OrthoQuery: A Tripal Database Module to Assess & Visualize Gene Family Evolution

Authors: Sumaira Zaman¹, Emily Grau¹, Sean Buehler¹, Risharde Ramnath¹, Stephen Ficklin², Jill Wegrzyn¹ 1. University of Connecticut, Storrs, CT 2. Washington State University, Pullman, WA

Abstract:

The abundance of transcriptomic resources for non-model organisms has enabled researchers to study comparative genomics on a larger scale. Generation of orthologous gene families facilitates comparative genomics by examining gene family evolution events related to selection pressures. Applications developed to study gene homology among species do not allow users to query data directly from external databases hosting resources that are not associated with a reference genome. Furthermore, real time computation of orthogroups for user selected subsets paired with interactive visualizations is lacking. OrthoQuery, a web-based Tripal module, provides a semi-automated analytical framework to enable comparisons among curated proteins and interactive visualizations in context of the resulting species tree. OrthoFinder, optimized with Diamond, is leveraged for protein level comparisons, and the Tripal database framework, coupled with Galaxy integration, supports various workflows and visualization options. OrthoQuery processes unigenes and stores a pre-computed set of orthogroups based on available species' resources in the local database. The module provides researchers with options to navigate the resulting species tree, identify ancestral/species-specific groups of genes, and associate orthogroups with functional annotations. We demonstrate OrthoQuery's application when studying conifer genomes. Although the development of this module is intended for forest tree research, any of the other 30+ clade or organism specific databases supported Tripal will have access to OrthoQuery as part of Tripal's open source project. This robust and flexible Tripal module aims to enable researchers in conducting comparative genomics analysis for user selected species, with an emphasis on pre-processing transcriptomic resources to include non-model organisms.

OrthoQuery Architecture





OrthoQuery User Interface



OrthoQuery Overview

Orthofinder is developed for comparative genomic analysis using orthogroups. Users can discover novel orthogroups (i.e. gene families) by uploading a transcriptome or a proteome and selecting species in the TreeGenes database for a comparative genomics analysis. Otherwise, users can upload a small set of protein sequences and determine which gene families these sequences belong by selecting species that reside in TreeGenes

Orthogroup Analysis

Lipload a proteome

Select a type of input sequence proteome

optoad a proteome	
Choose File No file chosen	UPLOAD
Sequences in a proteome will be directly s	sent to OrthoFinder for analyses.
SUBSET DATABASE OR SELECT ENT	IRE DATABASE
Select Entire Database	
Select Specific Species	
DEFAULT PARAMETERS FOR ORTH	OFINDER EXECUTION
Sequence Similarity Search Program * Me	hod of aligning protein sequences.
Diamond	
Inflation Parameter * Inflation parameter u	sed to create orthogroups
1.5	
Galaxy server * UConn Galaxy @ CAM	\$

OrthoQuery Interactive Visualization



United States Department of Agriculture