Phylogenomics

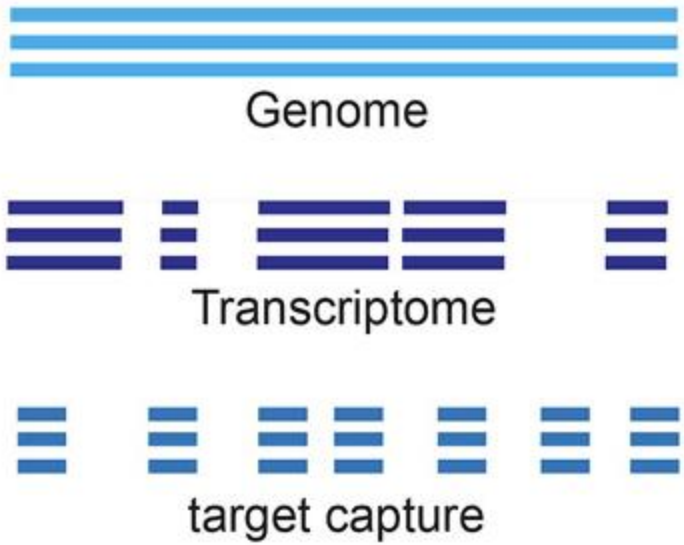
Ruiqi Li, Ph.D.
University of Southern California

Contents

Why phylogenomics?

Genome Tree vs Single Gene Tree

1. Data Type



Whole genome, genome skimming

Exon capture, ultra-conserved element (UCE) capture

Others: SNPs (RAD-seq)

How to choose the “best” method

- Morphology
- Single gene
- Multiple genes
- Target capture
- Whole genome sequencing

Sample availability

Morphology > Single gene

≈ Multiple genes

≈ Target capture

> RNA-seq

> Whole genome sequencing

Budget

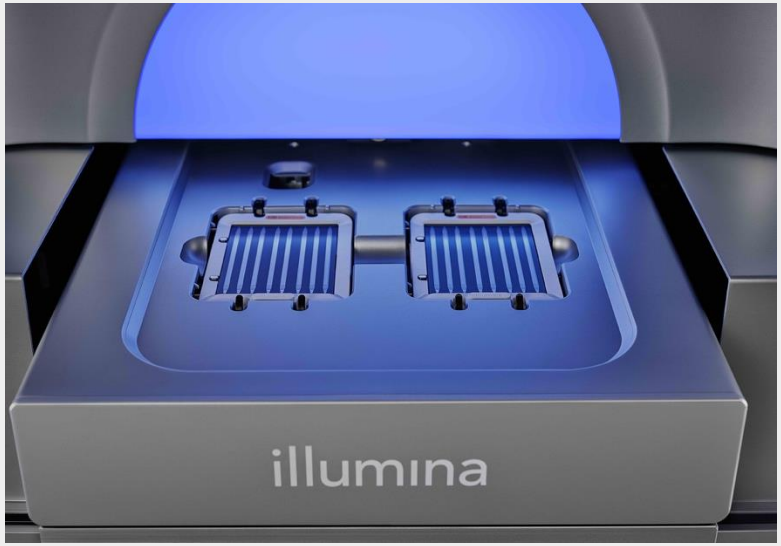
Morphology > Single gene

≈ Multiple genes

≈ Target capture

> RNA-seq

> Whole genome sequencing



Time sensitivity

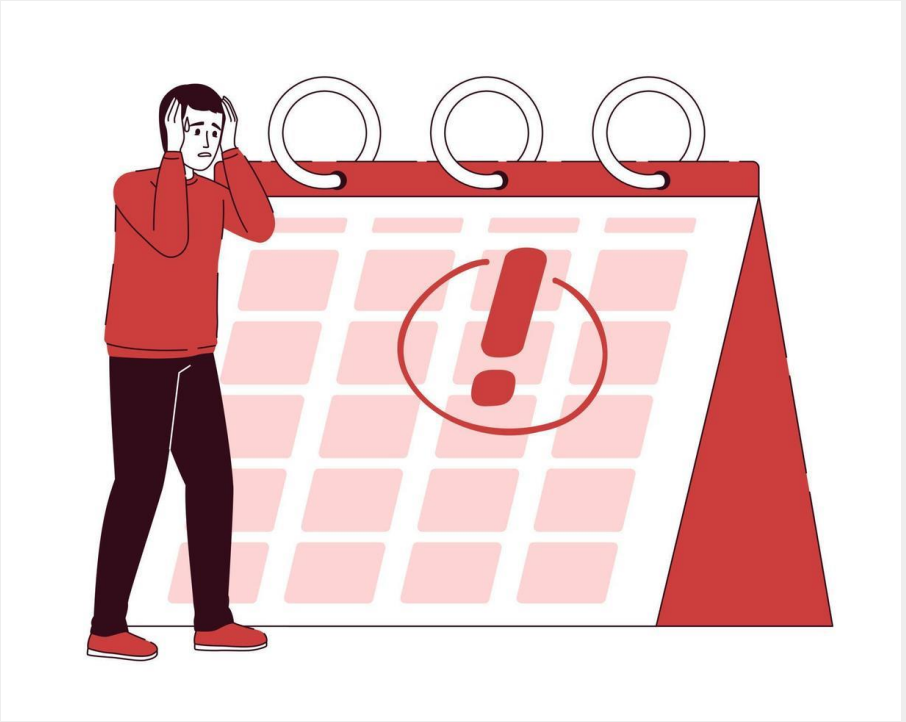
Morphology > Single gene

≈ Multiple genes

> Target capture

≈ RNA-seq

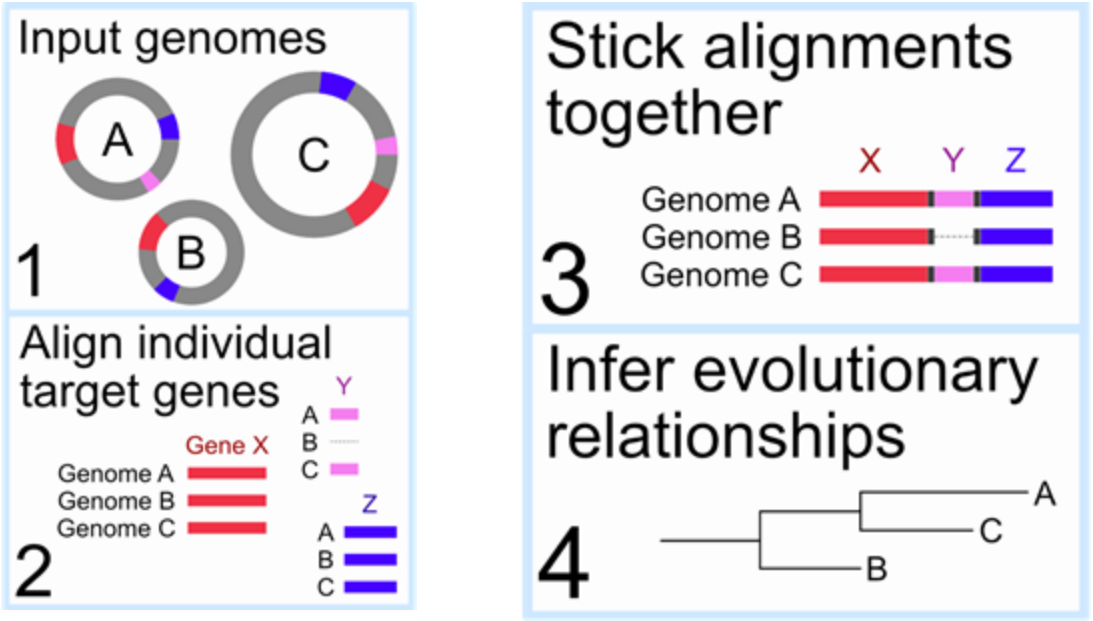
> Whole genome sequencing



Will the data answer your questions?

Morphology < Single gene
< Multiple genes
< Target capture
< RNA-seq
< Whole genome sequencing

2. Phylogenomics Pipeline

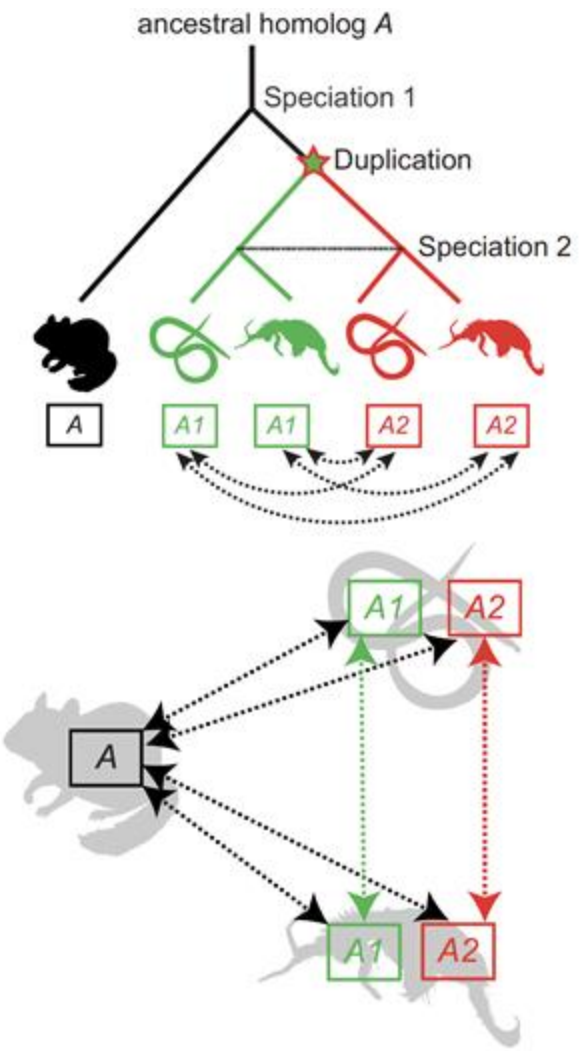


2. Phylogenomics Pipeline

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

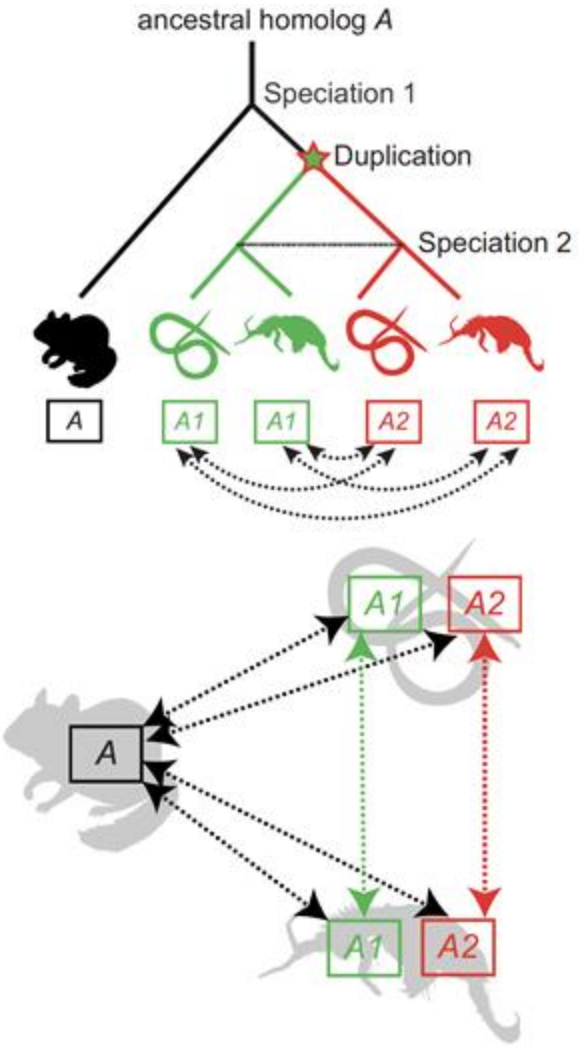
2.1 Finding Orthologs

The green A1s are _____.

A1 and A2 are _____.

A and A1 are _____.

A and A2 are _____.



2. Phylogenomics Pipeline

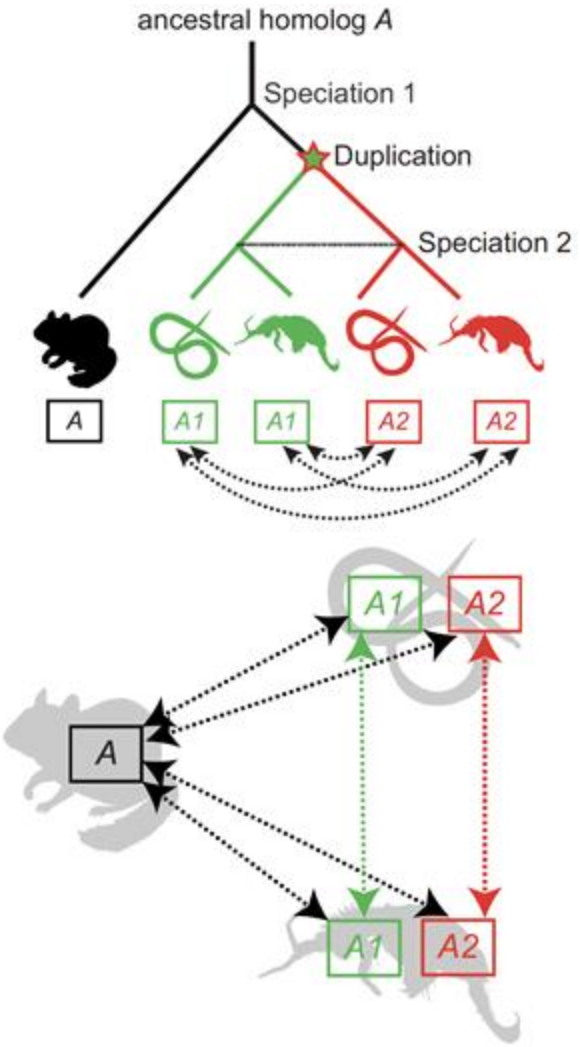
2.1 Finding Orthologs

The green A1s are orthologs.

A1 and A2 are paralogs.

A and A1 are orthologs.

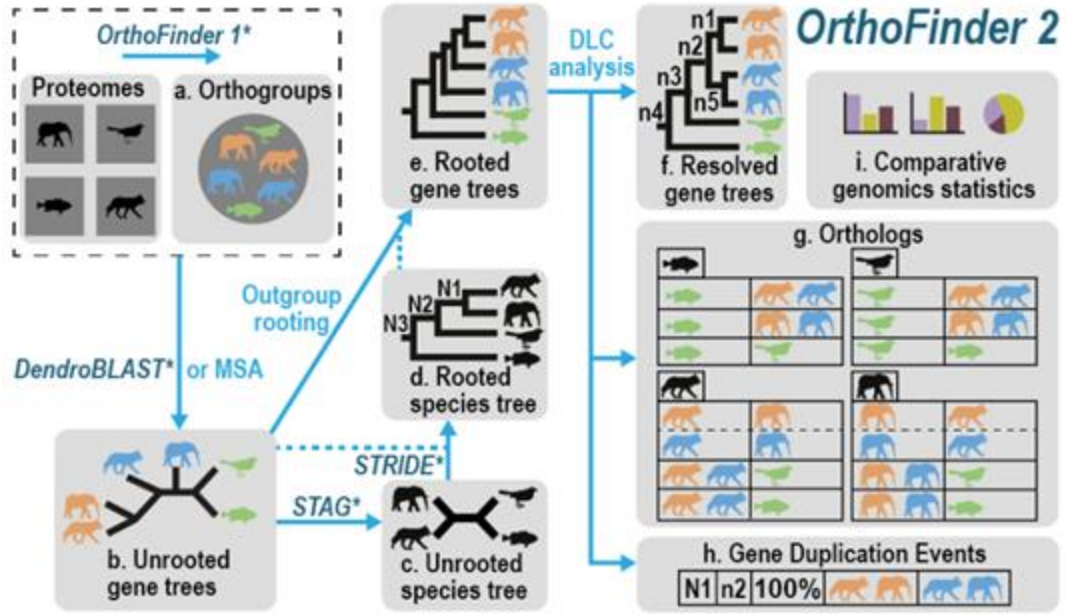
A and A2 are orthologs.



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

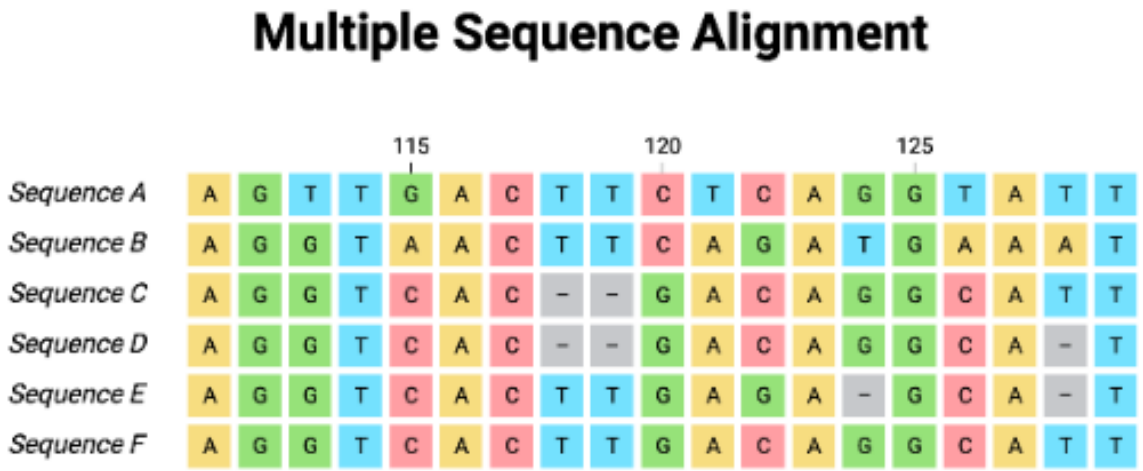
Design the probes with existing genome/transcriptome data

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

2. Phylogenomics Pipeline

2.2 Align individual genes

software: mafft, muscle, etc.



2. Phylogenomics Pipeline

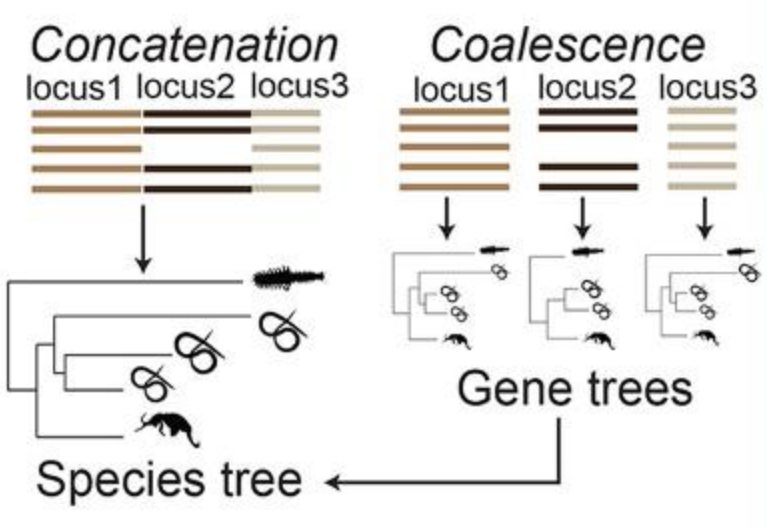
2.3 Trim Alignments

software: trimAL, etc.

2. Phylogenomics Pipeline

2.4 Phylogeny Inference

concatenation and coalescence



Tutorial 1:

Genome/Transcriptome Data - OrthoFinder (coalescence method)

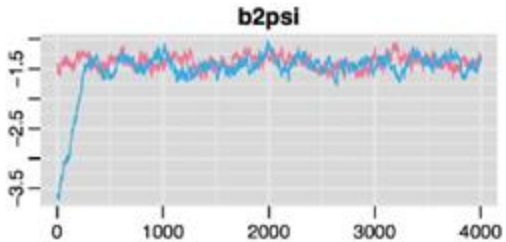
Tutorial 2:

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 P(Data|Tree)	search for the tree that maximizes the chance of seeing the tree given the data P(Tree|Data)
computing	Low	High



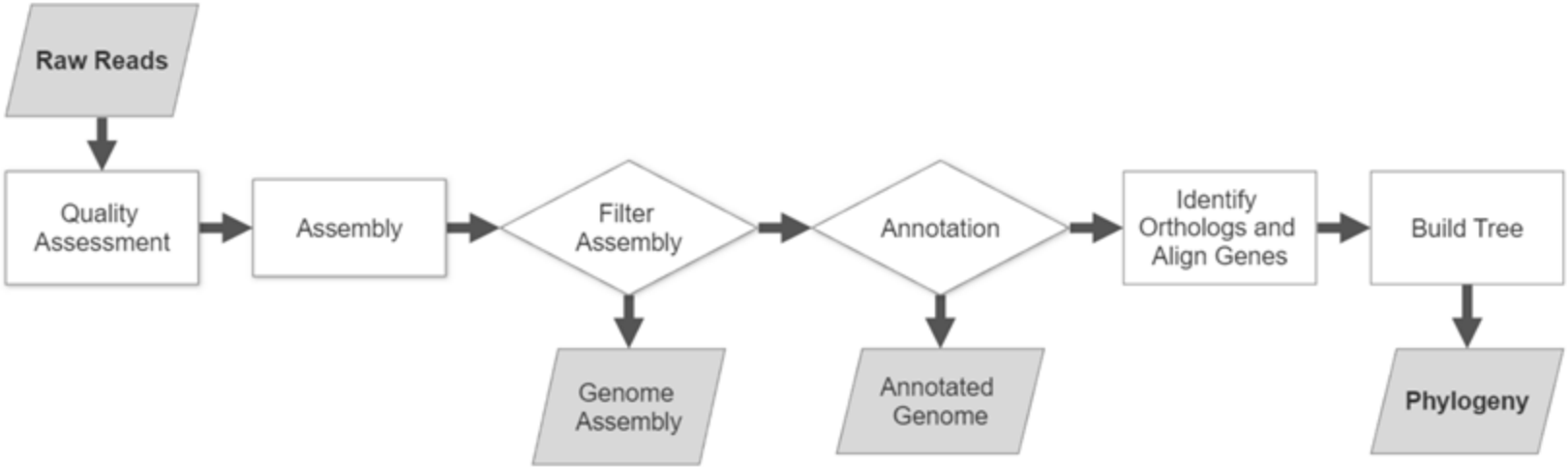
Bayesian MCMC chain convergence

Software Installation

Please follow instructions on the tutorial 4. Software Installation

Practice

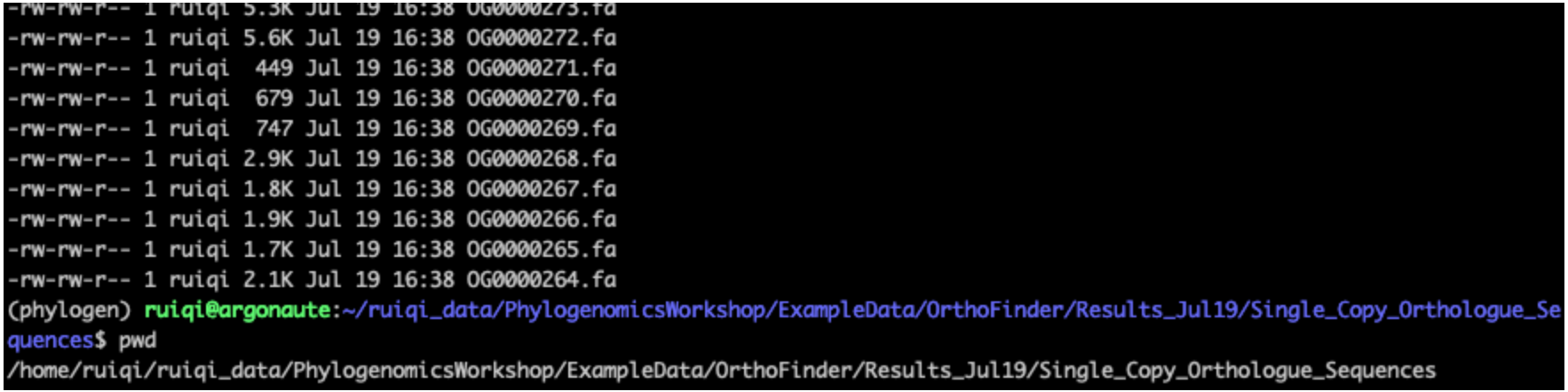
Tutorial 1. Simplified Workflow (Skip this time)



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



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 *catfasta2phyml*

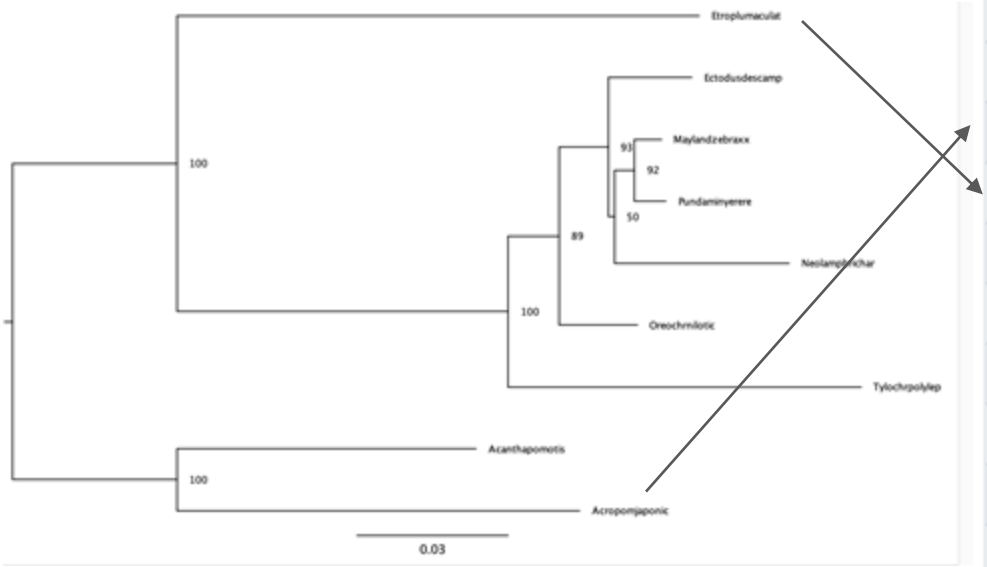
2.5 Phylogeny Inference with *raxml*

Practice

Tree visualization:

https://beta.phylo.io/viewer/#

Final ML tree



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