TUTORIAL FOR PRIMER DESIGN WITH SEQEDITOR

1. Downloading SeqEditor and getting familiar

- SeqEditor can be freely downloaded at https://gpro.biotechvana.com/download/SeqEditor
- For general use indications, please refer to the SeqEditor manual available at https://gpro.biotechvana.com/tool/seqeditor/manual

2. Case study: Primer Search for Singleplex PCR using the Primer3[1] implementation of SeqEditor.

2.1 Data requirements:

To reproduce the examples, the following data are required

- **target_sequences.fasta**: this file is available in the data folder called sample-data and contains rDNA sequences for 5 *Candida* species as summarized in Table S1. These target sequences belong to ribosomal gene domains such as 28s and Internal Transcribed Sequences (ITS). The presence of sufficient polymorphism within these loci allow the design of species-specific primers. Being a multicopy gene, one might expect higher PCR amplification when compared to a single copy gene, usually protein-coding genes such as Actin, Beta-tubulin, Elongation factor, RPBII, etc
- **non_targets.fasta**: this file is available in the data folder called sample-data and contains sequences for other species which are closely related to the target species as summarized in Table S2. This was created by blasting the target sequences in the NCBI database (Sayers et al. 2019) and selecting those subjects with the highest similarity available.

Table S1: Target species and sequences

Species	Sequence Name	NCBI Accessions
Candida albicans	Candida_albicans_CBS_1949_NS1_LR5_53910_Lodderomyces_clade	CP025165.1 CP025157.1
Candida glabrata	Candida_glabrata_CBS_138_NS1_LR5_58190_Saccharomycetaceaea	CR380958.2 MK394140.1
Candida tropicalis	Candida_tropicalis_CBS_8072_NS1_LR5_53938_Lodderomyces_clade	MK394119.1
Candida parapsilosis	Candida_parapsilosis_CBS_604_NS1_LR5_81205_Lodderomyces_clade	HE605209.1

[1] Primer Search implementation in SeqEditor is powered by a modified java implementation of the command line interface (CLI) tool Primer3 (Untergasser et al. 2012)

Candida dubliniensis	Candida_dubliniensis_CBS_7987_NS1_LR5_53918_Lodderomyces_clade	FM992695.1

Table S2 : Non target species and sequences

Species	NCBI ccessions
Homo sapiens	NG_054875.1; MF164260.1
Saccharomyces cerevisiae	KJ806314.1
Debaryomyces hansenii strain ATCC	GQ458041.1
Pichia norvegensis	NG_063278.1
Pichia norvegensis	AY497674.1
Pichia kudriavzevii	CP028774.1
Candida lusitaniae	M55526.1
Clavispora lusitaniae	JQ698900.1; KY106935.1; KY106931.1
Candida haemulonis	NG_063413.1; JN941107.1
Candida rugosa	KT336717.1; AB013502.1
Candida inconspicua	EF152417.1; KY106513.1
Diutina rugosa	KY563206.1
Aspergillus fumigatus	MF379664.1; KJ809565.1

2.1 Primer search for target region detection.

The objective of this example was to show how to use SeqEditor for the design of those singlePlex PCR primers, which will be able to detect the target DNA sequences. To do this, first open the target_sequences.fasta file with SeqEditor and then go to the Singleplex path via the Editor Top menu (Figure S1).

[**Primers -> Singleplex Primers**].

There are two different ways to run the Singleplex Primers design tool:

Batch/Single run mode

- To run the search in batch mode (i.e. for all sequences in the file), select Singleplex Primers while the fasta sequences file summary view is active.

- If a single sequence fasta is opened in the Sequence Browser, you can still run Singleplex Primers. However, please keep in mind the browser will operate only on the sequence that has been opened.

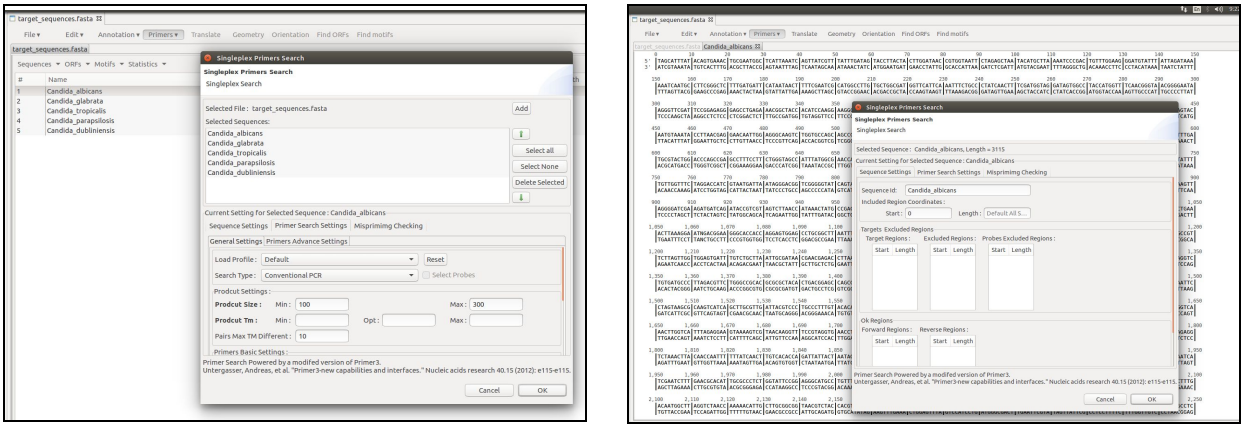

Figure S1. Singleplex Primer Search Dialog. Left batch mode. Right Single Mode.

Once you have selected the correspondent primer search mode from the menu, a search dialog will be opened to configure the search parameters. For each input target sequence, you can edit the search parameters.There are three main tabs for the editing of these search parameters:

- Sequence Setting: target sequence related info including :
  - A custom ID for the search result.
  - **Included region**: A subregion in the given input sequence inside which the search for primers is performed.
    - All region inputs are defined by two parameters:
      - **Start**: index of the first base of the region in the target sequence.
      - **Length**: total length of the subregion starting at the **Start** parameter.
  - Targets and excluded regions:
    - **Target regions**: subregions of the input sequence within which a suitable primer pair must flank at least one of them.
    - **Excluded Regions**: subregions within the input sequence excluded from the primer search (i.e. rejecting all primers or probes that overlap any of those regions).
    - **Probes Excluded Regions**: same as described for the excluded regions but only applying for probes selection.
  - **OK Regions**: possible sub regions in the sequence the search is directed to when designing primers (i.e. constrained search regions). Each entry in this configuration includes two sub-regions, including one for the forward primers and one for the reverse primers. It is also possible to constrain a search region for one of the primers only (forward or reverse) and not the other.
- Primer Search Settings:
  - General Settings: This section includes parameters such as primer length, Tm and GC content as well as product size. PCR type and load can also be changed by selecting one of the predefined PCR configurations (Profiles).

- Primers Advanced Settings: These include thermodynamic calculations parameters as well as other primers properties and score calculation settings.
- Probes Advanced Settings: this section will be shown when probe selection is enabled. These are similar to the Primers Advanced Settings, however to be applied for probes only.

- Mispriming Checking: provides the search with a library of sequences to mispriming.

Once all parameters are set, click on the OK button. The primer search will be launched in the background and you will be able to track it or cancel it from the progress view.

Primers Result view:

The bottom left side panel of Figure S2 shows a tree-like view of the primer search result sorted by the custom ID/sequence name that is provided in the search dialog. Candidate PCR products and primer pairs are listed under each group. The bottom right side panel of the figure displays the result in different sections, including a search summary at the top, a PCR product section containing a table of all candidate PCR products and their properties, and a table of all primers/probes of all products and their properties at the bottom. This window also includes a ”save the result to file” option where the user can export the results in either CSV, Fasta or GFF file format.

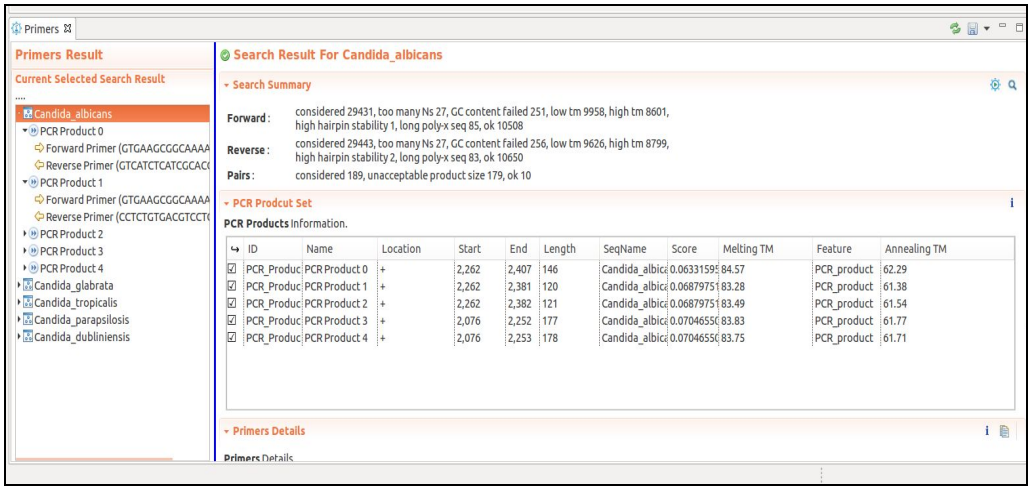

Figure S2. Primers Result View

2.2 Species-specific primer search

The aim of this example is to describe the process that SeqEditor follows to search for species-specific PCR primers. That is, those primers that will allow the amplification of DNA fragments via PCR experiments in the target species but not in any others. To do so, SeqEditor allows the user to include a mispriming library containing sequences belonging to the non-target species in the search. This process is shown in Figure S3 as follows:

Open both “***non_targets.fasta***” and “***target_sequences.fasta***” fasta files with SeqEditor. From the ***target_sequences.fasta*** file, select [**Primers -> Singleplex Primers**] in the menu toolbar. Next, adjust

the input parameters in the search dialog as described in previous example steps. From there, select any of the input sequences to select their parameters (e.g. select *C. albicans*) In the "Mispriming checking" tab -> non target library section, choose "construct library" in the source field. From the table context menu select "add new". In the selection dialog, select "all sequences" in the non-target fasta files. To avoid including any of the other target species, select all sequences in the same file except the input target sequence (e.g. *C. albicans* in this case). Repeat the same steps again for each target species.

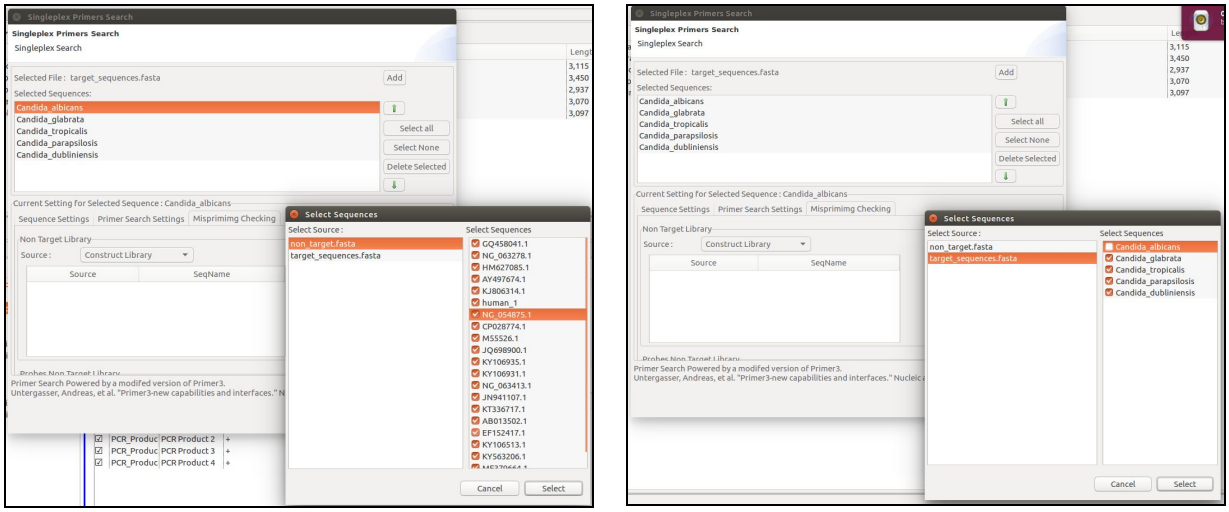

Figure S3. Adding a mispriming library to the the primer search parameters. (Left) Adding all sequences in the none_targets.fasta file. (Right) Adding all target sequences except *Candida albicans* as it is the currently selected sequence.

After configuring all search parameters, click on the "OK" button. The Primers result view as depicted in Figure S4 will be shown upon completion. Note that the results obtained in this case will be different from those in the previous example since the search summary for *C. albicans* that the tool considered in this search was **1291360** different primer pair combinations *versus* only **189** combinations obtained in the previous example shown in Figure S2. This is because most of the forward and reverse primers have a high similarity to one or more of the non-targets sequences.

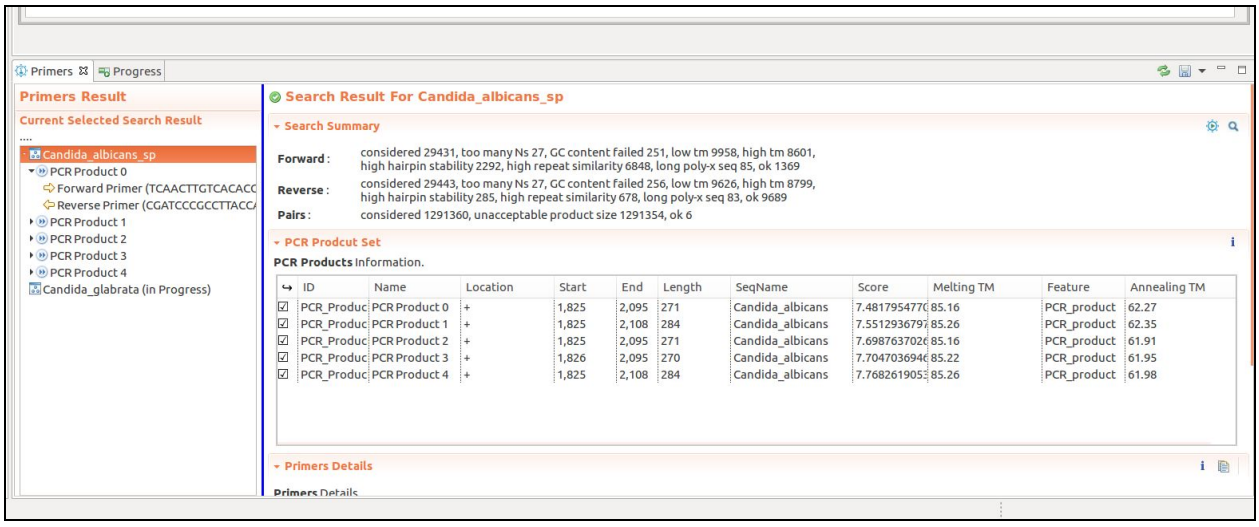

Figure S4. Primers Result View of the second example with mispriming checking.

3. Case study: Primer Search for Multiplex PCR.

The example shown above described the design of species-specific primers. These primers however, could not be used in a multiplex setting as some of the PCR products share the same length, making them indistinguishable from each other. One way to adopt this Singleplex search to multiplex PCR is to change the desired product size range for each species in the search to have both distinguishable product sizes in the PCR for each targeted to predict any dimer formation between primers. This might require running the search multiple times until a suitable set of primers is found. Taking this into consideration, SeqEditor provides a Multiplex search option that runs this process automatically yielding a valid primer set for either conventional PCR or qPCR.

To run this test, you will need the two files used in the previous example.

- In the ***target_sequences.fasta*** file tab, select [**Primers -> Multiplex Primers**] from the top toolbar menu.

Adjust the search parameters in the search dialog shown in Figure S5. In this case, all input parameters in both "Primer Search Settings" and "Mispriming Checking" will remain the same for all input sequences, however the "Sequence Settings" parameters can be different for each target as those parameters will control the search regions in the provided sequence.

  - If Conventional PCR is selected, adjust the product size so that the final PCR product can be distinguishable from the others. In this particular example, a range of 100-600bps would be suitable for 5 targets. However, broadening the size range will allow the tool to find a higher number of suitable solutions.

If intercalating dye-based qPCR is selected (e.g. SYBR Green), the preferred PCR product length should range from70-150bps up to a maximum of 200bps. The most important criterum here will be the product Tm, for which the tool will try to return PCR products with Tm differences of at least 1ºC.

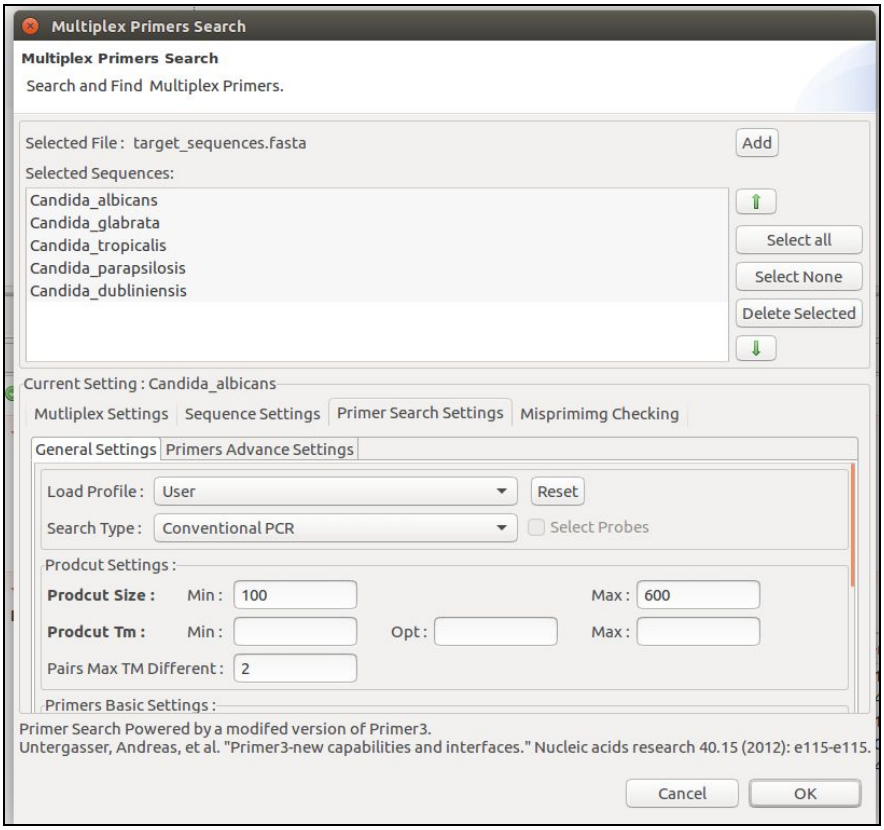

Figure S5. Multiplex Primer search dialog.

For Mispriming checking -> non-target library section, choose “construct library” in the source field, then select “add new” from the table context menu. In the selection dialog, select all sequences included in the non-target fasta files. Do not add any sequence from our targets as mispriming check against the target sequences will be performed in the multiplex search by default.

After configuring all search parameters, click on the “OK” button and wait for the search to complete. Once the search is finished, the Primer result view (Figure S6) will be shown. The result view in this case will be different from the Singleplex result. The left side panel shows the result primers grouped by candidate multiplex sets (valid Multiplex PCR primers). The right hand side panel of the results view will display the same information shown in the Singleplex case. From this point, the user can navigate the search results by selecting any of the sets in the tree view to show only those primers included in the selected set.

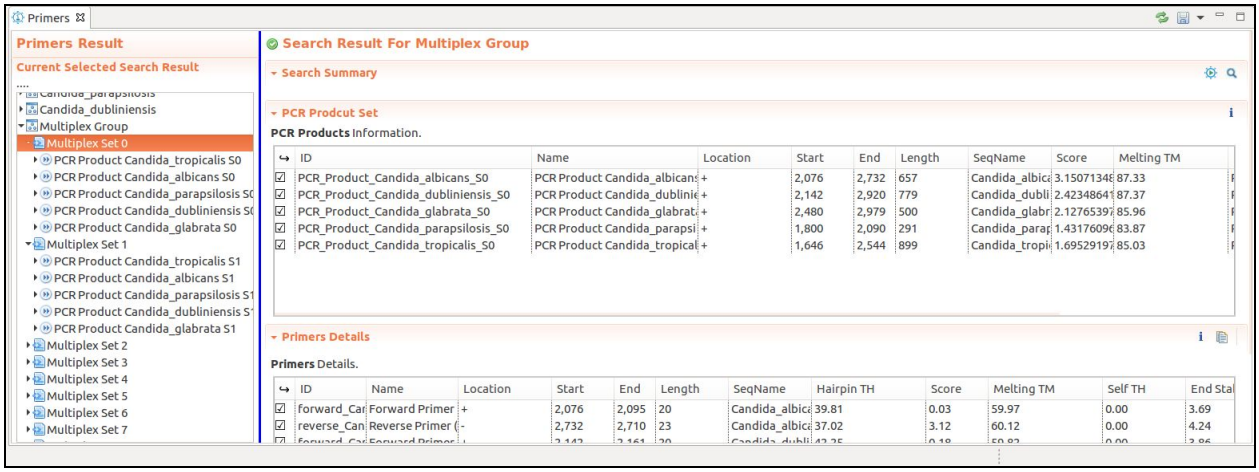

Figure S6. Primer Result View for a Multiplex PCR example.

3.1. Validating Multiplex primers sets:

For the experimental validation of the primer set results obtained in the test described above, a conventional PCR experiment was performed using the multiplex primers set shown in Table S3.

Table S3.- Multiplex primer set used in the validation PCR experiment

Species name	Primer sequence	PCR product length (bps)
C. albicans	forward: CCAAAAACATTGCTTGCGGC	613
	reverse: CAGAGGCTATAACACACAGCAG	
C. glabrata	forward: CGACTCCACTTCAGAGCGG	290
	reverse: ACACTCCCAGGTCTTTGTCG	
C. tropicalis	forward : GAGCAATCCTACCGCCAGAG	363
	reverse: TGGTGGCCACTAGCAAAATAAG	
C. parapsilosis	forward: GGTAGGCCTTCTATATGGGGC	977
	reverse: GCCAACATCCTAGGCCGAA	
C. dubliniensis	forward: CACCACATGTGTTTTGTTCTGG	412
	reverse : CCAGAGACCGCCTTAGCAAT	

The correspondent gel electrophoresis is shown in Figure S7. The correspondent PCR experimental conditions were as follows:

- Gel electrophoresis conditions: 6µl of each PCR product were run in a 2% agarose gel in TAE buffer (40 mM Tris-Acetate, 1mM EDTA) for 50 min at 130V.
- PCR mix (per sample): 2.5U of Biotaq DNA polymerase (Bioline), 10X Polymerase reaction buffer, 1.5mM MgCl2, 0.2mM of dNTP mix, 5 pmol of each primer, 1 ng of genomic DNA, miliQ-H_2O up to 50µl of final reaction volume.
- PCR program: 95 C for 5min

95ºC for 30 sec (x35)
60ºC for 30 sec
72ºC for 1min
72ºC for 8 min.

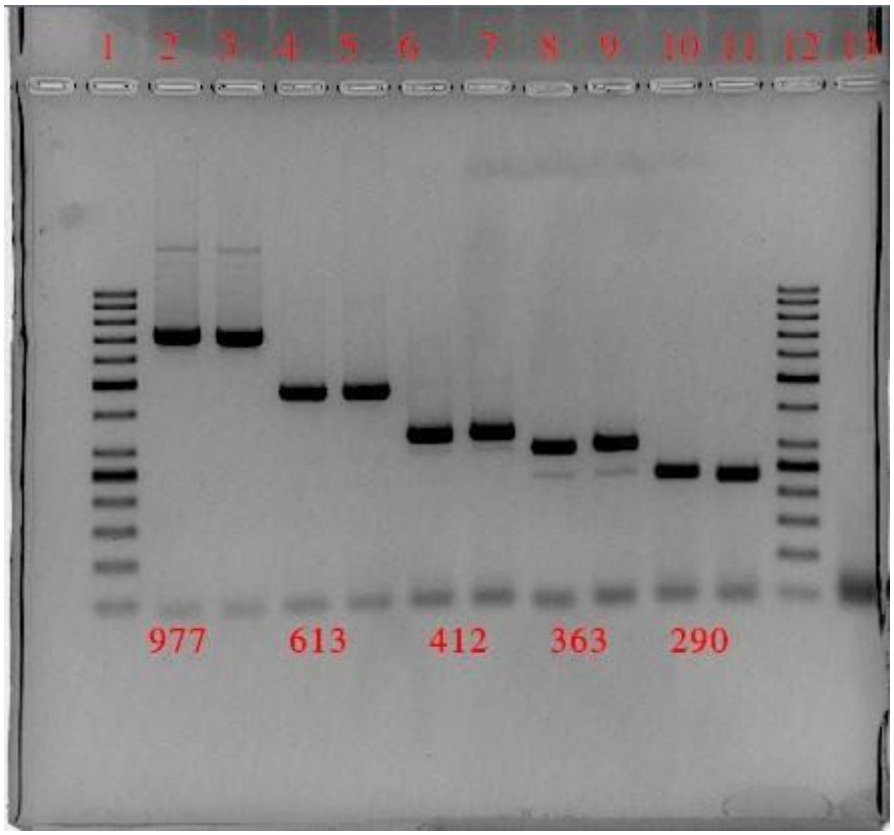

Figure S7. Gel electrophoresis of the PCR products amplified through the use of primer pairs designed with "Multiplex primer search" in SeqEditor. Lane 1. 50 bp DNA ladder (ThermoFisher). Lanes 2-3; *C. parapsilosis*, Lanes 4-5*: C. albicans*, Lanes 6-7: *C. dubliniensis*, Lanes 8-9: *C. tropicalis*, 10-11; *C. glabrata*, Lane 12 (Ladder) 13: Negative control (-genomic DNA).

4. Case study: Splitting Sequencing primers to pools using the PrimerPooler[2] implementation of SeqEditor.

The objective of this example is to show how to use SeqEditor to either split or distribute amplicons between multiple pools in order to reduce dimer formation.

Prerequisites to run the following example are as follows:

1. Download the pooler_example.fasta file available in the data folder included in this tutorial within the folder called sample-data. The file contains a list of sequencing primers for the human genome (taken from(Brown et al. 2017).

[2] PrimerPooler implementation in SeqEditor is powered by modified java implementation of the CLI tool PrimerPooler (Brown et al. 2017).

2. Download the human genome hg38.2bit available athttp://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.2bit . You may also download other genomes in 2bit as provided in the UCSC Genome Browser (http://hgdownload.cse.ucsc.edu/downloads.html).

Once you have the two files, open pooler_example.fasta in SeqEditor.

Data input format

Please note that input files should be a fasta file containing primer sequences, and that each sequence name should end up in either F or R (standing for forward and reverse respectively).Primer pairs should all have the same name prefix (e.g. "primer"). Degenerate bases are allowed when using the usualDNA code letters.

Example of valid input data:

```
>primer1-F
CGCCGTCTTCCACCAACCA
>primer1-R
GGTAGGCGCTGCGGTT
>primer2-F
TCACAAAACACTTCATCTTTACTCAT
>primer2-R
CTCCAGTCCTCTCAGCCT
```

The "examples" folder contains a valid data input example denoted as "pooler.example.fasta" which contains primer sequences for the human genome.

The user may also add tags to the primers using >tagF and >tagR. These primers will then be referred to as "tailed primers" and this process is known as barcoding) . The primer tags can be changed part-way through the file). Should the user have either any Taq probes or other primers that do not yield amplicons, these can end with other letters (e.g. >probe1-P). Any set of names differing in only the last letter will be kept in the same pool.

3. From the directory browser or File Menu open the example file "pooler.example.fasta"). Once open, select Primers -> Primer Pooler from the Editor Top menu. A Primer Pooler dialog will be displayed as shown in Figure S8. In the Input Primers/Tags List, show all input sequences and check that the input contains the required information.

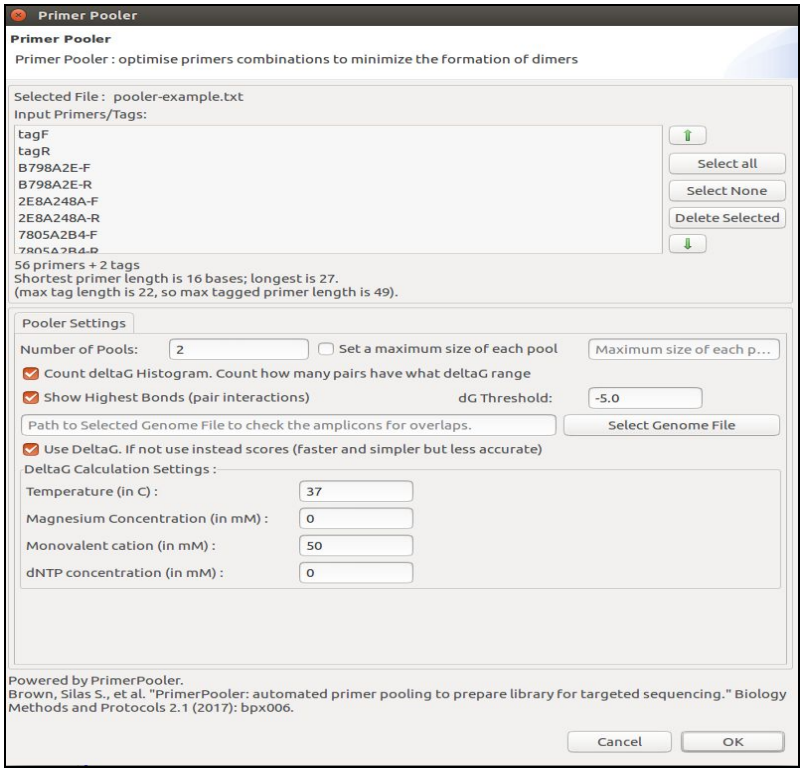

Figure S8. PrimerPooler dialog.

From the PrimerPooler dialog you can the configure the following tool parameters:

- Number of pools: number of pools to divide the input into.
- Set Max size of pools: setting a maximum size of each pool can make the pools more even.
- Count deltaG/Score: Create histograms from the deltaG/Score of all pairwise interactions of all primers.
- Show Highest Interaction: shows a summary and the dimer structure formed from the pairwise interaction with the highest interaction.
    - Threshold: sets either a maximum or a minimum dG score for the display of the pairwise interaction.
- Select Genome: select a genome file in 2bit format to check the amplicons for overlaps. Primer pairs that amplify overlapping regions of the genomes can produce an unwanted shorter amplicon if used in the same pool. (Fasta file use support will be added in future releases of the tool).
    - For this particular example, may can select hg38.2bit from the examples folder.
- Use DeltaG: use thermodynamic principles to calculate the correspondent ΔG for the pairwise interaction. 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 but less accurate way. To use deltaG you will need to input the following parameters:
    - Temperature.
    - Concentration of magnesium.
    - Concentration of monovalent cation
    - Concentration of deoxynucleotide (dNTP).

Upon parameter selection, click “OK” to launch the task. You can track its progress or cancel it from the progress view. To open the progress double click on the progress bar at the bottom of the screen (Figure S9). Alternatively, open the progress view from the views menu.

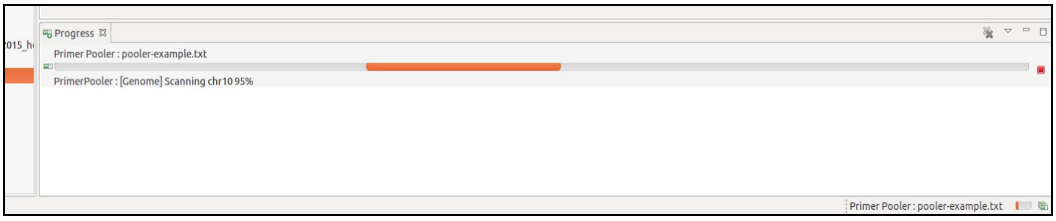

Figure S9. Progress View showing the progress of a task launched with the PrimerPooler tool.

Upon completion, a new PrimerPooler view will appear containing a full report of the results as shown in Figure S10. The left side of the view will display a tree view of the pools and primers assigned to each pool. The right-hand side panel will contain different sections of the result depending on the options selected by the user in the run dialog.

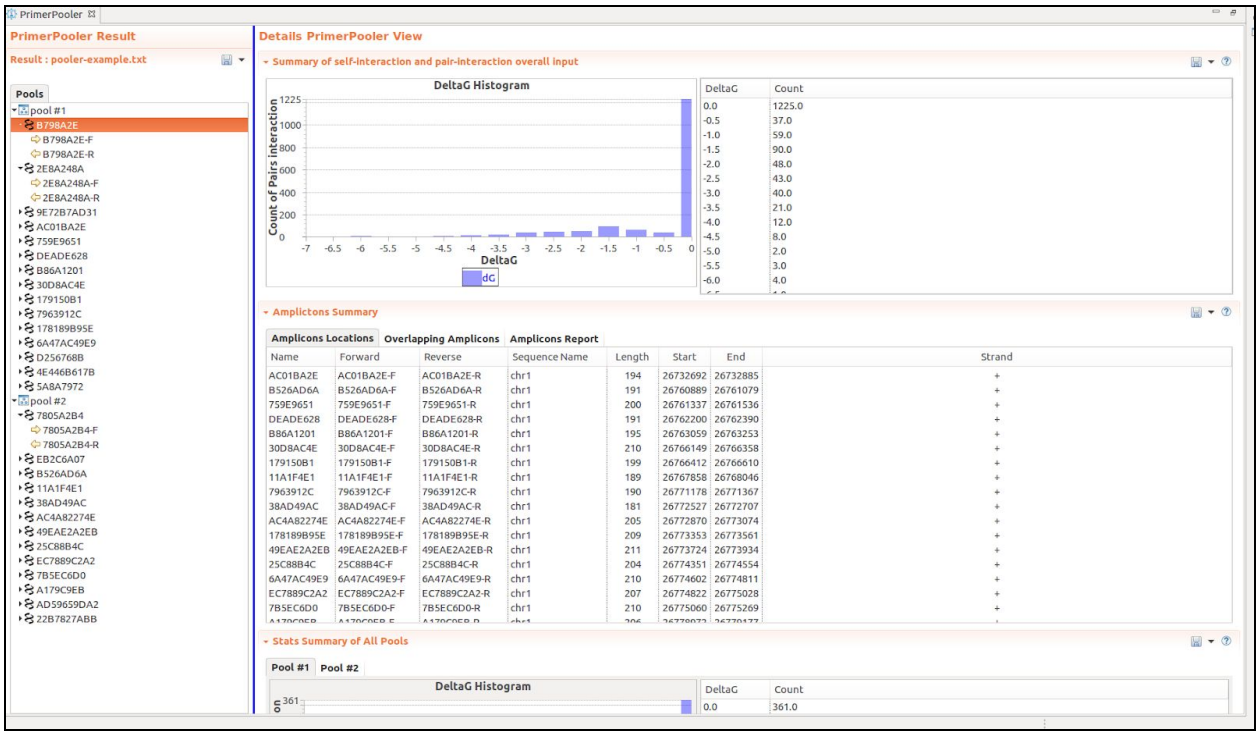

Figure S10. PrimerPooler result View.

If “Count deltaG/Score” is selected, a count summary of such interactions will be displayed in both a graph and a table as shown in Figure S11. The user may export these results from the correspondent “save” menu.

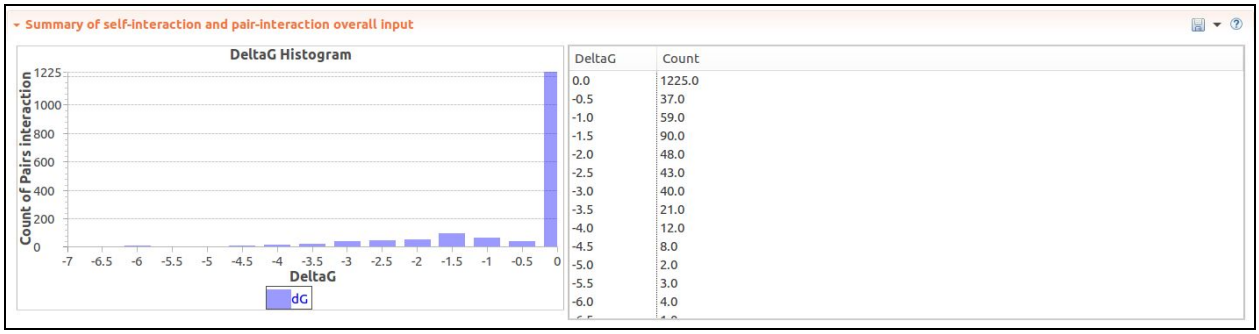

Figure S11. Summary of the "Count of pair interactions" overall input.

If a genome file is provided, an amplicons summary section will be displayed as shown in Figure S12, containing three tabs for amplicon locations in the genome, overlapping amplicons details and an amplicon summary report in text format. These results may also be exported via the "save" menu. In the Amplicon locations table, those amplicons that cannot be located in the genome will show no corresponding location.

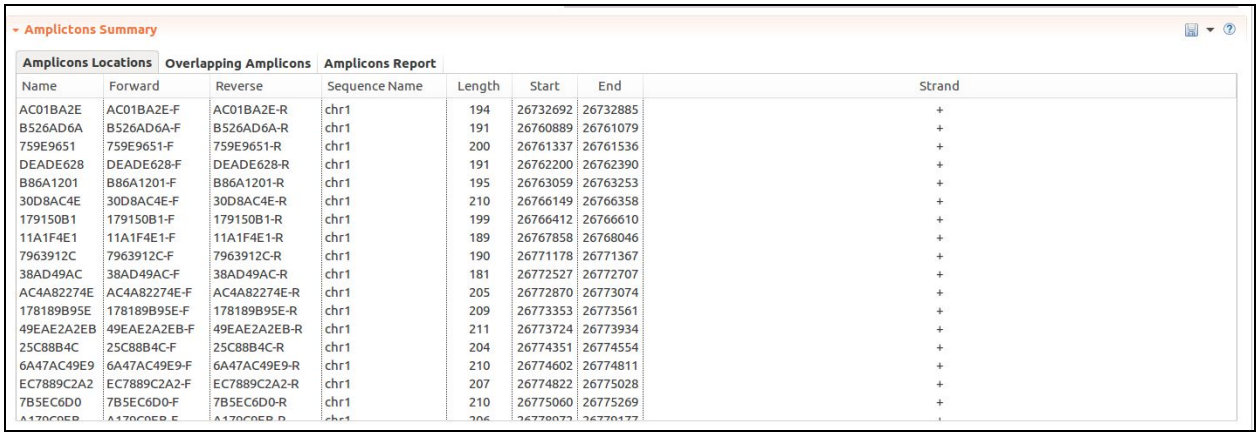

Figure S12. Amplicon location and overlap summary.

Figure S13. shows the "Stats Summary of All Pools", showing the same information as that included in "count of pair interaction" but this time at the level of each pool (e.g. only consider interaction between primers within the same pool). In this section a tab for each pool will be added, so the user can navigate the results by selecting the pool in the corresponding tab.

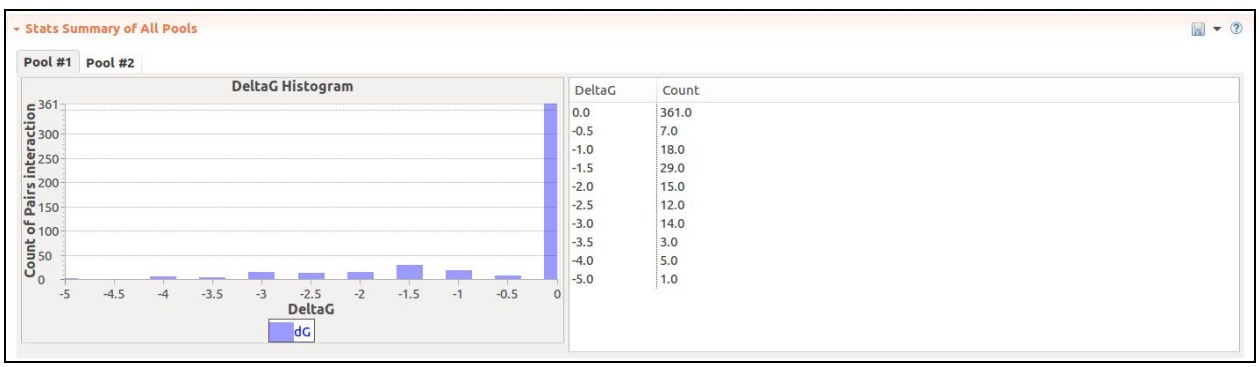

Figure S13. Pools Stats Summary.

Lastly, if "Show Highest Interaction" is selected, a section for the interaction will be displayed showing the highest interaction and highlighting the dimer structure formed between the two primers (Figure S14). From the section menu the user may then re-run this task selecting a different score/dG threshold value, then export the result to text files if desired.

Figure S14. Pools Stats Summary.

References

Brown, Silas S., Yun-Wen Chen, Ming Wang, Alexandra Clipson, Eguzkine Ochoa, and Ming-Qing Du. (2017). "PrimerPooler: Automated Primer Pooling to Prepare Library for Targeted Sequencing." Biology Methods and Protocols. https://doi.org/10.1093/biomethods/bpx006.

Sayers, Eric W., Jeff Beck, J. Rodney Brister, Evan E. Bolton, Kathi Canese, Donald C. Comeau, Kathryn Funk, et al. (2019). "Database Resources of the National Center for Biotechnology Information." Nucleic Acids Research, October. https://doi.org/10.1093/nar/gkz899.

Untergasser, Andreas, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C. Faircloth, Maido Remm, and Steven G. Rozen. (2012). "Primer3—new Capabilities and Interfaces." Nucleic Acids Research. https://doi.org/10.1093/nar/gks596.

Citing SeqEditor

If you need to cite SeqEditor please, use the following reference

Hafez, A. Arastehfar, A. Daneshnia, F. Miguel, A. Roig,F.J. 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 (Submitted)