**Workflow for reanalysis of UCE phylogeny for Hirundinidae**

Part 7a – Notes on pilot maximum likelihood analysis in RAxML

I ran a pilot concatenated analysis in RAxML using 100 randomly sampled UCEs with 95% complete data matrices using 122 taxon dataset (tissues + 21 skin samples). I wanted to be able to compare the results to Clare’s pilot tree detailed in her ‘7g-excluded-samples.docx’ file (also see Fig. 1).

The results are shown in Fig. 2, where skin sample branches are highlighted in red.

The results are remarkably consistent with Clare’s previous inference. The two most notable features are the incorrect placement of *Progne modesta* in the outgroup clade and *Riparia congica* within the *Progne* clade, and the persistence of long branch lengths for the skin samples. It does seem that the branch lengths are relatively shorter, so the correction step may have helped somewhat. Still, if (some of) the skin samples are to remain in the dataset, we’ll need to adjust the terminal branch lengths using the means of close relatives, as before.

At the equivalent previous step, Clare ended up removing skin samples with an initial DNA concentration < 20μL, as the concentration scaled reliably with UCE median length (positive correlation) and terminal branch lengths (negative correlation). The exception was keeping the *Pseudochelidon* samples, since these fall out in the right place and are generally very important taxa in the tree.

**This would remove the following samples (leaving 12 remaining skin samples):**

**Progne\_cryptoleuca\_Cuba\_LSUMZ\_142940\***

Hirundo\_nigrorufa\_Angola\_AMNH\_SKIN\_707943

**Progne\_cryptoleuca\_Cuba\_LSUMZ\_142944\***

Petrochelidon\_rufigula\_Angola\_AMNH\_SKIN\_707945

**Riparia\_congica\_Congo\_FMNH\_213543\***

Ptyonoprogne\_obsoleta\_Egypt\_UMMZ\_224076

Hirundo\_megaensis\_Ethiopia\_AMNH\_SKIN\_348889

**Progne\_modesta\_Ecuador\_MVZ\_130129\***

Progne\_sinaloae\_Mexico\_KU\_40045

**However,** only the samples with the asterisks have low DNA concentrations AND low numbers of assembled UCEs; the exception is *Progne modesta*, which has an intermediate DNA concentration and many UCEs, but consistently comes out in the outgroup clade, suggesting contamination is an issue. The remaining samples come out in reasonable places in the phylogeny, so I am going to try another pilot analysis with these samples included (see ‘README-5-format-UCE-data-matrix.md’ for details).



Figure. 1. Results from previous concatenated RAxML analysis, with skin samples highlighted in red, indicating long branches for these samples and incorrect placement of *Progne modesta* and *Riparia congica*.



Figure 2. Results from current pilot analysis of the full dataset (skin samples highlighted in red) after the phyluce correction step for skin samples. These results very closely resemble Clare’s previous inference, with the incorrect placement of *P. modesta* and *R. congica* and long branch lengths for most skin samples.