv

```
source <name of environment>/bin/activate
```

6. Upgrade pip and install dependencies from the same folder by executing the command:

```
pip3 install --upgrade pip &&  
pip3 install -r requirements.txt
```

7. Run the program by executing the command:

```
python3 main.py
```

Note: before running CASPER, Mac users may have to give permission to Casper_Seq_Finder and OT executables. This can be done by following this [tutorial](#).

8. Follow the instructions on screen to begin using CASPER. For more information on navigating CASPER, please refer to the **Navigating the Application** and **YouTube Tutorials** sections.

NAVIGATING THE APPLICATION

The CASPER GUI is composed of 4 primary modules, each of which carry their own features and functions:

- **New Genome**
- **Main Window**
- **Multitargeting**
- **Population Analysis**

This section will provide a basic understanding of what each of these 4 modules accomplishes, how to navigate them, and what inputs to provide them with. Detailed, step-by-step walkthroughs are deferred to the corresponding **YouTube Tutorials** at the end of this documentation.

The first window seen when starting CASPER is the launch window (Figure 2). The user starts by specifying a default directory for all CASPER-related files to be stored in. All files generated during CASPER analyses will be placed here, including CSPR database files (CSPR and DB files), genomic DNA files (FASTA/FNA), annotation files downloaded using the NCBI tool (GBFF), and exported gRNA files (CSV). If the user wishes to save files to a different location, they must either change the directory in the **Launch Window** or manually move the files outside of CASPER.

Next, the user is prompted to either proceed to the Main Window or New Genome. First-time users will have to generate a CSPR database in New Genome before continuing to Main Window. Once a CSPR file has been generated, the user may proceed directly to the Main Window when launching CASPER, if they so choose.

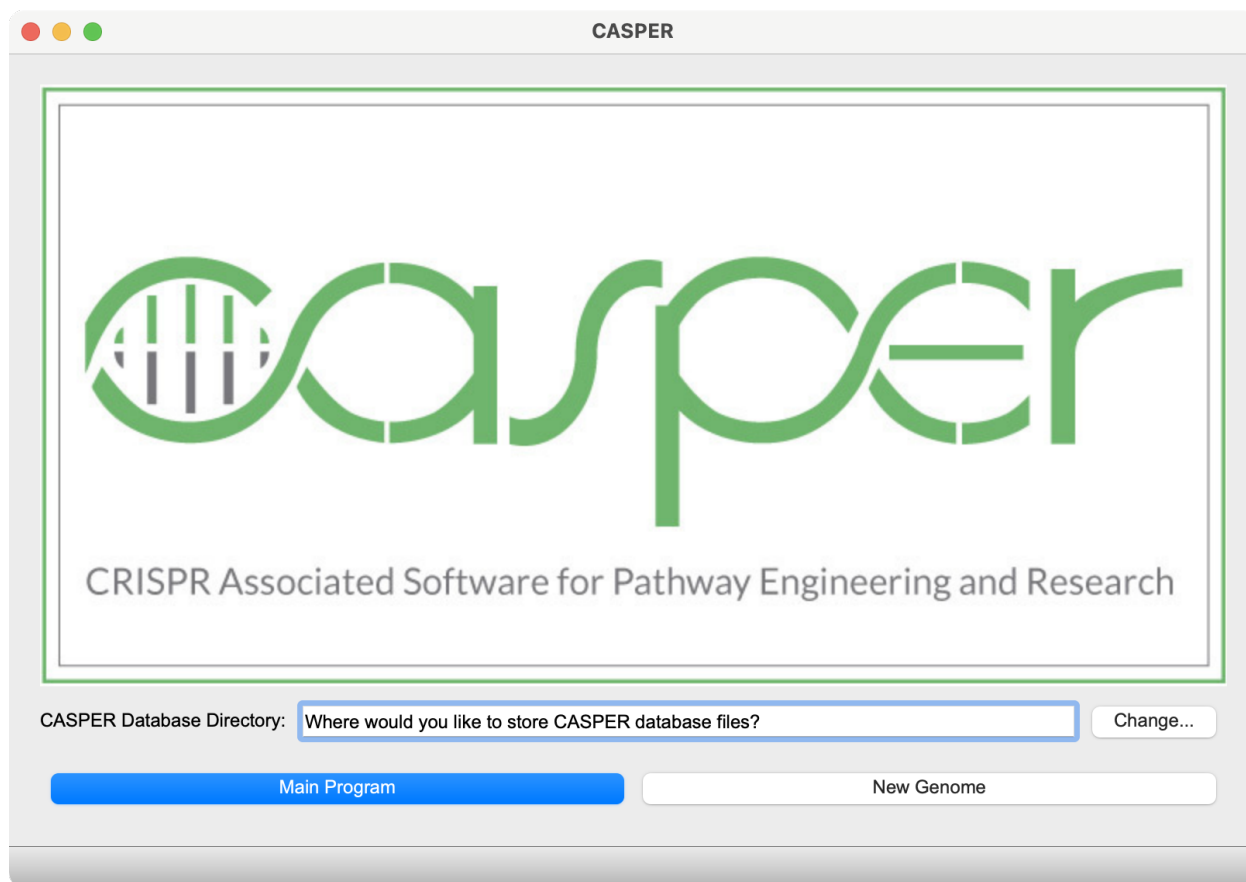


Figure 2: CASPER launch window.

New Genome:

New Genome is the module in which you may create new CASPR database files from an input FASTA/FNA file for the evaluation of unique and repeated gRNAs. It can be accessed via the **Launch Window** or the **Main Window** button: "Analyze New Genome". The window itself is essentially an entry form for the creation of a new CASPR file (Figure 3).

New Genome * Required

Step 1: Input Organism Info

* Organism Name: * Seed Length:

* Strain: 5' Length:

* Endonuclease: 3' Length:

* Organism Code: ☒ Generate Repeats ☒ Multi-threading

Step 2: Select File and Add Job to Queue

* Select Local FASTA/FNA File:

Step 3: Run Jobs

Job Queue	Job in Progress	Completed Jobs

Progress Output

Waiting for program initiation...

Figure 3: New Genome window.

The workflow is divided into 3 main steps:

1. Input Organism Info.
2. Select File and Add Job to Queue
3. Click "Run All".

The process of creating CSPR databases is covered in **Workflow 2** of the YouTube videos.

Note: for very large FASTA/FNA files (>1Gb), the runtime and memory requirements for CSPR database generation may become too great for the computer, depending on the robustness of hardware. In this case, we recommend breaking the file into multiple smaller files (~100Mb) of constituent parts (chromosomes/scaffolds/contigs), or only creating CSPR databases for the parts of the organism required for analysis.

Terminology:

- Organism Code: Manner in which CASPER will label its database files and references for the organism imported here. Ex: sau for *S. aureus* will yield files sau_<endonuclease>.cspr and sau_<endonuclease >_repeats.db
- Endonuclease: Cas enzyme to generate gRNAs for. Each enzyme has its own distinct targeting requirements (PAMs). Several have been preloaded into the program, but novel endonucleases may be added using the **New Endonuclease** feature.
- Seed Length: Length in nucleotides (nt) of the gRNA region in which binding is most crucial for Cas activity. CASPER's **SeqFinder Algorithm** uses this seed length and the endonuclease-specific PAM length to search for repeated gRNAs across the genome of interest.
- 5' Length: Length in nt of gRNA sequence upstream (5') of the seed sequence.
- 3' Length: Length in nt of gRNA sequence downstream (3') of the seed sequence.
- Multi-threading: Boolean (T/F) value for using multiple threads during the creation of CSPR files. It is highly recommended to leave multi-threading on, especially for larger genomes.

Main Window:

The Main Window (Figure 4) is opened by clicking "Main Window" in the **Launch Window**. It is the heart of CASPER, serving as the starting point for several different analyses and workflows, including:

- Monoplex/Multiplex Gene Targeting
- Multitargeting Analysis
- Population Analysis
- gRNA Library Generation

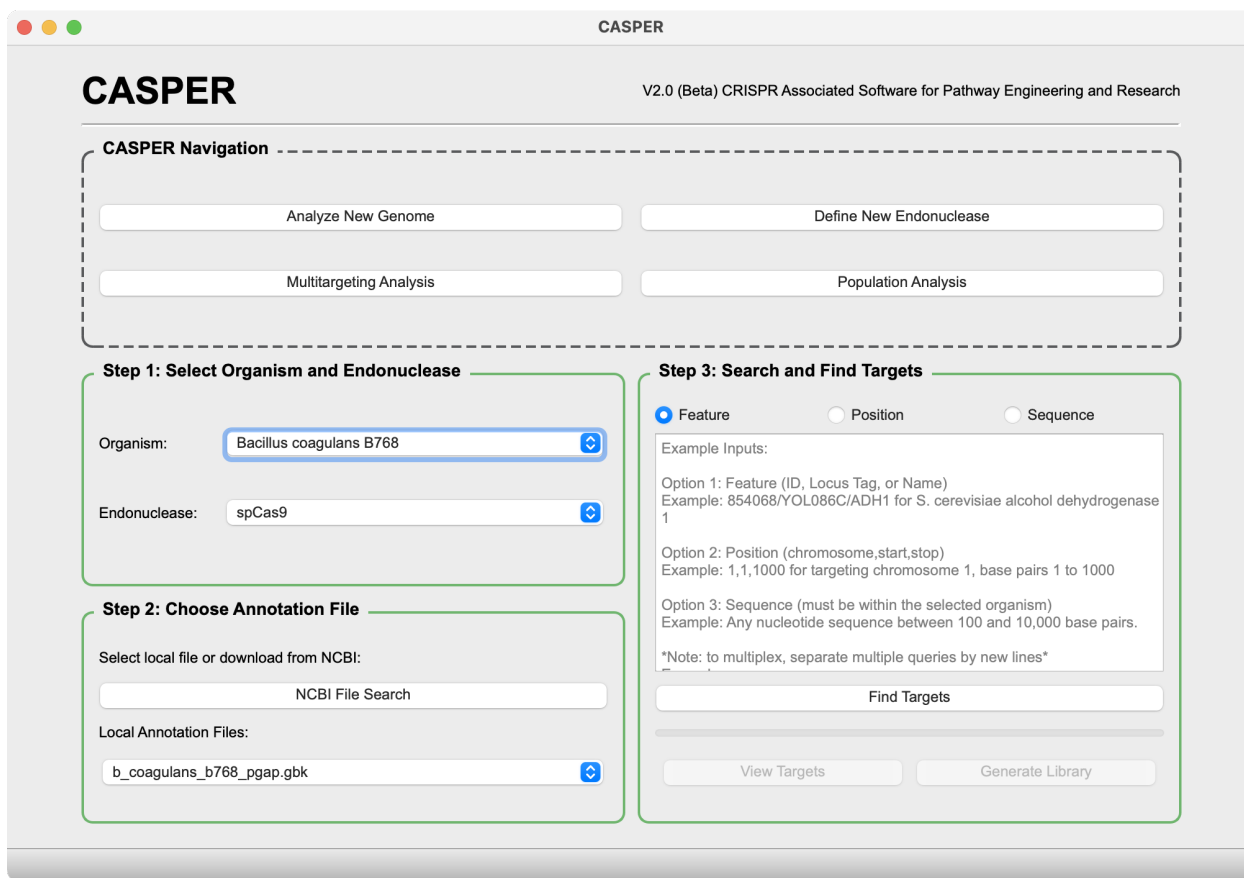


Figure 4: Main window.

The headlining feature of the CASPER software is the ability to do gene-based, location-based, or sequence-based gRNA generation using only an organism's genomic sequence file (FASTA/FNA) and annotation file (GBFF). It is important to note that the annotation file used for gene searches in this section must belong to *exactly* the same species and strain as the FNA/FASTA file used to create the CSPPR file. Having an annotation file mismatched on even the strain level could eliminate all gene targeting functionality, since annotation files are arranged corresponding to the layout of their cognate genomic sequence files, which may be organized in a variety of ways (chromosomes/scaffolds/contigs).

The basic steps for generating gRNAs for a gene or locus of interest are:

1. Select Organism and Endonuclease
 - All previously generated CSPPR files will automatically be populated from within the directory specified in the **Launch Window**.
2. Choose Annotation File
 - An annotation file may be downloaded manually using a web browser or by using the **NCBI Tool** provided in the Main Window/New Genome. Annotation files are automatically populated from within the directory specified in the **Launch Window**.

Note: the Feature and Sequence search options rely on GenBank-formatted annotation files, so if one is not available, users may only search using Position. To do this, simply select "None" from the list of Local Annotation Files and type the desired loci in the specified format (chrom,start,stop) with each new entry separated by a new line.

3. Search and View Targets

- Searches can be in the form of either gene names, location ranges, or sequence. Details of how to select appropriate genes and view gRNAs are included in **Workflow 3**. Once genes have been selected, gRNAs can be generated in one of two ways:

(a) **View Targets:** In-depth analysis and user-specified export to CSV.

(b) **Generate Library:** User-specified library settings with batch export to CSV.

In addition to the main gene targeting functionality, Main Window also contains smaller, useful features, including:

- **New Endonuclease**
- **NCBI Genome Browser**
- **NCBI Tool**

These features are discussed in greater detail below.

New Endonuclease: The New Endonuclease (Figure 5) feature allows the user to define a custom endonuclease, including its name, classification, PAM requirements, and gRNA sequence details (seed length, 5' length, 3' length, see **New Genome** section for definitions of these terms). Once submitted, the new endonuclease entry is saved into the CASPERinfo file that is located within the app directory. This tool is opened from the Main Window by clicking on the "Define New Endonuclease" button.

CASPER's on- and off-target algorithms can incorporate empirically-derived scoring matrices on an endonuclease-specific basis. For example, when scoring off-targets against spCas9 gRNAs, CASPER uses "HSU MATRIX-spCas9-2013" that is defined in CASPERinfo. New on- and off-target scoring matrices may be added to CASPERinfo manually, as long as they are formatted like the other datasets. When uploading a new endonuclease, the user is able to select on- and off-scoring matrices from dropdown menus automatically populated from the CASPERinfo file. This allows CASPER to retain its high-level predictive capabilities for on- and off-target scoring as new and improved datasets become available.

The user simply fills out the form, clicks "OK," and the new endonuclease definition is added to the CASPERinfo file. The user may then use their custom endonuclease in the creation of new CSPP files and subsequent analyses. This feature allows CASPER to nimbly adapt to emerging information as novel and engineered Cas enzymes are made available.

For a detailed demonstration of this process, please refer to **Workflow 4**.

New Endonuclease

Define New Endonuclease

* Required

Cas Details

* Organism:

Ex. Streptococcus pyogenes

* Abbreviation:

Ex. spCas9

* CRISPR Type:

Ex. II-A

gRNA Details

* Seed Length:

Ex. 16

* 5' Length:

Ex. 4

* 3' Length:

Ex. 0

* PAM Sequence:

Ex. NGG

* PAM Directionality:

☒ 3' PAM

☐ 5' PAM

Scoring Algorithms

* On-Target Matrix:

CRISPRSCAN

* Off-Target Matrix:

HSU MATRIX-spCas9-2013

Cancel

Submit

Figure 5: New Endonuclease window.

NCBI Genome Browser: The Genome Browser (Figure 6) feature allows users to visualize the architecture and orientation of the genes they are targeting by use of NCBI's genome viewer. Simply select the desired annotation file from the dropdown menu in Main Window and navigate to the Genome Browser via the toolbar: Genome → Open Genome Browser.

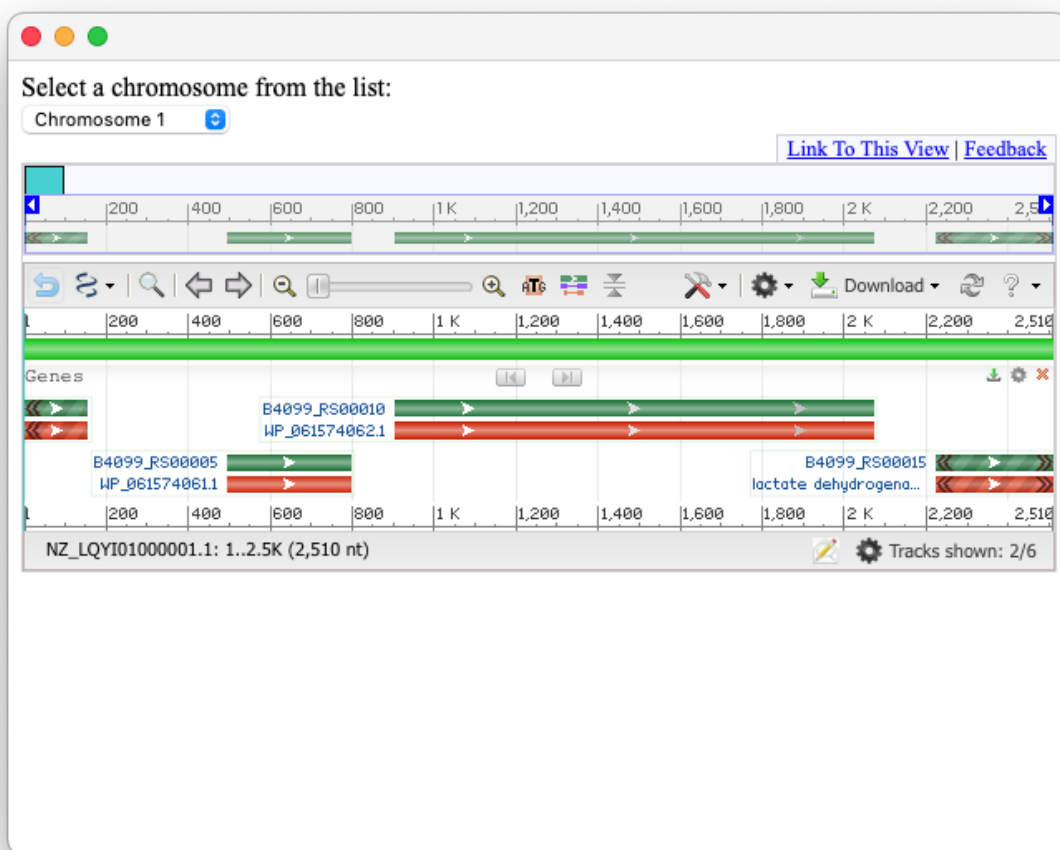
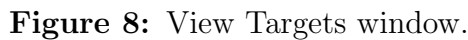


Figure 6: Genome Browser window.

NCBI Tool: The last and perhaps most useful side tool of the Main Window is the NCBI Tool, which is used for the automated retrieval of annotation and genomic sequence files from NCBI's databases (Figure 7). This allows users to query the vast NCBI GenBank and RefSeq collections without ever having to leave CASPER. A detailed demonstration for navigating this tool is included in **Workflow 2**.

It is *highly* recommended that users download GBFF files from the RefSeq collection if available, as these files tend to be better curated and contain more information. The gene feature and sequence search capabilities of CASPER cannot be guaranteed with annotation files from the GenBank collection.



Note: the Gene Viewer feature is populated based on the currently selected annotation file. If the annotation file selected does not match the organism selected, the Gene Viewer will display the wrong sequence. Also, if an annotation file isn't provided or is the wrong file type, the Gene Viewer will not work.

For example, because spCas9 and saCas9 have the 3' PAMs NGG and NNGRTT, respectively, genomic sequences adjacent to NNGRTT will be able to satisfy the PAM requirements of both endonucleases. These sequences can be said to be co-targets.

Page 14 of 22

all of the (compatible) endonucleases they wish to co-target with. Make sure all of the information entered into New Genome is the same for each organism, except for the endonuclease, so that CASPER will recognize the organism as available for Co-Targeting. Then, simply target a gene as normal and click on the "Co-Targeting" button in the View Targets window. CASPER will populate a table with all of the available Cas enzymes for that organism that are amenable to Co-Targeting. Lastly, highlight the desired enzymes and click "Submit" and guide RNAs will be populated in the View Targets table for use with each endonuclease as well as multiple endonucleases. To only view gRNAs that are compatible with multiple endonucleases, click "Filter Options" and check the "Co-Targets Only" box. For a detailed demonstration of Co-Targeting, please refer to **Workflow 5**.

Export to CSV: Export to CSV allows users to easily export gRNAs to a CSV file from the gRNA table in View Targets, as well as tables in **Multitargeting** and **Population Analysis**. To export, the user simply highlights the rows of the table they wish to export, specifies the destination directory and filename, and clicks "Submit". Optionally, users may choose to add a 3' or 5' motif to each of the gRNAs, such as a tRNA sequence to facilitate multiplexing. CSV files may easily be ported into a sheet-based editor like Excel or Google Sheets to facilitate facile manipulation or ordering of gRNAs.

Off-Target Analysis: Off-Target Analysis starts by highlighting the rows of the guides that the user wishes to analyze and then is launched by clicking the "Off-Target" button. A small window will pop up providing the user with various options for performing off-target analysis (Figure 9). Our proprietary Off-Target Algorithm is described in detail in our [previous publication](#).

Once all of the fields have been filled out, click "Run", wait till the progress bar shows that the scoring has been completed, and click "Submit" to return to View Targets, where the off-target scores will be shown in the table.

The image shows a software window titled "Off-Target Analysis". It is divided into three main sections, each enclosed in a green rounded rectangle:

- Step 1: Select References**
 - Organism: A dropdown menu with "Escherichia coli K-12" selected.
 - Endonuclease: A dropdown menu with "spCas9" selected.
- Step 2: Set Parameters**
 - Tolerance: A slider set to 0.05.
 - Max No. Mismatches: A dropdown menu with "4" selected.
 - Average Output: An unchecked checkbox.
 - Output File Name: A text input field containing "e_coli_offtarget".
- Step 3: Run Off-Target**
 - A large "Run" button.
 - A progress bar below the button, currently showing about 25% completion.

At the bottom of the window are two buttons: "Cancel" and "Submit".

Figure 9: Off-Target Analysis window.

Terminology:

- Organism: Reference to be examined for off-target hits.
- Endonuclease: Endonuclease whose PAM will be used to locate off-target hits. (Ex. spCas9 = NGG)
- Tolerance: Minimum off-target score for an off-target hit (lower means less chance of activity). Sites with off-target scores less than this value will not be categorized as off-target hits, meaning they will be treated as 0.
Note: computational burden increases as this value is decreased.
- Max No. Mismatches: Maximum number of mismatches relative to the gRNA within a putative off-target site that will be considered. All off-target sites with more mismatches than this value will not be considered for off-target analysis.
Note: computational burden increases as this value is increased.

-
- Average Output: Outputs averaged off-target scores only. Default output format is detailed and gives an average score, along with a breakdown of each contributing putative off-target site.
 - Check Non-Canonical PAMs: Allows users to search for off-target hits at sites with different PAMs than the canonical PAM(s) for the selected endonuclease.

Generate Library:

In Generate Library (Figure 10), users can quickly produce bespoke libraries of guides that adhere to specified criteria for a given gene or selection of genes. The criteria include library size and location, quality filtering options, and targeting restrictions. To run Generate Library, the user must first search for and select one or several genes in Step 3 of the **Main Window**. After submitting the selection, users may then click on "Generate Library" to continue to the Generate Library window. From there, the user simply fills out the fields with their desired parameters and clicks "Submit." A detailed video of this process can be found in **Workflow 6**.

Terminology:

- Number of Guides per Gene: Number of gRNAs that will attempt to be returned for each of the selected gRNAs. (Ex. if this option is set to 2 and 2 genes were selected, the library will contain 4 gRNAs, 2 for each gene.)
Note: if the "Modify Parameters If No Target Found" option is not checked, CASPER may return less than the requested number of guides due to there not being enough that fit the specified conditions.
- Targeting Range: Percentage of the gene from which guides will attempt to be pulled. (Ex. if wanting only gRNAs from the back half of the gene, the user would target from 50 to 100.)
- Find Off-Targets: Check this to filter library guides based on off-target score.
- Minimum On-Target Score: Minimum on-target score for guides included in library. (Higher is better. Higher on-target score means higher predicted on-target activity)
- Maximum Off-Target Score: Maximum off-target score for guides included in library. (Lower is better. higher off-target score means higher predicted off-target activity)
- Space Between Guides: Minimum amount of space that must exist between guides included in library.
- 5' End Specificity: Specifies the 5' characters of the guides included in library. For example, to optimize gRNA expression under some promoters, 5' G residues are needed. Typing "G" in the box returns only guides starting with G. Input must be "ACGT" alphabet.

- Modify Parameters If No Target Found: Checking this box will relax the specified library guidelines in order to always fill the Number of Guides per gene requested.

Figure 10: Generate Library window.

Multitargeting:

Users may navigate to Multitargeting by clicking the "Multitargeting Analysis" button in the CASPER Navigation pane of the **Main Window**. Multitargeting (Figure 11) allows users to view repeated gRNA motifs across an entire genome of interest at both global and granular levels. This is accomplished by searching for gRNA seeds, which are the portions of gRNAs most imperative for target acquisition and cleavage. In spCas9, the seed is the first 16 nt after the PAM, leaving a variable "tail" region of 4 nt for a 20 nt gRNA. Searching for seeds effectively expands the search space for multitargeting and population analyses.

On the left side of the interface, users are able to interact with specific gRNA seeds by clicking and highlighting rows of the table. Figures are automatically populated for the seed corresponding to the selected row, revealing 1.) The number of repeats per chromosome/scaffold, and 2.) The location of the repeats within the selected genome. In the bottom right corner of the Multitargeting window, genome-scale statistics and figures are included to give users a high-level overview of the repeat landscape in the selected genome. A breakdown of all the figures is included below:

- Seed Analysis:

Note: If multiple rows are selected, only data from the last row selected will be used to populate these figures.

- Scaffold/Chromosome Viewer: Red lines show the location of repeats for the selected seed. The black boxes represent either chromosomes, scaffolds, or contigs, depending on how the FASTA file was formatted when the CSPR files were created. Detailed information about each repeat is shown above the viewer upon hover.
- Seed Distribution: Shows how many repeats are on each chromosome/scaffold/-contig for the selected seed. If the number of chromosomes/scaffolds/contigs exceeds 10, some axis labels are removed to avoid labels overlapping, but the tick marks remain consistent with chromosome number.
- Global Analysis:
 - Statistics Overview: Provides basic, genome-wide multitargeting statistics for the selected organism.
 - Repeats per ID Number: Shows the total number of repeats for every seed in the selected genome. Seeds are arbitrarily assigned IDs.
 - Seeds per Number Repeats: Shows the number of unique seeds per repeat number. Y axis is scaled logarithmically for readability.
- Seed Search Table: Summary of gRNA seeds that are present in the selected organism. To further analyze a given seed, simply highlight the appropriate row in the seed table and the appropriate figures are populated.

Terminology:

- Total Repeats: Total number of times the seed sequence is repeated across all organisms analyzed
- Avg. Repeats/Scaffold: Average number of times the seed sequence is repeated per scaffold/chromosome/contig (varies depending on how the organism's FASTA file is arranged)
- Consensus Sequence: Full length guide RNA sequence that appears most commonly amongst all occurrences of the given seed.
- % Consensus: Percentage of all seed repeats that have the tail of the consensus sequence.

For a detailed demonstration of how to perform and interpret multitargeting analyses, please refer to **Workflow 7**.

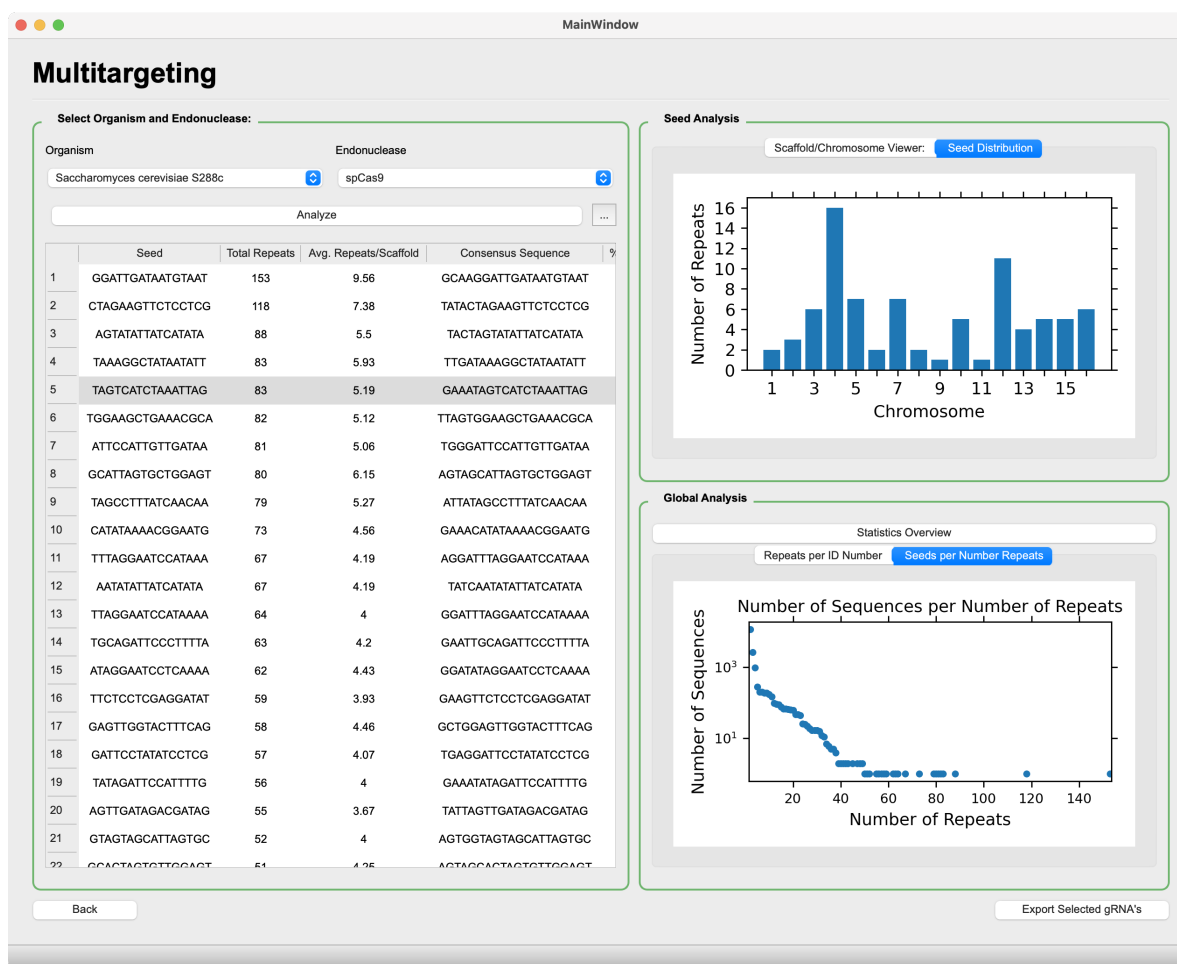


Figure 11: Multitargeting window.

Population Analysis:

Users may navigate to Population Analysis by clicking the "Population Analysis" button in the CASPER Navigation pane of the **Main Window**. Population analysis (Figure 12) allows users to select multiple organisms that have had CASPR files created for them and analyze the repeated gRNA seeds across the community. In this analysis, users will find:

- Shared Seed Heatmap: Pairwise distribution of repeated gRNA seeds for the selected organisms and endonuclease. Diagonals do not match color scale due to the high number of self-contained repeats relative to shared repeats, which skews the color scale. Hovering the cursor over tiles will show data callouts.

Note: Axis indices correspond to the index of the selection table in the top left of the window.

- Seed Search Table: summary of gRNA seeds that are present in all of the organisms analyzed. To investigate the individual occurrences of a seed sequence, simply highlight the appropriate row in the seed table and click "Find Locations."

Terminology:

- % Coverage: Percentage of analyzed organisms covered by the given seed sequence. (Ex. a coverage of 75% for an analysis of 4 organisms means that the seed appears in 3/4 of the organisms)
- Total Repeats: Total number of times the seed sequence is repeated across all organisms analyzed
- Avg. Repeats/Scaffold: Average number of times the seed sequence is repeated per scaffold/chromosome/contig (varies depending on how the organism's FASTA file is arranged)
- Consensus Sequence: Full length guide RNA sequence that appears most commonly amongst all occurrences of the given seed.
- % Consensus: Percentage of all seed repeats that have the tail of the consensus sequence.

To perform population analysis, select 1 or more CSPR databases from the table that are automatically populated from genomes previously analyzed in **New Genome**. Then, simply press "Analyze Organism(s)" to perform Population Analysis.

Note: Organisms must be analyzed with the same endonuclease within **New Genome** in order to be included in the same population analysis.

For a detailed demonstration of how to perform and interpret population analyses, please refer to **Workflow 7**.

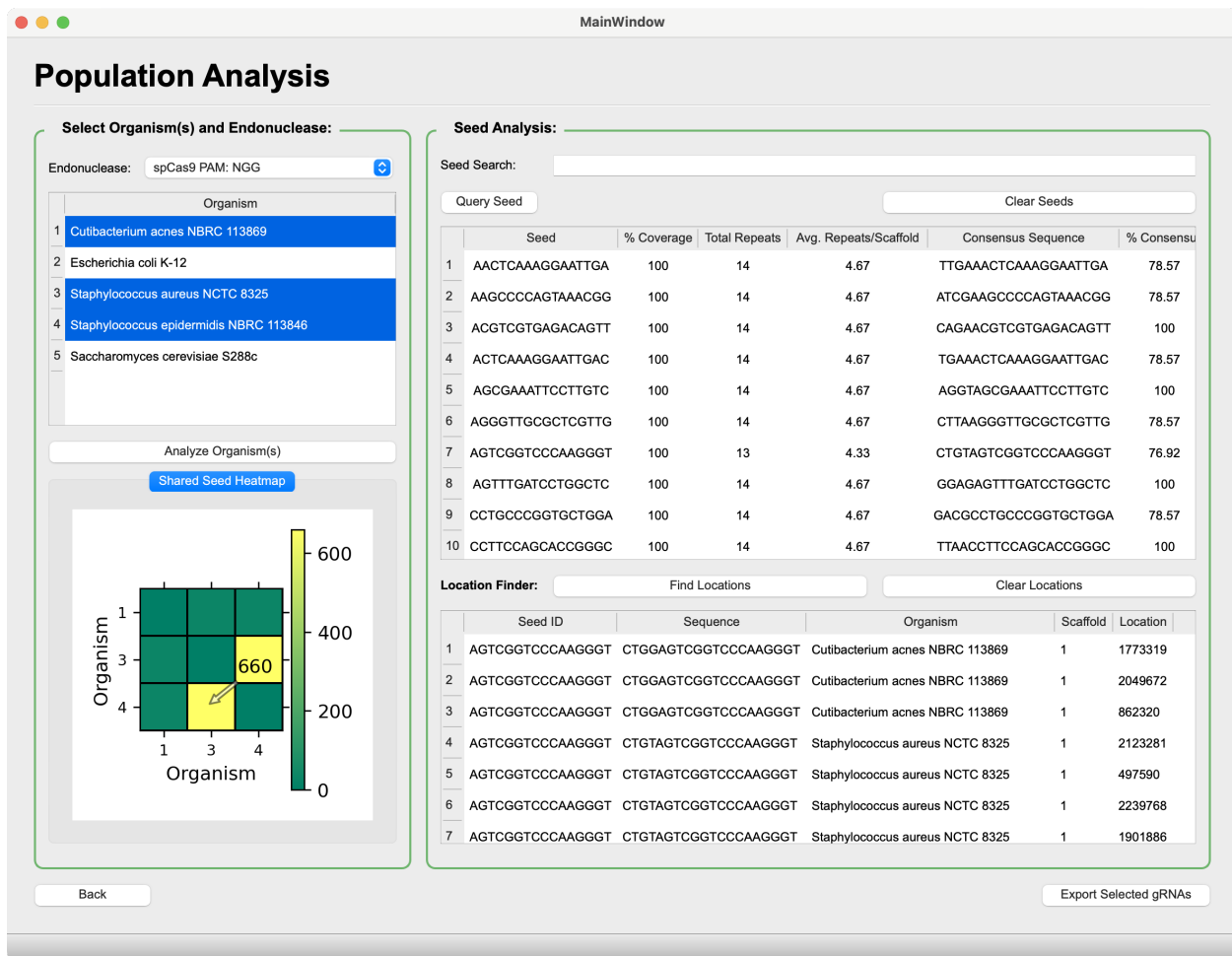


Figure 12: Population Analysis window.

YOUTUBE TUTORIALS

[Workflow 1: Installing CASPER](#)

[Workflow 2: Analyze a New Genome](#)

[Workflow 3: Perform Single Gene/Locus gRNA Generation](#)

[Workflow 4: Define New Endonuclease](#)

[Workflow 5: Perform Co-Targeting Analysis](#)

[Workflow 6: Perform gRNA Library Generation](#)

[Workflow 7: Perform Multitargeting/Population Analyses](#)