Step-by-Step Guide: Introduction to Phylogenetic Analysis with MEGA

Author: Itunuoluwa Isewon PhD

Email: itunu.isewon@covenantuniversity.edu.ng

You have been provided with a fasta file called Aspergillus18S.fasta

Step 1 : Import sequences into MEGA

1. Launch **MEGA X**.
2. **File → Open a File/Session…** → select Aspergillus18S.fasta.
3. When prompted, choose **Align**.
4. MEGA opens the **Alignment Explorer** showing your unaligned sequences.

Step 2 — Multiple Sequence Alignment (ClustalW)

In **Alignment Explorer**:

1. **Align → Align by ClustalW** .
2. Click **Options** and set/confirm the parameters below, then **Compute**.

ClustalW parameters (DNA) — what they mean & good defaults for 18S

Parameter	**What it controls**	**Typical default (MEGA)**	**When to change (18S guidance)**
Gap Opening Penalty	Cost of introducing a new gap in the alignment. Higher = fewer gaps.	15	If sequences are highly divergent, reduce (10–12) to allow more gaps; for conserved 18S regions, leave at default.
Gap Extension	Cost of extending an existing	6.66	Increase slightly (7–8) to discourage long gaps when

Penalty	gap once opened.		aligning 18S, since ribosomal RNA is generally conserved in length.
DNA Weight Matrix	Scoring system for matches/mismatches (e.g., IUB, ClustalW).	IUB (default)	For 18S, leave as IUB (standard). If working with closely related species
Transition Weight	Weighting of transition vs transversion mismatches.	0.5	If you suspect a high transition bias (common in rRNA), increase towards 0.7–0.8; otherwise, leave default.
Delay Divergent Sequences (%)	Excludes highly divergent sequences until later in alignment to avoid bias.	30%	Keep at default for most 18S datasets; reduce to 20% if sequences are very divergent across taxa.
Gap Separation Distance	Minimum separation between gaps to be treated independently.	4	Leave at default for 18S; only adjust if long conserved motifs are split by multiple nearby gaps.
Use Negative Matrix	Whether to penalize divergent positions heavily (rarely used).	Off	Usually leave off for 18S; may be tested if aligning very divergent protist sequences.

After alignment:

- Scroll through; 18S has conserved stems and variable loops—expect gaps mostly in variable regions.
- If the first/last ~10–30 bases are gappy, **select** and **Edit → Delete Selected Sites** (or **Mask**), so they don't add noise.

Save the alignment: **Data → Export Alignment → MEGA format (.meg)** and also **FASTA** for records.

Step 3 — Build a Neighbor-Joining (NJ) tree

1. Close Alignment Explorer (save when prompted).

2. In the main MEGA window, choose **Phylogeny → Construct/Test Neighbor-Joining Tree**.

3. In **Analysis Preferences** (the panel like your screenshot), set:

Analysis

- **Scope** → *All Selected Taxa* (or choose a subset beforehand).
- **Statistical Method** → **Neighbor-joining**.

Phylogeny Test (support)

- **Test of Phylogeny** → **Bootstrap method**.
- **No. of Bootstrap Replications** → **1000**.
  Rule of thumb: ≥70% = moderate support, ≥90% = strong.

Substitution Model (Distances)

- **Substitutions Type** → **Nucleotide**.
- **Model/Method** → **Maximum Composite Likelihood (MCL)**.
  Alternatives: **Tamura-Nei (TN93)** or **Kimura 2-parameter (K2P)**; try these if you want sensitivity analysis.
- **Substitutions to Include** → **d: Transitions + Transversions** (include both).

Rates and Patterns

- **Rates among Sites** → **Uniform Rates** (appropriate for conserved 18S).
  If analyzing more variable loci (e.g., ITS), consider **Gamma Distributed**; MEGA will estimate the shape (α).
- **Pattern among Lineages** → **Same (Homogeneous)**.

Data Subset to Use

- **Gaps/Missing Data** → **Pairwise deletion** (keeps more sites; good for rRNA with localized gaps).
  Complete deletion is stricter but can remove lots of signal.
- **Select Codon Positions** → Tick **Noncoding Sites** (coding positions have no effect for 18S).

4. Click **Compute** to infer the NJ tree.

Step 4 — Inspect, root, and annotate

- **Rooting**: If your file contains a clear outgroup (e.g., a non-Aspergillus fungus), **Tree Explorer → Tree → Root** on that taxon. If not, use **Midpoint rooting**.
- **Show bootstrap values**: In Tree Explorer, **View → Display Option → Show → Bootstrap values**.
- **Tidy labels**: **Edit → Replace tip labels** can shorten long headers; you kept accession+species, which is perfect.
- **Collapse weak nodes**: Optionally collapse branches with support <50–70% to simplify.

Step 5 — Export & document

- **File → Export Current Tree (Newick)** → Aspergillus18S_NJ.nwk.
- **Image export**: **File → Export Image** (PNG/PDF/SVG) at 300–600 dpi.
- Save your **.meg** project so you can reopen without re-aligning.

Now try to repeat this process using the yeast18S.fasta