Local Phylogenetics with MAFFT, PhyML, Seqret, and IQ-TREE using Conda

This document describes a cross-platform procedure to set up a reproducible environment for multiple sequence alignment, format conversion, and maximum-likelihood phylogenetic inference using MAFFT, Seqret, PhyML, and IQ-TREE.

---

1. Install Miniconda

Obtain the Miniconda installer from:

https://docs.conda.io/en/latest/miniconda.html

Install with default settings, then restart the terminal.

---

2. Verify Conda availability

```
conda --version
```

---

3. Configure Conda channels

```
conda config --add channels defaults
conda config --add channels conda-forge
conda config --add channels bioconda
```

---

4. Create an environment including MAFFT, PhyML, and EMBOSS (Seqret)

```
conda create -n phyloenv mafft phyml emboss
conda activate phyloenv
```

---

5. Prepare input sequences

Place unaligned protein sequences in a FASTA file named:

```
sequences.fasta
```

---

6. Produce a multiple sequence alignment using MAFFT

Standard mode:

```
mafft --auto sequences.fasta > alignment.fasta
```

---

7. Convert FASTA alignment to PHYLIP using Seqret (strict formatting)

```
seqret -sequence alignment.fasta -outseq alignment.phy -osformat2 phylip
```

Seqret may not insert the required spacing between the sequence name and the aligned sequence. Fix this using:

```
sed -E 's/^(.{10})([^ ])/\1 \2/' alignment.phy > alignment_spaced.phy
```

---

8. Infer a maximum-likelihood phylogenetic tree using PhyML (with optional bootstrap)

Tree without bootstrap:

```
phyml -i alignment_spaced.phy -d aa -m LG
```

Optional bootstrap analysis (very slow):

```
phyml -i alignment_spaced.phy -d aa -m LG -b 100
```

---

9. Install IQ-TREE (optional)

```
conda install -c bioconda iqtree
```

Run IQ-TREE:

```
iqtree -s alignment.fasta -m LG -bb 1000 -alrt 1000
```