

## Genome analysis

## Primaclade—a flexible tool to find conserved PCR primers across multiple species

Michael D. Gadberry, Simon T. Malcomber\*, Andrew N. Doust\* and Elizabeth A. Kellogg†

Department of Biology, University of Missouri-St. Louis, One University Boulevard, St. Louis, MO 63121, USA

Received on August 29, 2004; revised on November 1, 2004; accepted on November 2, 2004

Advance Access publication November 11, 2004

### ABSTRACT

**Summary:** Primaclade is a web-based application that accepts a multiple species nucleotide alignment file as input and identifies a set of polymerase chain reaction (PCR) primers that will bind across the alignment. Primaclade iteratively runs the Primer3 application for each alignment sequence and collates the results. Primaclade creates an HTML results page that recaps the original alignment, provides a consensus sequence and lists primers for each alignment area, with primers color-coded to reflect the level of degeneracy in the primer.

**Availability:** Primaclade can be accessed freely at <http://www.umsl.edu/~biology/Kellogg/primaclade.html>

**Contact:** [tkellogg@umsl.edu](mailto:tkellogg@umsl.edu)

### INTRODUCTION

Comparative studies of genes and genomes—including studies of molecular evolution, organism evolution and genetic mapping—rely on the polymerase chain reaction (PCR) to amplify orthologous genes among related organisms. Such studies require efficient methods to design primers from a nucleotide alignment. Currently, most available primer design software accepts only a single nucleotide sequence. Software that does design primers from a multiple species alignment, such as Primer Premier (Premier Biosoft International, Palo Alto, CA, USA) is limited by input format and is only available commercially. Thus we have developed a free, web-based, primer prediction application, Primaclade, to design minimally degenerate primers for comparative studies of multiple species.

### ALGORITHM AND IMPLEMENTATION

Primaclade employs a BioPerl-based executable file, which runs as a typical CGI script on an Apache-based web server (Fielding and Kaiser, 1997). Running the application requires a standard Perl 5.8.0 installation, a few Comprehensive Perl Archive Network (CPAN) Perl modules, the BioPerl 1.4.0 set of modules (Stajich *et al.*, 2002) and version 0.9 of the Primer3 software (Rozen and Skaletsky, 2000; Code available at [http://www-genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/genome_software/other/primer3.html)).

Primaclade accepts as input a multiple alignment file saved in Clustal (Thompson *et al.*, 1997), NEXUS (Maddison *et al.*, 1997),

EMBOSS (Rice *et al.*, 2000), PHYLIP (Felsenstein, 2004) or numerous other alignment formats. Users can specify the maximum number of degenerate base pairs per primer (up to five), the number of gapped sequence lines in the alignment file to ignore and a single region of the alignment to exclude. The last feature is most useful in excluding areas that are so conserved that they would be shared by many paralogous genes. Melting temperatures and percent GC content can also be input for each run, or default values can be used.

To determine primers for a set of sequences, the alignment file is read and a consensus computed using the 'consensus\_iupac' method from BioPerl AlignIO.pm (pm = Perl Module). The alignment is then split into individual sequences. To find as many unique primers as possible, the script runs Primer3 11 times for each sequence of the alignment starting with a search for an 18-mer primer, and incrementing, each time by 1 bp, up to a 28-mer. The output file from each run of Primer3 is then parsed, and both upstream and downstream primers are saved into a unique array for each line of sequence data in the alignment. After any gaps are accounted for, the primer starting location and length are calculated, and the primer sequence is compared to the corresponding nucleotides in the alignment consensus sequence. If the consensus sequence contains the correct number or fewer degenerate nucleotides then the primer is saved; otherwise it is discarded. Primers that pass the test for degeneracy are screened to determine the number of gap sequences that occur at their positions within the alignment, and primers that meet the input criteria are saved into a final array. The array is sorted, any duplicates are removed, and a final result HTML document is generated.

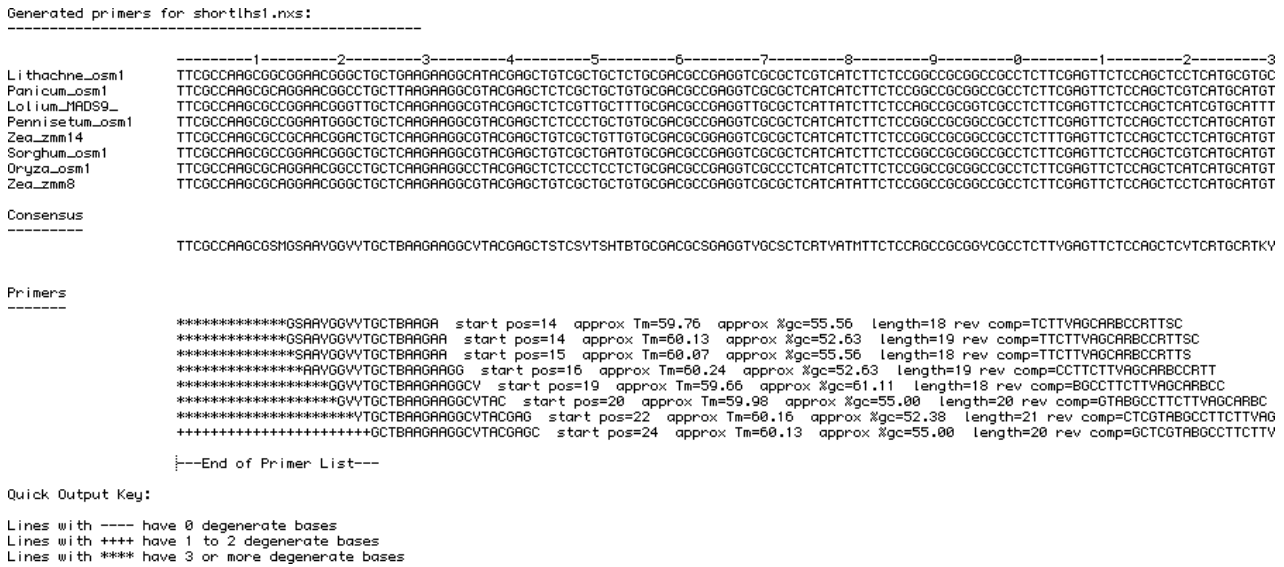
A typical output page contains the original alignment file followed by a single line showing the consensus sequence (black and white version, Fig. 1), with highly conserved nucleotides in colored text and less-conserved bases in black. At the bottom of the page, the list of primers is printed under their correct position within the alignment display. The primer list is color-coded, with green for primers with no degenerate base pairs, orange for primers with one or two degenerate bases, and red for primers with three or more degenerates. The reverse complement for the 3' primers is provided, as are T<sub>m</sub> and %GC. The output page can also be saved in plain text format.

### TESTS OF THE PROGRAM

We have used Primaclade successfully to design primers in alignments of 200–2128 bp comprising 2–17 sequences. Primaclade generally performed best with alignments of up to about eight

\*To whom correspondence should be addressed.

†These authors contributed equally to this work.



**Fig. 1.** Primaclade output screen (converted to black and white) showing primers identified in the 5' region of the *LEAFY HULL STERILE 1 (LHS1)* gene from a selection of grass species. Upper block consists of aligned sequences, middle block the consensus. The bottom four lines list primer sequences, followed by the number of the start position, Tm, %GC content and length.

sequences and up to 29.0% sequence divergence (see Primaclade webpage). Including more sequences causes the program to run more slowly, but the precise effect depends on the quality of the alignment. Input of a good alignment is vital, as the software is not effective in finding primers in ambiguously aligned regions, or in alignments with poor consensus. For very divergent alignments we partition the alignment into several smaller files and run Primaclade independently on each file. In general, an iterative approach works well, starting with input of an entire alignment and using the default settings. If suitable primers are not found, we then increase the allowable number of degenerate sites, range of melting temperatures, range of GC percentages and the number of alignment gaps to skip. If this still is unsatisfactory, we sequentially remove the most divergent sequences and/or divide the file into two more homogeneous subfiles.

In summary, Primaclade provides a quick, easy, powerful and freely available solution for researchers who want to design PCR primers across multiple species. It can greatly simplify the design of PCR primers for any comparative molecular study.

**ACKNOWLEDGEMENTS**

We thank Rosa Ortiz-Gentry and Jill Preston for test data sets, and Patrick Sweeney for help with the web interface, and two

anonymous reviewers for comments on the manuscript. This project was supported by NSF grants MCB-01 10809 and DBI-01 10189 to EAK.

**REFERENCES**

Fielding,R.T. and Kaiser,G. (1997) The Apache HTTP Server Project. *IEEE Internet Comput.*, **1**, 88–90.

Felsenstein,J. (2004) *PHYLIP (Phylogeny Inference Package) version 3.6*. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Maddison,D.R., Swofford,D.L. and Maddison,W.P. (1997) NEXUS: an extensible file format for systematic information. *Syst. Biol.*, **46**, 590–621.

Rozen,S. and Skaletsky,H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In Krawetz,S. and Misener,S. (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 365–386.

Rice,P. Longden,I. and Bleasby,A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.

Stajich,J.E., Block,D., Boulez,K., Brenner,S.E., Chervitz,S.A., Dagdigian,C., Fuellen,G., Gilbert,J.G.R., Korf,I., Lapp,H. et al. (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.*, **12**, 1611–1618.

Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **24**, 4876–4882.