

03__ancestral_states

September 2, 2025

```
[10]: #Need ancseq kernel

from pathlib import Path

#Define paths for current project
# --- Centralized paths ---
ROOT = Path("..")
DATA = ROOT / "data"
LOGS = ROOT / "logs"
SCRIPTS = ROOT / "scripts"
RESULTS = ROOT / "results"
ALIGN_DIR = RESULTS / "align"
TREE_DIR = RESULTS / "trees"
FIGURES = RESULTS / "figures"
ANCESTORS = RESULTS / "ancestors"
```

Note:

For long-running jobs (like **ancseq**), do **not** run them in a Jupyter notebook if you are worried about losing your connection.

Instead, use a terminal multiplexer such as **tmux** or **screen** in a regular terminal:

1. Open a terminal.
2. Start a tmux session: **tmux**
3. Activate your environment and run your command:

```
conda activate ancseq
ancseq -s ../results/align/combined_plus_hasegawa24_ALN.fasta -m AA -o ../results/ancestors
```

4. Detach from tmux with **Ctrl+b** then **d**.

This ensures your job continues running even if you disconnect.

```
[11]: from Bio import SeqIO

# Paths
HITS = ALIGN_DIR / "pumphits_ALN.fasta"
ANCESTOR_FASTA = ANCESTORS / "cph24/30_result/ancestral_state_result.fasta"
FINAL = DATA / "final_with_ancestors.fasta"
```

```

# Ancestor node names
Ancestors = ("Node36", "Node73", "Node37", "Node6",
             "Node87", "Node4", "Node117", "Node38",
             "Node40", "Node44")

# Read original alignment
aln_records = list(SeqIO.parse(HITS, "fasta"))

# Read ancestor sequences
ancestor_records = [rec for rec in SeqIO.parse(ANCESTOR_FASTA, "fasta") if rec.
                    ↪id in Ancestors]

# Concatenate and write to FINAL
with open(FINAL, "w") as out_handle:
    SeqIO.write(aln_records + ancestor_records, out_handle, "fasta")

print(f"Combined alignment and ancestor sequences written to {FINAL}")

# Align with MAFFT
ALN_FASTA = ALIGN_DIR / "final_with_ancestors_ALN.fasta"
!mafft --auto --thread -1 --quiet "{FINAL}" > "{ALN_FASTA}"
print("Aligned ->", ALN_FASTA)

```

```

Combined alignment and ancestor sequences written to
../data/final_with_ancestors.fasta
Aligned -> ../results/align/final_with_ancestors_ALN.fasta

```