

NRC Publications Archive Archives des publications du CNRC

An enrichment model for wheat gene annotations using phylogeny, orthology and existing gene ontologies in 9 plant species Tulpan, Dan; Léger, Serge; Tchagang, Alain B.; Pan, Youlian

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. /
La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
<https://nrc-publications.canada.ca/eng/view/object/?id=dd263013-adbd-4878-ae67-379efbf7e787>
<https://publications-cnrc.canada.ca/fra/voir/objet/?id=dd263013-adbd-4878-ae67-379efbf7e787>

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.

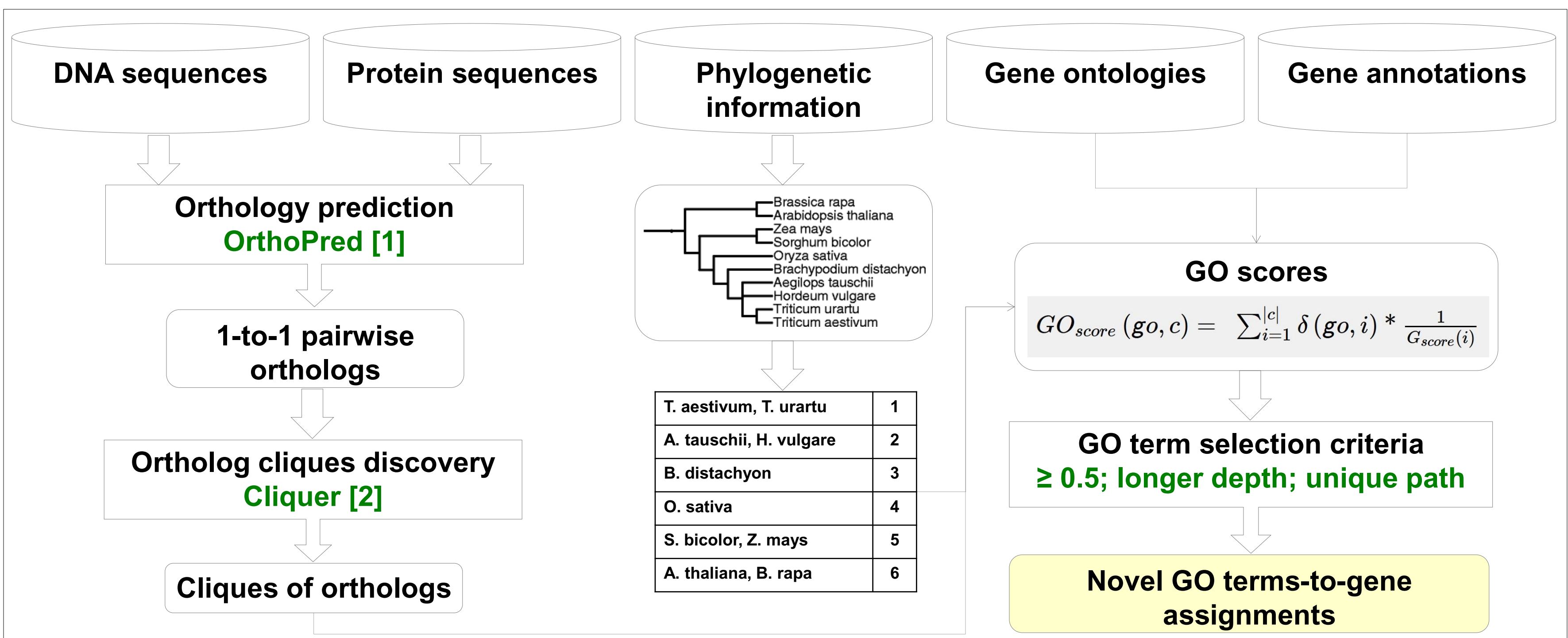
An enrichment model for wheat gene annotations using phylogeny, orthology and existing gene ontologies in 9 plant species

Introduction

Genome sequencing efforts for the *Triticum aestivum* genome produce massive amounts of contigs, preliminary assemblies and putative genes/proteins, nevertheless their annotation is still in its infancy. Given the much larger percentage of annotated genes in other previously sequenced plant genomes such as *Arabidopsis thaliana* and *Oryza sativa* and the known phylogenetic and orthology relationship among these plant species and their corresponding genes, we propose an enrichment model that will further expand the horizon of wheat gene annotations.

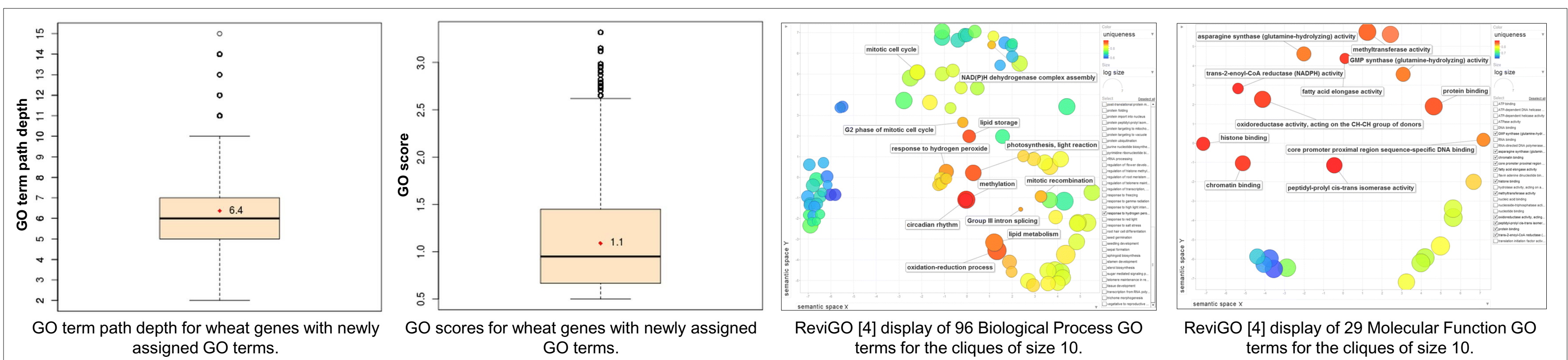
The gene annotation enrichment model

Our sequences and annotations base includes data from Ensembl Plants for wheat and 9 other plant species: *Aegilops tauschii*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Brassica rapa*, *Hordeum vulgare*, *Oryza sativa subsp. japonica*, *Sorghum bicolor*, *Triticum urartu* and *Zea mays*. Orthology relationships between wheat genes and each of the 9 plant species are predicted using an in-house software package. Next, ortholog cliques are identified such that each set of genes within a clique represents pairwise orthologs. Using the phylogenetic distances between wheat and each plant species to quantify the level of confidence for gene ontology assignments within each ortholog clique, new gene annotations are assigned to wheat genes such that either novel or more specific GO terms are associated with those genes.



Results

Overall, based on clique size equal to or larger than 3, our model enriched the existing gene-GO term associations for 7,838 (8%) wheat genes, of which 2,139 had no previous annotation. For the particular case of ortholog cliques of size 10 (13 in total) where all 10 genes within a clique are tightly connected via pairwise orthology, 85 new and more specific GO terms were identified, which represent a 65% increase compared with the previously 130 known GO terms. These observations are further supported for 4 out of the 10 plant species considered in this work by experimental evidence using expressologs [3].



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] Niskanen, S., Östergård, P.R.J. **Cliquer User's Guide, Version 1.0.** *Communications Laboratory, Helsinki University of Technology, Espoo, Finland*, Tech. Rep. T48, 2003.
- [3] Patel, R.V., Nahal, H.K., Breit, R., Chen, Y., Provart, N.J. **BAR Expressolog identification: expression profile ranking of "orthologous" genes in plant species.** *The Plant Journal* 2012, 71:1038–1050.
- [4] Supek, F., Bošnjak, M., Škunca, N., Šmuc, T. **REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms.** *PLoS ONE* 2011, 6(7):e21800.

OUR PARTNERS

