GENEIOUS SEQUENCING ALIGNMENT, CLEANING, AND TREE BUILDING

1. <u>Import sequences</u>
   a. Open geneious and import sequences to a "raw sequences" folder.
   b. Check the percentage of untrimmed bases from each sequence (%HQ) > 30% is good, >50% is optimal.
2. <u>Alignment *de novo* to generate contigs (set of DNA segments or sequences that overlap in a way that provides a contiguous representation of a genomic region).</u>
   a. Select pairs of sequences to align
   b. Click on "align/assemble"
   c. Click on "de novo assemble"
      i. Options Data box -> Assemble by "1st" part of name, separated by "_(Underscore)", or choose as appropriate. Keep all other options as default.
3. <u>Save aligned sequences into folder "aligned sequences"</u>
4. <u>Clean aligned sequences by:</u>
   a. Delete regions with pink bars if "small". If pink regions are too large check and assess whether delete a large section is feasible.
      i. Select "allow editing" to delete pink sections
      ii. If pink trim regions are not annotated already, select "annotate & predict" next to "allow editing"
      iii. Select "trim ends" keep default values -> then delete pink trim regions
   b. Check gaps and ambiguities
      i. On the right icon tabs, select the monitor icon -> highlighting -> drop down select gaps/ambiguities -> go <> "in any sequence"
      ii. On ambiguities or gaps highlighted sections, check line graph with highest peak (Red-Adenine, Blue-Cytosine, Green-Thymine, Yellow-Guanine) to make informed decisions about base changes/filling gaps.
      iii. Save. Check "NO" when asked "*Do you wish to apply changes to the original (unaligned) sequences?*
      If you choose no, this alignment will permanently discard references to the unaligned sequences and you will no longer be able to apply changes to them"
      iv. Copy and paste clean contigs into a new folder named "clean contigs".
5. Perform a multiple alignment with all clean contigs
   a. Click on align/assemble -> multiple align -> consensus align -> create alignment of consensus sequences only
   b. Copy clean contigs alignment into a new folder "alignment genebank"
6. Perform multiple alignment with genebank sequences
   a. Download sequences from genebank (find relevant sequences from the literature) into the "alignment genebank" folder
   b. Select genebank sequences and clean contig alignment
   c. Click on align/assemble -> consensus align -> create alignment of all sequences-> check box "automatically determine direction"

      d.   Trim aligned sequences into smallest fragment
7.   Copy genebank alignment into new folder "tree"