Bioinformatics, 32(15), 2016, 2382–2383
doi: 10.1093/bioinformatics/btw134
Advance Access Publication Date: 9 March 2016
Applications Note

OXFORD

Sequence analysis

GSP: a web-based platform for designing genome-specific primers in polyploids

Yi Wang[1,2], Vijay K. Tiwari[3], Nidhi Rawat[3], Bikram S. Gill[3], Naxin Huo[1], Frank M. You[4], Devin Coleman-Derr[2,*] and Yong Q. Gu[1,*]

[1]USDA-ARS, Western Regional Research Center, Crop Improvement and Genetics Research Unit, Albany, CA 94710, USA, [2]USDA-ARS, Plant Gene Expression Center, Albany, CA 94710, USA, [3]Wheat Genetic Resource Center, Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA and [4]Cereal Research Centre, Agriculture and Agri-Food Canada, Morden, MB R6M 1Y5, Canada

*To whom correspondence should be addressed.
Associate Editor: Janet Kelso

Received on September 25, 2015; revised on February 17, 2016; accepted on March 4, 2016

Abstract

Motivation: The sequences among subgenomes in a polyploid species have high similarity, making it difficult to design genome-specific primers for sequence analysis.
Results: We present GSP, a web-based platform to design genome-specific primers that distinguish subgenome sequences in a polyploid genome. GSP uses BLAST to extract homeologous sequences of the subgenomes in existing databases, performs a multiple sequence alignment, and design primers based on sequence variants in the alignment. An interactive primers diagram, a sequence alignment viewer and a virtual electrophoresis are displayed as parts of the primer design result. GSP also designs specific primers from multiple sequences uploaded by users.
Availability and implementation: GSP is a user-friendly and efficient web platform freely accessible at http://probes.pw.usda.gov/GSP. Source code and command-line application are available at https://github.com/bioinfogenome/GSP.
Contacts: yong.gu@ars.usda.gov or devin.coleman-derr@ars.usda.gov
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Developing genome-specific primers for polyploids not only will facilitate the study of sequence diversity and the association mapping of genes, but also the development of gene-based functional markers (Huang and Brule-Babel, 2010). However, the polyploid genome contains two or more homeologous subgenomes with high sequence similarities, especially in the coding regions, which presents a particular problem with respect to the design genome-specific primers that amplify individual genes. To our knowledge, only one pipeline named PolyMarker can generate genome-specific primers (Ramirez-Gonzalez *et al.*, 2015). However, PlolyMarker was specifically developed for SNP assay with the KASP™ (Kompetitive Allele Specific PCR) system, a very unique system using a three primers system in PCR reactions. PolyMarker requires users to prepare sequence files with SNP information and the target chromosome input. Therefore, the software has limited utility in a variety of other PCR-based applications. Ma *et al.* described a method for detecting locus-specific polymorphism in hexaploid wheat (Ma *et al.*, 2015). In this method, users are required to do step-wise processes using provided multiple perl scripts for blast search, extracting homeologous sequences, local alignment among the retrieved sequences to identify polymorphism sites. However, users still need to manually design primers based on the alignment. Currently, no good web-based tool is available that allows the user to design genome-specific primers by only inputting a query sequence. Here, we present a web-based, user-friendly primer design tool named GSP, which employs a BLAST search and multiple sequence alignment (MSA) to identify

sequence variant sites among the homeologous sequences retrieved and then designs specific primers.

2 Methods

Our genome-specific primer design server consists of steps (Supplemental Fig. S1A) including:

2.1 Blast based searches against a polyploid genome

The increasing availability of genome sequence data from polyploid species now makes it possible to conduct homeologous sequence comparative analyses between subgenomes in polyploids. Bread wheat (*Triticum aestivum* L.) is a well-known, hexaploid, polyploid species, with genome size estimated at ~17 Gbp and composed of three closely related and independently maintained genomes (International Wheat Genome Sequencing, 2014). In the current version of GSP, users can select genome sequence databases from three polyploid species (bread wheat, upland cotton and switchgrass) to design genome-specific primers. Because genome sequences for polyploid species are expected to be increased and improved, we will update the GSP dataset semiannually. In GSP, homeologous sequences in the subgenomes are extracted by BLAST using query sequences. The coordinates of homeologous sequence in the genome was retrieved by BLAST hits using the method described by Ma *et al.* (2015).

2.2 Aligning and identifying variant sites between homeologous sequences

GSP uses the MUSCLE (Edgar, 2004) program to align homeologous sequences extracted from polyploid genomes. Sequence variations, such as nucleotide polymophisms, insertions and deletions, are identified in the MSA. The position of sequence variations among sequences are defined as variant sites. All variant sites in the alignment can be accounted for in designing specific primers. Users can set the number of variant sites in the primer or only extract primers that contain the variant site in the 3' end.

2.3 Analyzing primers

The primers in each sequence are generated by the Primer3 core program (Rozen and Skaletsky, 2000). GSP uses primers and variant sites information to design specific primers and sort the primers based on two criteria: the number of variant sites and their location in the primer sequence. In general, the higher number of variant sites and their location in the 3′ region will give a high weight in ordering primers. Users can view the ordered primers, indexed in numerical order on the primer results page.

2.4 Designing specific primers based on multiple sequence alignment

GSP not only allows users to BLAST the existing polyploid genome sequences to design genome-specific primers, but also allows the input of multiple sequences to generate MSA and design specific primers that can distinguish individual sequences. This function provides a user-friendly way to design specific primers from various taxa, gene families and other closely related sequences.

3 Results

The GSP web platform provides a user-friendly graphical view that displays primer design output. The main results of the GSP analysis are: MSA of the homeologous sequences (Supplemental Fig. S1B-1), a distribution of primers and products (Supplemental Fig. S1B-2), a detail of the primers information (Supplemental Fig. S1B-3) and a primer-pair pool panel that contains the selected primer pairs (Supplemental Fig. S1B-4). GSP provides an interactive view for selecting and analyzing primers. Users can easily find their primers of interest, obtain the sequence products and generate a virtual electrophoresis gel. GSP also allows users to manually select a specific region of a sequence in the alignment to verify its use as a good primer. In addition, GSP can check the specificity of primers input by users.

We designed 53 primer pairs for group 2 homoeologous chromosomes for hexaploid wheat and tested their genome specificity using PCR assays. A total of 85% primers showed genome specificity and identified A, B and D-genome homoeologs (Supplemental Table S1 and Supplemental Fig. S2).

4 Conclusion

With the rapid increase of polyploid genome sequence data from high-throughput sequencing technologies, effective primer design tools for homeologous sequences from the subgenomes of polyploids has become an important component of gene and genome analysis. GSP is a time-saving, user-friendly and easy-to-use web tool for designing genome-specific primers to distinguish sequences among the subgenomes of polyploid species. This web application tool can be easily extended to other polyploids when their genome sequences become available.

Funding

This work has been supported in part by U.S. National Science Foundation (Grant number IOS 0822100) and by United State Department of Agriculture Research Service CRIS project 5325-21000-021. VKT was supported by NSF-IUCRC grant contract (IIP1338897).

Conflict of Interest: none declared.

References

Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.

Huang,X.Q. and Brule-Babel,A. (2010) Development of genome-specific primers for homoeologous genes in allopolyploid species: the waxy and starch synthase II genes in allohexaploid wheat (*Triticum aestivum* L.) as examples. *BMC Res. Notes*, **3**, 140.

International Wheat Genome Sequencing, C. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*, **345**, 1251788.

Ma,J. *et al.* (2015) A high-throughput pipeline for detecting locus-specific polymorphism in hexaploid wheat (*Triticum aestivum* L.). *Plant Methods*, **11**, 39.

Ramirez-Gonzalez,R.H. *et al.* (2015) PolyMarker: a fast polyploid primer design pipeline. *Bioinformatics*, **31**, 2038–2039.

Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.