

NRC Publications Archive Archives des publications du CNRC

Interactive visualization of orthologous genes and their annotations in plants

Tulpan, Dan; Léger, Serge

For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

<https://doi.org/10.4224/21276406>

NRC Publications Archive Record / Notice des Archives des publications du CNRC :

<https://nrc-publications.canada.ca/eng/view/object/?id=8ba53e18-7d26-47cf-9478-548b71a0d35d>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=8ba53e18-7d26-47cf-9478-548b71a0d35d>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

Interactive visualization of orthologous genes and their annotations in plants

Introduction

Fully sequenced genomes of model organisms and staple crops such as *Aegilops tauschii*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Brassica rapa*, *Hordeum vulgare*, *Oryza sativa*, *Sorghum bicolor*, *Triticum aestivum*, *Triticum urartu* and *Zea mays* provide a unique opportunity to visually compare their gene order conservation via orthology. The conservation in gene order provides information about gene function and interactions of proteins that are encoded by these genes. While gene order is conserved and well preserved at phylogenetic distances for closely related plant species, it is low or lost during evolution in distantly related species.

A good understanding and visual perspective of gene content and gene order similarities and differences across plant species allows researchers and experimentalists to design breeding strategies that lead to resilient next generation crops (e.g. current GM crops), to nutritionally enhanced foods (e.g. golden rice) or to pharmaceuticals synthesization (e.g. somatotropin in tobacco plants).

The Plant Orthology Browser (POB)

Here we propose The Plant Orthology Browser (POB) - a web-based orthology and annotation visualization tool which supports 10 plant species and provides highly intuitive and interactive pairwise comparison and visualization of genomic traits via gene orthology. POB currently includes three interconnected modules: (i) a gene order visualization module implementing a highly interactive environment for exploration of gene order between any pair of chromosomes in two different plant species, (ii) a synteny visualization module providing interactive dot-plot representations of orthology relations between a pair of chromosomes in two distinct plant species, and (iii) a search module, which realizes the interconnectivity among all modules via a Google-like customized free-text search capability with online "as-you-type" suggestions and highlighting that allows users to explore various facets of the underlining information without being constrained by interface-dependent search fields.

Orthology prediction

We use a Reciprocal Best BLAST Hit (RBBH) approach for 1-to-1 orthology prediction. Two BLAST runs are executed for each pair of plant species and for each sequence type (DNA, protein) to identify reciprocal best hits (RBBHs). A 1-to-1 orthology relationship is assigned for those pairs of genes that are bidirectional hits within a confidence interval ($e\text{-value} \leq 10^{-5}$).

Orthologous DNA / Protein Prediction	A. tauschii	A. thaliana	B. distachyon	B. rapa	H. vulgare	O. sativa	S. bicolor	T. aestivum	T. urartu	Z. mays
A. tauschii		2147	12984	2385	12442	10527	11384	13002	15024	9493
A. thaliana			3054	15258	2740	3190	3544	3395	2733	2855
B. distachyon				3058	13141	14219	15209	5087	12653	12611
B. rapa					2348	2704	2990	1228	2301	2485
H. vulgare						10911	11722	5192	12071	9905
O. sativa							14340	7282	10512	11851
S. bicolor								7549	11058	10059
T. aestivum									11645	6402
T. urartu										9250
Z. mays										

<http://nrcmonsrv01.nrc.ca/pob/>

Plant Orthology Browser (POB) 3.0
 an Orthology and Gene Order Visualizer for Plant Comparative Genomics

Home Search Gene order view Synteny view About Features Contact

Home Page

The Plant Orthology Browser (POB) is currently being developed as part of the **Canadian Wheat Improvement Flagship** which represents the **National Research Council's** contribution to the **Canadian Wheat Alliance (CWA)**, a strategic collaboration with Agriculture and Agri-Food Canada (AAFC), the University of Saskatchewan's Crop Development Centre and the province of Saskatchewan.

Aiming for accelerating variety development of wheat through Genomics Assisted Breeding, POB uses comparative genomics to aid with the identification of high quality orthologous genes in wheat and other species such as *Brachypodium distachyon* and *Sorghum bicolor*. You can interactively search and visualize the **order of orthologous genes** and their **synteny conservation** using our visualization tools.

Currently, POB includes information related to the following plant species:

- A. tauschii
- A. thaliana
- B. distachyon
- B. rapa
- H. vulgare
- O. sativa
- S. bicolor
- T. aestivum
- T. urartu
- Z. mays

Please register or login with your existing account to access POB today.

References

Tulpan, D., Leger, S., Tchagang, A., Pan, Y. (2015) "Enrichment of *Triticum aestivum* gene annotations using ortholog cliques and gene ontologies in other plants", BMC Genomics, 16:299.

Tulpan, D., Leger, S., Lund, A., Pan, Y., Famili, F. (2014) "Prediction of flowering time orthologous genes in *A. tauschii*, *B. distachyon*, *S. bicolor* and *T. urartu*", Plant Genomics Congress Asia, Kuala Lumpur, Malaysia.

Tulpan, D., Leger, S., Cuperlovic-Culf, M., Pan, Y. (2013) "Improved sequence-based orthologs identification using genomic context information and their impact on pathway analysis in plants", Plant and Animal Genome conference (PAG Asia), Singapore.

Copyright (C) 2014 NRC.

Synteny blocks

Blocks of orthologs with conserved gene order are identified using *DAGChainer* [2]. Lists of gene pairs representing orthologs and their physical location on chromosomes are provided as input. *DAGChainer* identifies maximally scoring blocks of ordered gene pairs using a scoring function which accounts for BLAST similarity scores and distances between neighbouring genes.

EnsemblPlants release 25
 EnsemblPlants Biomart release 25
 Gene Ontology Consortium release 1.2
 Gramene release 44
 NRC OrthoPred release 2.0

Figure 1. Synteny between *S. bicolor* and *B. distachyon* chromosomes.

Multiple highly conserved blocks of genes can be visualized interactively [1].

Each dot represents a pair of orthologous genes. The x-y coordinates for each dot correspond to the relative order of each gene on their corresponding chromosomes.

The color for each block is proportional to the number of orthologs included within.

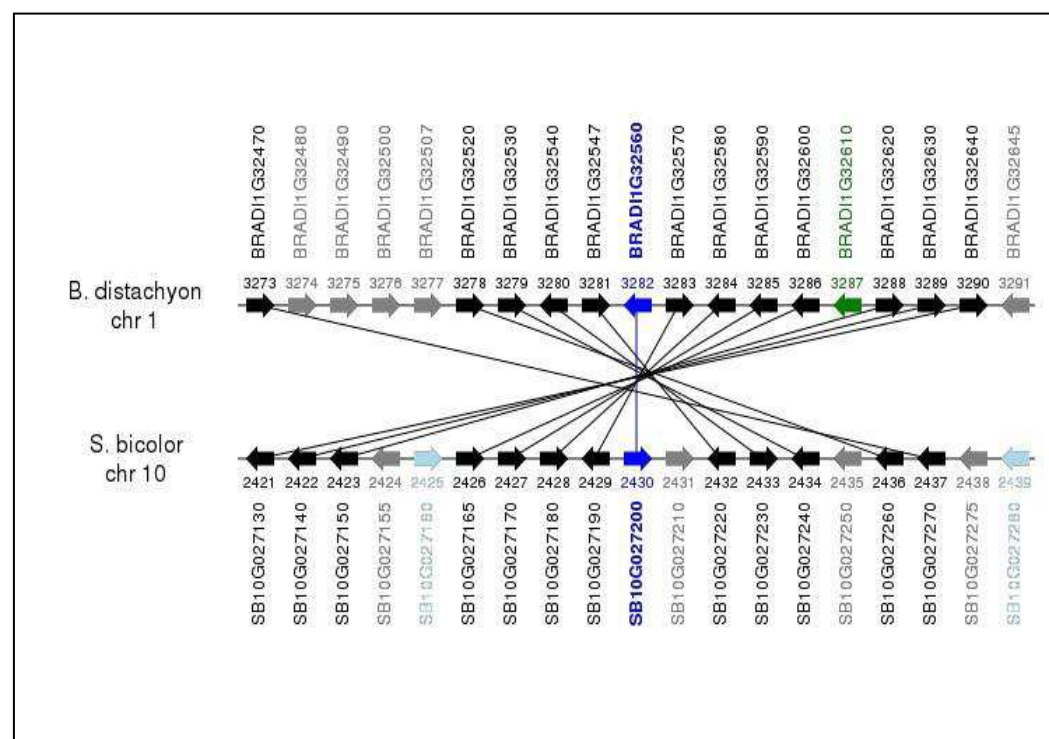
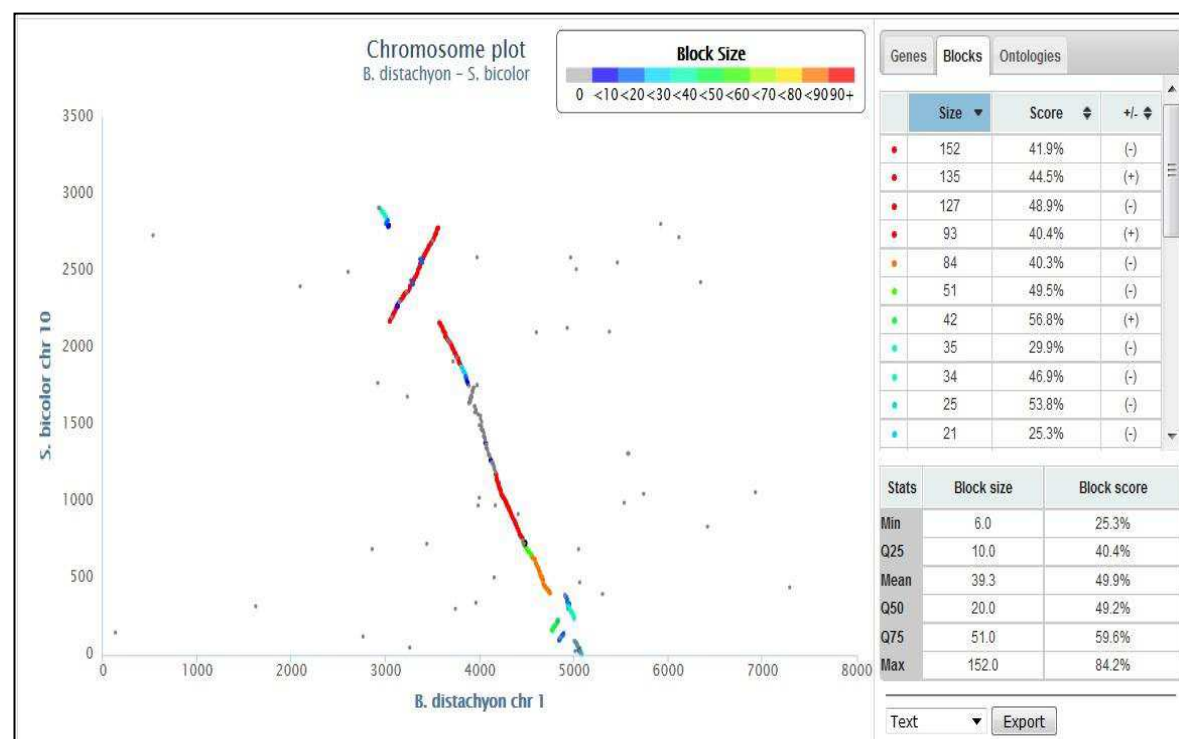


Figure 2. Gene order view of *S. bicolor* and *B. distachyon* genes.

A close-view of a portion from a highly conserved block of genes is displayed. The blue genes represent an orthologous pair of genes of interest.

Each gene is represented with an arrow. The direction of the arrow represents the strand orientation. Right arrows denote positive strand orientations while left arrows depict negative strand orientations.

Results

Aiming for accelerating variety development of wheat through Genomics Assisted Breeding, the Plant Orthology Browser (POB) uses comparative genomics to aid with the identification of high quality orthologous genes in wheat and other plant species. For example, the complete set of 380,329 1-to-1 orthologs that powers POB was used for the GO term association enrichment of 7,838 (8%) wheat genes, of which 2,139 had no previous annotation [1].

References

[1] Tulpan, D., Leger, S., Tchagang, A., Pan, Y. **Enrichment of *Triticum aestivum* gene annotations using ortholog cliques and gene ontologies in other plants.** BMC Genomics 2015, 16:299.

[2] Haas, B.J., Delcher, A.L., Wortman, J.R., Salzberg, S.L. **DAGchainer: a tool for mining segmental genome duplications and synteny.** Bioinformatics 2004; 20(18):3643-3646.

OUR PARTNERS

Canada UNIVERSITY OF SASKATCHEWAN Government of Saskatchewan