OrthoQuery: A Tripal Module to Assess Gene Family Evolution in Sequenced Gymnosperms

Sumaira Zaman

@Sumaira_Zaman
Why Gymnosperms?

- Gymnosperms are HARD!
  - Evolution?
  - Large & Complex Genomes
    - Ranges from 4 to 32 Gbps

- Traditional Analyses:

- TreeGenes
  - Powered by Tripal
    - Supports ≥12 databases
    - OrthoQuery IS extensible (i.e. NOT limited to gymnosperms)
    - Can Integrate with Galaxy
What are Orthogroups?

- An orthogroup is the set of genes that are descended from a single gene in the last common ancestor of all the species being considered.

Application of Orthogroups

- Understand Evolution
- Discover Novel Genes
- Comparative Genomics

Discover Orthogroups

OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy

David M. Emms and Steven Kelly
Genome Biology 2015 16:157
Received: 23 December 2014 | Accepted: 8 July 2015 | Published: 6 August 2015
OrthoQuery as a Tripal Module

Input → Species → Galaxy → Results

1. Single Protein Sequences
2. Transcriptome
3. Proteome

Build Proteome

Execute Sequence Similarity Search (Diamond)

Functionally Annotate Execute Orthofinder
OrthoQuery as a Tripal Module

Input → Species → Workflows → Results

Transcriptomes → Filter & Frame Selection → Functionally Annotate

Resides in TreeGenes Database – UniGenes (40 species)
OrthoQuery as a Tripal Module

Tripal Galaxy API

blend4php v0.1
A PHP API for interacting with Galaxy
OrthoQuery as a Tripal Module

- Input
- Species
- Workflows
- Results

Diagram showing a flow from Input to Species to Workflows leading to Results, with a tree structure representing proteome data.
Future work: Attach functional annotation (stored in TreeGenes) to sequences in selected orthogroups.
Thank you!

Acknowledgements: Emily Grau, Sean Buehler, Risharde Ramnath, Stephen Ficklin, Jill Wegrzyn

Questions?

Douglas fir