

Finding set of "core genes" for Nematoda, Trematoda and Cestoda

John Martin (✉ jmartin@wustl.edu)

Mitreva Lab (MGI, Washington University in St. Louis)

Makedonka Mitreva

MGI, Washington University in St. Louis

Method Article

Keywords: core genes, orthologous groups, protein families, helminths, nematoda, platyhelminthes, trematoda, cestoda

Posted Date: May 17th, 2018

DOI: <https://doi.org/10.1038/protex.2018.043>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Sets of core genes were defined for nematodes, trematodes and cestodes for use as training sets for gene prediction.

Introduction

"Core genes" are defined as genes shared by all species in a taxonomic group. Using the protocol, sets of core genes can be identified for such a group. These can then be used as training sets for gene prediction. This method describes the construction of sets of Nematode Orthologous Groups (NOGs)¹, Trematode Orthologous Groups (TROGs) and Cestode Orthologous Groups (CEOGs).

Reagents

1. an installation of HMMER (v2.3.2) 2. prerequisites for OrthoMCL² (e.g. BLAST, MySQL)

Equipment

compute cluster for parallel processing of multiple steps in OrthoMCL.

Procedure

Building the NOGs and TROGs: 1. The full proteomes of twelve nematode species, four trematode species, one cestode and several eukaryotic outgroups are clustered using OrthoMCL². 2. NOGs are defined as clusters containing at least one member from each of the twelve included nematode species. 3. TROGs are defined as clusters containing at least one member from each of the four included trematode species. 4. An HMM is built for each cluster using HMMER (v2.3.2)³ which can be used for training gene prediction software. For the associated publication, the CEOGs were defined using a different set of cluster input since the OrthoMCL input for the steps above only include a single cestode: 1. The full proteomes of four cestode species and several eukaryotic outgroups were clustered using OrthoMCL. 2. CEOGs are the groups containing at least one member from each of the four included cestode species. 3. An HMM is built for each clustering using HMMER (v2.3.2)³ which can be used for training gene prediction software.

Anticipated Results

Sets of HMM models for core genes of Nematoda, Trematoda and Cestoda which can be used for training gene prediction software.

References

1. Mitreva, M. et al. The draft genome of the parasitic nematode *Trichinella spiralis*. Nat Genet 43, 228-235, doi:10.1038/ng.769 (2011). 2. Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13, 2178-2189, doi:10.1101/gr.1224503 (2003). 3. "HMMER URL":<http://hmmer.org/>