**Completed so far**

* Sequence filtering, QC, variant calling and consensus file generation
  + Max-missing=50%, max-meanDP=60, min-meanDP=3, minDP=2
* 10kb sliding windows
  + Gene tree inference
  + ASTRAL species tree inference

**In progress/planned**

* 10kb sliding windows
  + Concatenated ML tree
  + Discordance metrics & targeted discordance summaries
    - RF distance
    - ASTRAL annotations
    - gCF and sCF
    - gene-wise log-likelihoods (Shen et al. 2021)
    - Quartet sampling (Pease et al. 2018)
* Scaffold/scaffold analysis
  + D statistics, including Dfoil (MVFtools)
  + Pixy
  + Histogram of genic regions along chromosomes
* Functional annotations

*Bams*

* + Pull mapped sequences in genic regions and generate *de novo* (?) sequences for these genes from the bams. Then realign on a gene-by-gene basis. This helps identify actual indels and generates gene sequence files. (samtools consensus? – simple basecalling -- <https://github.com/samtools/bcftools/issues/1459>)

**Measuring discordance among gene trees**

**NOTE – that incongruence statistics can be done for a specific triplet topology. This can still be confusing when there are aberrant individuals, but nonetheless…**

1. -t ASTRAL branch annotations. And/or gCFS annotations in IQtree.
2. Quartet sampling (pease et al. 2018)
3. Robinson-Foulds distance, can be done in ete3
4. Gene-wise log likelihoods for alternate topologies (how strongly does gene support topology of interest?) -> Shen et al. 2021