

Research article

Open Access

PlasmoDraft: a database of *Plasmodium falciparum* gene function predictions based on postgenomic data

Laurent Bréhélin*, Jean-François Dufayard and Olivier Gascuel

Address: Projet Méthodes et Algorithmes pour la Bioinformatique, LIRMM, Univ. Montpellier 2, CNRS, 161 rue Ada, 34392 MONTPELLIER, France

Email: Laurent Bréhélin* - brehelin@lirmm.fr; Jean-François Dufayard - Jean-francois.Dufayard@lirmm.fr; Olivier Gascuel - gascuel@lirmm.fr

* Corresponding author

Published: 16 October 2008

Received: 13 June 2008

BMC Bioinformatics 2008, 9:440 doi:10.1186/1471-2105-9-440

Accepted: 16 October 2008

This article is available from: <http://www.biomedcentral.com/1471-2105/9/440>

© 2008 Bréhélin et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Of the 5 484 predicted proteins of *Plasmodium falciparum*, the main causative agent of malaria, about 60% do not have sufficient sequence similarity with proteins in other organisms to warrant provision of functional assignments. Non-homology methods are thus needed to obtain functional clues for these uncharacterized genes.

Results: We present PlasmoDraft <http://atgc.lirmm.fr/PlasmoDraft/>, a database of Gene Ontology (GO) annotation predictions for *P. falciparum* genes based on postgenomic data. Predictions of PlasmoDraft are achieved with a *Guilt By Association* method named Gonna. This involves (1) a predictor that proposes GO annotations for a gene based on the similarity of its profile (measured with transcriptome, proteome or interactome data) with genes already annotated by GeneDB; (2) a procedure that estimates the confidence of the predictions achieved with each data source; (3) a procedure that combines all data sources to provide a global summary and confidence estimate of the predictions. Gonna has been applied to all *P. falciparum* genes using most publicly available transcriptome, proteome and interactome data sources. Gonna provides predictions for numerous genes without any annotations. For example, 2 434 genes without any annotations in the Biological Process ontology are associated with specific GO terms (e.g. Rosetting, Antigenic variation), and among these, 841 have confidence values above 50%. In the Cellular Component and Molecular Function ontologies, 1 905 and 1 540 uncharacterized genes are associated with specific GO terms, respectively (740 and 329 with confidence value above 50%).

Conclusion: All predictions along with their confidence values have been compiled in PlasmoDraft, which thus provides an extensive database of GO annotation predictions that can be achieved with these data sources. The database can be accessed in different ways. A global view allows for a quick inspection of the GO terms that are predicted with high confidence, depending on the various data sources. A gene view and a GO term view allow for the search of potential GO terms attached to a given gene, and genes that potentially belong to a given GO term.

Background

Malaria is one of the most prevalent disease in the world, infecting 400 million people every year, and causing 2.7

million deaths, mainly children under 5 years [1]. *Plasmodium falciparum*, the main causative agent of this parasitic disease, develops drug resistance and no effective vaccine