

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

Sumaira Zaman

January 16, 2018

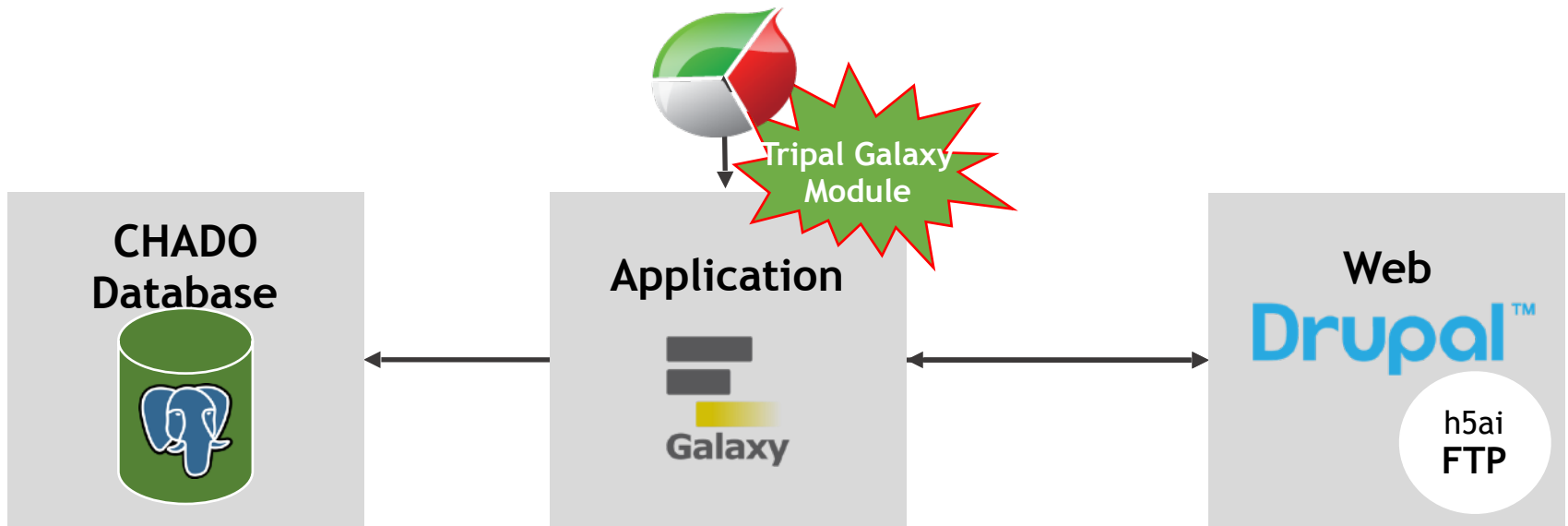
 @Sumaira_Zaman_



Tripal Database



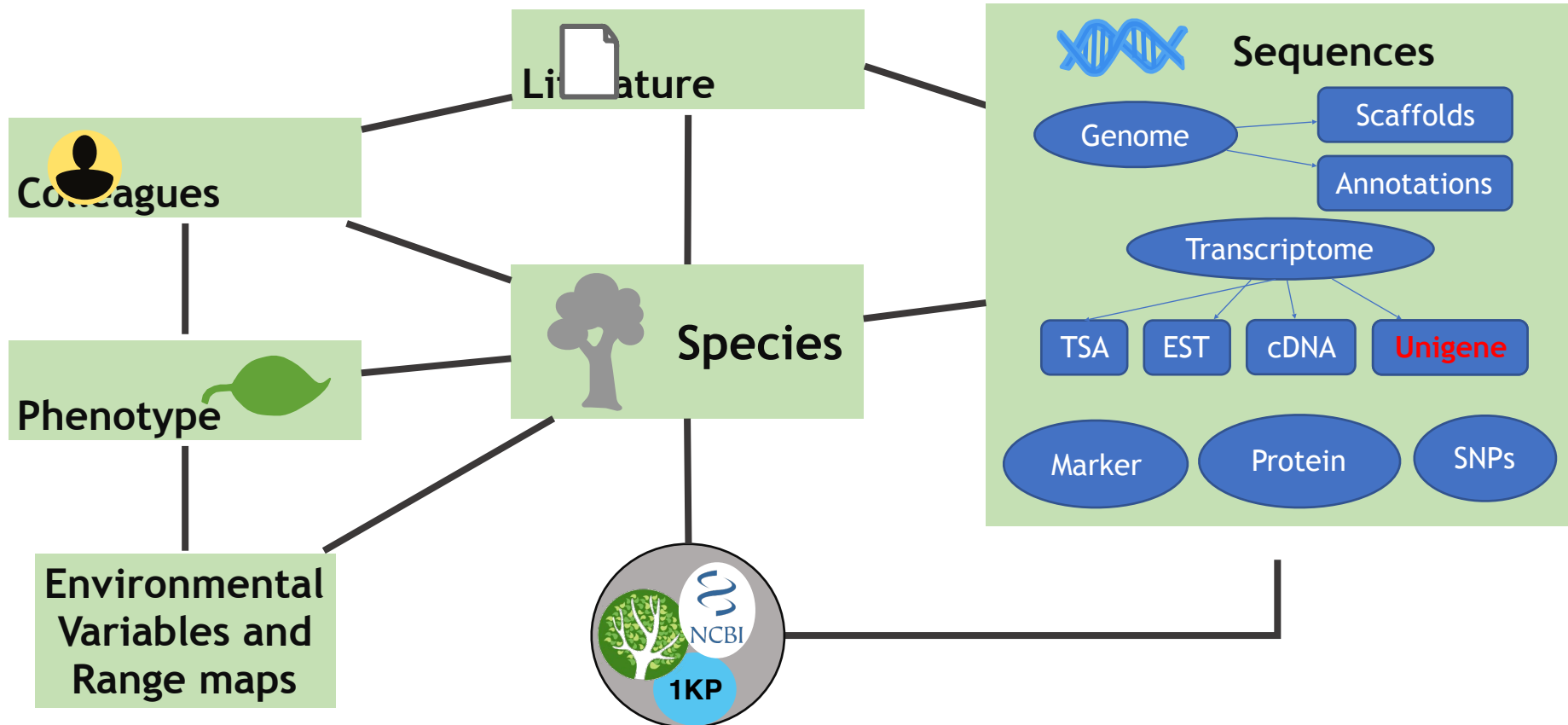
Tripal Overview



30+ clade organism databases → including TreeGenes!



Data in TreeGenes



Unigenes stored in TreeGenes

Sequence	Species	Centroid Sequence	Source of Centroid	Functional Annotation
TG98.Pagl.v1.1	pagl	GW737853.1	EST	

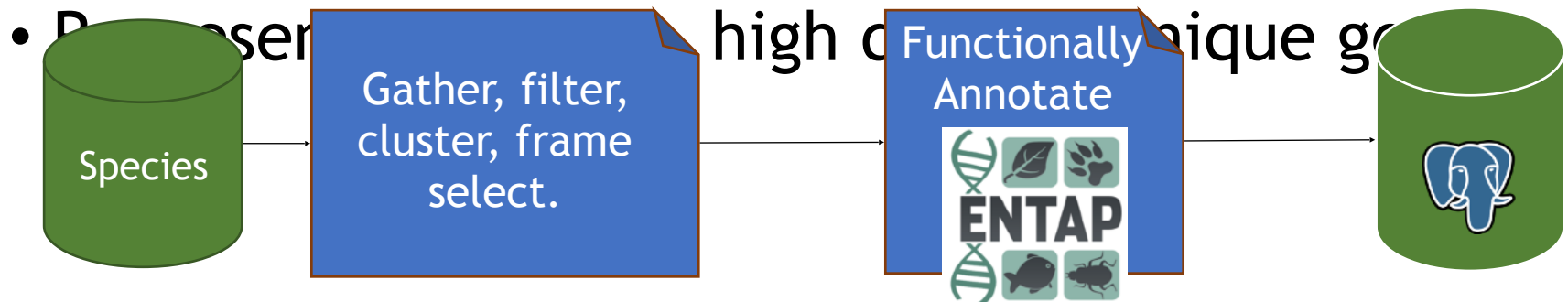
Sequence	Sequence Similarity Hit	Gene Ontology Hit
TG98.Pagl.v1.1		

Sequence	Query, % Identity, Alignment length, ..., Database
TG98.Pagl.v1.1	

Sequence	Egglog Description, KEGG Terms, Gene Ontology terms
TG98.Pagl.v1.1	

Unigene Data

- Originates from multiple repositories (i.e. NCBI, 1kp)
- These sequences are derived from multiple libraries
 - Transcriptome shotgun assembly (TSA) & Expressed sequenced tags (ESTs)
- Includes multiple tissue types



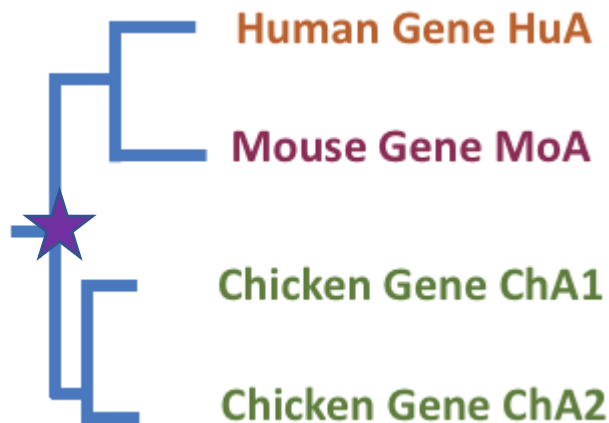
Gene Family Evolution



Background: Orthogroups

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

Comparative Genomics

Understand evolution

Discover novel genes

Discover Orthogroups

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

David M. Emms and Steven Kelly ✉

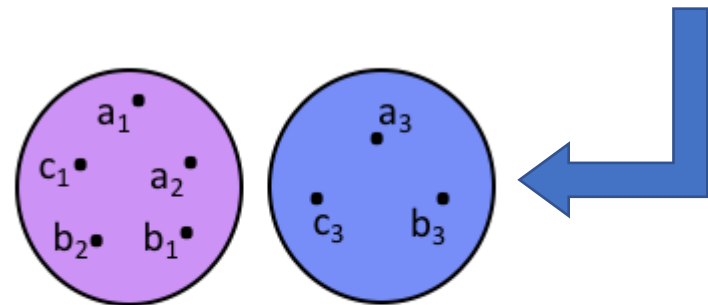
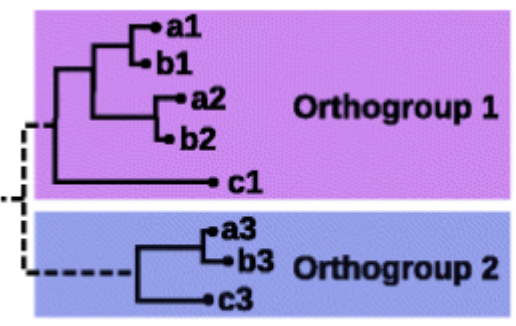
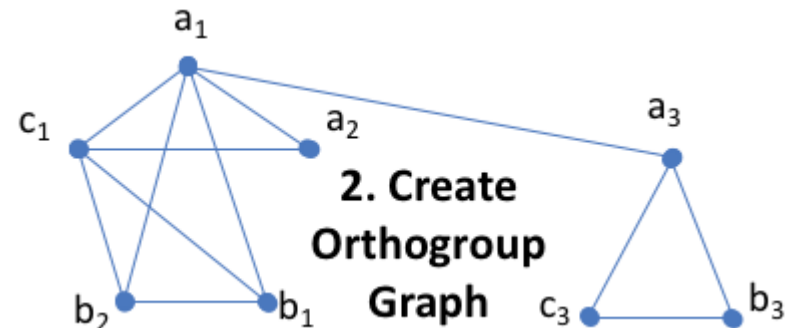
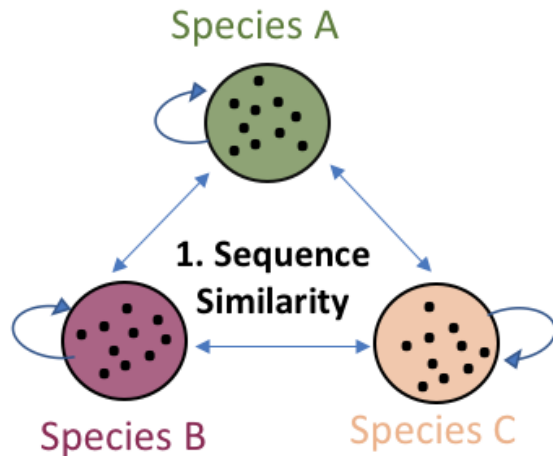
Genome Biology 2015 16:157

<https://doi.org/10.1186/s13059-015-0721-2> | © Emms and Kelly, 2015

Received: 23 December 2014 | Accepted: 8 July 2015 | Published: 6 August 2015



OrthoFinder Summary



4. Infer Relationship

3. Orthogroups



OrthoFinder Results

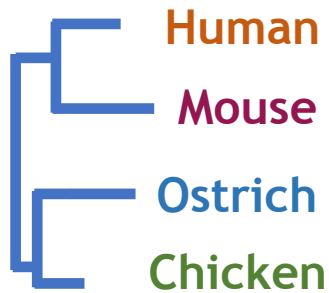
PLETHORA of information

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	proteome.ath	proteome.atr	proteome.bdi	proteome.bra	proteome.cae	proteome.can	proteome.cer	proteome.cga	proteome.cgl	proteome.cim	proteome.cle	proteome.cma	proteome.cmi	proteome.cml	proteome.cmm	proteome.cmu	proteome.cmu	proteome.cmu	proteome.cmu	proteome.cmu	proteome.cmu	
2	OG00000001	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
3	OG00000002	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	OG00000003	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	OG00000004	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	OG00000005	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	OG00000006	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	OG00000007	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	OG00000008	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	OG00000009	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	OG00000010	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	OG00000011	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	OG00000012	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	OG00000013	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	OG00000014	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	OG00000015	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	OG00000016	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	OG00000017	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
19	OG00000018	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	OG00000019	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	OG00000020	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
22	OG00000021	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
23	OG00000022	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	OG00000023	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
25	OG00000024	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	OG00000025	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
27	OG00000026	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
28	OG00000027	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
29	OG00000028	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
30	OG00000029	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
31	OG00000030	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
32	OG00000031	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
33	OG00000032	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
34	OG00000033	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
35	OG00000034	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
36	OG00000035	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

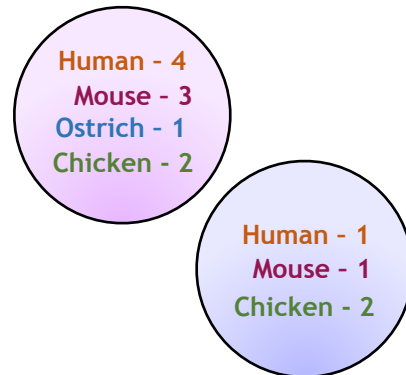


OrthoQuery - Tying Orthofinder Results

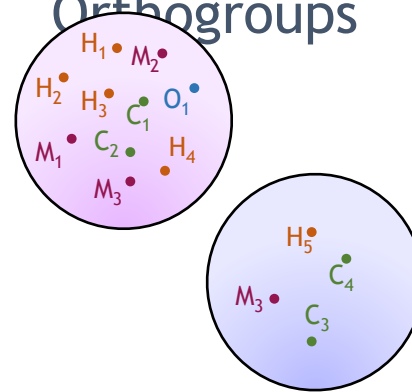
1. Species Tree



2. Orthogroup's Gene Count



3. Genes constituting Orthogroups



4. Functional Annotation

Back-end Database

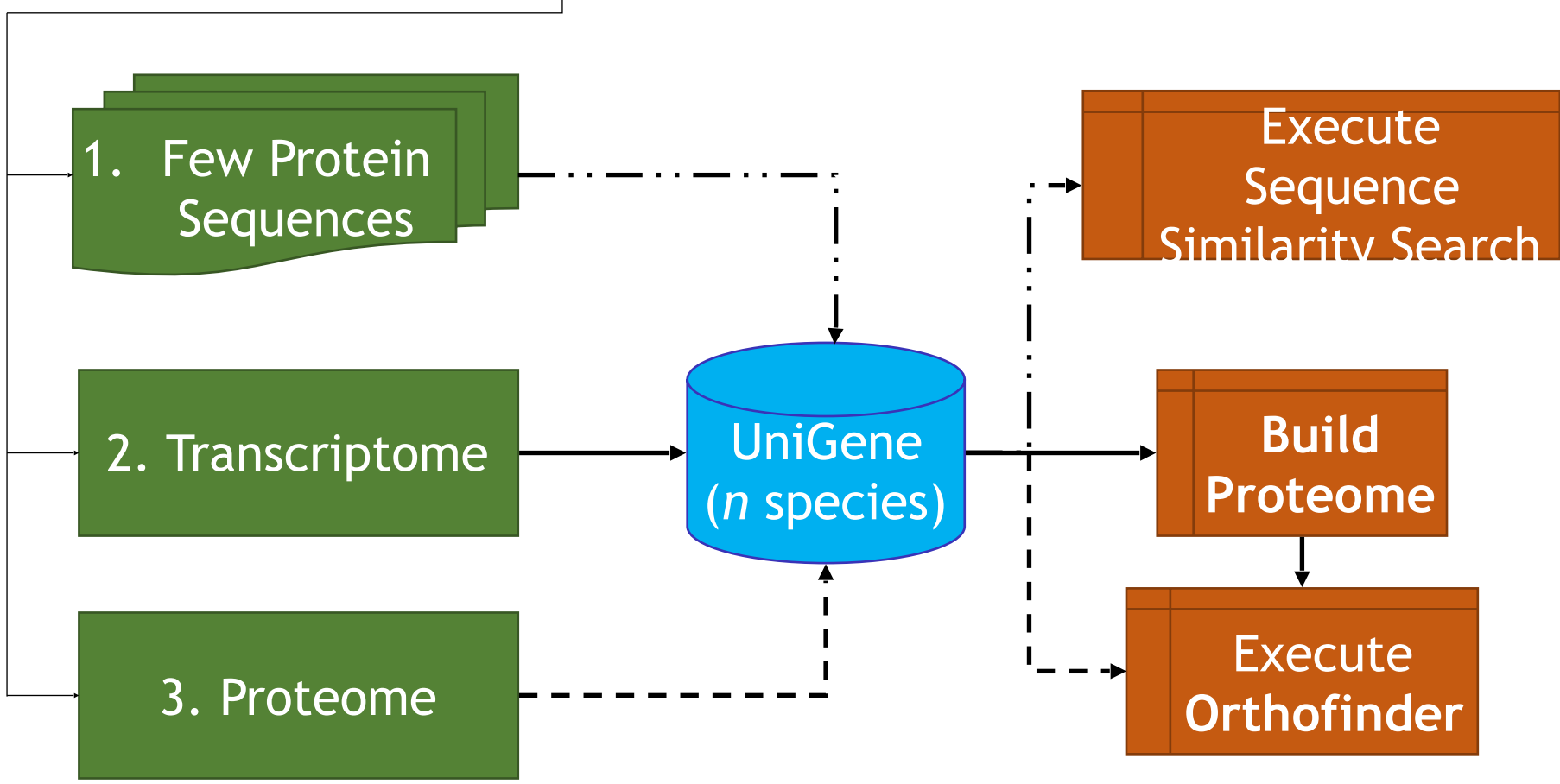
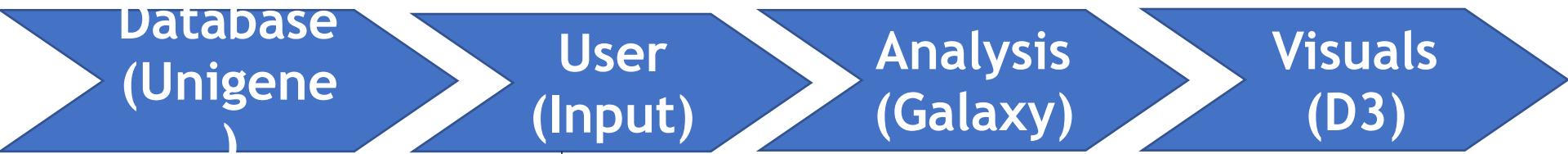
Gene	Function
H ₄	Disease Resistance
M ₃	Disease Resistance
O ₁	Metabolite
C ₄	Metabolite

Question: How are orthogroups evolving in specific parts of tree and what is their function?



OrthoQuery





Query target database (Protein)

Query target database (Protein)

Query target database (Protein)

Query target database (Protein)

SUBMIT

- Abies alba UniGene, version 1
- Abies sachalinensis UniGene, version 1
- Acacia auriculiformis UniGene, version 1
- Acacia koa UniGene, version 1
- Acacia mangium UniGene, version 1
- Araucaria cunninghamii UniGene, version 1
- Betula papyrifera UniGene, version 1
- Betula platyphylla UniGene, version 1
- Callitris intratropica UniGene, version 1
- Callitris sulcata UniGene, version 1
- Cephalotaxus hainanensis UniGene, version 1
- Corylus avellana UniGene, version 1
- Cryptomeria japonica UniGene, version 1
- Cunninghamia lanceolata UniGene, version 1
- Fagus sylvatica UniGene, version 1
- Jatropha curcas UniGene, version 1
- Juniperus ashei UniGene, version 1
- Juniperus pinchotii UniGene, version 1
- Juniperus virginiana UniGene, version 1
- Larix kaempferi UniGene, version 1
- Millettia pinnata UniGene, version 1
- Picea glauca UniGene, version 1
- Picea sitchensis UniGene, version 1
- Pinus albicaulis UniGene, version 1
- Pinus canariensis UniGene, version 1
- Pinus flexilis UniGene, version 1

USER MENU

[Edit Profile](#) >

[Log out](#) >

[My account](#) >

[Add new job posting](#) >

[Add new meeting posting](#) >

MEETINGS

**Ecological Society of America
Annual Meeting**
August 5 to 10 2018
New Orleans, USA

**Tree Resistance Workshop
2018**
August 5 to 10 2018
Mt. Sterling, USA

[View more meetings](#)



Executing Workflows in Galaxy



Tripal Galaxy Module

Database
(Unigene)

User
(Input)

Analysis
(Galaxy)

Visuals
(D3)

Galaxy / Treegenes

Analyze Data Workflow Visualize Shared Data Admin Help User

Using 87.3 MB

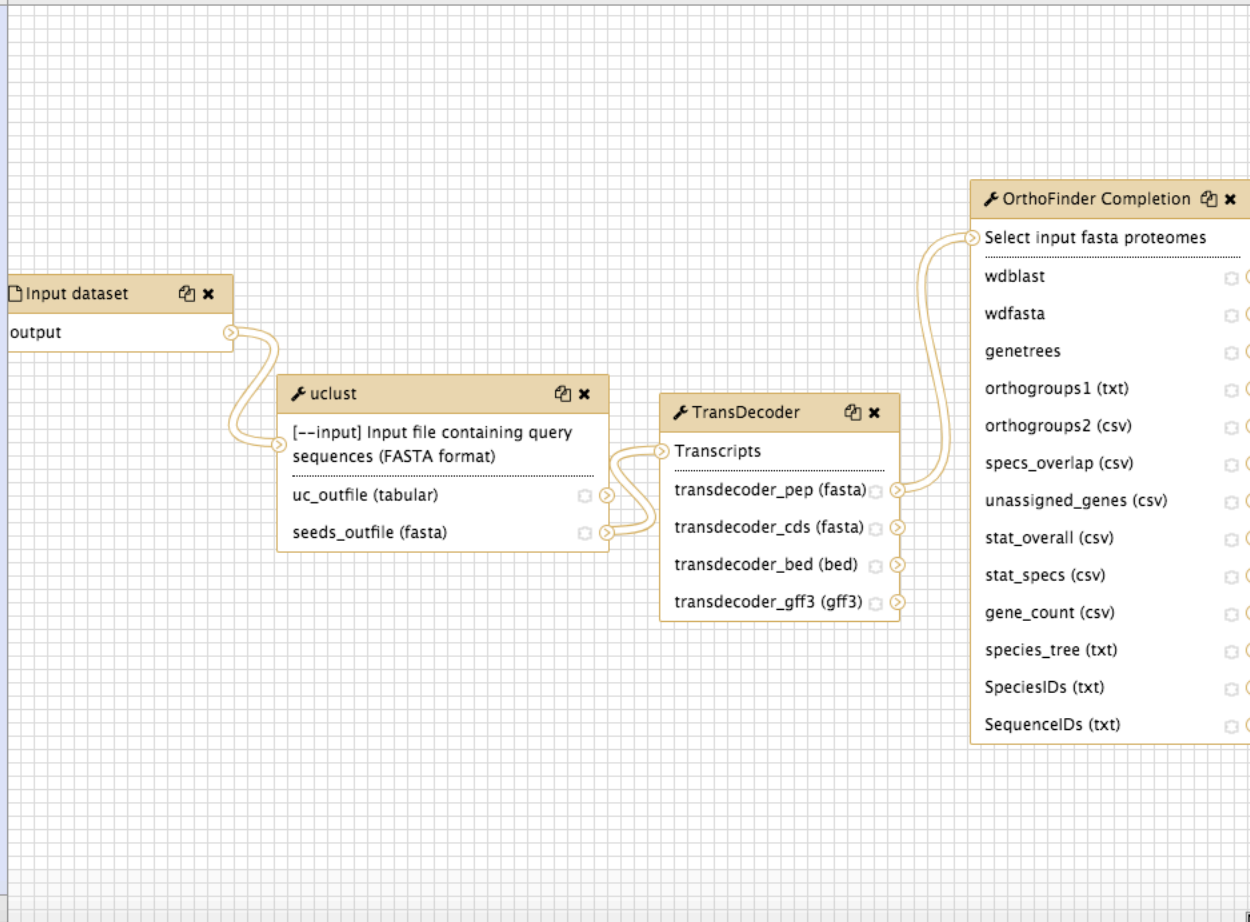
Tools

Workflow Canvas | OrthoQuery_transcriptome

Details

search tools

- Inputs
- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Phenotype Association
- NGS: QC and manipulation
- AMTools
- Comparative Genomics
- Statistics
- Graph/Display Data
- Association Mapping
- Sequence Search
- Annotation
- testing
- Statonlab Tools
- Aurora Galaxy Tools
- usearch
- vsearch
- Data Manager Tools



Edit Workflow Attributes

Name:
OrthoQuery_transcriptome

Tags:

Apply tags to make it easy to search for and find items with the same tag.

Annotation / Notes:
Workflow executed when user provides a full transcriptome for any given species
Add an annotation or notes to a workflow; annotations are available when a workflow is viewed.

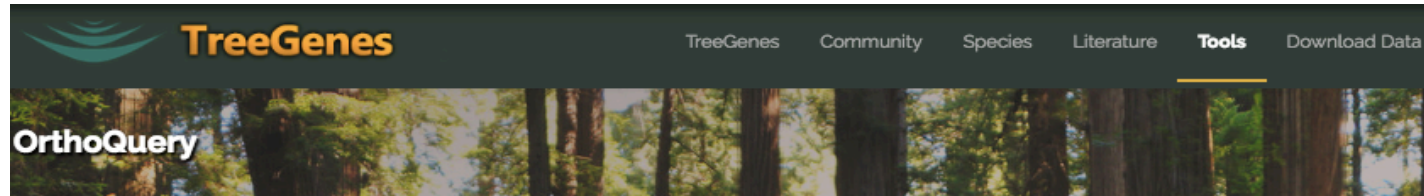


Database
(Unigene)

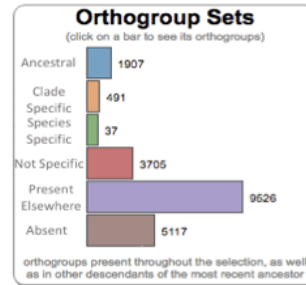
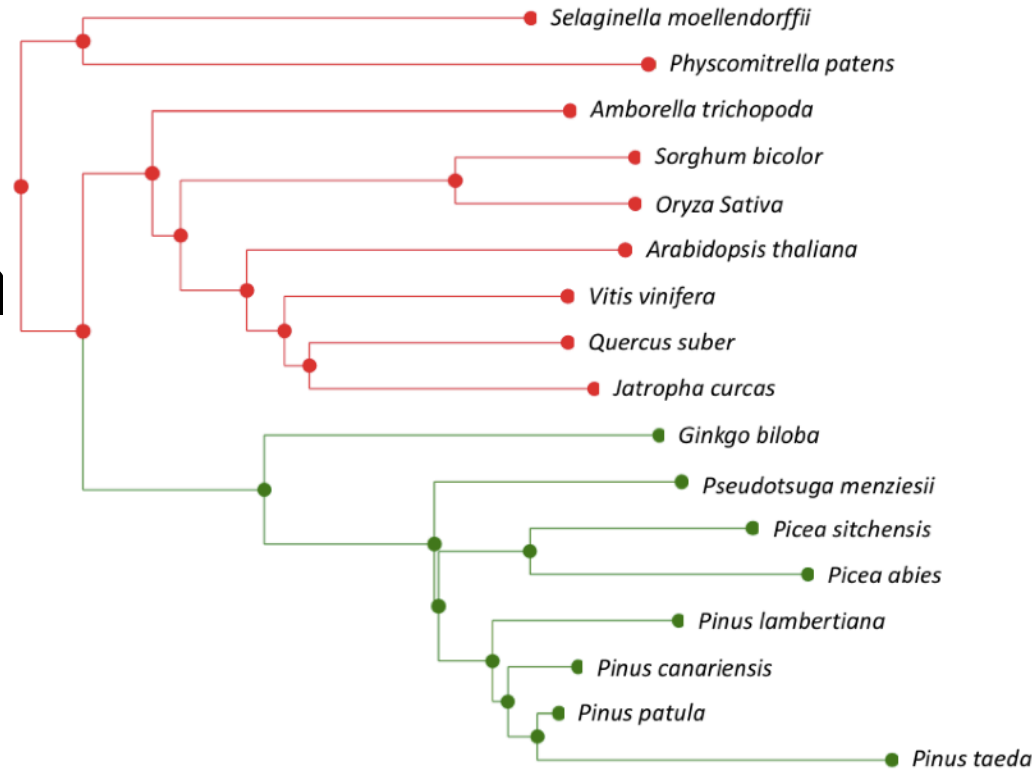
User
(Input)

Analysis
(Galaxy)

Visuals
(D3)



Show branch length



Interactive
Visualization



[See Summary](#)

[Click on a node for more details...](#)

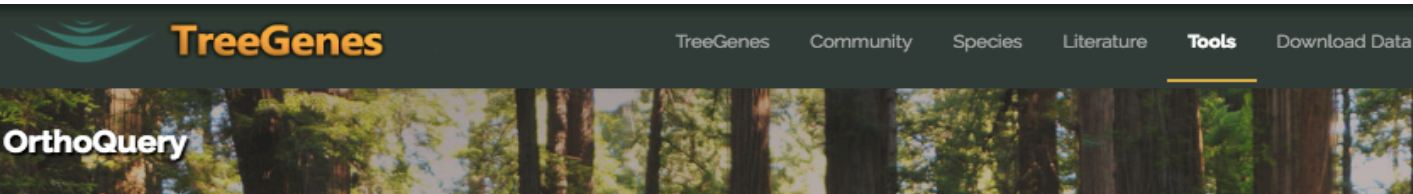


Database
(Unigene)

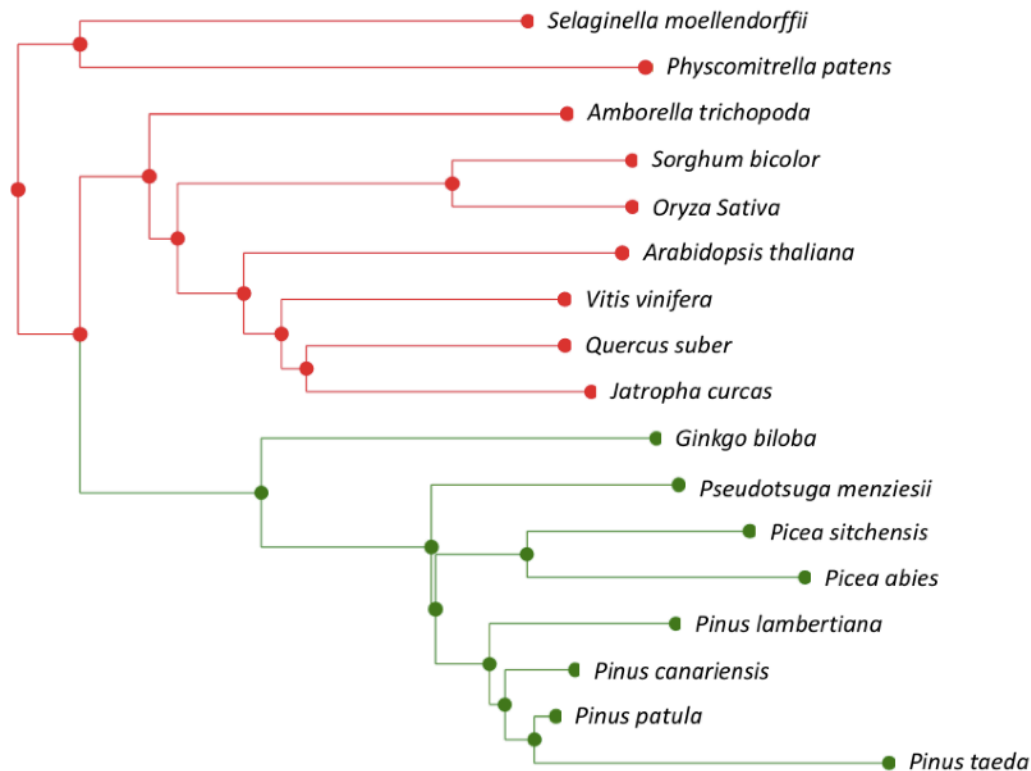
User
(Input)

Analysis
(Galaxy)

Visuals
(D3)

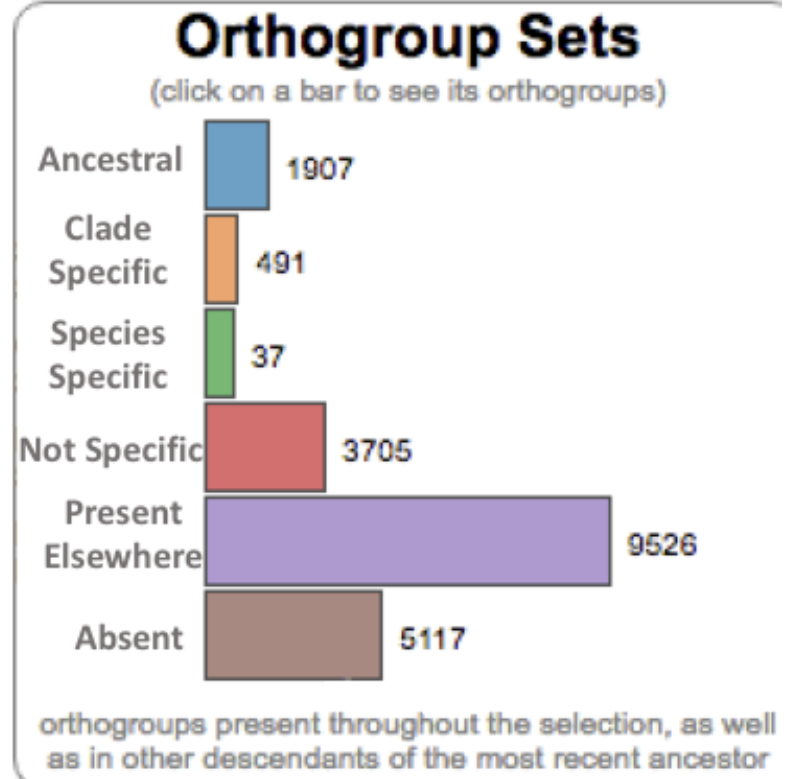


Show branch length



Click on a node for more details...

[See Summary](#)



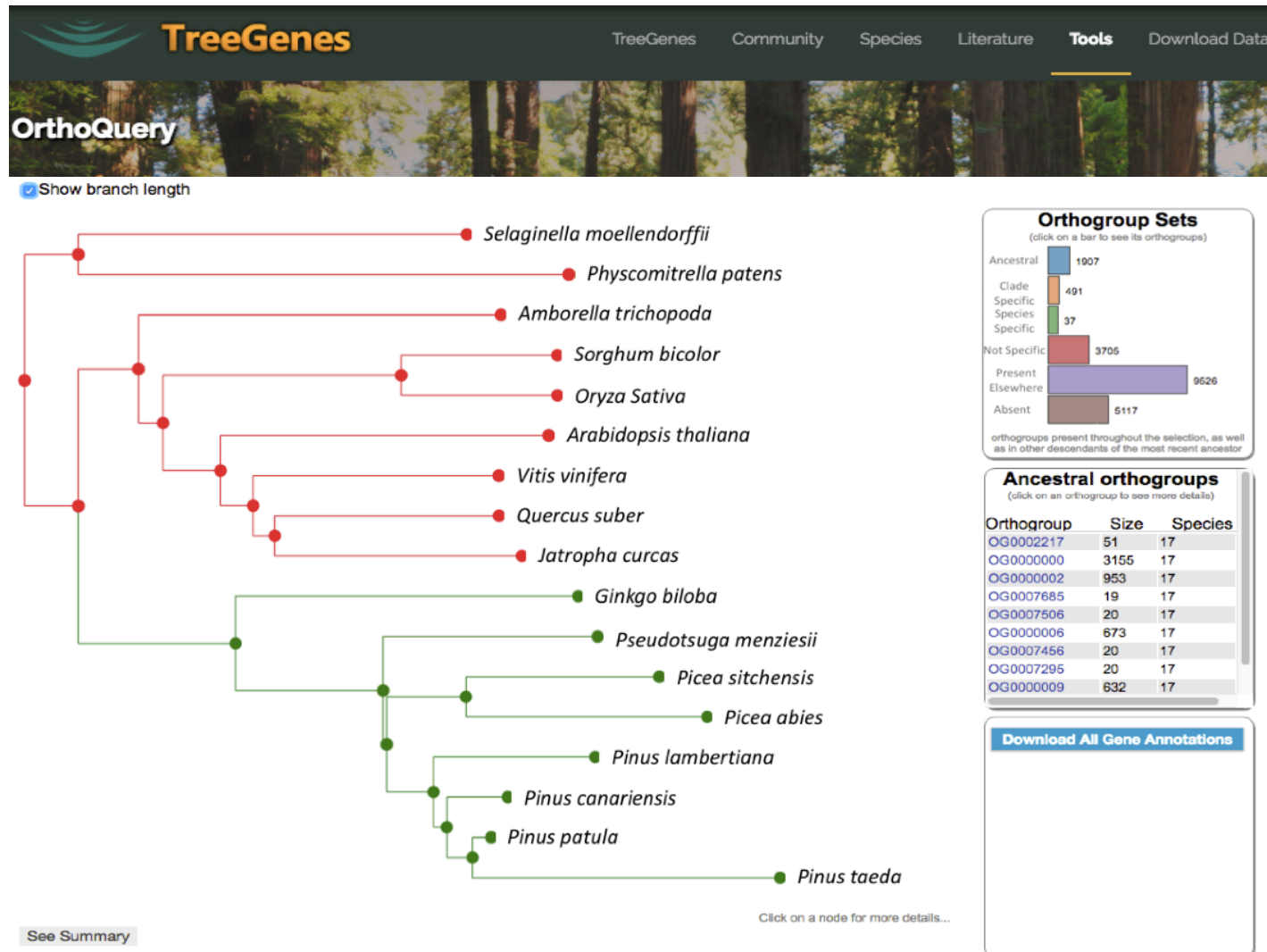
Database
(Unigene)

User
(Input)

Analysis
(Galaxy)

Visuals
(D3)

Select
Relationship
of
Interest

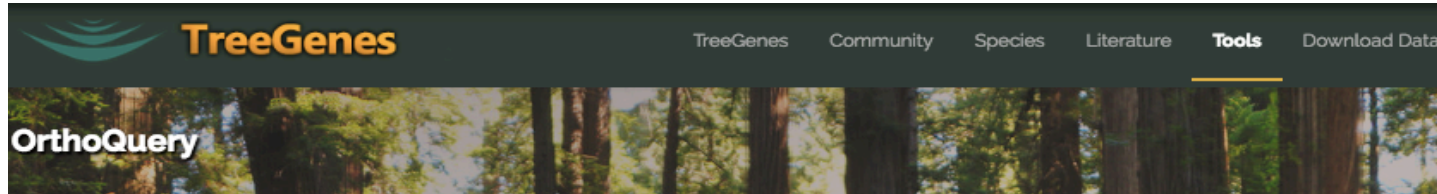


Database
(Unigene)

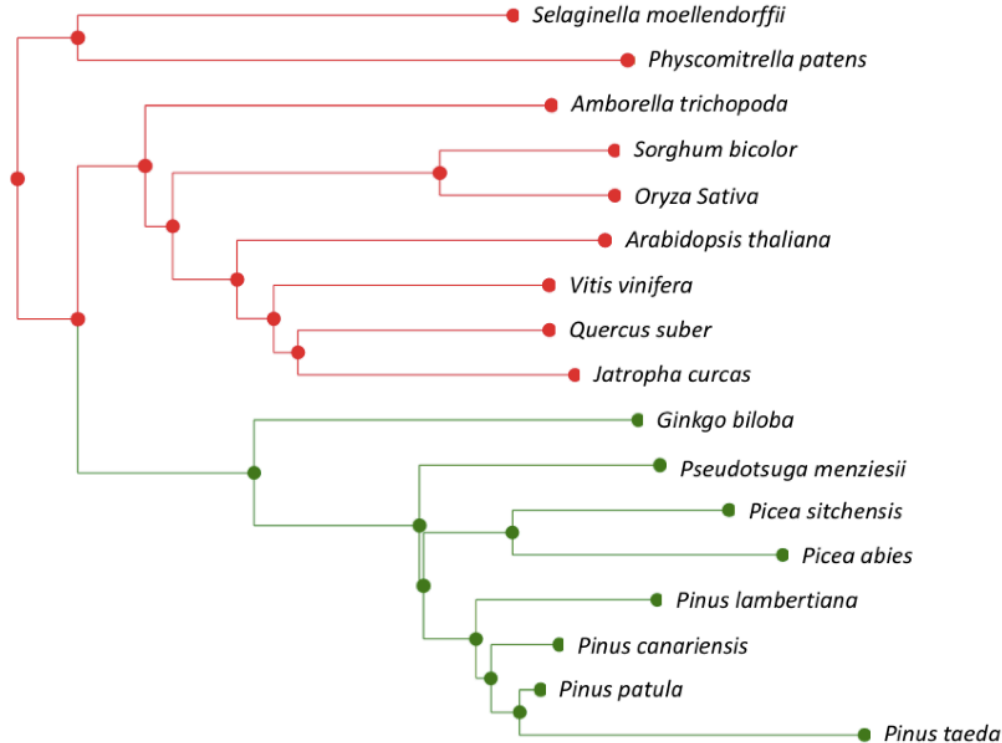
User
(Input)

Analysis
(Galaxy)

Visuals
(D3)

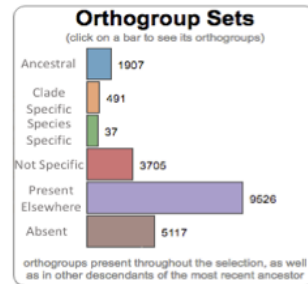


Show branch length



[See Summary](#)

[Click on a node for more details...](#)



Ancestral orthogroups
(click on an orthogroup to see more details)

Orthogroup	Size	Species
OG0002217	51	17
OG0000000	3155	17
OG0000002	953	17
OG0007685	19	17
OG0007506	20	17
OG0000006	673	17
OG0007456	20	17
OG0007295	20	17
OG0000009	632	17
OG0000010	631	17

Top OG0000000 gene

Gene
PITA_00442

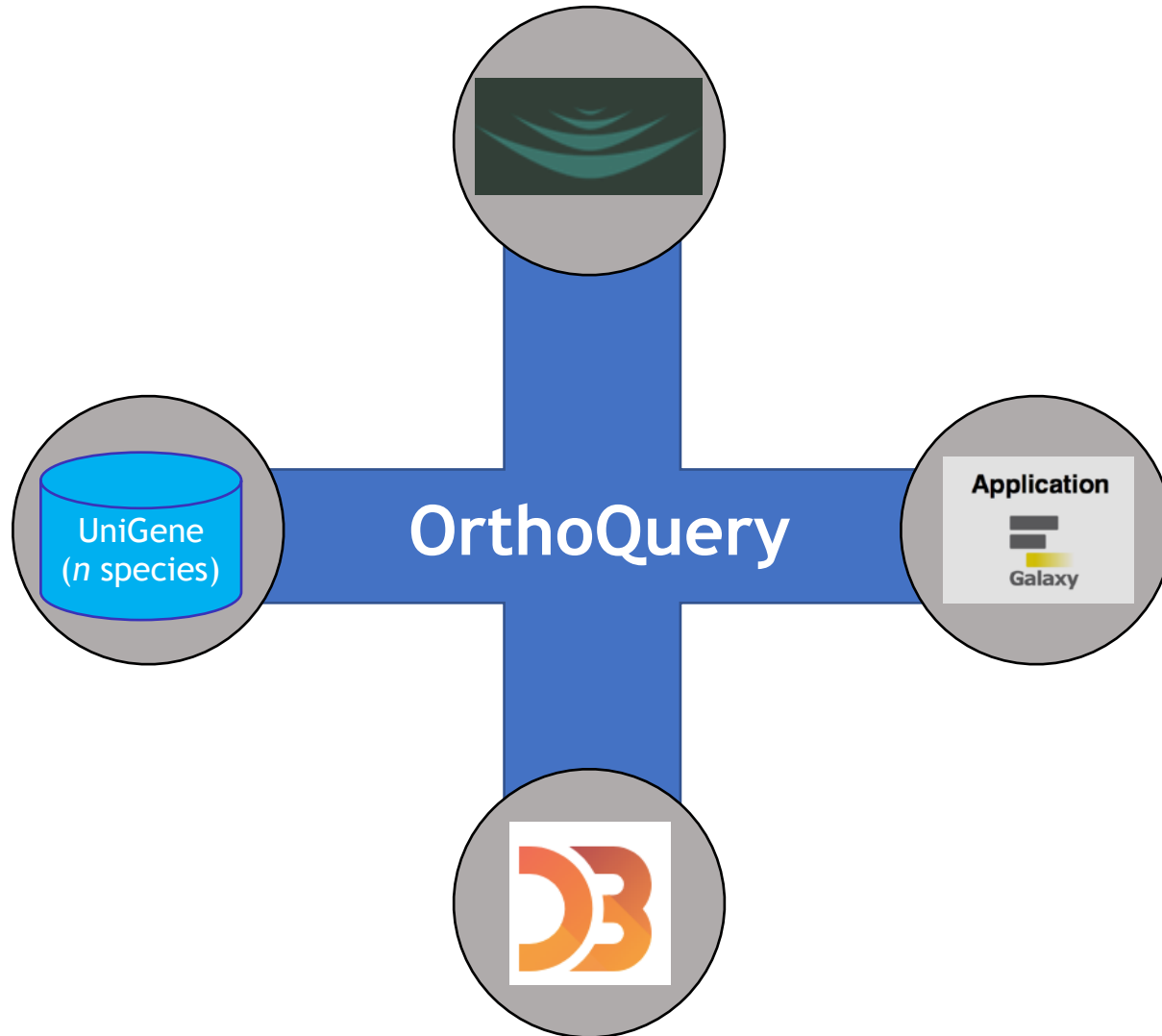
Egglog Description
dynammin-related protein

[Download All Gene Annotations](#)

Select
Orthogroup
of Interest



Conclusion





Thank you!

Questions

Acknowledgements: Emily Grau, Sean Buehler , Nic Herndon, Risharde Ramnath, Shawna Spoor, Stephen Ficklin, Jill Wegrzyn

