Building Phylogenetic Trees with phangorn in RStudio

Neighbour Joining & Maximum Likelihood in Comparison

Milena Krennmayr, Jasmin Le, Alina Lisa Marie Riedl

What is a Phylogenetic Tree?

- Represents evolutionary relationships
- Nodes = common ancestors
- Tips/leaves = current species or sequences
- Built from genetic sequence data

Neighbour Joining

- Based on pairwise distances between sequences
- Iteratively joins the closest "neighbors"
- Tree construction minimizes total branch length
- Does not require an explicit evolutionary model
- Fast and scalable for large datasets
- Produces an **unrooted** tree

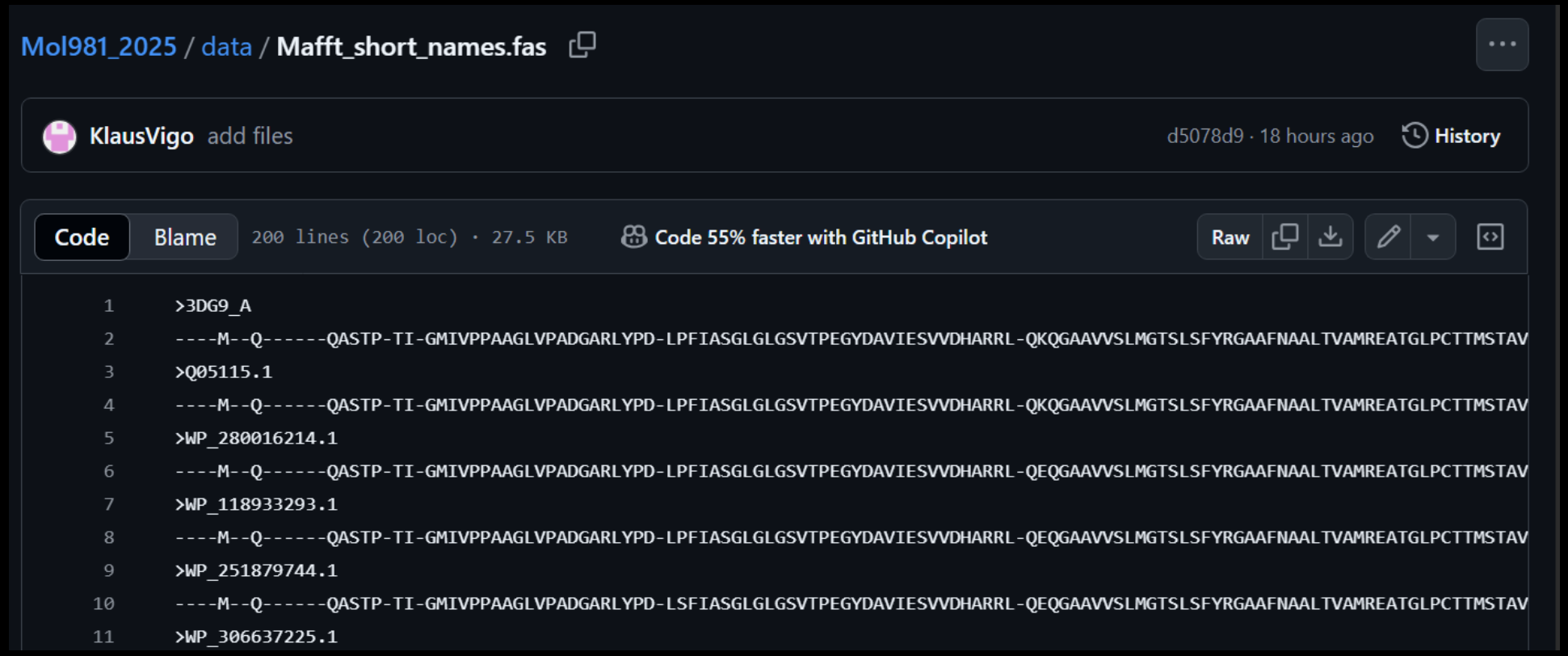
Maximum Likelihood (ML)

- Uses a statistical model of sequence evolution
- Calculates the likelihood of the data for each possible tree
- Searches for the tree with the **highest likelihood**
- Incorporates substitution models (e.g. Jukes-Cantor, GTR)
- More accurate, but computationally expensive
- Can include rate variation (e.g. gamma distribution)
- Produces a **rooted or unrooted** tree, depending on input

Key Differences

	Neighbour Joining (NJ)	**Maximum Likelihood (ML)**
Type	Distance based	Model based (statistical)
Speed	fast	slower
Evolutionary Model	not used	required
Initial Trees	Yes	optimization
Handles rate variation	No	Yes
Output	fixed tree	Tree with likelihood value

Dataset - aligned Arylmalonate decarboxylase Sequences



Summary of Workflow in RStudio

Step	What happens
1-2	Installs and loads libraries
3	Loads a protein sequence alignment
4	Calculates genetic distances between sequences
5	Builds and plots a Neighbor Joining tree
6	Cleans tip labels (needs fixing!)
7	Builds a Maximum Likelihood tree using model testing
8	Cleans tip labels in ML tree (also needs fixing!)
9	Compares both trees for agreement

Thank you for your attention!