

Accepted Manuscript

Title: Matching Algorithms for Assigning Orthologs after Genome Duplication Events

Author: Guillaume Fertin Falk Hüffner Christian Komusiewicz Manuel Sorge



PII: S1476-9271(18)30161-0
DOI: <https://doi.org/doi:10.1016/j.compbiolchem.2018.03.015>
Reference: CBAC 6820

To appear in: *Computational Biology and Chemistry*

Received date: 8-3-2018
Accepted date: 13-3-2018

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

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Matching Algorithms for Assigning Orthologs after Genome Duplication Events

Guillaume Fertin^{a,1}, Falk Hüffner², 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)

¹GF was supported by PHC Procope (project 17746PC).

²FH was supported by the DFG, project ALEPH (HU 2139/1).

³CK was supported by the DFG, project MAGZ (KO 3669/4-1), and by the DAAD Procope program (project 57317050).

⁴MS was supported by the DFG, project DAPA (NI 369/12), the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 631163.11, and the Israel Science Foundation (grant no. 551145/14).

Download English Version:

<https://daneshyari.com/en/article/6486919>

Download Persian Version:

<https://daneshyari.com/article/6486919>

[Daneshyari.com](https://daneshyari.com)