BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 16 2010, pages 2053–2054
doi:10.1093/bioinformatics/btq312

Sequence analysis

Advance Access publication June 14, 2010

A new method for designing degenerate primers and its use in the identification of sequences in *Brachiaria* showing similarity to apomixis-associated genes

Eduardo Gorrón[1,2], Fausto Rodríguez[1], Diana Bernal[1], Luis Miguel Rodriguez-Rojas[2], Adriana Bernal[2], Silvia Restrepo[2,*] and Joe Tohme[1,*]

[1]Centro Internacional de Agricultura Tropical (CIAT), A.A. 6713, Cali and [2]Laboratorio de Micología y Fitopatología, Universidad de los Andes, Carrera 1 No 18A-10, Bogotá, Colombia

Associate Editor: John Quackenbush

ABSTRACT

Motivation: We developed a technique and a tool for degenerate primer design based on multiple local alignments employing the MEME algorithm supported with electronic PCR. The objective is to find adequate primers starting from sequences with poor global similarity. We show an example of its application in our laboratory to find sequences in *Brachiaria* with similarity to ESTs related to apomixis.

Contact: srestrep@uniandes.edu.co; j.tohme@cgiar.org

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on March 14, 2010; revised on May 17, 2010; accepted on June 8, 2010

1 INTRODUCTION

Classical methods for degenerate primer design include software applications such as CODEHOP (Rose *et al.*, 1998) or PrimaClade (Gadberry *et al.*, 2005). These methods usually rely on the identification of clear blocks of conserved regions in multiple global alignments. Therefore, alignment quality should be very high. However, in many cases, primer design using multiple global alignments is unsuccessful (Kwok *et al.*, 1994) due to poor conservation among the sequences under study. To overcome this problem, we propose a new method to design primers, which could be used for sequences with poor global similarity or without well-conserved blocks in multiple global alignments. Instead, the technique is based on multiple local alignments, using the Multiple Expectation – Maximization for Motif Elicitation (MEME) algorithm (Bailey *et al.*, 2006), in order to search for conserved regions long enough to serve as primers. The results of MEME are then verified by electronic PCR (e-PCR) (Schuler, 1997) and standard primer design programs.

We employed this technique to design degenerate primers for sequences related to apomixis in the tropical forage grass *Brachiaria*. We demonstrate that the amplification of sequences with similarity to genes reported in other species (possibly homologues) is possible, even when the species in question is poorly characterized at the molecular level.

*To whom correspondence should be addressed.

2 IMPLEMENTATION

Primer design: A database of expressed sequences of *Brachiaria*-related species was constructed using EST sequences from plants in the families Alliaceae and Poaceae. Sequences were retrieved from the TIGR Gene Indices database (http://compbio.dfci.harvard.edu/tgi/) and TIGR Plant Transcript Assemblies database (http://plantta.tigr.org, both consulted on September 2007). Input sequences can be full-length cDNAs if available or genomic sequences.

Twenty-seven sequences related to apomixis (Albertini *et al.*, 2004; Chen *et al.*, 2005; Guerin *et al.*, 2000; Gustine *et al.*, 1993; Pessino *et al.*, 2001; Rodrigues *et al.*, 2003; Vielle-Calzada *et al.*, 1996) (accession numbers AB000809, AF242539, AF475105, AJ786393, AJ810708, AJ810709, AJ810710, AJ841698, AY375366, D37938, D37939, D37940, EF517497, EF517498, EF530198, EF530199, M32653, M38357, U40219, U65082, U65383, U65384, U65385, U65386, U65387, U65388, U65389) (Supplementary Table 1) were employed as BLAST queries against the constructed database, in order to obtain groups of sequences similar to the query. The maximum accepted E-value in this test was 10^{-6} and sequences showing values over this threshold were discarded. We could obtain more than two similar sequences for only 24 accession numbers, and the remaining three (u65386, u65387 and M32653) were not further analyzed.

Each original query and its group of related sequences were aligned using T-COFFEE (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi?stage1=1&daction=TCOFFEE::Regular&referer0=embnet) to determine global similarity. A lack of block conservation was observed in all cases in these multiple global alignments. Alignments were then performed with MEME (http://meme.sdsc.edu/meme4_1/cgi-bin/meme.cgi) (Bailey *et al.*, 2006). The conditions of these last alignments included a minimum motif length of 18 nt and a maximum of 25, and minsites = 2/3 of the total number of sequences. Sequences with low similarity in the MEME alignment and redundant sequences were discarded, and a second MEME alignment was performed with the remaining ones. In this case, more than half of the results exhibited a common arrangement of motifs between the analyzed sequences, indicating a global similarity that could not be well resolved with common algorithms of multiple global alignments (Fig. 1 and Supplementary Figs 1 and 2). This proved the existence of well-conserved regions

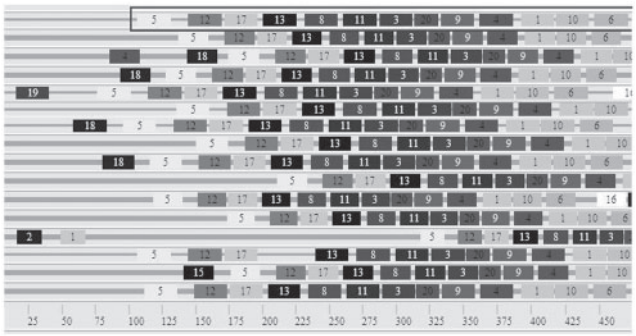

Fig. 1. MEME results showing general arrangement of motifs for the original query sequence aj810708 (shown in the rectangle at top); the other sequences, which were taken from its corresponding BLAST results, are represented below. Every numbered box represents a specific motif.

and similarity in the analyzed sequences, which could be useful for primer design.

The first 20 motifs obtained in each MEME run were extracted, ordered according to the score given in the program, and paired in all possible ways, in order to test each motif pair as a primer pair. Every pair was tested in the group of its corresponding sequences (the sequences used in the second MEME alignment) in the e-PCR program (Schuler, 1997) under the condition that amplicon size must be at least 200 bp in length. In many cases, the first eleven motifs had an acceptable level of degeneracy, and they were preferred to make subsequent analyses.

The presence of the original query and the number of degenerations were verified in MEME alignments of both motifs of the pair, beginning with the pair with the highest sum of motif scores (referred to as the highest score). If the query was absent in the alignments and/or the number of degenerations was greater than 12 in the primer pair, it was discarded and the test continued with the next one. When one motif pair was acceptable, it was tested in NetPrimer (http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html), in order to verify annealing temperatures, secondary structure formation and primer dimers. If the pair had a T_m that was too high, a T_m difference $>5°C$, or formed dimers or secondary structures with a ΔG less than -8 KJ/mol, bases of one or both motifs were eliminated, until the conditions were acceptable. If these adjustments could not improve them, the pair was discarded and the test continued with the next pair. Finally, once one pair was accepted, the degenerations were put in their positions and it was synthesized as a degenerate primer pair. If neither motif pair had the required conditions, the analysis for this particular sequence stopped. Taking into account these exclusion criteria, we were able to design primers for only 22 out of the 24 original queries.

We developed a software application for the primer design that can be accessed at http://bioinf-mac.uniandes.edu.co/design.

Primer assays: PCR reactions with obtained primers were standardized and assayed on cDNA from pistils (reproductive tissue) and a bulk of leafs, stems and roots (vegetative tissue) (Supplementary Fig. 3). Amplicons could be obtained in 15 out of the 22 designed primer pairs. With the exception of sequences aj786393, ef530198 and pcu402019, no PCR gave a multiple band product. Sequence analysis of the products showed a good correspondence with the original query in 12 out of the 15 cases. The three false positive results and the cases where an amplicon could not be obtained were associated to low similarity patterns in MEME analysis; i.e. cases where no clear common arrangement of motifs could be seen. This indicates that primer design should be avoided in those cases where MEME does not show a good level of conservation in terms of motif organization. These results are considered very positive, because many times our input sequences consist on few similar sequences.

3 CONCLUSION

In conclusion, we were able to develop a new bioinformatics method and a corresponding tool to amplify genes associated with a very poorly characterized phenomenon at the molecular level such as apomixis. This method could help to obtain amplifications in other poorly understood biological events, in which there are few related sequences. In general terms, the technique can be used for sequences that have very few known homologues, or to confirm them, and to design degenerate primers when the classical methods do not work. The method, however, needs to be improved, for instance trying to automatize the time-consuming steps and avoiding sequences with low similarity patterns. Our results will allow the analysis of all those proposed candidate genes for apomixis in a unique plant species.

Funding: This research was partially supported by the Faculty of Sciences at Universidad de los Andes.

Conflict of Interest: none declared.

REFERENCES

Albertini,E. *et al*. (2004) Isolation of candidate genes for apomixis in *Poa pratensis* L. *Plant Mol. Biol*., **56**, 879–894.

Bailey,T.L. *et al.* (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.,* **34(Web Server issue)**, W369–W373.

Chen,L. *et al*. (2005) Developmental expression of ASG-1 during gametogenesis in apomictic guinea grass (*Panicum maximum*) *J. Plant Physiol.*, **162**, 1141–1148.

Gadberry,M.D. *et al*. (2005) Primaclade—a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics*, **21**, 1263–1264.

Guerin,J. *et al*. (2000) A DEFICIENS homologue is down-regulated during apomictic initiation in ovules of *Hieracium*. *Planta*, **210**, 914–920.

Gustine,D.L. *et al*. (1993) A strategy for cloning apomixis associated cDNA markers from buffelgrass. In *Proceedings of the XVII International Grassland Congress*, New Zealand Grassland Association, Palmerston North, New Zealand, pp. 1033–1034.

Kwok,S. *et al*. (1994) A guide to the design and use of mismatched and degenerate primers. *PCR Methods Appl*., **3**, S39–S47.

Pessino,S.C. *et al*. (2001) Isolation of cDNA clones differentially expressed in flowers of apomictic and sexual *Paspalum notatum*. *Hereditas*, **134**, 35–42.

Rodrigues,J.C. *et al*. (2003) Identification of differentially expressed cDNA sequences in ovaries of sexual and apomictic plants of *Brachiaria brizantha. Plant Mol. Biol*., **53**, 745–757.

Rose,T.M. *et al*. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.*, **26**, 1628–1635.

Schuler,G.D. (1997) Sequence mapping by electronic PCR. *Genome Res*., **7**, 541–550.

Vielle-Calzada,J.P. *et al*. (1996) Comparative gene expression in sexual and apomictic ovaries of *Pennisetum ciliare* (L.) Link. *Plant Mol. Biol*., **32**, 1085–1092.