GPRO SUITE
BY BIOTECHVANA

SeqEditor

User Guide

1. OVERVIEW

1.1 - About SeqEditor

SeqEditor is a Sequence Browser for the management and analysis of nucleotide and protein sequences of up to 300 megabases. This enables it to manage the largest chromosomes currently known. SeqEditor is an updated version of TIME, a former sequence editor (Muñoz *et al.* 2011) also implemented as a sequence editor in the former release of GPRO (Futami *et al.* 2011). The new version provides improved updates of the functions (translation of nucleotide sequences into proteins, editing of sequences, changes in the geometry and orientation of the sequences etc.) and also provides new utilities for the searching and filtering of sequences, ORFs and motifs, managing multiple files simultaneously, primer design, metrics inference, and more.

Citing SeqEditor:

Hafez A, Futami R, Arastehfar A, Daneshnia F, Miguel A, Roig FJ, Soriano B, Perez-Sánchez J, Boekhout T, Gabaldón T, Llorens C. 2020. SeqEditor: an application for primer design and sequence analysis with or without GTF/GFF files. Bioinformatics, btaa903, https://doi.org/10.1093/bioinformatics/btaa903

Citing the GPRO suite:

Futami R, Muñoz-Pomer A, Viu JM, Dominguez-Escribá L, Covelli L, Bernet GP, Sempere JM, Moya A, Llorens C. 2011. GPRO: the professional tool for management, functional analysis and annotation of omic sequences and databases. Biotechvana Bioinformatics: 2011-SOFT3, http://biotechvana.uv.es/bioinformatics/index.php/article/35

1.2 - Version and Downloads

SeqEditor is distributed as an installer for Windows 7 or later (64 bit), a self-extracting disk image for Mac OS X 10.6 or later (64 bit), and a compressed tarball file for Linux 2.6 kernel series or later (64 bit). You can download the latest version of these executables at this link:

https://gpro.biotechvana.com/download/SeqEditor

1.3 - Installing SeqEditor in your PC

Every software application of the GPRO suite is a standalone Java app that requires a minimum 2GB of RAM or higher as well as the installation of a Java JRE (Java Runtime Environment) version 6 or superior.

To check if you already have a JRE installed on your computer:

Open a command line interface.

Type: java -version

Assuming that you have the java version 1.8.0_xx you should see a message like the following:

```
$ java -version
java version "1.8.0_92"
Java(TM) SE Runtime Environment (build 1.8.0_92-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.92-b14, mixed mode)
```

If you see a “Command not found” error message, it means that JRE is not installed. To install JRE, go to the official JRE repository at:

http://www.oracle.com/technetwork/java/javase/downloads/index.html

Download the version required for your operating system. Once installed, check again the output of java-version command on a command line interface as described above. Remember that sometimes even after JRE is installed, it is not set at the correct path.

To install the Windows version

Download the SeqEditor-win32.win32.x86_64.zip file and unzip it

Then browse to the executable file “SeqEditor.exe” and execute/run it.

To install the Mac version

Extract the archive to the desired destination using:

Download the SeqEditor-macosx.cocoa.x86_64.zip file and unzip it

Then browse to the executable binary file “SeqEditor.app” and execute/run it.

To install the Linux version

Download the SeqEditor-linux.gtk.x86_64.zip file and unzip it

Then browse to the executable binary file “SeqEditor” and execute/run it.

1.4 - RAM assignation to your PC

The amount of RAM memory assigned to SeqEditor can be tweaked in a configuration file named “SeqEditor.ini” that is located in the SeqEditor application folder on both Linux and Windows computers.

On MacOS computers, this configuration file is found by right-clicking on SeqEditor app → Show package contents → Contents → MacOS → SeqEditor.ini.

Open the file with any plain text editor and locate these two following parameters:

- xms1024 (minimum allocated memory)

- xmx2048 (maximum allocated memory)

Change the values according to the amount of physical memory available on your computer. For example, if your computer has 8GB of physical memory, it is recommended assigning Xms2048m and Xmx4096m for a better performance. If required, Xmx can be increased up to Xmx6144m for handling large datasets. Avoid using values close to the maximum available memory as it can lead to instability of the computer's operating system.

1.5 – Getting Familiar

SeqEditor can work as a fasta manager or as a graphical sequence browser. Figure 1 shows a screenshot of SeqEditor, whose layout is composed of four main sections: "directory browser" (controlling the information that displayed in a directory listing), "top menu" (allowing the user to access the sequence browser or to work with several sequences or fasta files at the same time), "sequence browser" (to graphically manage the sequence analysis), and "browser menu" allowing the users to analyze the current opened sequence at the sequence browser.

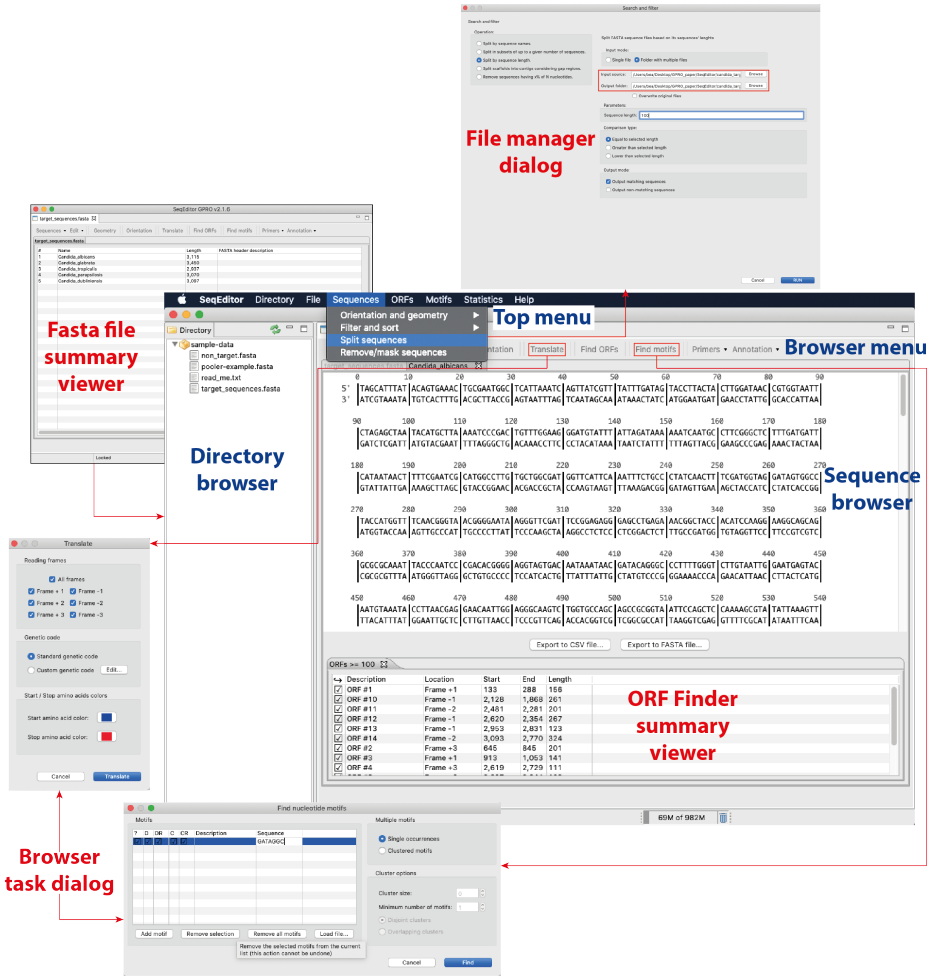


Figure 1. Overview of the SeqEditor application layout.

Additional interfaces are also available to run the analyses or display summaries for each browser's task. These include a "fasta manager dialog" that can be opened with any of the Top menu options and allows the user to analyze multiple sequences in a file or

several fasta files, the “fasta file summary viewer” that shows the sequences included in a fasta file and their lengths, the “ORF Finder summary viewer”, that shows the found ORFs in a specific sequence and their characteristics, and the ”Browser task dialog”, that can be accessed through the browser menu options and varies in accordance to those functions.

1.6 - Mouse actions

SeqEditor allows different functions by clicking on the directory browser or the Sequence Browser. By right clicking anywhere in the directory space, a dialog will appear enabling all the standard actions that allow the creation and management of files and folders. From this space, files can be cut, copied, pasted, deleted, and renamed. By right clicking on the sequence browser you can unlock sequences for further editing or copy and paste a sequence.

2. USAGE

2.1.- Top Menu

The **TOP MENU** provides access to the sequence browser and to other interfaces through which the fasta manager tasks can be executed. This allows the user to process and analyze all the sequences included in one or multiple fasta files simultaneously without the need of open the files separately.

The Top Menu is organized as follows:

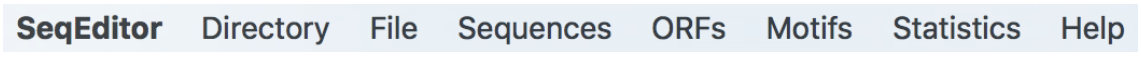


Figure 2. Main menu of the SeqEditor application.

TOP MENU: DIRECTORY

- [Directory → Select directory folder] **:** It selects the root folder shown in the Directory Browser.
- [Directory → Show] : view the Directory Browser.
- [Directory → Hide] **:** hide the Directory Browser.

TOP MENU: FILE

- [File → New]: Creates a new nucleotide or protein sequence file.
- [File → Open file]: Opens a new sequence file.
- [File → Save]: Saves the currently active file.
- [File → Save as]: Saves the currently active file as a new file.
- [File → Save All]: Saves all opened files in SeqEditor.

- [File → Close]: Closes the currently active file.
- [File → Save All]: Closes all opened files in SeqEditor.

TOP MENU: SEQUENCES

- [Sequences → Orientation and geometry]

  [Orientation and geometry → Change seq orientation in fasta files]: Changes the orientation of the sequences in one or more fasta files by selecting Reverse, Complement or Reverse complement option.

- [Orientation and geometry → Change seq geometry in fasta files]: Changes the geometry of the sequences in one or more fasta files by selecting either DNA or RNA.

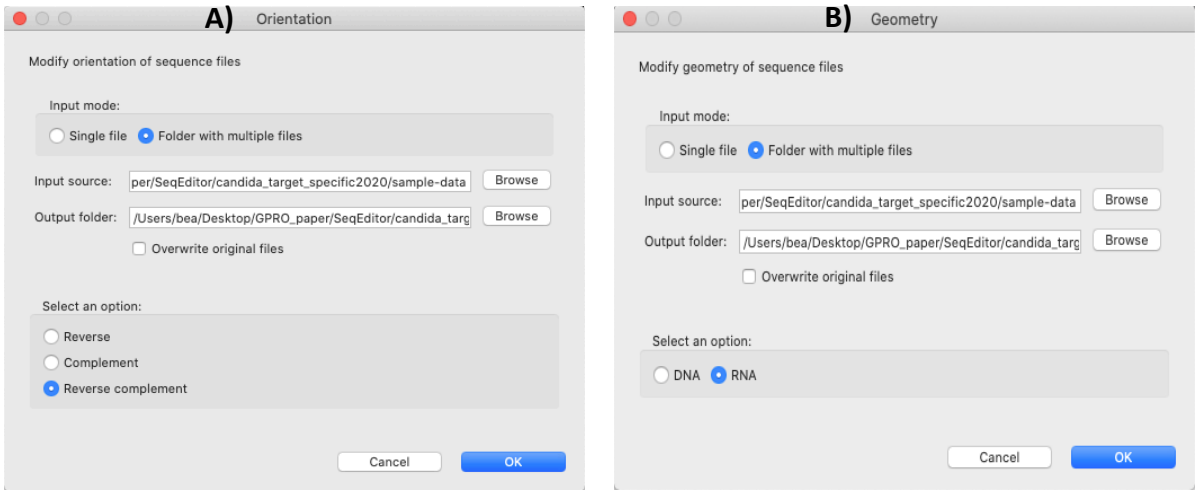


Figure 3. Orientation and Geometry windows layout. **A)** The orientation dialog allows the user to change the orientation of the sequences to their reverse, complement or reverse complement sequences; **B)** The geometry dialog allows the user to change sequences from DNA to RNA or vice versa.

- [Sequences → Filter and sort]

  [Filter and sort → Filter sequences in fasta files]: Filters sequences in fasta files using search terms provided in either a list of terms or a CSV file, selecting a matching criterion: Exact match, Partial match or Regular expression.

  [Filter and sort → Sort sequences in fasta files]: Sorts sequences in fasta files using either an alphanumeric criterion or in a forced manner by giving search terms contained in a CSV file.

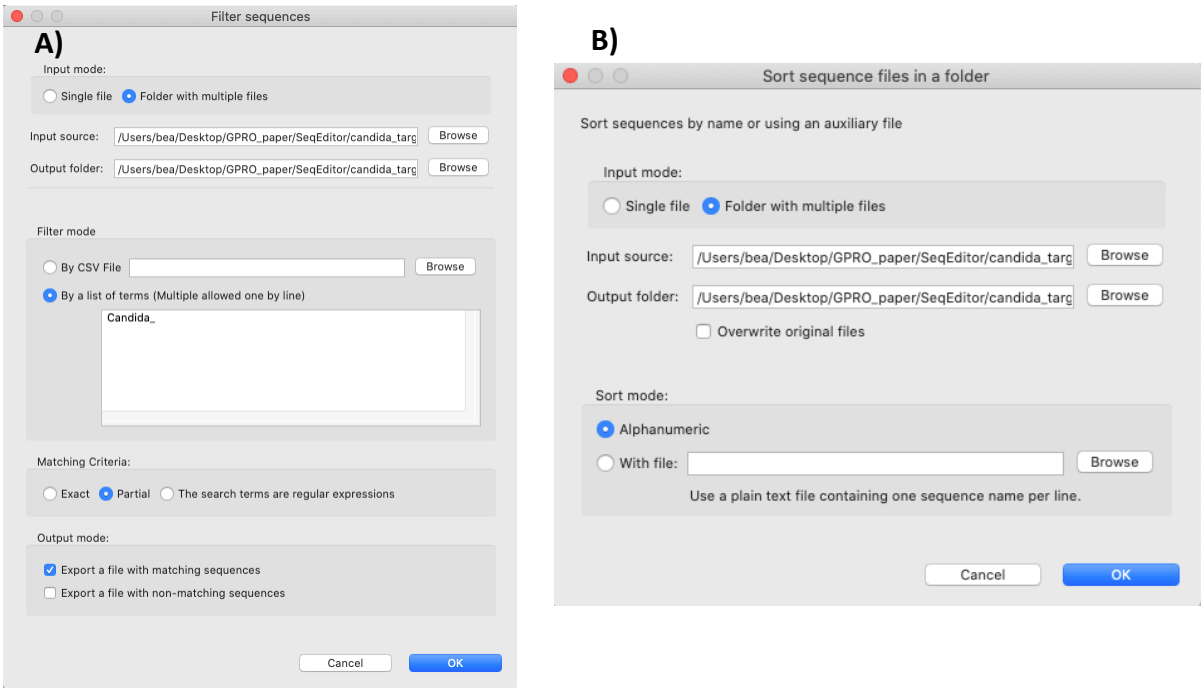


Figure 4. Filter and Sort windows. **A)** The ”filter” dialog allows the user to filter sequences by given search terms and matching criteria. **B)** The ”sort” dialog allows the user to sort sequences either alphanumerically or by a given term in a CSV file.

- [Sequences → Split sequences]: To split sequences of a given file according to different criteria.

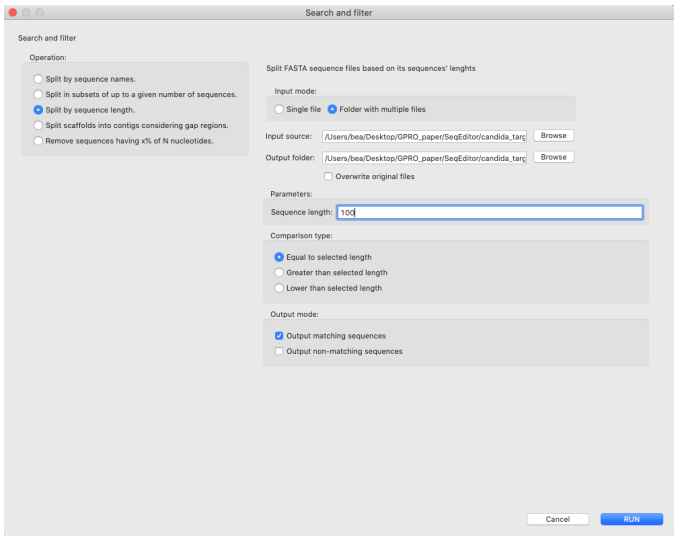
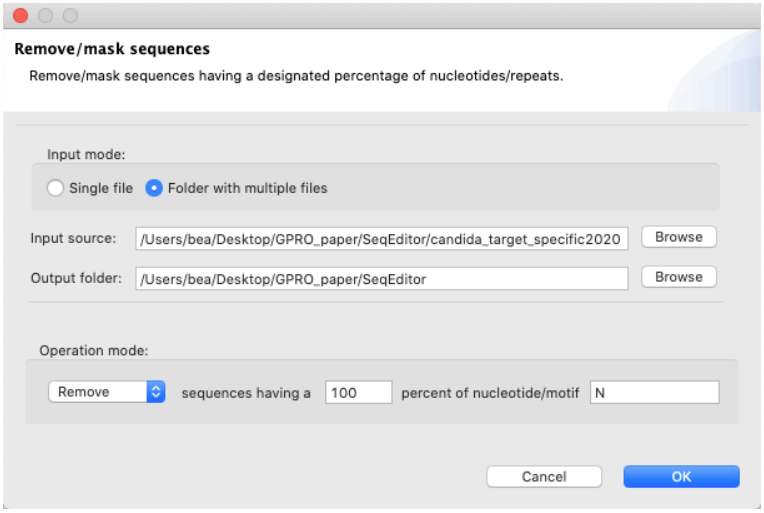


Figure 5. Split sequences dialog. The tool allows the user to: i) Split the sequences of your files by sequence names introducing list of names or a text file, including them in a select and match mode (exact or partial); ii) split in subsets of up to a given number of sequences introducing the number of sequences per file; iii) Split by sequence length introducing the desired length and select the comparison type: equal to selected length, greater than selected length or lower than selected length; iv) split scaffolds into contigs considering gap regions introducing sequence size and, if it is the case, a padding value.

- [Sequences → Remove/mask sequences]: Removes or masks sequences having a certain percentage of a specific nucleotide or motif.

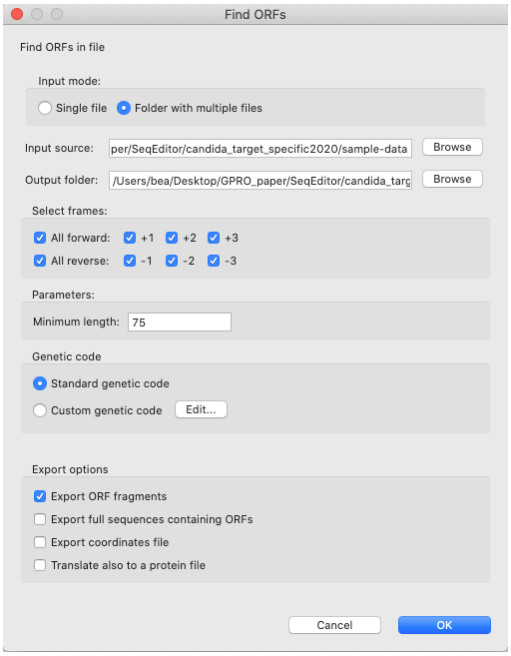
Figure 6. Remove/masks sequences from your fasta files. Using this parameter, the function will remove every sequence which is composed by a 100% of N.



TOP MENU: ORFs

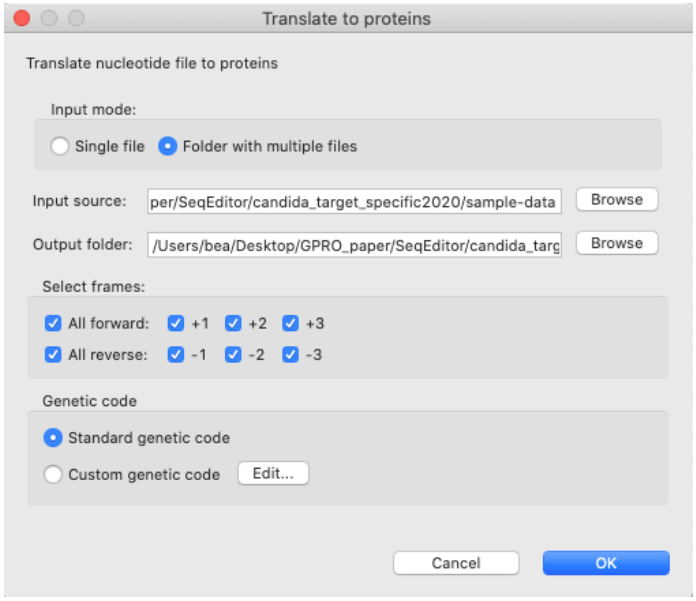
[ORFs → Find ORFs in fasta files]: Simultaneously searches and finds ORFs in one or more fasta files with multiple sequences.

Figure 7. "Find ORFs" searches and finds ORFs, simultaneously, in one or more fasta files with multiple sequences just specifying a minimum length and the open reading frames (both forward and reverse). Detected ORFs can be selected and exported or translated and exported as protein sequences. The tool also exports annotation files with the coordinates of the ORFs.



- [ORFs → Translate nucleotide sequences in fasta files to proteins]: Simultaneously translates to proteins all sequences contained in the fasta files.

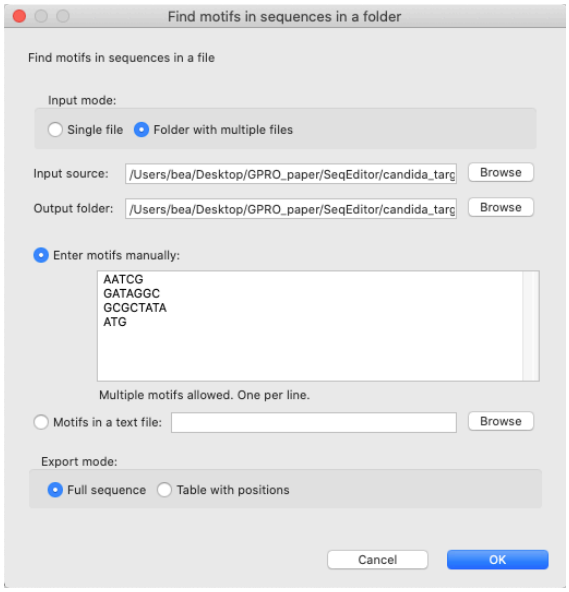
Figure 8. Translate to proteins dialog. Sequences contained in the fasta file can be translated into proteins by selecting the desired frame and genetic code. The output is a results file that contains the correspondent protein sequences obtained from each selected frame.



TOP MENU: MOTIFS

- [Motifs → Sequence motifs in fasta files]: Simultaneously searches and finds sequence motifs in all the sequences contained in the fasta files.

Figure 9. "Find motifs in sequences in a folder" dialog. By either introducing the motifs that need to be searched for either manually (one in each line) or by uploading them in a text file with the same format, this tool automatically searches for and finds these motifs in the selected fasta files.



TOP MENU: STATISTICS

- [Statistics → Infer sequence sizes in fasta files]: Simultaneously infers the size of all sequences contained in the selected fasta files.

B)

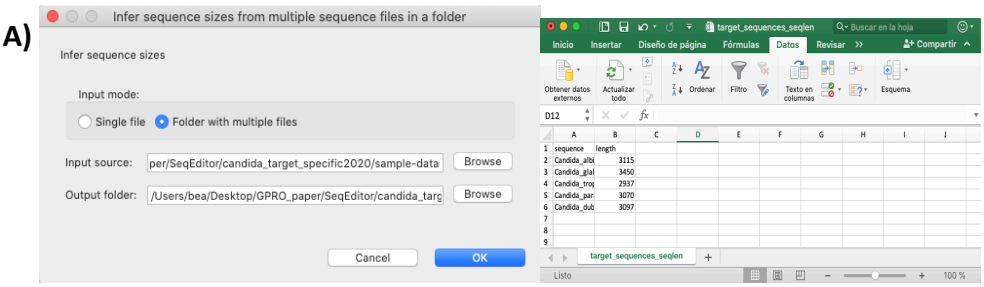


Figure 10. Infer sequence size tool. **A)** Infer sequence sizes dialog. **B)** Example of results file, containing the ID of the sequences and their correspondent length.

[Statistics → Infer overall metrics in nucleotide fasta files] : Registers a given set of metrics from the sequences contained in the fasta files, namely the number of sequences as well as the size of the largest and shortest sequence, N50 and L50 using a Java implementation of the Assemblathon Stats script (Bradnam 2011). This option is only available for nucleotide sequences.

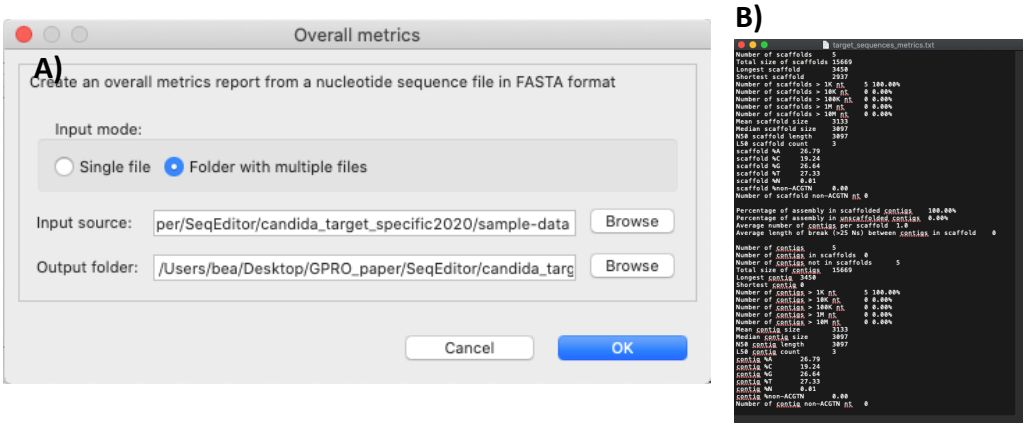


Figure 11. Sequence metrics inference dialog. **A)** "Overall metrics" dialog. **B)** Typical text file example of results obtained from this search, listing numbers of scaffolds/contigs, largest size found, N50 and L50 among others.

TOP MENU: HELP

- [Help → Manual] : A link to this manual online.
- [Help → About Worksheet] : Other technical details and copyright of the worksheet.

2.2- SEQUENCE BROWSER MENU

The SEQUENCE BROWSER is a GUI that allows the users to navigate, edit, and analyze sequences under visual control. It automatically opens when opening a fasta file by clicking on the top menu -> File -> Open File. It can also be called by right-clicking on an input sequence file in the specified directory. The menu of the sequence browser in the latest version of SeqEditor (v2.1.6) compiles all the different functions of the sequence browser and two new implementations that interact with the browser. These

are a GTF/GFF viewer tool that allows the user to mine and export sequence features using specific annotations of GTF/GFF files; and a set of tools for PCR primer design based on an interface adaptation of two CLI tools - Primer3 (Untergasser, *et al.* 2012) and PrimerPooler (Brown, *et al.* 2017). These tools include newly optimized and validated search processes for the design of multiplex and target-specific primers. The screen of the Sequence Browser is interactive, allowing users to select, copy, cut and paste sequence traits by right-clicking on the sequence. You can lock or unlock manual sequence editing by selecting Locked/Unlocked in the “Edit” tab of the sequence browser menu.

The graphic capacity of the sequence browser relies on the RAM power of the user's hardware; however, it has been optimized to manage large sequences, such as contigs, scaffolds and chromosomes without experiencing significant slowdown. This means that the users can work with sequences up to 300Gb with only 25Gb of RAM assigned to SeqEditor and will not experiencing any delays in processing times. These sizes efficiently cover the largest chromosomes known to date. Only one sequence is allowed per browser screen. However, when opening a fasta file with multiple sequences, a file summary view is opened at the bottom of the browser summarizing the sequences included in the file. Additional sequences can then be selected and analyzed in other screens of the browser. Lastly, the fasta file summary screen view presents a submenu that includes other tools for the editing and management of the listed sequences.

The Sequence Browser can manage both nucleotide and protein files and can automatically tell them apart. This means that if you create a new nucleotide layout in the sequence browser and try to copy a protein sequence (or *viceversa*) in it, you will be not allowed to do so.

The sequence browser manages the sequences one by one. If you are interested in managing files with multiple sequences or multiple files, you must use the top menu. The menu of the sequence browser is organized as follows:

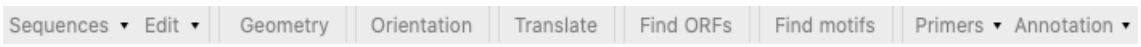


Figure 12. Sequence Browser Menu.

BROWSER MENU: SEQUENCES

- [Sequences → Add New sequence]: Creates a new nucleotide/protein sequence.
- [Sequences → Delete selected sequence]: Selects sequences in the sequence browser.
- [Sequences → Prints]: Prints the browsed sequence.
- [Sequences → Sort]: Sorts options for multi-sequence files (only available when uploading a multi sequence file).

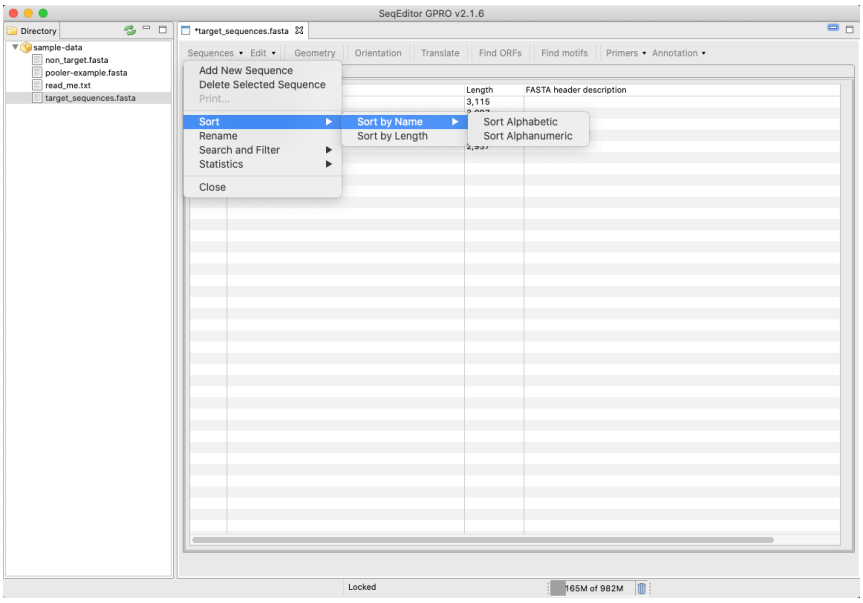


Figure 13. Sort sequences tool of the sequence browser menu. Using this option, the user can sort either by length or by name the sequences of a multiple fasta file opened in SeqEditor. The sequences will be listed either alphabetically or alphanumerically.

- [Sequences → Rename]: Allows the renaming of a selected sequence name.
- [Sequences → Search and Filter]: Allows for the searching and filtering the sequences contained in the input file.

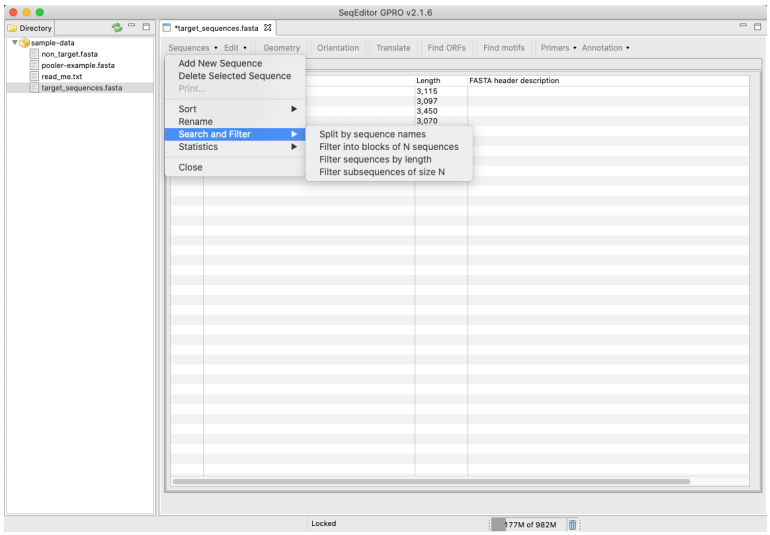


Figure 14. "**Search and filter" option of the sequence browser menu**. i) The "split by sequence names" option allows the user to split the sequence file into other files sorted by sequences names that are selected by the user. These names can either be inputted via a text file or manually in the dialog, selecting a matching mode (exact or partial.; ii) The "split into blocks of N sequences" option allows the users to do so by introducing the number of sequences per file. iii)The "filter sequences by length" option does so when introducing the sequence length and the comparison of choice (equal, greater than or lower than). iv) The "filter subsequences of size N splits at gap regions, if present, by selecting the sequence size and a padding value when applicable.

- [Sequences → Statistics]: Displays the overall metrics of the sequences contained in the selected fasta file as shown in Figure 11

BROWSER MENU: EDIT

- [Edit → Undo]: Undoes an edit made to the sequence.
- [Edit → Redo]: Redoes an edit made on the sequence.
- [Edit → Cut]: Allows the user to cut a sequence from the sequence browser.
- [Edit → Copy]: Allows the user to copy a sequence from the sequence browser.
- [Edit → Paste]: Allows the user to paste a sequence from the sequence browser.

[Edit → Locked]: Locks or unlocks sequence editing.

BROWSER MENU: GEOMETRY

- [Geometry]: It allows the change from DNA to RNA and *vice versa*, as well as the view of the sequence as a single or double strand.

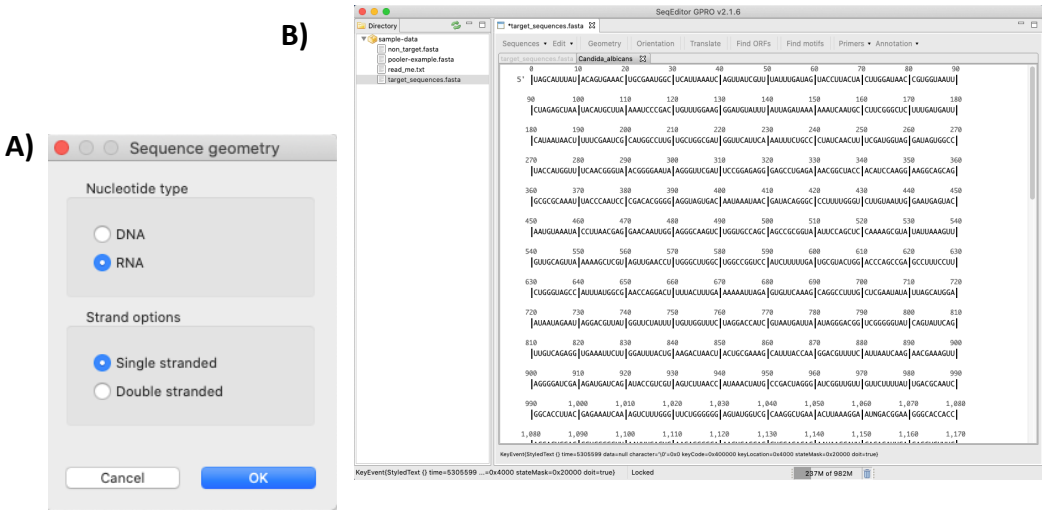


Figure 15. Changing the geometry of a sequence via the "Geometry" option in the sequence browser menu. **A)** Sequence geometry selection dialog. **B)** Example of change in geometry from double stranded DNA to single stranded RNA.

BROWSER MENU: ORIENTATION

- [Orientation]: Switches the sequence orientation as 1) reverse, 2) complementary or 3) reverse-complementary.

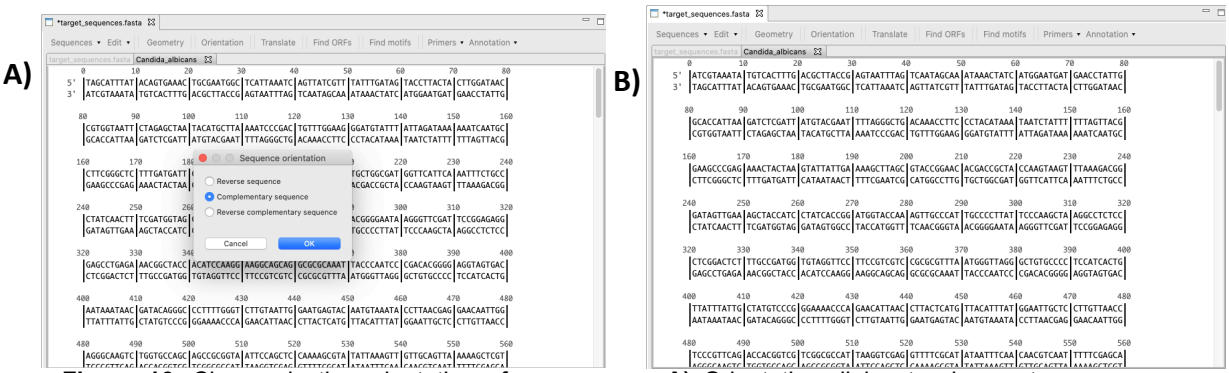


Figure 16. Change in the orientation of a sequence. **A)** Orientation dialog to change to reverse, complementary or reverse complementary sequence. **B)** Example of changing the sequence orientation to the complementary sequence.

BROWSER MENU: TRANSLATE

[Translate]: Translates nucleotide sequences into proteins by selecting the desired frame(s) – on the forward or reverse strand starting from the +1, +2 or +3 nucleotide positions. Either the standard genetic code or a custom one can be selected for reference. The translate tool can open Gene Runner's translation table format (.trt files) as well as a native plain text format files (.txt), which can be created in any text editor of choice. Please note that in these files, lines that start with the hash symbol ("#") are interpreted as comments and thus will be ignored (with the one exception of the first line that will hold the code's name. However, this line is not mandatory). Each following line is composed of a codon (RNA and DNA are allowed), a hyphen and a 'greater than' symbol (">") followed by the correspondent amino acid symbol according to the 1-letter IUPAC code.

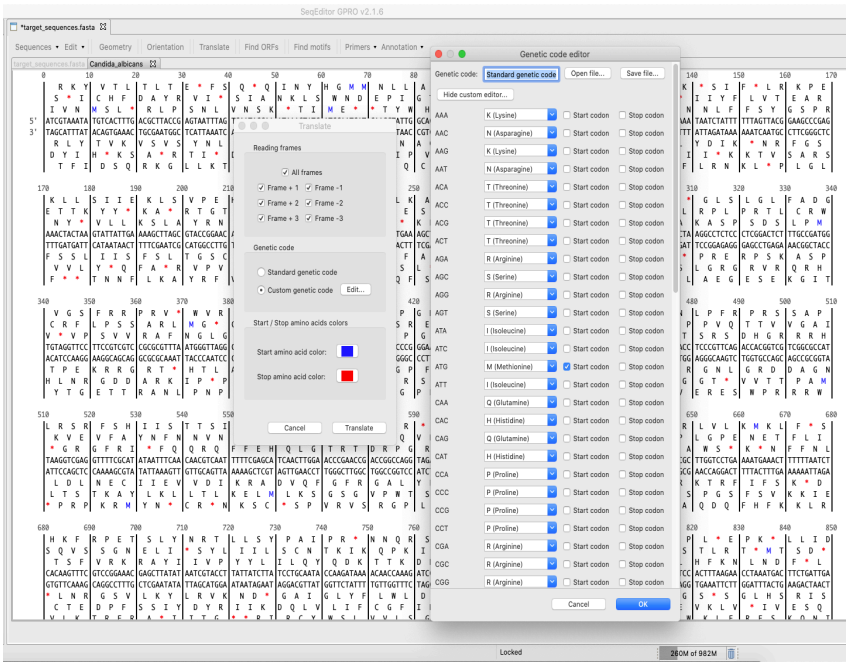


Figure 17. "Translate" dialog and Genetic code editor dialog. The tool calls a pop-up dialog allowing users to translate nucleotide sequences into proteins by selecting the desired translation frame (+1, +2, +3). The standard genetic code is used by default. However, a custom genetic code can also be selected. Once translated, codons can be edited, renamed and saved. Previously saved custom codes can also be opened. The default colors for the start and stop codons are set in blue and red respectively but can be changed using the color palette in the dialog.

BROWSER MENU: FIND ORFS

- [Find ORFs]: Searches and retrieves ORFs contained both the forward and the reverse strands by specifying a minimum ORF length.

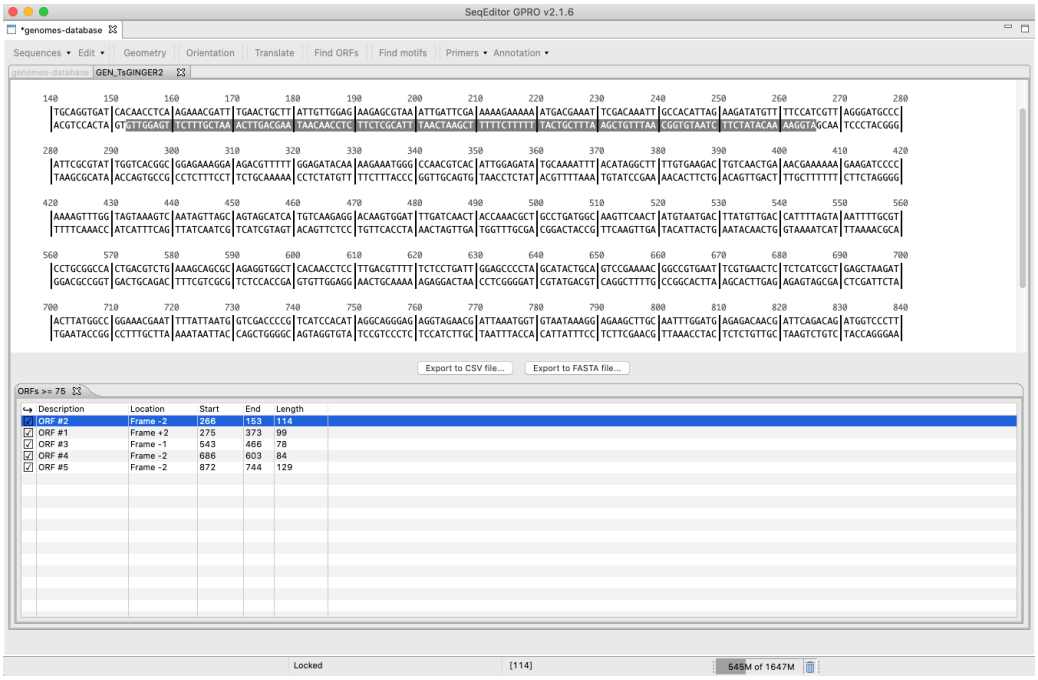


Figure 18. Find ORFs tool. By selecting a length of the ORF, a search is performed and the corresponding report is produced listing the found ORFs. ORF coordinates are specified on the summary browser. By clicking on a given ORF, it will be selected and highlighted in the Sequence Browser. Results can be then exported as sequences or annotations. In the first case, the tool will allow the sequence to be exported as either a nucleotide or a protein sequence.

BROWSER MENU: FIND MOTIFS

- [Find Motifs]: This option performs searches for a specific protein or nucleotide motif (binding sited, restriction sites, etc.). Once the search is completed, results will be shown in the summary browser. You are also allowed to search for multiple motifs as single occurrences or as clusters of motifs by accessing the “Multiple Motif Editor” located in the “Find motif” interface.

BROWSER MENU: PRIMERS

SeqEditor implements a set of three tools for singleplex, multiplex PCR primer design and primer pooling powered by an interface adaptation of two CLI tools: Primer3 (Untergasser, *et al.* 2012) and PrimerPooler (Brown, *et al.* 2017). This tool also incorporated a newly optimized search process that is based on two algorithms for multiplex and target-specific primer design (Hafez *et al.* 2020). These three tools (designated as SinglePlexPCR, MultiPlexPCR and PrimerPooler respectively) are organized in three separate interactive interfaces that are accessible through the “Primers” tab of the browser menu.

- [SinglePlex Primers]: Searches for primers for the amplification of one or more sequences. This option can be run in two modes:

- **Batch run mode**: searches for primers suitable for all sequences contained in a given file.
- **Single run mode**: searches for primers that are suitable for a given sequence that has been opened only.

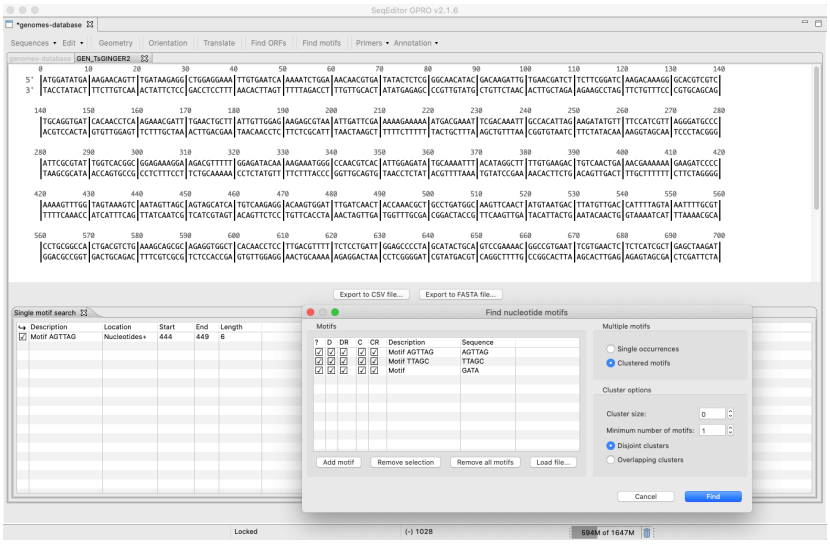


Figure 19. Multiple Motif Editor dialog. The search for motifs can be performed either as single occurrences if the motifs are searched as independent terms or as clustered motifs withmotifs falling together in a sequence frame. In that case, some search parameters will need to be selected, namely i) Minimum cluster size (for instance, a frame of 500 nucleotides); ii) Minimum number of motifs within the cluster; iii) Select whether clusters are allowed to overlap (overlapping clusters option) or not (disjoint clusters option). Note that you can also add motifs from a file using the tab “Load File” in the Multiple Motif Editor.

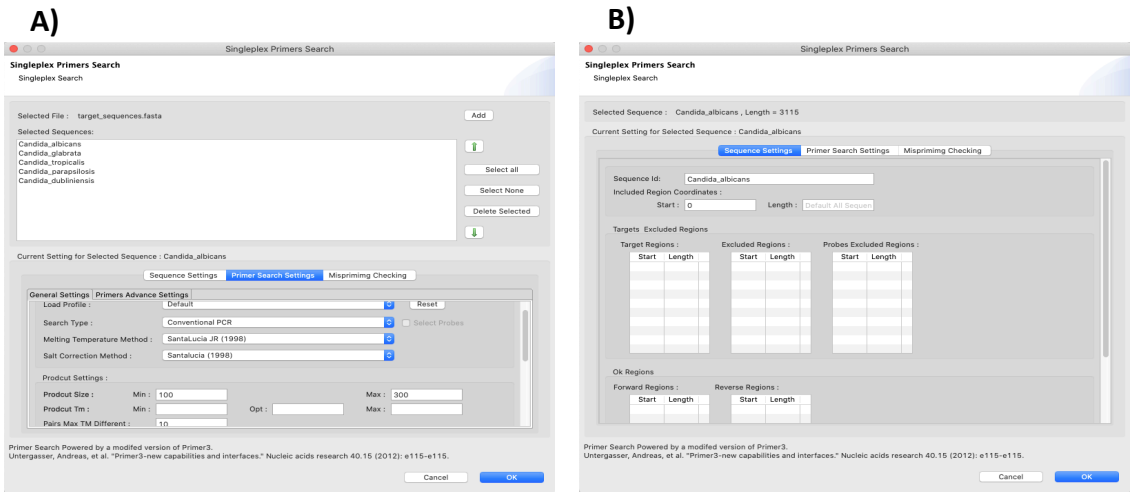


Figure 20. Singleplex Primer Search dialogs. A) SinglePlex primer search in batch mode. The file upload window is displayed at the top. B) SinglePlex primer search in single mode. Once this mode has been selected the search parameters can be set at the bottom panel. There are three main tabs to enable this: i) Sequence setting: The ‘sequence ID’ for which primer search will be launched on can be inputted here. Once the search has been performed, the bottom panel will display those regions that correspond to on target, off target (excluded), and probe excluded regions obtained in the search. ii) Primer Search Settings: allows the following search parameters to be set: 1) general settings such as primer length, Tm,GC content and product size; 2) Advanced settings such as thermodynamic calculations and other primer and score calculation settings; 3) Advanced settings -similar to the advanced primer settings- but to be applied for probes only; iii) Mispriming Cheking: Allows the user to input a library of mispriming sequences to be used during the search

- [MultiPlex Primers]: It allows the search of MultiPlex and species-specific primers for multiple target sequences using different settings.

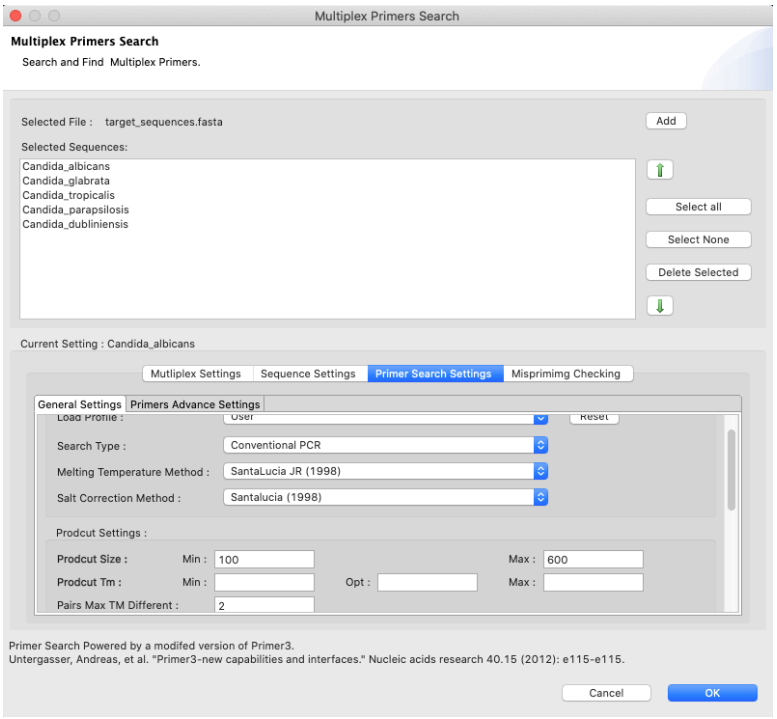


Figure 21. Multiplex Primer Search dialog. The search parameters can be set through four main tabs displayed at the bottom, namely 'Sequence Settings', 'Primer Search Settings', 'Mispriming Settings' and 'Multiplex Settings'. The first three are identical to those included in the SinglePlex Primer Search and their parameters are described in Figure 20. Multiplex Settings allows the user to assign a name for the group search.

- [PrimerPooler]: Optimizes the multiplex PCR input by dividing the primers into different pools.

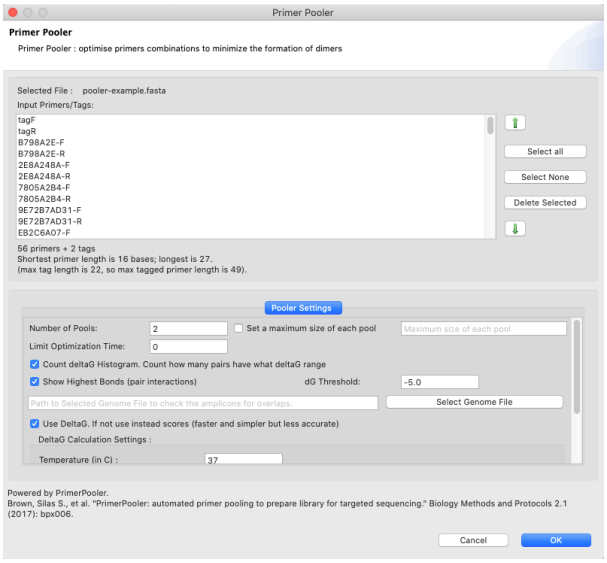


Figure 22. PrimerPooler tool dialog. The following parameters can be set when using this tool: i) Number of pools to divide the input into ("Number of pools" option); ii) Setting a maximum size for each pool to make the pools more even ("Set max Size of Pools"); iii)Creating histograms from the deltaG/Score for all pairwise interactions of all primers ("Count deltaG/Score"); iv)Showing a summary and the dimer structure formed from the pairwise interaction with the highest interaction ("Show Highest Interaction"); v) Selecting a genome file in 2bit format to check the amplicons for overlaps (Select Genome) Primer pairs that amplify overlapping regions of the genomes can produce an unwanted shorter amplicon if used in the same pool; vi)Use thermodynamic principles to calculate the correspondent ΔG for the pairwise interaction (Use DeltaG). If not selected, a score will be automatically calculated based on alignment. However please note that such a score will be calculated in a faster yet less accurate way. To use deltaG additional parameters such as

temperature, concentration of magnesium, concentration of monovalent cation and concentration of deoxynucleotide (dNTP) will need to be inputted.

To provide some examples how to use PrimerPooler, MultiPlexPCR and SinglePlexPCR as well as to validate the new search strategy for multiple species-specific primers, we have prepared a case study tutorial. In this case study, the MultiplexPCR tool use was used the design of multiplex species-specific primers for five *Candida* species (OPATHY Consortium and Gabaldon 2019). The target DNA sequences selected for this example were the ribosomal DNA sequences of the five *Candida* species as retrieved from the NCBI databank (Sayers, *et al.* 2019). This tutorial is available online at the following link: https://gpro.biotechvana.com/tool/seqeditor/candida_target_specific2020.

BROWSER MENU: ANNOTATION

SeqEditor supports the analysis of genomes and transcriptomes that have a reference annotation file (GTF/GFF). Thus, we have incorporated a GTF/GFF viewer that reads the reference annotation file allowing the users to search, filter, and extract sequence features (such as chromosomes, genes, exons or introns) contained in the assembly file using the annotation as a reference.

The GTF/GFF viewer has been tested with several different assemblies and GTF/GFF files available in the Candida Genome Database (Skrzypek *et al.* 2017), Ensembl (Cunningham *et al.,* 2019) and the NCBI (Sayers *et al*.2019), as well as a with a tailor-made GTF file that was created based on the gene prediction generated by the AUGUSTUS 3.3 software (Stanke *et al*. 2008) for the *Sparus aurata* genome (Pérez-Sánchez *et al*. 2019). The GTF/GFF viewer also accepts .bed files and other files in plain format.

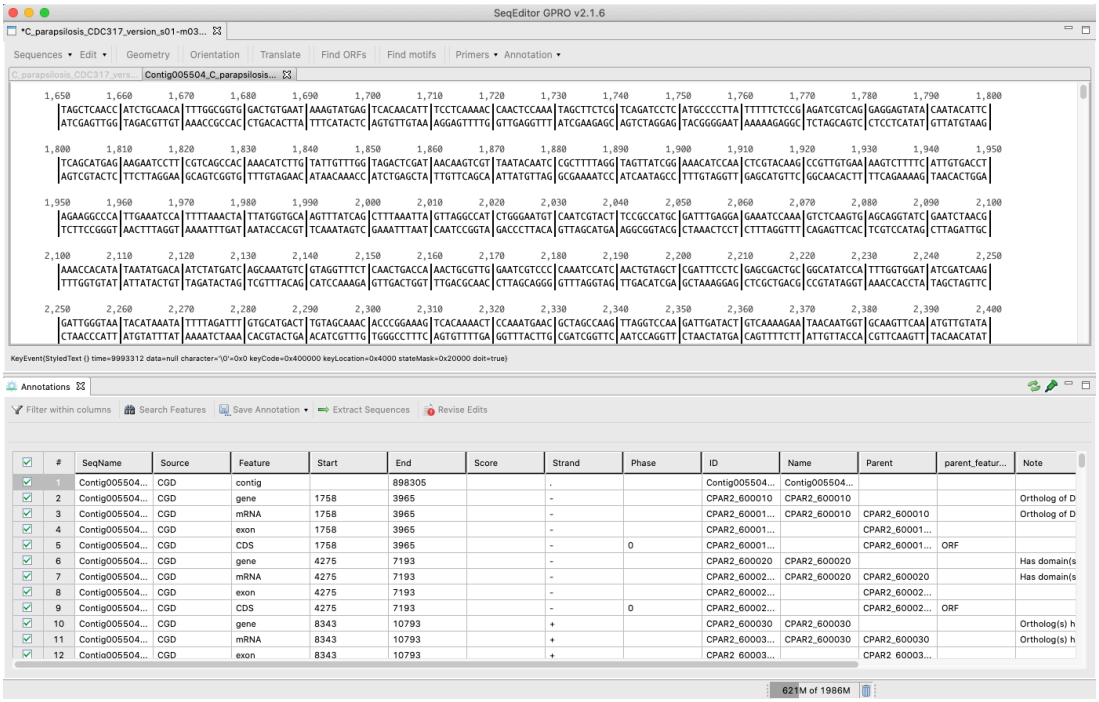


Figure 23. View of the sequence browser (top) and the GTF/GFF viewer (bottom) when opened simultaneously. The annotation viewer shows the features (such as gene, exon, CDS, etc.) that are contained in the sequence of interest. The viewer has five additional options for the management of these annotations: i) Filter within columns: Shows/hides the filter option. When the filter option is shown, the user can filter by any of the terms contained in any column displayed in the annotation file. For example, when writing “gene” on the filter section of the Feature column, the annotation viewer will only display those rows that correspond to a gene. ii) Search Features: Searches and checks a feature according to different criteria; III) Save Annotation: Saves, Saves As or Exports (just for rows); iv) Extract sequences: Extract features from

the reference sequences through a pop-up dialog; v) Revise Edits: Revises and corrects the features of the coordinates that are affected by the edits of the reference sequence.

The viewer will open when double-clicking on a sequence file in the directory browser or via the “Annotation” tab of the browser menu. The “Annotation” tab will show the following options:

- [Open Annotation File]: Opens and attaches the annotation file (GTF/GFF) to the previously opened fasta file.
- [View/Hide]: Views/hides the annotation viewer.
- [Close Annotation Set]: Closes/saves the opened annotation file.

Once the GTF/GFF file is loaded, users will visualize all the data as an annotated grid of rows and columns that will be organized based on the information contained in the annotation file.

The GTF/GFF viewer allows the user to perform different tasks. Some of them can be executed with the mouse while others can be implemented via a menu.

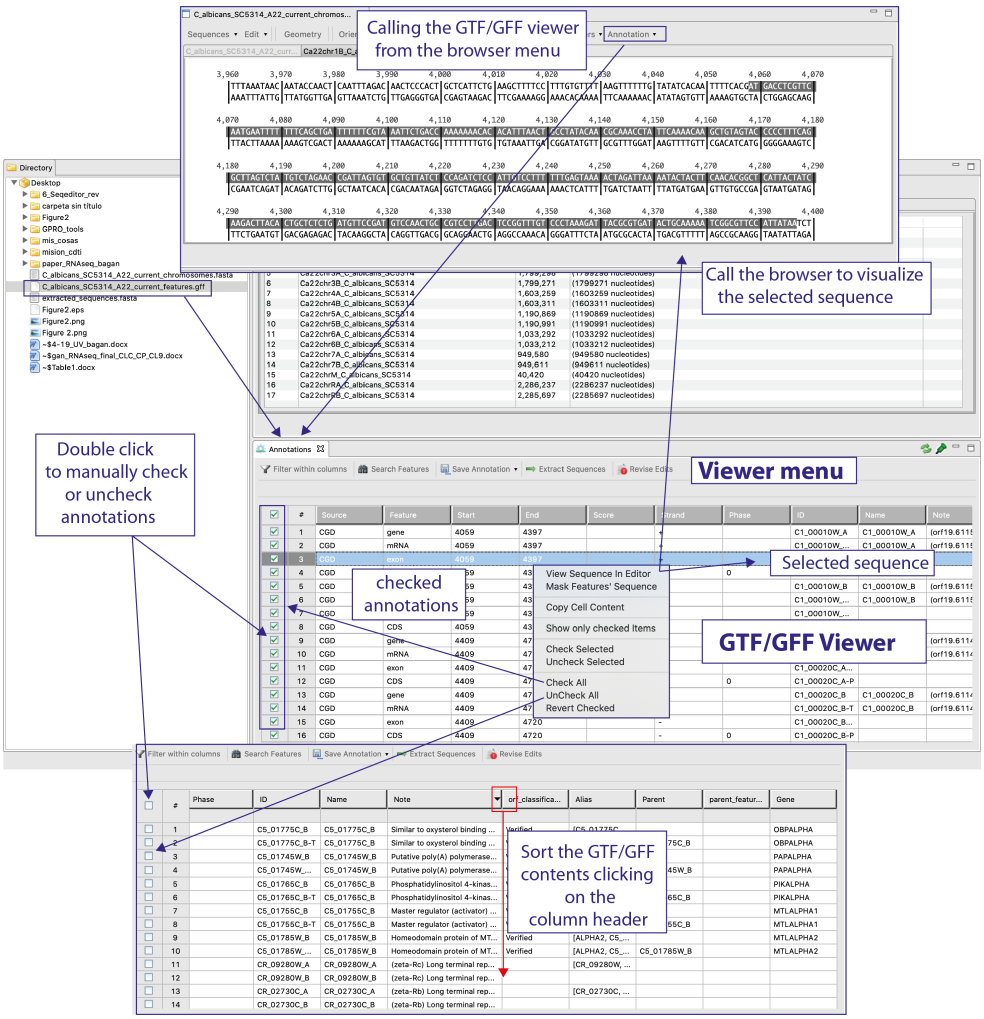


Figure 24. GTF/GFF viewer and different options for mouse-dependent tasks in the viewer. Annotations can be manually selected or deselected by clicking (twice) with the mouse to check/uncheck rows or by right-clicking anywhere to call a context menu providing distinct checking options or for visualizing a feature in the browser. By clicking one or two times on the column headers of the viewer, users can sort annotation file contents. Users can also select the texts of rows and from the column cells shown in the viewer using the mouse.

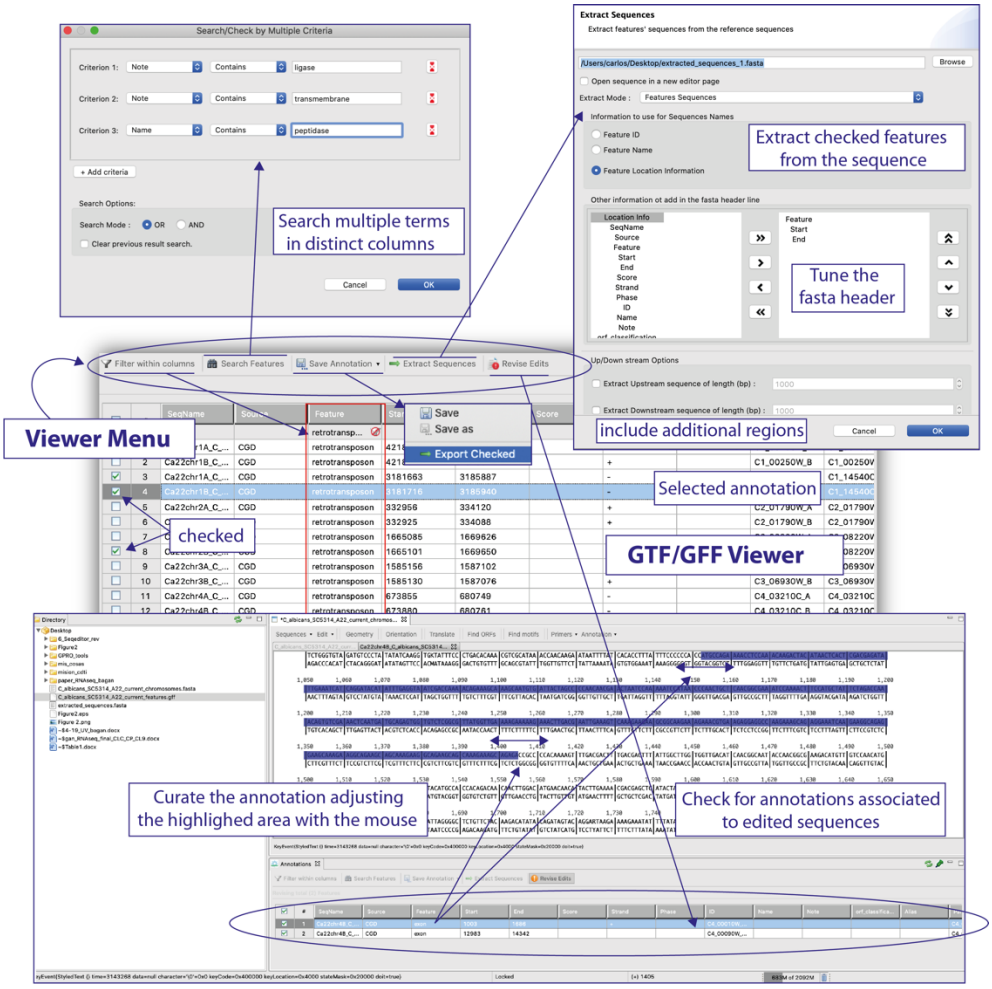


Figure 25. Different task options provided by the menu of the GTF/GFF viewer. The tab “Filter within columns” let users filter annotations in a specific column to show only those that match a key word. The tab “Search features” gives access to a context dialog allowing to specify one or more key words to search and check a subset of annotations matching these criteria (the options “or” and “and” can be used to improve the search). “Save Annotations” enables the saving any edit in the GTF/GFF or to export only the checked annotations in a new GTF or GFF file. “Extract Sequences” calls another context dialog that permits to extract sequence features indicated as checked in the viewer; the dialog offers additional exporting options to name the fasta headers of exported sequences or for exporting the sequences with upstream and downstream nucleotide extensions of a size determined by the user. Finally, “Revise edits” permits the editing of the GTF/GFF file to correct or curate the annotation of any sequence if it has been previously edited with the browser. That is, if a user opens a sequence file and the associated GTF or GFF with the sequence browser and the GTF/GFF viewer, the user can edit the sequence in the browser. To update the GTF/GFF file according to this change the user only needs to click on the tab “Revise edits”. In doing so, the viewer detects and shows in the GTF/GFF viewer the annotations of those sequences that have been edited in the browser. Then, if clicking on the row for the edited sequence, the browser is called again, and the region affected by the edit is highlighted in the browser. Finally, the user only can use the mouse to manually adjust the highlight of the edited region (for example an exon) by dragging the highlight until the correct coordinate of that feature in the browser. Next, the coordinates of that feature are corrected in the GTF/GFF viewer according the final highlight stated in the sequence browser. Then the user only needs to save the new GTF or GFF file using the options provided by the tab “Save Annotation”.

3. ACKNOWLEDGEMENTS AND CITATIONS

3.1- Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 642095 for the OPATHY consortium, by the pre-doctoral research fellowship from Industrial Doctorates of MINECO (Grant 659 DI-17-09134); by the State Plan for Scientific and Technical Research and Innovation 2017-2020 under the Grant TSI-100903-2019-11 from the Secretary of State for Digital Advancement from Ministry of Economic Affairs and Digital Transformation, Spain; and by Expedient IDI-2021-158274-a from Ministry of Science and Innovation, Spain

3.2- Literature cited

- Bradnam K. Assemblathon Stat Script. 2011 [http://korflab.ucdavis.edu/datasets/Assemblathon/Assemblathon2/Basic_metrics/assemblathon_stats.pl]
- Brown, S.S. Chen, Y. Wang, M. Clipson, A. Ochoa, E. and Du, M. 2017 PrimerPooler: Automated Primer Pooling to Prepare Library for Targeted Sequencing. Biology Methods and Protocols 2(1), bpx006. https://doi.org/10.1093/biomethods/bpx006.
- Cunningham, F. (and 66 co-authors). 2019. Ensembl 2019. Nucleic acids research 47(D1), D745-D751. https://doi.org/10.1093/nar/gky1113
- Futami R, Muñoz-Pomer A, Viu JM, Dominguez-Escribá L, Covelli L, Bernet GP, Sempere JM, Moya A, Llorens C. 2011. GPRO: the professional tool for management, functional analysis and annotation of omic sequences and databases. Biotechvana Bioinformatics: 2011-SOFT3
- Hafez A, Futami R, Arastehfar A, Daneshnia F,Miguel A, Roig FJ, Soriano B, Perez-Sánchez J, Boekhout T, Gabaldón T, Llorens C. 2020. SeqEditor an application for primer design and sequence analysis with or without GTF/GFF files Bioinformatics, btaa903, https://doi.org/10.1093/bioinformatics/btaa903
- Muñoz-pomer A, Futami, R, Covelli L, Domínguez-Escribà, L, Bernet, GP, Sempere JM, Moya A, and llorens C. 2011. TIME: a sequence editor for the molecular analysis of large DNA and protein sequence samples. Biotechvana Bioinformatics.2011-SOFT2.
- OPATHY Consortium. 2019. Recent trends in molecular diagnostics of yeast infections: from PCR to NGS. FEMS Microbiol Rev 43:517-547.
- Pérez-Snchez, J. Naya-Català, F. Soriano, B. Piazzon, M.C. Hafez, A. Gabaldon, T. Llorens, C. Sitjà-Bobadilla, A. Calduch-Giner J.A.2019. Genome Sequencing and Transcriptome Analysis Reveal Recent Species-Specific Gene Duplications in the Plastic Gilthead Sea Bream (Sparus aurata). Front. Mar. Sci., doi: https://doi.org/10.3389/fmars.2019.00760
- Sayers, E.W. Agarwala, R. Bolton, E.E. Brister, J.R. Canese, K. Clark, K. Connor, R. Fiorini, N. Funk, K. Hefferon, T. Holmes, J.B. Kim, S. Kimchi, A. Kitts, P.A. Lathrop, S. Lu, Z. Madden, T.L. Marchler-Bauer, A. Phan, L. Schneider, V.A. Schoch, C.L. Pruitt, K.D. and Ostell, J. 2019. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 8;47(D1):D23-D28. doi: https://doi.org/10.1093/nar/gky1069.
- Skrzypek, M.S. et al. 2017. The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high

throughput sequencing data. Nucleic Acids Research, 45, issue D1, D592-D596.
- Stanke, M. Diekhans, M. Baertsch, R. and Haussler, D. 2008. Using native and syntenically 1013 mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24, 637-644. doi: 1014 10.1093/bioinformatics/btn013
- Untergasser, A. 2012. Primer3--new capabilities and interfaces. Nucleic Acids Res 40:e115.

4. LICENSING

4.1 License of use

The GPRO license is a free, proprietary license for academic and commercial researchers to use of the six software applications constituting the GPRO suite. Please note that all the terms and conditions specified in this license apply exclusively to these six applications, including SeqEditor. This means that we hold intellectual property rights over the source code of the suite, but not over the pipelines, scripts, databases, and/or server dependencies. These instead include not only the scripts developed by Biotechvana but also other free software and databases developed by third parties under conventional open source licenses. Where applicable, the correspondent third party developers are cited in both the GPRO applications and in their correspondent manuals. For the specific terms and conditions for the use of any third-party materials, please refer to the website of the correspondent developer.

The use of GPRO suite including SeqEditor is subject to the terms and conditions stated in this license. This means that by downloading, installing, running, or otherwise using the software you agree to comply with the terms and conditions specified herein. These terms and conditions will also apply to any updates, upgrades, supplements, and support services for the software unless specified otherwise.

To provide online technical support for the use of GPRO, we host a user forum that may be used at any time at https://forum.biotechvana.com. We appreciate any feedback you may provide us with, including suggestions for improvement and the report of any issues encountered with the software, if any. Please note that no personalized technical support services are freely included in the GPRO suite package. However, we can provide them upon demand. Should you be interested in receiving personalized technical support, please contact us at biotechvana@biotechvana.com.

We recommend that any analysis of the results obtained using the software is performed and interpreted by qualified professionals. For any questions about the use of any application of the GPRO Suite including SeqEditor or the information contained in this license, please contact us at biotechvana@biotechvana.com.

4.2 Definitions

“This license” refers to version 2.0 and higher of the GPRO Suite Open Access License. "Product" and “software” refer to either the entire suite or to any of the six GPRO applications constituting it.

“Update” refers to a newer minor version (an increase in any digits after the first period of the version number).

“Upgrade” refers to a newer major version (identified by an increase in the digits before the first period of the version number).

“We” refers to Biotechvana SL.

“User” and "You" refer to the user, either an individual or entity that downloads, installs or uses GPRO and has agreed to all terms and conditions included in this license.

To “modify” a software means to copy from or adapt all or part of it in a fashion requiring copyright permission, other than the making of an exact copy. The resulting software is called a “modified version” of the earlier software or a software “based on” the earlier version.

4.3 Purpose

This license gives you the right to use GPRO Suite as expressly permitted herein, free of charge for both academic and industrial researchers and without any associated license fees. GPRO Suite is therefore licensed, not sold.

4.4 Intellectual property

Biotechvana SL holds intellectual property rights over the use of the six desktop applications constituting the Suite including SeqEditor. These applications are protected under copyright and intellectual property laws, including international copyright treaties and other intellectual property treaties. The design, trademark, and logos of GPRO are copyright of Biotechvana and this material cannot be copied or modified without prior written permission of Biotechvana.

4.5 Permissions

4.5.1. You may not sell, charge for, sub-license, rent, lease or loan the product without our prior written consent. You may however use the product for your research or for delivering research services to a third party or other commercial business.

4.5.2. You may not decompile, disassemble, reverse engineer or otherwise attempt to discover the source code of the product except to the extent that you may be expressly permitted by us.

4.5.3. You may not repackage, modify, vary, adapt, alter in any way, create any software that derives from, or integrate any other computer programs with, the product in whole or in part.

4.5.4. You are not permitted to grant any sub-licences of the product without our permission.

4.5.5. You may not use the product to engage in or allow others to engage in any illegal activity.

4.5.6. You acknowledge and agree that all rights, title, and interest in and to the Suite, including associated intellectual property rights, are and shall remain with Biotechvana.

4.5.7. You agree that any feedback, suggestions, comments, improvements, modifications and other information that you provide to Biotechvana relating to the product or its performance may be used to improve the software.

4.5.8. The user may back up the software installation files as needed. The files may be used for reinstallation purposes only.

4.6 License expiration

This license does not expire.

4.7 Termination

Biotechvana reserves the right to cancel any GPRO license at any given time without prior notice. Should the user breach the terms and conditions stated herein, the license to use GPRO will automatically terminate.

4.8 Disclaimer of warranty

4.8.1. We provide the software “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 risk associated with the quality and performance of the program is accepted by the user. We will not be held responsible for any malfunction of the software, neither for any secondary issues that may arise from its use. However, any reported issues will be carefully studied, and we will aim to amend them on future software updates.

4.8.2. Any material obtained using the software is produced at your own risk. You will be solely responsible for any damage to your computer system or loss of data that results from the download or use of such material.

4.8.3. No advice or technical support information, whether oral or written, provided directly by Biotechvana or via the use of the software shall imply any warranty or liability.

4.9 Limitation of liability

Biotechvana is not liable to you, unless required by applicable law or agreed to in writing, for any damages to your computer, including any general, special, incidental or consequential damages arising out of the use or inability to use the software (including but not limited to loss of data or data being rendered inaccurate, losses sustained by you or third parties or failure of the software to operate with any other programs), even if either you or other party has been advised of the possibility of such damages.

Except as expressly stated in this license, the user shall also not be held liable for any direct, indirect, incidental, or consequential damages that may occur through the use of the software.