**Variant calling and file generation**

* Call variants without indels. Generate consensus fastas from these.

*Fastas*

* + Estimate sliding window trees
    - ~~Generate Astral species tree~~
    - RF distance of each FULL gene tree to the species tree
    - Cloudogram (DensiTree)
    - Gene tree subsampling and topology/concordance inferences
      * Subsampling? Pease et al. 2016
  + Generate concatenated IQtree species tree
  + Read into MVFtools and do D statistic stuff

*VCFs*

* + Pixy

*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>)

*Genome*

* + Histogram of genic regions along chromosomes
  + Functional annotations

**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