Accepted Manuscript

Title: Matching Algorithms for Assigning Orthologs after Genome Duplication Events

Author: Guillaume Fertin Falk Hüffner Christian Komusiewicz Manuel Sorge

Please cite this article as: Guillaume Fertin, Falk H*ddotu*ffner, Christian Komusiewicz, Manuel Sorge, Matching Algorithms for Assigning Orthologs after Genome Duplication Events, <![CDATA[Computational Biology and Chemistry]]> (2018), https://doi.org/10.1016/j.compbiolchem.2018.03.015

Matching Algorithms for Assigning Orthologs after Genome Duplication Events

Guillaume Fertin[a,1], Falk Hüffner[2], Christian Komusiewicz[b,3],
Manuel Sorge[c,d,4]

[a]*LS2N UMR CNRS 6004, University of Nantes, Nantes, France.*
[b]*Fachbereich für Mathematik und Informatik, Philipps-Universität Marburg, Marburg, Germany.*
[c]*Ben-Gurion University of the Negev, Beer Sheva, Israel.*
[d]*Technische Universität Berlin, Berlin, Germany.*

Abstract

In this paper, we introduce and analyze two graph-based models for assigning orthologs in the presence of whole-genome duplications, using similarity information between pairs of genes. The common feature of our two models is that genes of the first genome may be assigned two orthologs from the second genome, which has undergone a whole-genome duplication. Additionally, our models incorporate the new notion of *duplication bonus*, a parameter that reflects how assigning two orthologs to a given gene should be rewarded or penalized. Our work is mainly focused on developing exact and reasonably time-consuming algorithms for these two models: we show that the first one is polynomial-time solvable, while the second is NP-hard. For the latter, we thus design two fixed-parameter algorithms, i.e. exact algorithms whose running times are exponential only with respect to a small and well-chosen input parameter. Finally, for both models, we evaluate our algorithms on pairs of plant genomes. Our experiments show that the NP-hard model yields a better cluster quality at the cost of lower coverage, due to the fact that our instances cannot be completely solved by our algorithms. However, our results are altogether encouraging and show that our methods yield biologically significant predictions of orthologs when the duplication bonus value is properly chosen.

Email addresses: `guillaume.fertin@univ-nantes.fr` (Guillaume Fertin),
`komusiewicz@informatik.uni-marburg.de` (Christian Komusiewicz), `sorge@post.bgu.ac.il` (Manuel Sorge)