- manual -

1. Scope and preface

easyPAC is a software that facilitates primer design based on multispecies alignments or consensus sequences. It can be typically used to design primers for the amplification of orthologous genes in comparative gene and genome studies. Currently, easyPAC is the only freely available primer design software that allows testing of degenerate primers by separate testing of each possible primer sequence. Testing involves assessment for adequate conservation, T_m, GC-content, 3′-end complementarity and stability. The probability of mispriming can be reduced by providing an arbitrary number of reference files like repeat/transposon databases, lists of paralogous genes or even whole genomes, thus obviating the need for subsequent primer BLAST.

2. Installation

Download, unpack and copy the easyPAC folder including all subdirectories to the desired location on your computer. easyPAC will create output files and temporary files and will optionally run SeqMap for Primer mapping. This means that you should have the administrative rights to read/write/create and execute files in this directory while running easyPAC. We provide easyPAC as executable (.exe) file that runs on Windows 64bit systems without further requirements. Alternatively, we provide the original platform independent PERL script containing the easyPAC source code. Running easyPAC as PERL script requires the installation a freely available PERL distribution like Strawberry PERL (http://strawberryperl.com/) or ActivePerl (http://www.activestate. com/activeperl). PERL is preinstalled on most Macintosh and Linux systems. However, the installation of additional PERL modules may be required if they are not already part of the installed PERL distribution: Cwd, File::Copy, GD, Iterator::Array::Jagged, Tk, Tk::BrowseEntry, Tk::JPEG, Tk::Pane, Tk::PNG, Tk::StayOnTop. Additional PERL modules are freely available at the Comprehensive Perl Archive Network (CPAN, http://www.cpan.org/). easyPAC optionally searches primer candidates in reference files via SeqMap (Jiang and Wong 2008). SeqMap executables are provided within the easyPAC_files folder and easyPAC will automatically choose the right file for your operating system (Windows, Linux or Macintosh OS). If you are on a different operating system, SeqMap search is not supported. If you use SeqMap for primer search, please also cite the Jiang and Wong paper (see below).

3. Input and PCR conditions

The input can be either a multi-species (multi-gene) alignment in FASTA format or one single (consensus) sequence. easyPAC supports the full IUPAC nucleotide code (ATGCRYSWKMBDHVN). The input can be imported from a separate file or the alignment/sequence can be pasted into the box on the left. You can adjust the maximum number of gaps (n) at a certain position of the alignment that will be tolerated. Loci with a gap in less than n sequences will serve as putative primer target. You can also use predefined forward/reverse primers to find adequate reverse/forward primers. By default, T_m will be calculated via the nearest neighbor base stacking

method using thermodynamic parameters from SantaLucia (1998). Alternatively you can choose to calculate T_m using the following equation: $T_m = 64.9°C + 41°C \times$ (number of G's and C's in primer sequence – 16.4) / (primer length).

4. Primer settings

This box provides all customary primer design options. Most of the settings are self-explanatory. The maximum 3′-complementarity describes mutual- (primer pair) as well as self-complementarity (scoring: match=1, mismatch=-1, gap=-2). The maximum 3′-stability describes the 3′-stability for the last five 3′-bases (high 3′-stability will promote mispriming). The value refers to the maximum $|\Delta G|$ [kcal/mol] for duplex disruption using Nearest-Neighbor thermodynamic parameters. You also can demand a specific 3′-end of the primer (e.g. S [G/C], if a higher 3′-end stability is desired). Finally, you can set the maximum size of poly-X and poly-XY motifs within each primer (e.g. ATATWTAY → poly-XY value = 4).

5. Reference and program performance

easyPAC optionally compares all putative primer sequences with one or several reference files. Primers that match to a reference sequence will not be rejected. A canny strategy is to supply one file that contains paralogs of the gene in question (if exist) and another file that contains repetitive elements from the taxonomic order in question. Repeat sequences can be downloaded at RepBase: http://www.girinst.org/repbase/ (Jurka 2000). Alternatively, you can supply a file that contains the whole genome of the species in question and click the option 'Allow primer to match once in reference'. This will ensure primer specificity while increasing computation time. To compare primers to reference, easyPAC holds two different search algorithms. The first is the SeqMap algorithm which is very fast and allows up to five mismatches and three insertions/deletions, but maps only ATGC. Though, if your reference file(s) contains degenerate sequences (e.g. transposon consensus sequences) it might be better to search with the internal easyPAC algorithm which allows only one mismatch (including insertion/deletion) and works notably slower but will e.g. map A to ARWMDHVN. Since this would also map a primer sequence to a stretch of N's, easyPAC will initially remove poly-N from your reference file(s). The minimum number of N's to be removed from reference can be specified (the default value is 6).

Generally, it is recommended to use SeqMap when your reference is very large (e.g. a whole genome) and contains no or very few degenerate positions. If your reference is small and degenerated (e.g. transposon consensus sequences), you should use the easyPAC algorithm.

easyPAC will produce a graphical output including an annotation of your alignment (or single consensus sequence) which highlights the PCR target, excluded regions, regions out of range (max. product size), simple repeats and internal duplications. You can optionally also mark matches to the reference file(s) but this is not recommended if you use large reference files (genomes) since this will take long computation time.

In addition, the indication of sequence conservation and the display of the primer position on the alignment will make primer suggestions easily comprehensible. The number of suggested primer pairs can be affected by adjusting 'Minimum primer pair output' and 'Take best n for./rev. primers to find proper pairs'. First, the best n forward and reverse primers are combined to proper combinations. If the number of proper combinations is less than the value of 'Minimum primer pair output', this step will be repeated taking the best n x 2 primers and so on, until the desired number of proper combinations or the maximum possible number of proper combinations is found.

6. Results / Output

When primer search is finished and proper primer pairs are found, you can display the results in text- or image form. The image will show you the annotated alignment including a consensus sequence and the suggested forward (green) and reverse (blue) primers (5′-3′) sorted by their quality. The conservation of the alignment is indicated by a color coded bar (see below). The textual output provides a list of proper primer pairs in order of descending quality. Forward and reverse primers are stated in 5′-3′ orientation. Beneath each reverse primer you can find the reverse complement in brackets to facilitate searching the reverse primer within the consensus sequence. Further information involves resulting product size, primer size, coordinates in alignment, melting temperature, GC content and degeneracy (ambiguity factor).
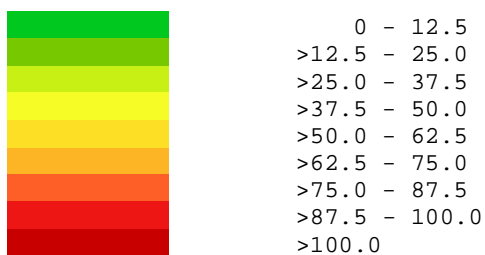
```
                                                50                                                100
        --------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--
Sequence 1   ACTAGCGAGCTGTGTGGCTACTGTAGCATCTGTGCTGAAGATGACGGGCTCTGNGCAGTGGAAGAGCAGTAGCACAGCGATCCCTGCAGTACCCCCCCCCC
Sequence 2   AC-TGCGAGCTCTGTGGCAACTGTAGCATCTGTGCTGAAGCTGACGG-CTCTGAGCAGTTGAAGAGGAGTAGCACAGCGATCCCTGCAGTACCCCCTCCCC
Sequence 3   AC-TGCGAGCTGTGTGGCAACTGTAACATCTGTGCTGAAGCTGACGGGCTCTGAGCAGTTGAAGTGCAGTAGCACAGCGATTCCTGCGGTACCCCTCCCC

Consensus    ACTWGCGAGCTSTGTGGCWACTGTARCATCTGTGCTGAAGMTGACGGGCTCTGNGCAGTKGAAGWGSAGTAGCACAGCGATYCCTGCRGTACCCCCYCCCC

Conservation
Annotation                                                         target    target    target                                        simple
Primer Pair #1               > ACTGTARCATCTGTGCTGA >                                                  < GGGGTACYGCAGGRATCG <
Primer Pair #2               > ACTGTARCATCTGTGCTGA >                                          < YGCAGGRATCGCTGTGCT <
Primer Pair #3               > ACTGTARCATCTGTGCTG >                                                    < GGGGTACYGCAGGRATCG <


PRIMER PAIR                        size        coordinates      Tm[°C]         %GC             ambiguity factor

#1  -> product size: 76 nt
forward:  ACTGTARCATCTGTGCTGA     19 nt       20  - 38        47.92 - 50.25   42.10 - 47.36    2
reverse:  GGGGTACYGCAGGRATCG      18 nt       78  - 95        51.45 - 56.16   61.11 - 72.22    4
          (CGATYCCTGCRGTACCCC)

#2  -> product size: 69 nt
forward:  ACTGTARCATCTGTGCTGA     19 nt       20  - 38        47.92 - 50.25   42.10 - 47.36    2
reverse:  YGCAGGRATCGCTGTGCT      18 nt       71  - 88        53.89 - 57.13   55.55 - 66.66    4
          (AGCACAGCGATYCCTGCR)

#3  -> product size: 76 nt
forward:  ACTGTARCATCTGTGCTG      18 nt       20  - 37        46.21 - 48.64   44.44 - 50       2
reverse:  GGGGTACYGCAGGRATCG      18 nt       78  - 95        51.45 - 56.16   61.11 - 72.22    4
          (CGATYCCTGCRGTACCCC)
```
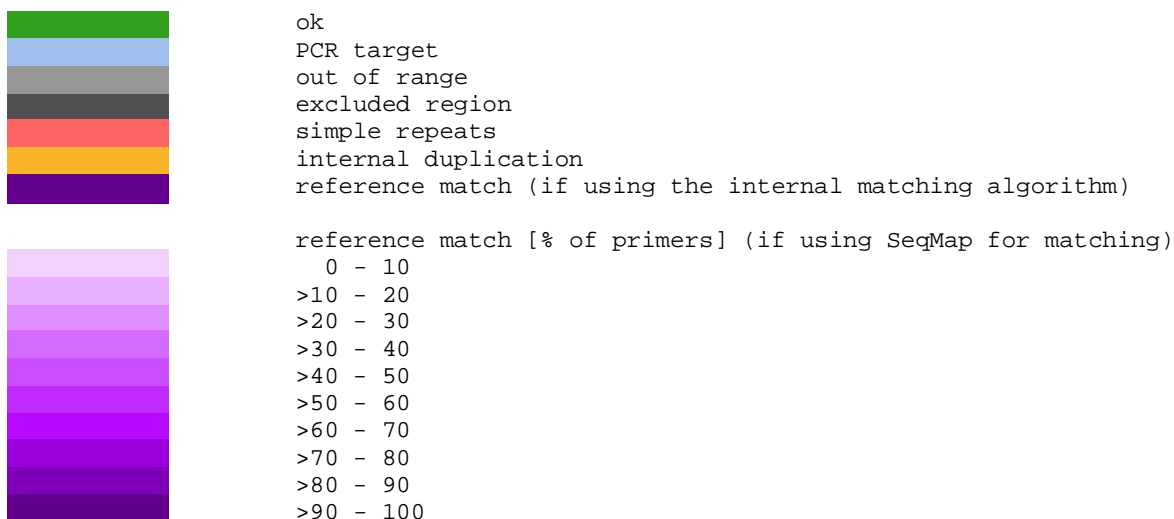
Sequence conservation color code [% of maximum allowed degeneracy]:

```
           0 - 12.5
        >12.5 - 25.0
        >25.0 - 37.5
        >37.5 - 50.0
        >50.0 - 62.5
        >62.5 - 75.0
        >75.0 - 87.5
        >87.5 - 100.0
        >100.0
```

Annotation:

```
        ok
        PCR target
        out of range
        excluded region
        simple repeats
        internal duplication
        reference match (if using the internal matching algorithm)

        reference match [% of primers] (if using SeqMap for matching)
          0 - 10
        >10 - 20
        >20 - 30
        >30 - 40
        >40 - 50
        >50 - 60
        >60 - 70
        >70 - 80
        >80 - 90
        >90 - 100
```

References:

JIANG, H. & WONG, W. H. 2008. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* 24(20): 2395-2396.

JURKA, J. 2000. Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.* 9:418-420

ROSENKRANZ, D. 2012. easyPAC: A Tool for Fast Prediction, Testing and Reference Mapping of Degenerate PCR Primers from Alignments or Consensus Sequences. *Evolutionary Bioinformatics* 8:151-159

SANTALUCIA JR, J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA.* 95(4): 1460–1465

Please report any bugs, trouble or suggestions for improvement to:

David Rosenkranz
Johannes Gutenberg-University Mainz
Institute of Anthropology
55099 Mainz, Germany
Anselm-Franz-von-Bentzel-Weg 7
rosenkrd@uni-mainz.de