BLAST Workflow 2015-09-03

1. Clean/assemble raw sequences in Geneious. Color code sporophytes sequences as follows:

Blue = single best Moorea species for ID

Pink = single best Tahiti species for ID

Black (no color) = good quality, but don’t use for ID

Green = outgroup species

Red = bad seq

2a. Export clean sporophyte sequences from Geneious as CSV file (*marker*\_clean\_*date*.csv) to R/moorea/data. When exporting, include the following fields: name, sequence, color.

2b. Export clean gametophyte sequences from Geneious as FASTA file (*marker*\_clean\_gametos­\_*date*.fasta) to Analysis/moorea/data. Search and replace any ? with N in FASTA file.

3. Run **csv\_to\_fasta.R** to convert CSV file (*marker*\_clean\_*date*.csv) to FASTA file (*marker*\_*date*.fasta), copy to Analysis/moorea/data.

4. Run **blast.sh** to build local blast library, search gametophyte sequences against it. Copy output (blast\_out\_*marker*\_id\_*date*.txt) to R/moorea/data. Blast output is in tab-delimited table form (outfmt 6), includes top two hits for each query.

(optional: also query library against itself for quality check)

5. Run **blast\_ID.R** on BLAST output to resolve species identifications, remove duplicates, merge with collection data, and produce species x plot matrix.