

DPPrimer – A Degenerate PCR Primer Design Tool

Shachi Gahoi¹, Arya L¹, Rai Anil² & Marla SS^{1*}

¹National Bureau of Plant Genetic Resources, Pusa campus, New Delhi-12; ²Indian Agricultural Statistics Research Institute, New Delhi-12; Marla SS - Email: soma.marla@nbpgr.ernet.in; *Corresponding author

Received March 03, 2013; Revised October 02; Accepted October 19, 2013; Published November 11, 2013

Abstract:

Designed degenerate primers unlike conventional primers are superior in matching and amplification of large number of genes, from related gene families. DPPrimer tool was designed to predict primers for PCR amplification of homologous gene from related or diverse plant species. The key features of this tool include platform independence and user friendliness in primer design. Embedded features such as search for functional domains, similarity score selection and phylogenetic tree further enhance the user friendliness of DPPrimer tool. Performance of DPPrimer tool was evaluated by successful PCR amplification of ADP-glucose phosphorylase genes from wheat, barley and rice.

Availability: DPPrimer is freely accessible at http://202.141.12.147/DGEN_tool/index.html

Keywords: Degeneracy, PCR, Primer design.

Background:

Polymerase Chain Reaction (PCR) is a fundamental technique widely employed in current molecular biology for amplification of a given DNA sequence to exponential copies. In course of gene flow and subsequent species formation, evolution dictates that coded amino acid sequences unlike nucleic acids are conserved. While a particular amino acid of a protein can be conserved between two species, the corresponding codons may differ and have all the alternatives (to choose from four nucleotide bases) and is called degeneracy [1]. Thus by having nucleotides in a primer set at alternative positions it becomes possible to identify all those closely related homologous genes and amplify them in a PCR reaction. By designing primers using degenerate positions all the possible variants of the target sequence can be amplified.

Implementation:

Degeneracy is a critical factor in defining sensitivity in PCR since a highly degenerate primer will have a few matches that precisely match the template. In the initial cycles of PCR more homologous primers are likely to be included in the amplicon. The

efficiency with which amplification proceeds in next cycles depends on similarity of the remaining primers in the selected pool. Also primers with least number of degenerate positions always have high success rate. Given a set of DNA sequences, our goal is to design a pair of degenerate primers, so that the primers match and amplify as many of the sequences as possible [2]. To design right primers that match and amplify large number of genes, we need to use highly degenerate primers. At the same time, to cut short the chances of amplifying non-related sequences, the degeneracy should be defined and limited. The goal of primer design strategy is to achieve a balance between the coverage of the variant sequences with the negative effects of degeneracy. This tradeoff can be resolved by defining and optimization of various primer properties and thermodynamic parameters essential for a successful PCR. Hence we consider degenerate primer designing a problem involving optimization of various parameters i.e. degeneracy at most P and length l of the designed primer K , that define the suitability of primers and amplification success of desired DNA fragment.

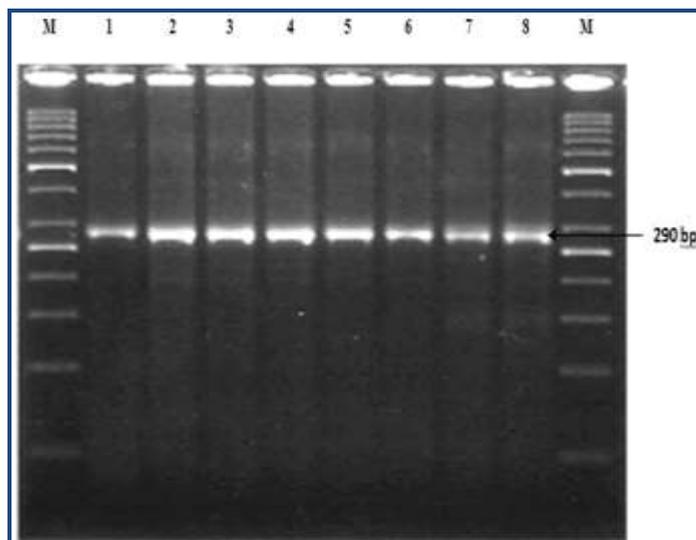


Figure 1: PCR amplification in Wheat and Barely using degenerated primer for ADP-glucose phosphorylase [1:C 306 (*Triticum aestivum*), 2:HD 2307 (*Triticum aestivum*), 3: A-9-30-1 (*Triticum durum*), 4: WH912 (*Triticum durum*), 5: DDK 1001 (*Triticum dicoccum*), 6: DDK 1009 (*Triticum dicoccum*), 7: Dolma (*Hordeum vulgare*, Hullles), 8: BH331 (*Hordeum vulgare*, Hulled), M-50bp molecular weight marker (MBI Ferments)].

DPPrimer designing includes three major steps:

Collection of related input sequence set from sequence alignment, defining, scoring, filtering and selection of a primer set with desired functions.

Collection of related input sequence set

Firstly a candidate protein sequence was submitted to InterProScan [3] to check presence of desired functional domains. The sequence was subsequently searched against blastp [2] and the resulting output containing homologous sequences obtained using 60 percent similarity threshold. Choice of user defined similarity scores in blastp search helps in limiting the number of related sequences especially from distant homologues. A phylogenetic tree was constructed to visualize the evolutionary relationships existing among these homologous sequences. Collected homologous protein sequences were also subjected to multiple sequence alignment [4] at 90 percent sequence similarity level and consensus sequence was generated. Features such as functional domain presence, choice of similarity score selection and illustration of a phylogenetic tree further enhance the user friendliness of DPPrimer tool.

Scoring & filtering

The obtained consensus sequence was reverse translated and evaluated for primer properties such as degeneracy, T_m , GC content and potential amplicon size. Obtained primer sets with above properties (at a specified parameter threshold level) were searched and filtered. Computationally the obtained sets represent the desired primer quality and potentially assure coverage of sequences by the template primer. The next step involves determination of a minimal set of primers required to amplify all potential sequences in the aligned set (after filtration) thus avoiding all other unrelated homologues. Greedy set approximation algorithm described by Slavik [5] is employed to obtain a minimal set of primers and to select a primer pair that satisfies the chosen ideal parameters to cover

all the sequences. For alignment of $M_i, \dots, i+j$ with its matching primers $B^* = [B_1, B_2, \dots, B_n]$ and the set S B^* that minimizes $\sum K_i$ (where K_i is a cost function) and matches all sub sequences in $M_i, \dots, i+j$.

Primer count is taken as the primary cost function for primer number minimization. The cost function reflects approximation quality and a best primer is optimized, parameter threshold values are used to correct deviations. This permits to select one of the two primers that can potentially cover all sequences. This gives an answer yielding minimum set of primers for amplification of all given sequences in the alignment set of $M_i, \dots, i+j$.

Selection of minimum number of flanking primers: Both forward and reverse primers are selected by optimization of T_m , product length and cross reactivity. For a successful PCR it is required that the distance between the 5' - and 3' - primer match site is large (so that biologically valid region is amplified). Primer length can be extended as long as its degeneracy falls in the user pre-defined limits.

Performance of DPPrimer was tested using the degenerate primer pair with PCR amplification ADP-glucose phosphorylase genes [6] from wheat and barley. Primer pair used for PCR amplification was designed using ADP glucose phosphorylase amino acid sequence from oryza sativa (NCBI accession: ABL74524.1). Designed primer set: Forward primer - GGNTAYTGGGARGAYATH (degeneracy of 96), Reverse primer - ATGGNGCNGAYAYTAY (degeneracy of 128). The size of amplified DNA fragments was found to be identical in both wheat and barley (Figure 1) demonstrating robustness of DPPrimer in degenerate primer design and successful amplification of members of orthologous genes from various species.

Software requirements:

DPPrimer works best with Mozilla Firefox and Chrome browsers. The standalone version requires perl and Bioperl installed in the unix environment.

Input options:

DPPrimer accepts a protein NCBI accession number or amino acid sequence either in raw or fasta format.

Output options:

Output comprises chosen parameter information, pair of primers (forward and reverse), InperProScan output depicting presence of desired functional domain, homologous sequences to input sequence and a Phylogenetic tree.

Conclusion:

DPPrimer is more selective in defining degeneracy of input sequences from both distant and related gene families as compared to other primer designing tools. A popular tool Hyden [1] and CODEHOP [7] employed widely for designing degenerate primers for example does not offer the user choice to know the presence of desired functional domain in input query sequence. Besides all other degenerate primer design tools unlike DPPrimer do not let the user to set the required level of similarity score while analyzing homologous sequences using blastp and display of phylogenetic tree.

Acknowledgement:

We wish to thank National Agricultural Bioinformatics Grid of Indian Council for Agricultural Research for gracious support for conducting the experiments and payment of salaries to SG. Authors are thankful to Ajit Mishra for making DPPrimer available on World Wide Web.

References:

- [1] Linhart C *et al. Methods Mol Biol.* 2007 **402**: 221 [PMID: 17951798]
[2] Altschul SF *et al. J Mol Biol.* 1990 **215**: 403 [PMID: 2231712]

- [3] <http://www.ebi.ac.uk/Tools/pfa/iprscan/>
[4] Notredame C *et al. J Mol Biol.* 2002 **302**: 205 [PMID: 10964570]
[5] <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.4068>
[6] Rober BM *et al. The EMBO Journal.* 1992 **11**: 1229 [PMID: 1373373]
[7] Staheli JP *et al. Methods Mol Biol.* 2011 **687**: 57 [PMID: 20967601]

Edited by P Kanguane

Citation: Gahoi *et al.* Bioinformation 9(18): 937-940 (2013)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Workflow of Dp primer Tool:

