NbdLamprey summary:

-Note: this is just post-genotyping scripts, bioinformatics for genotyping is a separate pipeline

Big note about the data included: only 20,000 randomly selected SNPs are included as a test data set. The full data set with all mapped reads is included on NCBI. The pipeline for processing reads has been previously published by Sard et al 2020, and is summarized in a different post.

Definitions that should be included in the read me:

* Collection
* Location
* Cohort
* Cluster
* COLONY cluster
* SNP set

Scripts:

1. Filter reads and calculate summary statistics
2. COLONY SNP sets for each location
3. Length-based aging models for each collection
   1. Maximum Likelihood models & Bayesian models
   2. Combining aging models and reconstructed pedigrees for locations to generate cohorts
4. COLONY and NeEstimator sets for each cohort (pull COLONY sets from script 2 and just subset it into cohorts)
5. Nb/Ns calculations for each cohort
   1. PwoP calculations
   2. Ns calculations
6. Figure/Table scripts (separate folder?)
   1. Map of all locations
   2. Length-frequency histograms
   3. Boxplots of length distributions sorted by COLONY clusters
   4. Reconstructed pedigree visualizations
   5. Ns visualizations

Question to self: this is a lot of scripts and a lot of functions. Should this get broken up a different way? Also, is this the kind of repository that I could build into an R package?

Scripts that I’m not sure where to put (a different repository? Supplemental scripts?):

* Sample size experiments
  + Two scripts: one for creating pops and one for post-processing
  + Script for associated figure
* Variance and uncertainty testing
* ANOVA for brook lamprey