Using SeqEditor for primer design and sequence analysis with or without GTF/GFF files

Ahmed Hafez[1,2,3], Amir Arastehfar[4,5], Farnaz Daneshnia[5], Ana Miguel[1], Francisco J. Roig[1], Beatriz Soriano[1], Jaume Perez-Sánchez[6], Teun Boekhout[4,5], Toni Gabaldón[2,7,8,9], Carlos Llorens[1*]

[1]Biotechvana, Parc Científic Universitat de València

[2]Universitat Pompeu Fabra, Barcelona, Spain

[3]Faculty of Computers and Information, Minia University, Egypt

[4]Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands.

[5]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands.

[6]Institute of Aquaculture Torre de la Sal (CSIC), Castellon, Spain

[7]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

[8]Barcelona Supercomputing Centre (BSC-CNS). Barcelona, Spain

[9]Institute for Research in Biomedicine (IRB), The Barcelona Institute of Science and Technology, Barcelona, Spain

*corresponding author: carlos.llorens@biotechvana.com, Tel: +34 960 067 493

Abstract

Motivation

Sequence analyses oriented to investigate specific features, patterns and functions of protein and DNA/RNA sequences usually require tools based on graphic interfaces whose main characteristic is their intuitiveness and interactivity with the user´s expertise, especially when curation or primer design tasks are required. However, interface-based tools usually pose certain computational limitations when managing large sequences or complex datasets, such as genome and transcriptome assemblies.

Results

SeqEditor is a standalone desktop application for the analysis of nucleotide and protein sequences, which can either work as a file manager or as a graphical sequence browser. SeqEditor accepts the input of one or multiple fasta files and performs all typical tasks for sequence analysis. Its sequence browser is optimized for very long sequences, such as scaffolds and chromosomes. SeqEditor implements a GTF/GFF viewer that allows the user to mine reference genomes and transcriptome data to extract contents and features annotated in GTF/GFF files. Besides, SeqEditor includes a set of tools for the search and design of PCR primers, including singleplex, multiplex and target-specific primers, powered by a newly optimized search strategy based on two algorithms for constructing multiplex indexes and search of target-specific primers. This search strategy is experimentally validated here using a clinically-relevant case study aiming at the detection of five opportunistic fungal pathogens.

Availability

SeqEditor is publicly available at https://gpro.biotechvana.com/download. The SeqEditor' user manual is available at https://gpro.biotechvana.com/tool/seqeditor/manual.

Contact

carlos.llorens@biotechvana.com

Supplementary Information

Supplementary data are available at Bioinformatics online.

Introduction

Analysis of DNA, RNA and proteins at the sequence level is central for the understanding of their specific features, functions and evolutionary patterns. Most of the currently available software for sequence analysis relies on Command Line Interface (CLI) tools, such as BuddySuite (Bond et al. 2017) and FAST (Lawrence et al. 2015) among others. CLI tools are very efficient for the performance of a wide variety of tasks as they do not have graphical requirements for sequence representation and visualization, however CLI tools require the user to have basic computer literacy, which limits their use to bioinformaticians that are familiarized with command lines. Moreover, CLI tools do not allow direct graphical visualization of the analyzed sequences, which could be useful when investigating specific sequence regions and patterns, since they may escape detection via automated algorithms and are frequently identified by human expert abstraction. It is because of this, among other reasons, that biological researchers usually seek more interactive and responsive tools based on Graphical User Interfaces (GUI), such as GeneRunner (http://www.generunner.net), Geneious (https://www.geneious.com) and Sequencher (https://www.genecodes.com). Unlike CLI tools, GUIs only require basic informatics skills, making them easier to learn and manage than CLI tools. However, GUI tools pose certain limitations when managing large and multiple sequences, making CLI tools more efficient for the management of complex datasets as well as for the assembly of genomes and transcriptomes. With this in mind we have developed SeqEditor, a cross-platform desktop tool for the analysis of nucleotide and protein sequences. SeqEditor is an application of the GPRO suite (Futami et al. 2011) that integrates an updated version of TIME editor, a former sequence browser of that suite for sequence-to-sequence analysis (Mñnoz-Pomer et al. 2011), as well as a wide variety of new implementations. Here, we introduce SeqEditor providing a general overview of all its tools for sequence analysis making particular emphasis in describing the most relevant implementations.

General overview and functions

SeqEditor is a standalone desktop application for the analysis of DNA, RNA and protein sequences. The application has been developed in Java using Eclipse Rich Client Platform (Kornstadt and Reiswich 2010). SeqEditor can either work as a file manager or as a graphical sequence browser to perform all the typical sequence analysis tasks, including searching and filtering of sequences, Open Reading Frames (ORFs) and motifs; translation of nucleotide to protein sequences using either the universal genetic code or a user-defined one; changing sequence geometry and orientation; computing metrics for sequence datasets, and more. Figure 1 shows the layout of SeqEditor, which is composed of four main elements, namely a directory browser, top menu, sequence browser, and browser menu. Additional interfaces are used to either run the analyses or display summaries of both data management workflows and visualization of results. These include different interfaces for each browser's task and summary view interfaces, e.g. ORF finder summary screen view. The top menu provides access to both the sequence browser and other interfaces through which file manager tasks can be executed. This allows the user to process and analyze all the sequences included in one or multiple fasta files at the time without the need of opening the files, thus working with similar efficiency to that of CLI tools. The sequence browser is an improved version of the former TIME editor, a graphic screen allowing the users to navigate, edit and analyze the sequences under visual control. The browser can be called from the top menu or by right-clicking on an input sequence file in the directory browser (users can set any folder in their PCs as directory browser). The browser menu organizes the different functions of the sequence browser, allowing the user to edit sequences, search ORFs and motifs, translate nucleotide sequences to proteins, change the geometry and orientation of the sequences, design single- and multiplex PCR primers, and mine sequence features using annotations of GTF/GFF files, that will be described with more detail in the next sections of this article**.** The browser screen is interactive, allowing users to select, copy, cut and paste sequence traits by right-clicking on the sequence. The graphic capacity of the sequence browser relies on the RAM power of the user's hardware, yet it has been optimized to allow the management of large sequences, such as contigs, scaffolds and chromosomes. For example, in PCs with at least 25 Gigabytes of RAM assigned to SeqEditor, the browser can manage sequences of up to 300 megabases. 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 to summarize 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 edition of the sequences, such as sequence name editing directly within the fasta file or sequence sorting by name or size.

SinglePlexPCR, MultiPlexPCR and PrimerPooler: Set of tools for PCR primer design

We previously indicated in the Section above "General overview and Functions" that SeqEditor implements a set of three tools (designated as SinglePlexPCR, MultiPlexPCR and PrimerPooler) for PCR primer design based on an interface adaptation of two CLI tools - Primer3 (Untergasser et al, 2012) and PrimerPooler (Brown et al, 2017) - and a newly optimized search process for multiplex and target-specific primer design. These three tools are organized in separate interactive interfaces accessible through the tab *Primers* of the browser menu.

- SinglePlexPCR permits the user to search for primers in batch mode for one or more sequences. This is achieved through the use of an optimized search based on the general Primer3 search algorithm search in order to find favorable candidates during the early stage of the search. The search starts by populating a list of forward and reverse primer candidates. The tool then eliminates any primer that does not comply with the initial design parameters screening, such as primer length, CG content and melting temperature (Tm), among others. Next, SinglePlexPCR selects candidate primer pairs by producing virtual PCR products that satisfy the input design parameters. Subsequently, more complex computational evaluations are performed, such as the detection of potential formation of primer-dimer and hairpin structures. Lastly, SinglePlexPCR stores and sorts all suitable primer pairs based on a penalty score and finishes the search as soon as it finds a predefined number of accepted PCR primer pairs. This algorithm is applied to single targets and guarantees the finding of suitable pairs of primers if they exist, and the exhaustion of the search in which all possible primer combinations are explored. In this case, users are allowed to continue

interactively with the primer search in order to find more results without the need to restart the search from the beginning.

- MultiPlexPCR executes an efficient greedy strategy to search and find primers utilizing complex indexes and an optimized search process. Since finding multiplex primers is a complex computational problem in which a brute force search consumes excessive computational resources, the search strategy of MultiPlexPCR for multiplex search is different from that described above for singleplex searches. This is largely because the storage and checking of primers products that will not yield a suitable multiplex set is a waste of computational resources as the number of possible primer combinations grows exponentially with the size of the input. Hence, to ensure that only valid potential multiplex primer sets are stored and validated, MultiplexPCR uses a complex index to access and store the potential set of candidate primers according to two index keys, namely PCR product criteria (in terms of product length or Tm for conventional or Real time PCR, respectively) and penalty scores as an approximation of the quality of the primer sets. The index provides an optimized complex data structure to store and evaluate only those primers that populate the list of suitable multiplex primer sets, significantly reducing computational time and memory. Even though this index may significantly reduce the memory usage, it is worth noting that the process of building the index itself still consumes a lot of memory, particularly when analyzing large sequences and large numbers of targets. However, a more restricted design parameter set, such as primer length range and PCR product or other design criteria, can significantly reduce index memory usage as the more restrictive the parameters the more candidate primers will be rejected, and hence, less candidate primers will be stored in the index. With this in mind, MultiplexPCR is powered by an optimized search process that starts from building a complex index over potential multiplex primer sets (as outlined in Algorithm 1), to subsequently perform a brute force specific search using a greedy strategy that uses the complex potential multiplex sets index (outlined in Algorithm 2). In this way, MultiplexPCR first finds a list of forward and reverse candidate primers in a similar fashion to the basic singleplex search but also including additional criteria to ensure

that designed primers are specific to their target and do not bind anywhere else. Then, should the primers bind to multiple targets they are then marked as specific shared primers. As a result, a shared forward primer that perfectly binds to two different targets can be used in combination with two other specific reverse primers yielding two distinguishable PCR products for each target sequence.

- PrimerPooler allows the user to introduce a list of primers and divide them into different pools in order to optimize multiplex PCRs. Finally, it is worth to note that PrimerPooler, MultiPlexPCR and SinglePlexPCR present their results through an interactive summary interface of SeqEditor from which the users can manage and export the result in different file formats such as csv, fasta or GTF/GFF.

With the aim to provide some examples for using PrimerPooler, MultiPlexPCR and SinglePlexPCR as well as for validating the new search strategy for multiple species-specific primers we applied it to the design of diagnostic primers of five *Candida* species that collectively account for almost 90% of all Candida-derived bloodstream infections worldwide (Fuchs et al. 2019), namely *Candida albicans, Candida glabrata, Candida tropicalis, Candida parapsilosis* and *Candida dubliniensis*. Supplementary File 1 shows a case study tutorial with step-by-step indications on how to use the MultiplexPCR tool to design multiplex species-specific primers for these five fungal pathogens. The target DNA sequences selected for this example are the ribosomal DNA sequences of the five *Candida* species as retrieved from the NCBI databank (Sayers et al. 2019).

GTF/GFF viewer: Using GTF or GFF files to mine sequences for specific annotated contents

We have also previously noted that SeqEditor supports the analysis of genomes and transcriptomes with reference GTF/GFF file. SeqEditor implements a GTF/GFF viewer that reads the reference GTF or GFF file with the annotations for either the sequence loaded in the browser or with the genome or transcriptome assembly allowing users to search, filter and extract sequence features, e.g. chromosomes, genes, exons, introns, from the assembly using the annotations provided by GTF/GFF file. That viewer is accessible by

double-clicking on a sequence file placed at the directory browser or through the tab "Annotation" of the browser menu. Once the GTF/GFF file is loaded, users can visualize all the annotated features, as a grid of rows and columns in the viewer. In Figure 2, we show the different viewer's tasks that can be executed with the mouse, while the different tools implemented via menu in the GTF/GFF viewer for filtering, searching, saving and editing are shown in Figure 3. The GTF/GFF viewer has been tested with distinct assemblies and GTF/GFF files provided by 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 tailor-made GTF file created based on the gene prediction Generated by the software AUGUSTUS 3.3 (Stanke et al., 2008) for the *Sparus aurata* Genome (Pérez-Sánchez et al 2019). The viewer also accepts bed files and other files in plain format.

Discussion and Conclusion

In this article, we have introduced SeqEditor a new desktop application for browsing, editing, management and analysis of nucleotide and protein sequences. SeqEditor can work either as a file manager or as a graphical sequence browser thus combining the graphical versatility of GUI applications with a high efficiency for data processing almost comparable to that of CLI tools. SeqEditor is optimized for the analysis of large sequences and although this kind of analyses is rather dependent of the RAM power of each PC, with 25-30 Gigabases of RAM the sequence browser of SeqEditor is able to convincingly deal with sequences of up to 300 megabases, thus providing biological researchers with a user-friendly but efficient GUI-based application for analysis of the largest scaffolds and chromosomes (Note that the largest human chromosome is around 250 megabases). Besides, SeqEditor implements a set of tools for singleplex, multiplex primer design and primer pooling that has been introduced in this article making particularly emphasis in explaining the search strategy and algorithms for multiplex and target-specific primers. To exemplify, validate and highlight the utilities of this specific implementation, we have performed a comprehensive primer design test using five human fungal pathogens for which fast and accurate diagnostics is necessary (Consortium OPATHY and Gabaldon 2019). Here, it is worth to stress the

importance of target-specific primers, which are central for the identification of species in many microbiological processes or for determining antimicrobial susceptibility as well as infection load. In fungal research, e.g. target-specific primers are frequently used for the identification of a yeast pathogen that is responsible for a given infection and this is clinically relevant because, despite all currently available antifungal drugs, invasive fungal infections still have a mortality rate of ≥50% (Brown et al. 2012); in particular, *Candida*-related bloodstream infections have a mortality rate between 30%-60% (Hirano et al. 2015). These ranges imply high economic costs derived from longer hospital stays and the need for multiple analyses that account for over $7.2 billion in 2017 just in the USA (Benedict et al. 2019) and it is thus within this context where applications like SeqEditor become of utility for accurate diagnosis. The other implementation that makes SeqEditor a valuable tool for many biological researchers working with reference genomes and transcriptomes, is its GTF/GFF viewer. Indeed, reference genome and transcriptome are difficult (almost impossible) to manage with GUI-tools due the complexity of the genome/transcriptome assemblies and their associated GTF/GFF files thus limiting the analyses of this material to CLI tools and therefore to expert bioinformaticians. In contrast, the GTF/GFF viewer of SeqEditor can be easily managed by any researcher with bioinformatic skills at the user level in order to mine assemblies and extract information and data such as exons promoters, gene families.

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); and 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 of MINECO.

References

Benedict, K. Jackson, B.R. Chiller, T. and Beer, K.D. Estimation of Direct Healthcare Costs of Fungal Diseases in the United States. Clin Infect Dis 2019, 68(11):1791-1797. https://doi.org/10.1093/cid/ciy776.

Bond, S.R. Keat, K.E. Barreira, S.N. and Baxevanis, A.D. BuddySuite: Command-Line Toolkits for Manipulating Sequences, Alignments, and Phylogenetics Trees. Mol Biol Evol 2017 34(6):1543-1546. doi: 10.1093/molbev/msx089.

Brown, S.S. Chen, Y. Wang, M. Clipson, A. Ochoa, E. and Du, M. PrimerPooler: Automated Primer Pooling to Prepare Library for Targeted Sequencing. Biology Methods and Protocols 2017 2(1), bpx006. https://doi.org/10.1093/biomethods/bpx006.

Brown, G.D. Denning, D.W., Gow, N.A. Levitz, S.M. Netea, M.G. and White, T.C. Hidden killers: human fungal infections. Sci Transl Med 2012 4(165). https://doi.org/10.1126/scitranslmed.3004404.

Consortium OPATHY, Gabaldón, T. Recent trends in molecular diagnostics of yeast infections: from PCR to NGS. FEMS Microbiology Reviews 2019 43(5). https://doi.org/10.1093/femsre/fuz015.

Cunningham, F. (and 66 co-authors). Ensembl 2019. Nucleic acids research 2019 47(D1), D745–D751. https://doi.org/10.1093/nar/gky1113

Fuchs, S. Lass-Flörl, C. and Posch, W. Diagnostic Performance of a Novel Multiplex PCR Assay for Candidemia among ICU Patients. J Fungi 2019 5(3). https://doi.org/10.3390/jof5030086.

Futami, R.L. Muñoz-Pomer, A. Viu, J.M. Dominguez-Escriba, L. Covelli, L. Bernet, G.P. Sempere, J.M. Moya, A. and Llorens, C.GPRO: The Professional Tool for Management, Functional Analysis and Annotation of Omic Sequences and Databases. Biotechvana Bioinformatics 2011. 2011-SOFT3.

Hirano, R. Sakamoto, Y. Kudo, K. and Ohnishi, M. Retrospective analysis of mortality and *Candida* isolates of 75 patients with candidemia: a single hospital experience. Infect Drug Resist 2015 8:199-205. https://doi.org/10.2147/IDR.S80677.

Kornstadt, A. and Reiswich, E. Composing systems with Eclipse rich client platform plug-ins. IEEE Software 2010, 27(6):78–81.

Lawrence, T.J. Kauffman, K.T. Amrine, K.C.H. Carper, D.L. Lee, R.S. Becich, P.J. Canales, C.J. and Ardell D.H. FAST: FAST Analysis of Sequences Toolbox. Front. Genet. 2015 6:172. https://doi.org/10.3389/fgene.2015.00172.

Muñoz-Pomer, A. Futami, R. Covelli, L. Dominguez-Escriba, L. Bernet, G.P. Sempere, J.M. Moya, A. and Llorens, C. TIME: A Sequence Editor for the Molecular Analysis of Large DNA and Protein Sequence Samples. Biotechvana Bioinformatics 2011. 2011-SOFT2.

Pérez-Sánchez, J. Naya-Català, F. Soriano, B. Piazzon, M.C. Hafez, A. Gabaldon, T. Llorens, C. Sitjà-Bobadilla, A. Calduch-Giner J.A. Genome Sequencing and Transcriptome Analysis Reveal Recent Species-Specific Gene Duplications in the Plastic Gilthead Sea Bream (*Sparus aurata*). Front. Mar. Sci., 20 December 2019 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. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2019 Jan 8;47(D1):D23-D28. doi: https://doi.org/10.1093/nar/gky1069.

Skrzypek, M.S. Binkley, J. Binkley, G. Miyasato, S.R. Simison, M. and Sherlock, G. The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. Nucleic Acids Res 2017. 45 (D1); D592-D596.

Stanke, M. Diekhans, M. Baertsch, R. and Haussler, D. Using native and syntenically 1013 mapped cDNA alignments to improve de novo gene finding. Bioinformatics 2008 24, 637-644. doi: 1014 10.1093/bioinformatics/btn013

Untergasser, A. Cutcutache, I. Koressaar, T. Ye, J. Faircloth, B.C. Remm, M. and Rozen S.G. Primer3—new capabilities and interfaces. Nucleic Acids Res 2012; 40(15):e115. https://doi.org/10.1093/nar/gks596.

Figure legends

Figure 1: Layout and main sections of SeqEditor. The directory browser, the top menu, the sequence browser and the browser menu are indicated with blue labels. Examples of interfaces, summary views and task dialogs for ORFs, motifs, primers are highlighted with red labels. Views are customizable and the user can adjust each location of each part by dragging it to the desired location. Note that the default layout on startup will only show the directory browser. Other views and windows will be displayed when performing the corresponding task.

Figure 2: GTF/GFF viewer and different options for mouse-dependent tasks in that viewer. Annotations can be manually selected or deselected by clicking (two times) 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 3: Different task options provided by the menu of the GTF/GFF viewer. The tab "Filter within columns" let users to use a key word to filter annotations in a particular column and show only those that match the 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" permits to save any edit or change done 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 to edit 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 is allowed to 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 (as also shown in Figure 3). 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. After this action 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 need to save the new GTF or GFF file using the options provided by the tab "Save Annotation".

Supplementary files

Supplementary File 1: Case study tutorial for the design of multiplex species-specific PCR primers with SeqEditor. Ribosomal DNA sequences of five fungal pathogens namely *C. albicans, C. glabrata*, *C. tropicalis, C. parapsilosis* and *C. dubliniensis* were used in the test to validate the search algorithms.

Algorithms

Algorithm 1 : Building Potential Multiplex Set Index

```
Input : seqList : a list of n target sequences
        pArgs   : Design Parameters
Output: pIndex  : potential Multiplex Set Index
Result: Potential Multiplex Set Index

/* Potential Multiplex Set Index provide a fast and optimized complex     */
/* data structure to store and evaluate only primers that contribute to   */
/* acceptable multiplex primer set.                                       */
foreach targetSeq in seqList do
    /* Basic Primer3 search                                               */
    Populate forward/reverse primer list for targetSeq;
    Sort primers lists;
end
foreach primer in forward/reverse list do
    Check primer specificity;
    /* Specific primers bind only to any target sequences and do          */
    /* not bind to any sequences in not target library if provided        */

    Categorize/groups primers that bind to multiple target;
    /* Primers could bind to multiple targets, this way it can be         */
    /* used to minimize the number of primers as long any                 */
    /* combination produces distinguishable PCR products.                 */
end

pIndex = Initialize an empty potential Multiplex Set Index;

foreach possible pcrProduct formed by forward/reverse pairs do
    /* Check acceptable PCR product criteria (Length, Tm) */
    if pcrProduct is acceptable product then
        Insert pcrProduct into pIndex ;
    end
    /* pIndex is built with two keys :                                    */
    /* PCR product criteria and multiplex set score                       */
    /* Set score is an average score of all primers pairs                 */
    /* Score is approximated as it is calculated before evaluating primers */
    /* to avoid heavy computational if not needed                         */
end
return pIndex;
```

Algorithm 2.- Specific/Multiplex Primers search

Input : ***seqList*** : a list of *n* target sequences
pArgs : Design Parameters
pIndex : potential Multiplex Set Index (Returned by **Algorithm 1**)
Output: List of candidate multiplex primers set
Result: Candidate Specific multiplex primers set

```
mSets = Initialize an empty list for candidate multiplex set ;
while TRUE do
    potentialSet = Pull a potential multiplex set from pIndex with the best score;
     foreach pcrProduct in potentialSet do
        Evaluate Forward/Reverse primers in pcrProduct;
        if any does not satisfy any criteria in pArgs then
            Ignore potentialSet;
            Update pIndex;
            /* Index update will invalidate any set contains any          */
            /* of the invalidated primers                                 */
            continue;
        end
     end
    Evaluate primers;
    /* Evaluate all primers across different tagert to check for         */
    /* any primer dimer formations                                       */
    if potentialSet satisfy all criteria provided by pArgs then
        Add potentialSet to mSets;
    if mSets have enough solutions or max number of iterations reached then
        break;
    end
    if pIndex has not more candidates then
        /* Search exhausted no more possible solutions                   */
        break;
    end
end
return  mSets;
```

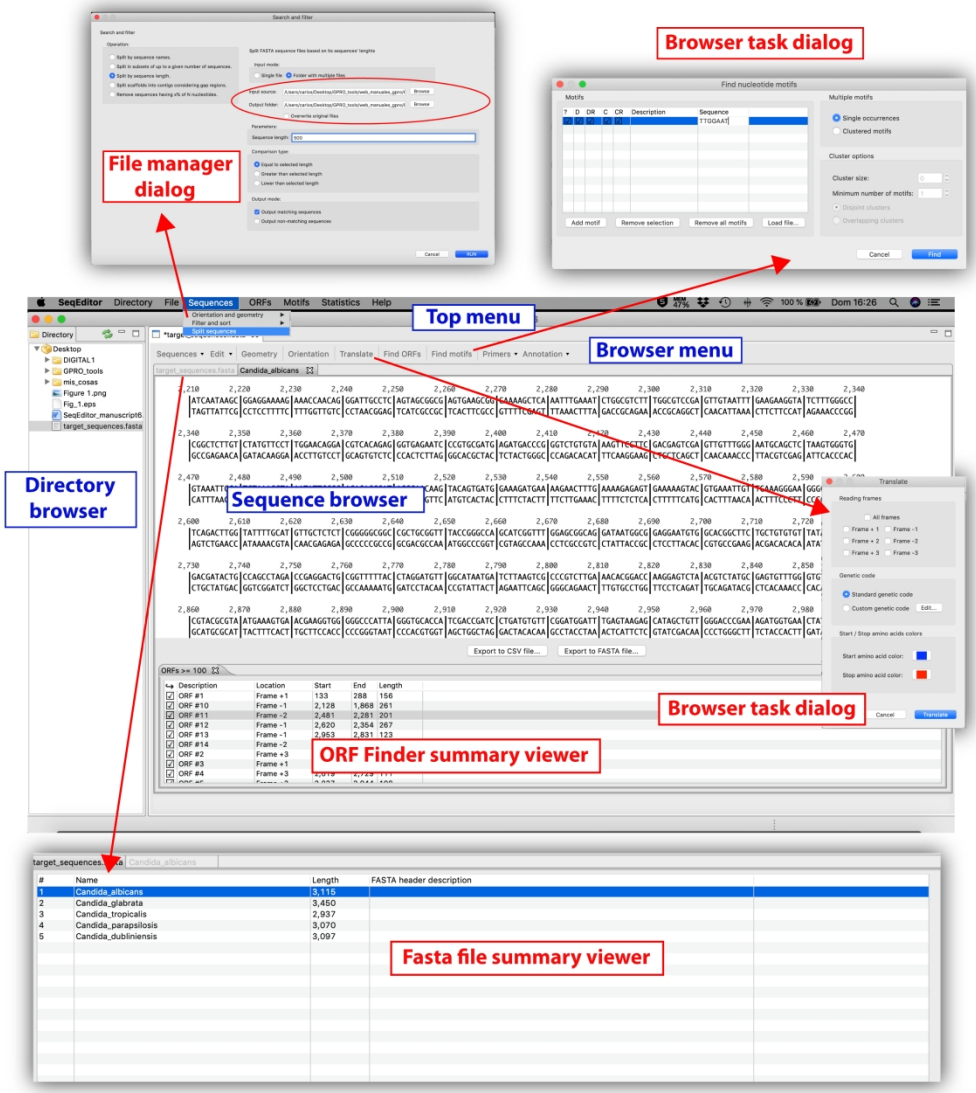

Figure 1: Layout and main sections of SeqEditor. The directory browser, the top menu, the sequence browser and the browser menu are indicated with blue labels. Examples of interfaces, summary views and task dialogs for ORFs, motifs, primers are highlighted with red labels. Views are customizable and the user can adjust each location of each part by dragging it to the desired location. Note that the default layout on startup will only show the directory browser. Other views and windows will be displayed when performing the corresponding task.

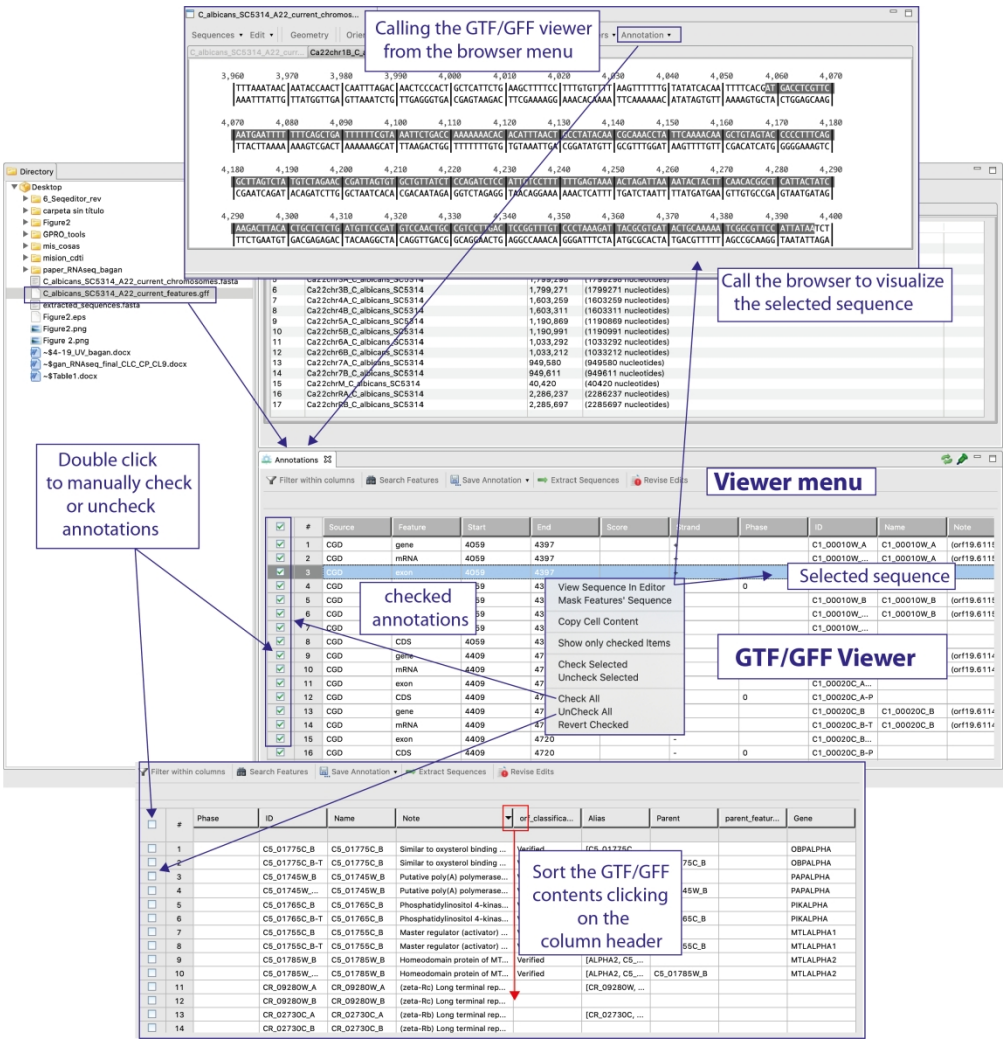

Figure 2 : GTF/GFF viewer and different options for mouse-dependent tasks in that viewer. Annotations can be manually selected or deselected by clicking (two times) with the mouse to check/uncheck rows or by right-clicking anywhere to call a context menu providing distinct checking options or for visualizing 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.

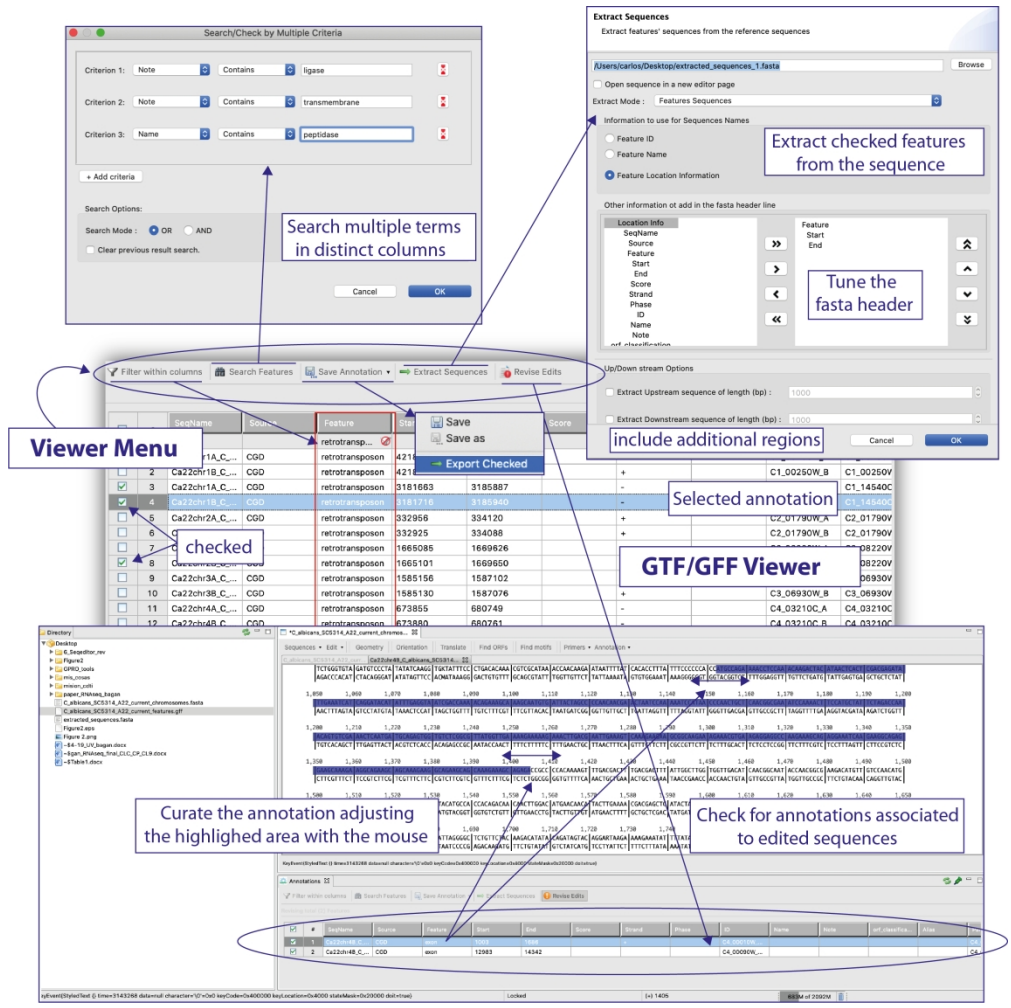

Figure 3 : Different task options provided by the menu of the GTF/GFF viewer. The tab "Filter within columns" let users to use a key word to filter annotations in a particular column and show only those that match the word. The tab "Search features" gives access to a context dialog allowing to specify 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" permits to save any edit or change done 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 to edit 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 is allowed to 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 (as also shown in Figure 3). 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. After this action 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 need to save the new GTF or GFF file using the options provided by the tab "Save Annotation".