EasyPrimer Online - Manual

Free tool available at: https://skynet.unimi.it/index.php/tools/easyprimer/

EasyPrimer is a user-friendly open-source tool developed to assists primer design in difficult contexts, e.g. on an alignment of hundreds of sequences and/or on hypervariable genes. The tool uses as input a nucleotide multi-fasta file and identifies the best regions for primer design: two low variable regions flanking a highly variable one.
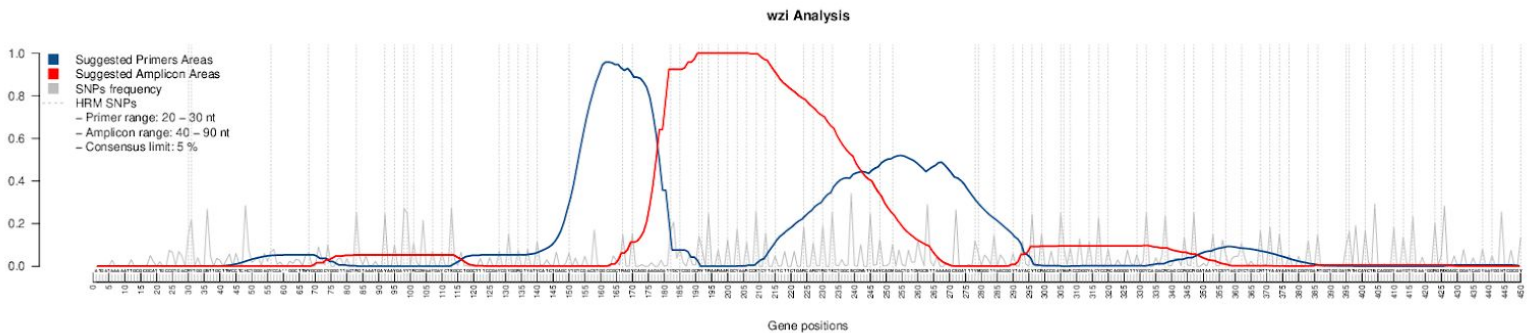
For questions or support please contact matteo.perini@unimi.it.

How to use the tool:

Mandatory INPUTS:
- a valid **email address**
- a gene **multi FASTA*** file either by uploading a FASTA file or by copy-pasting the FASTA file content in the textbox.
- Select **Analysis type** to "HRM" or "ALL" SNPs (explanation Below)

How to read the OUTPUT:

Example of a EasyPrimer output of a gene with 563 alleles. In the zoomed image below The blue arrows were manually added to show the positions of the two primer pairs designed on Perini et al. 2020 (https://doi.org/10.1038/s41598-020-57742-z).
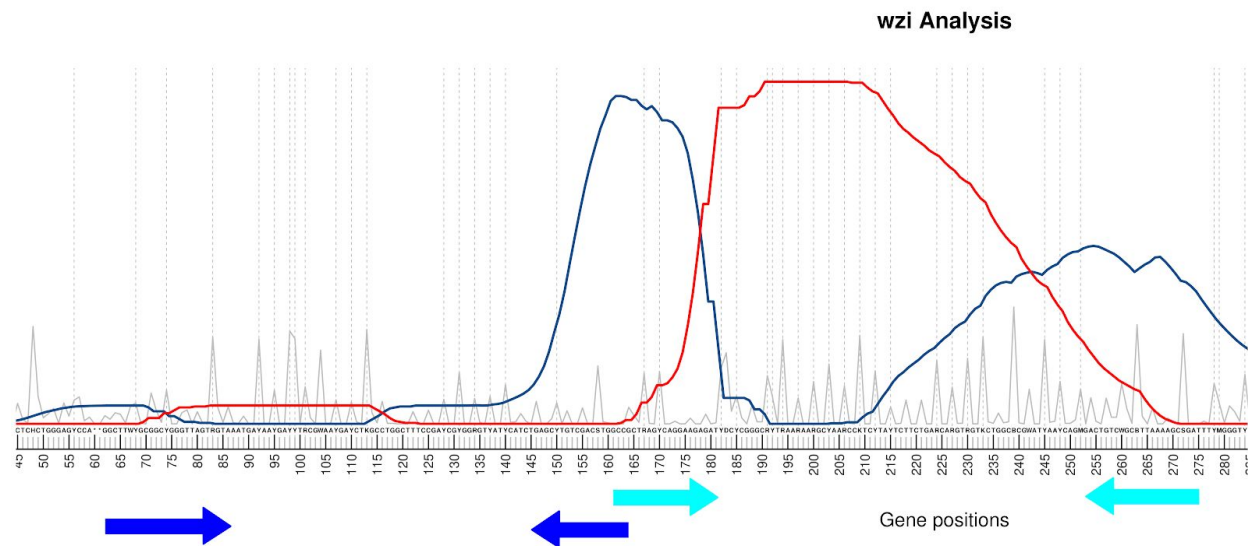
The consensus sequence calculated from the gene alignment is reported on the x-axis.
- Residues under the peaks of the **blue curve** are highly conserved and thus they are suitable for primer design.
- Conversely the **red curve** increases over the highly variable regions suggested to be

amplified. The grey peaks represent all the Single Nucleotide Polymorphisms (SNPs) with their own frequency.

- The **dotted grey lines** are used to highlight the "HRM-detectable" SNPs, i.e. the ones causing a change in the GC content.

- **Red residues** are added to indicate in the consensus sequence that gaps in the alignment in that position are greater than consensus threshold (more than 5% in default) and les than 1 - consensus threshold (less than 95% in default).

- **Asterisks (*)** indicate a gap present more than 1 - consensus threshold (more than 95% in defaults), this indicates an insertion in one or very few variances in the gene. The user will probably want to consider that position as 'not existing' during the primer design process.

IMPORTANT NOTICES ON THE FASTA SEQUENCES

- The tool was designed to give the **best result** with gene alignment of **hundreds of sequences**. Uploading alignments with only a few sequences do not give the best results.

- The FASTA file to upload **must not have space characters** in the file's name; accepted extensions are: .fasta, .FASTA, .fas, .ffn, .txt, .fna, .fa.

- The online tool **works on genes**, not on congis (the maximum sequence length is 2500bp). Please consider trimming your genomic sequences to the contig portion (or gene) you are interested in. You can also run EasyPrimer without length restrictions in your UNIX-based machine, downloading the stand-alone version of the tool here:

https://skynet.unimi.it/wp-content/uploads/easy_primer/EasyPrimer.tar.gz

ANALYSIS TYPE OPTION EXPLANATION

Analysis type can be set to:

-"**HRM**": (High Resolution Melting) the tool will take into account the HRM-detectable Single Nucleotide Polymorphism (SNPs). An HRM detectable SNP confers a change in CG content to de amplicon (i.e., any A↔G or T↔G or A↔C or T↔G count as a HRM detectable SNP). Use this setting to design primers for HRM analysis.

-"**ALL**": the tool will take into account all the possible SNPs, use this setting for any PCR or Real-Time PCR analysis.

FURTHER INFORMATION

EasyPrimer was developed to identify the best regions to design primers in an alignment. Indeed, the tool finds the best "primer-amplicon-primer" gene region, where two highly conserved gene regions (suggested for primer design) flank a highly variable region (the "amplicon region"). The algorithm scans the alignment to find - at least - one highly variable region (amplicon region) flanked by two conserved regions (primer regions). For the computation of the highly variable region, the user can decide to consider all the Single Nucleotide Polymorphism (SNPs) or only the HRM-detectable ones (HRM-SNP): a HRM-SNP is a SNP that confers a change in CG content (i.e., any A↔G or T↔G or A↔C or T↔G counts as a HRM-SNP). EasyPrimer will identify the most HRM discriminatory region in the alignment when HRM SNPs are used to score the variability of the amplicon region. Conversely, considering all the SNPs the tool will identify the most variable region in general. Additionally,

the user provides length ranges both for primers and amplicon, the tool suggests regions for primer design with a reasonable spacing, meaning that they always amplify a region of optimal length.

*FASTA Definition by https://blast.ncbi.nlm.nih.gov/:
"*A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line (defline) is distinguished from the sequence data by a greater-than (">") symbol at the beginning. It is recommended that all lines of text be shorter than 80 characters in length.*"
Example sequences in multi FASTA format is:
>HSBGPG_(BGP)_variant1
GGCAGATTCCCCCTAGACCCGCCCGCACCATGGTCAGGCATGCCCCTCCTCATCGCTGGG
CACAGCCCAGAGGGTATAAACAGTGCTGGAGGCTGGCGGGGCAGGCCAGCTGAGTC
>HSBGPG_(BGP)_variant2
GGCAGATTCCCCCTAGACCCGCCCGCACCATGGTCAGGCATGCCCCTCCTCATCGCTGGG
CACGGCCCAGAGGGTATAAACAGTGCTGGAGGCTGGCGGGGCAGGCCAGCTGAGTC
>HSBGPG_(BGP)_variant3
GGCAGATTCCCCCTAGACCCGCCCGCACCATGGTCAGGCATGCCCCTCCTCATCGCTGGG
CACGGCCCAGAGGGTATAAACAGTGCTGGAGGCTGGCGGGGCAGGCCAGCTGAGT