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 Interactive Visualization