**Phylogenetic divergence dating practical:**

**When did the Galápagos flycatcher colonize the Galápagos?**



In this short practical we are going to estimate the time of colonisation of the Galápagos islands by the ancestors of the endemic Galápagos flycatcher (*Myiarchus magnirostris*). To do this, we will reconstruct and date a phylogeny of genus *Myiarchus*, including multiple individuals of *Myiarchus magnirostris* as well its most closely related species from outside the Galápagos. For this practical you will need the following software: Geneious; BEAST 2 package; Tracer and FigTree.

We will use mitochondrial cytochrome B (CytB) and ND2 sequences from Sari & Parker 2012 (*Molecular Phylogenetics and Evolution*, 63: 244-254).

1. **Download sequences from Genbank**. Use Geneious (or your preferred software to deal with DNA sequences) to download CytB and ND2 sequences from Sari & Parker 2012 from Genbank.

In the NCBI search tool, type the following text to download accessions JQ004294 to JQ004346 (Cytochrome B sequences, Table 1 in Sari & Parker 2012)

JQ004294:JQ004346[accn]

Create a new folder called ‘CytB’ and drag the 53 CytB sequences to it.

Then download the ND2 sequences:

JQ004347:JQ004396[accn]

Create a folder ‘ND2’ to place the 50 ND2 sequences.

1. **Batch rename sequences.** Rename the sequence files using the information contained within the Genbank files. This ensures that the CytB and ND2 sequences have exactly the same names, allowing us to concatenate the alignments by putting together the sequences from the same individuals.

Select the sequences and use the Batch rename function under ‘Edit’ in Geneious. Follow this scheme:

Organism\_isolate\_specimenvoucher

Use the batch rename function to remove all hyphens (replacing ’–‘ by ‘’) and to replace all spaces with underscore (replacing ‘ ‘ with ‘\_’)

**Important:** The taxon names of two ND2 sequences on Genbank do not match those reported in the paper (the submitters may have made a mistake). These are the accessions JQ004370 and JQ004371. Once you are done with the batch rename, change the name *sagrae* to *stolidus* in these two ND2 sequences. **Lesson: Never fully trust Genbank!**

1. **Create alignments** for the ND2 and the CytB sequences (2 separate alignments). Select all sequences from each gene and use the ‘Muscle’ alignment function in Geneious (Tools: Align: Multiple Alignment: Muscle Alignment). This will create an alignment file within the folder where the sequences are located. The sequences and the alignment file are linked, which means changes you make directly in the sequences will automatically be made also to the alignment.
2. **Concatenate the alignments.** Drag the two alignments to the same folder. Use the concatenate sequence function (in Tools) in Geneious. Name your alignment ‘Myiarchus\_concatenated’. The concatenated alignment should have 58 sequences and 2010 nucleotide positions.
3. **Export the alignment:** Export as Nexus (\*.nex)
4. **Import the Nexus File into software BEAUti 2.**
5. **Set up Beast 2 XML file.** Select the following options in BEAUti 2:
   1. Site Model Tab: We’re going to use the HKY+I+G model for both genes (normally you would first find the best model for each partition using a software like jModeltest). Select Estimate substitution rate; 4 Gamma rate categories; Estimate Shape (Leave at default 1.0); Proportion of Invariant sites: 0.1; Estimate Proportion Invariant; HKY model; Estimate Kappa.
   2. Clock Model Tab: Select Strict Clock; Clock.rate: 0.01035 (This is equivalent to 2.07% pairwise divergence following Weir & Schluter *Mol Ecol*. 2008. 2.07/2/100=0.01035 per lineage rate)
   3. Priors Tab: Set Birth Death Model as tree prior. Leave other prior options with the default values.
   4. MCMC tab: Set chain length: 10 000 000; Tracelog file name: Myiarchus.log
6. **Export and save the XML file.** In File: Save as file “Myiarchus.xml”
7. **Run XML file in BEAST 2.** Open BEAST 2 and load Myiarchus.xml file that was created in BEAUti 2. This will run the MCMC chain – this will take a couple of minutes.

Note: Normally you should run multiple independent chains and combine the results from several runs using LogCombiner but for practical purposes we will only run 1.

1. **Check for convergence of MCMC run.** Once the BEAST run is completed load the Beast output ‘Myiarchus.log’ file to Tracer.
2. **Generate Maximum clade credibility tree.** Load Beast output ‘Myiarchus.trees’ file to TreeAnnotator to summarize the results of the posterior distribution of trees into a single tree (Maximum clade credibility tree) Select:
   1. Target tree type: ‘Maximum clade credibility’;
   2. Burn-in percentage: 20%
   3. Node heights: ‘Mean heights”
   4. Output file: “Myiarchus.tre’
3. **Open your maximum credibility tree in Figtree.** Here you can visualize the tree topology, the posterior probability of different nodes and the node ages.