

# Phylogenomics

Ruiqi Li  
PhD candidate  
University of Colorado Boulder

# Contents

1. Data Type
2. Phylogenomics Pipeline
  - 2.1. Finding Orthologs/Exon assemble
  - 2.2. Align individual genes
  - 2.3. Trim Alignments
  - 2.4. Infer Phylogeny
3. Practice

# Why phylogenomics?

Genome Tree vs Single Gene Tree

# 1. Data Type



Genome

Whole genome, genome skimming



Transcriptome

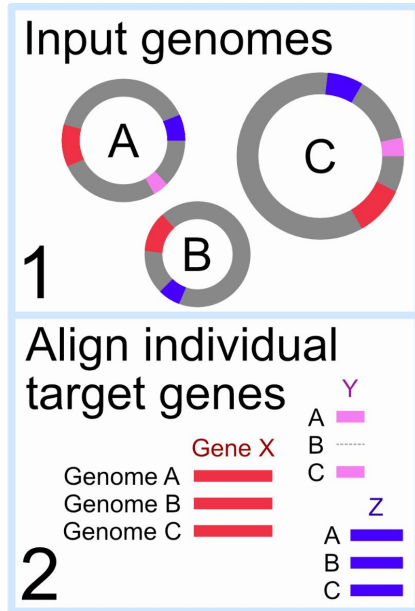


target capture

Exon capture, ultra-conserved element (UCE) capture

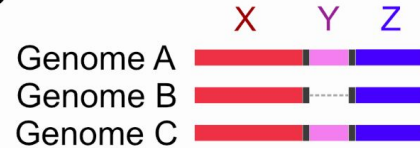
Others: SNPs (RAD-seq)

## 2. Phylogenomics Pipeline



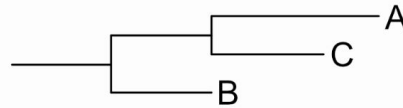
3

Stick alignments together



4

Infer evolutionary relationships



## 2. Phylogenomics Pipeline

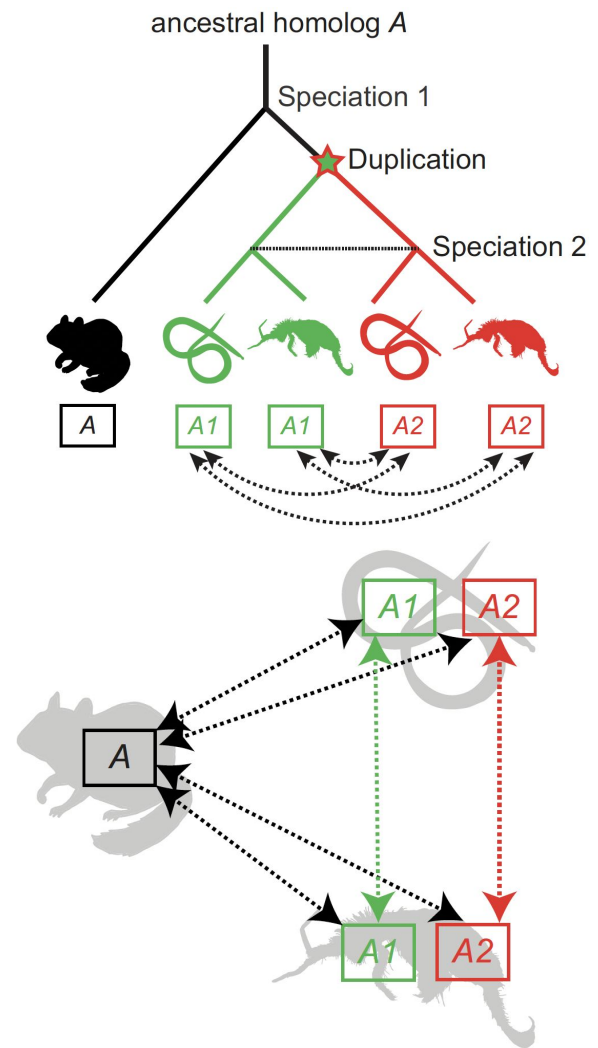
### 2.1 Finding Orthologs

The green A1s are \_\_\_\_\_.

A1 and A2 are \_\_\_\_\_.

A and A1 are \_\_\_\_\_.

A and A2 are \_\_\_\_\_.

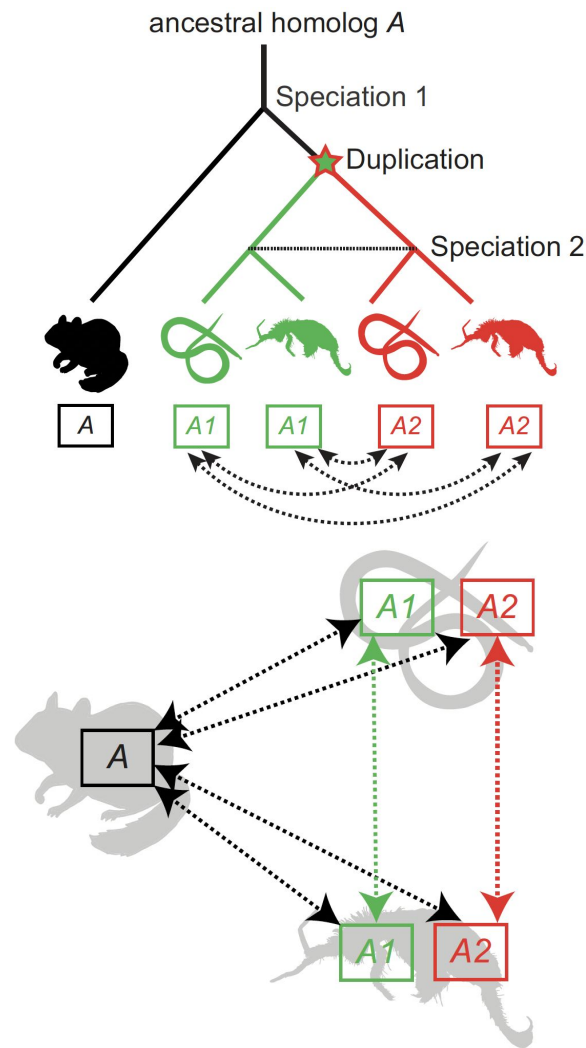


## 2. Phylogenomics Pipeline

### 2.1 Finding Orthologs

**Orthologs** are genes in different species evolved from a common ancestral gene by a **speciation event**.

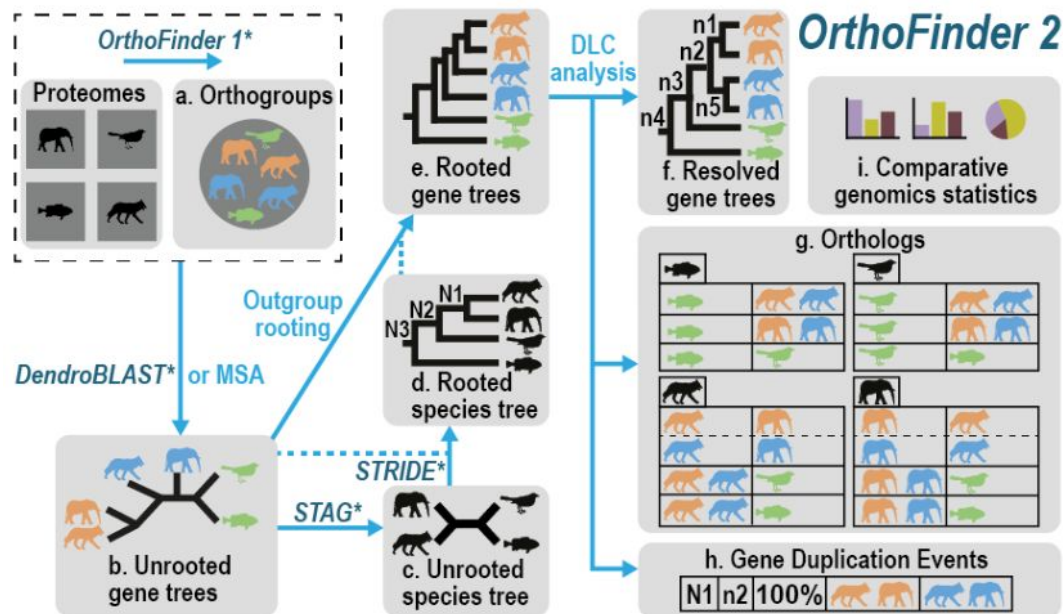
**Paralogs** are gene copies created by a **duplication event** within the same genome.



## 2. Phylogenomics Pipeline

### 2.1 Finding Orthologs from transcriptomes and genomes

software: OrthoFinder, OrthoMCL, OMA, OrthoFisher





## 2. Phylogenomics Pipeline

### 2.1 “Finding” Orthologs from target capture data

With targeted resequencing, **a subset of genes or regions of the genome** are isolated and sequenced. Target enrichment works by capturing genomic regions of interest by hybridization to target-specific biotinylated probes

Design the probes with existing genome/transcriptome data

Assembling with reference: HybPiper pipeline (reference-guided assembly)

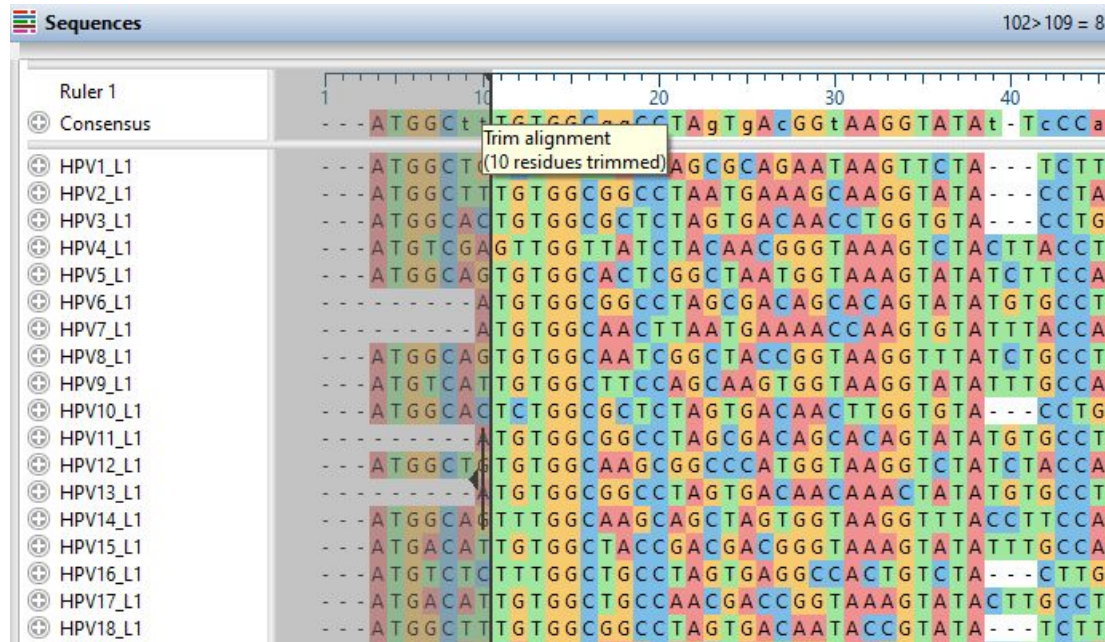
## 2. Phylogenomics Pipeline

### 2.2 Align individual genes

software: mafft, muscle, etc.

#### Multiple Sequence Alignment

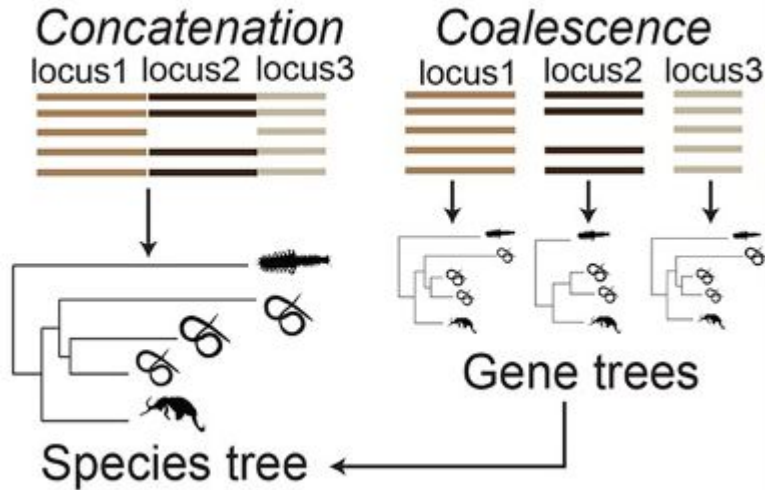
|            |   |   |   |     |   |   |   |   |     |   |   |   |   |     |   |   |   |   |   |
|------------|---|---|---|-----|---|---|---|---|-----|---|---|---|---|-----|---|---|---|---|---|
|            |   |   |   | 115 |   |   |   |   | 120 |   |   |   |   | 125 |   |   |   |   |   |
| Sequence A | A | G | T | T   | G | A | C | T | T   | C | T | C | A | G   | G | T | A | T | T |
| Sequence B | A | G | G | T   | A | A | C | T | T   | C | A | G | A | T   | G | A | A | A | T |
| Sequence C | A | G | G | T   | C | A | C | - | -   | G | A | C | A | G   | G | C | A | T | T |
| Sequence D | A | G | G | T   | C | A | C | - | -   | G | A | C | A | G   | G | C | A | - | T |
| Sequence E | A | G | G | T   | C | A | C | T | T   | G | A | G | A | -   | G | C | A | - | T |
| Sequence F | A | G | G | T   | C | A | C | T | T   | G | A | C | A | G   | G | C | A | T | T |



## 2. Phylogenomics Pipeline

### 2.4 Phylogeny Inference

concatenation and coalescence



Tutorial 1:

Genome/Transcriptome Data - OrthoFinder  
(coalescence method)

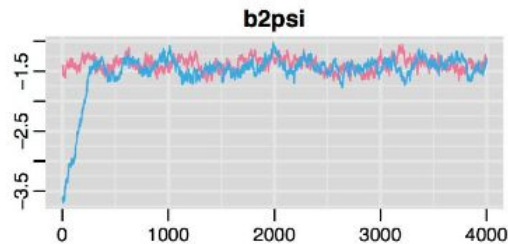
Tutorial 2:

Target Sequence Data - Alignment, Trim,  
concatenation, phylogeny (concatenation  
method)

## 2. Phylogenomics Pipeline

### 2.4 Phylogeny Inference

|            | Maximum Likelihood (ML)   | Bayesian  |
|------------|---|---|
| software   | raxml   | phylobayes  |
| Statistics | search for trees that maximizes the chance of seeing the data<br>$P(\text{Data} \text{Tree})$ | search for the tree that maximizes the chance of seeing the tree given the data<br>$P(\text{Tree} \text{Data})$ |
| computing  | Low   | High  |



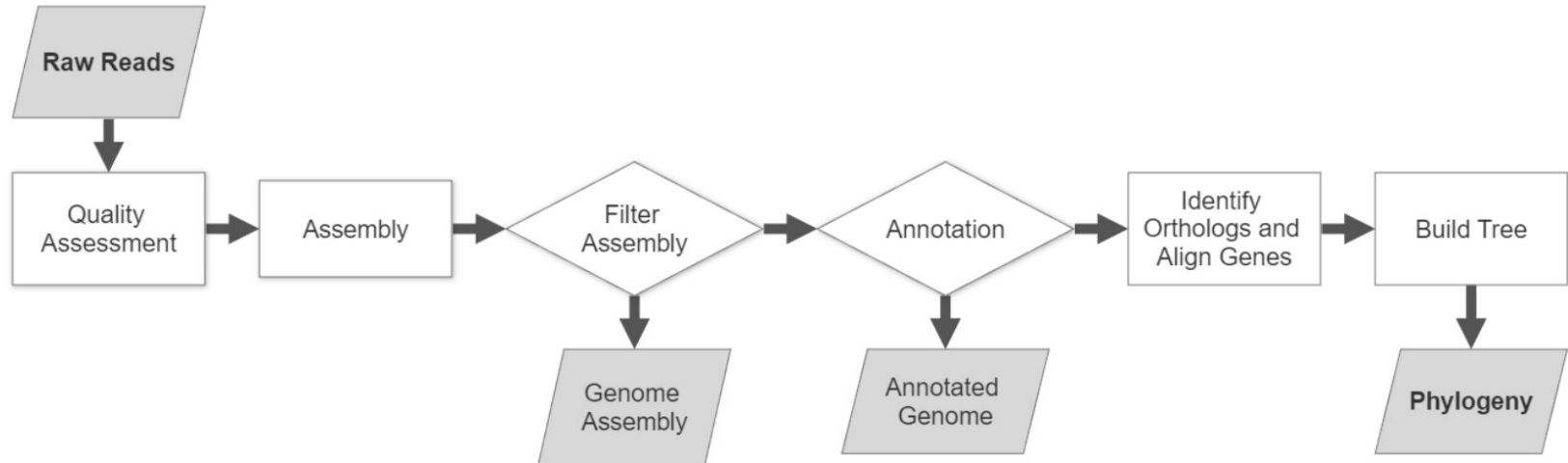
Bayesian MCMC chain convergence

# Software Installation

Please follow instructions on the tutorial [4. Software Installation](#)

# Practice

## Tutorial 1. Simplified Workflow



(Heidi's pipeline)

# OrthoFinder Output

```
total 132K
drwxrwxr-x 2 ruiqi 36K Jul 19 16:38 Orthogroup_Sequences
drwxrwxr-x 2 ruiqi 4.0K Jul 19 16:38 Orthogroups
drwxrwxr-x 2 ruiqi 12K Jul 19 16:38 Single_Copy_Orthologue_Sequences
drwxrwxr-x 2 ruiqi 4.0K Jul 19 16:38 Putative_Xenologs
drwxrwxr-x 2 ruiqi 4.0K Jul 19 16:38 Phylogenetic_Hierarchical_Orthogroups
drwxrwxr-x 2 ruiqi 4.0K Jul 19 16:38 Phylogenetically_Misplaced_Genes
drwxrwxr-x 2 ruiqi 20K Jul 19 16:38 Resolved_Gene_Trees
drwxrwxr-x 2 ruiqi 20K Jul 19 16:38 Gene_Trees
drwxrwxr-x 2 ruiqi 4.0K Jul 19 16:38 Gene_Duplication_Events
drwxrwxr-x 2 ruiqi 4.0K Jul 19 16:38 Comparative_Genomics_Statistics
drwxrwxr-x 3 ruiqi 4.0K Jul 19 16:38 Species_Tree
drwxrwxr-x 7 ruiqi 4.0K Jul 19 16:38 WorkingDirectory
drwxrwxr-x 6 ruiqi 4.0K Jul 19 16:38 Orthologues
-rw-rw-r-- 1 ruiqi 729 Jul 19 16:38 Log.txt
-rw-rw-r-- 1 ruiqi 2.5K Jul 19 16:38 Citation.txt
```



# Single Copy Ortholog from OrthoFinder Results

```
-rw-rw-r-- 1 ruiqi 5.3K Jul 19 16:38 OG0000273.fa
-rw-rw-r-- 1 ruiqi 5.6K Jul 19 16:38 OG0000272.fa
-rw-rw-r-- 1 ruiqi 449 Jul 19 16:38 OG0000271.fa
-rw-rw-r-- 1 ruiqi 679 Jul 19 16:38 OG0000270.fa
-rw-rw-r-- 1 ruiqi 747 Jul 19 16:38 OG0000269.fa
-rw-rw-r-- 1 ruiqi 2.9K Jul 19 16:38 OG0000268.fa
-rw-rw-r-- 1 ruiqi 1.8K Jul 19 16:38 OG0000267.fa
-rw-rw-r-- 1 ruiqi 1.9K Jul 19 16:38 OG0000266.fa
-rw-rw-r-- 1 ruiqi 1.7K Jul 19 16:38 OG0000265.fa
-rw-rw-r-- 1 ruiqi 2.1K Jul 19 16:38 OG0000264.fa
(phylogen) ruiqi@argonaute:~/ruiqi_data/PhylogenomicsWorkshop/ExampleData/OrthoFinder/Results_Jul19/Single_Copy_Orthologue_Sequences$ pwd
/home/ruiqi/ruiqi_data/PhylogenomicsWorkshop/ExampleData/OrthoFinder/Results_Jul19/Single_Copy_Orthologue_Sequences
```

# Practice

## Tutorial 2. Manual Workflow

### 2. Manual Workflow

---

2.1 Genome/Transcriptome/Target Capture Assembly

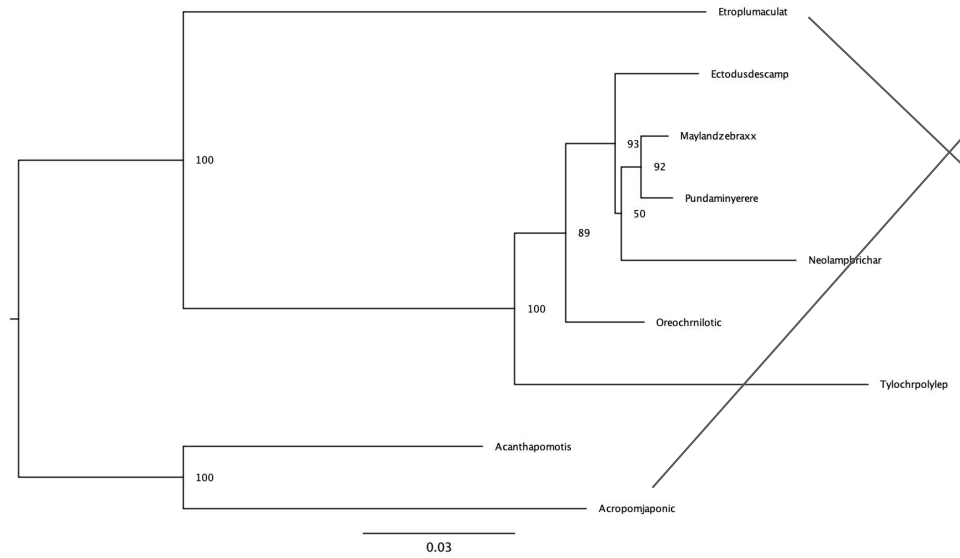
2.2 Alignment with *mafft*

2.3 Trimming with *trimal*

2.4 Concatenation with *catfasta2phym*

2.5 Phylogeny Inference with *raxml*

# Final ML tree



| ID             | Species                         | Group            |
|----------------|---------------------------------|------------------|
| Acanthapomotis | <i>Acantharchus pomotis</i>     | Non-cichlid      |
| Acropomjaponic | <i>Acropoma japonicum</i>       | Non-cichlid      |
| Etroplumaculat | <i>Etroplus maculatus</i>       | Indian cichlids  |
| Maylandzebraxx | <i>Maylandia zebra</i>          | African cichlids |
| Neolampbrichar | <i>Neolamprologus brichardi</i> | African cichlids |
| Oreochrnilotic | <i>Oreochromis niloticus</i>    | African cichlids |
| Pundaminyerere | <i>Pundamia nyererei</i>        | African cichlids |
| Tylochrpolylep | <i>Tylochromis polylepis</i>    | African cichlids |
| Ectodusdescamp | <i>Ectodus descampsii</i>       | African cichlids |