Chimp Pop Gen

Tram Nguyen

1/22/2020

## Chimp Population Genetics Anaylsis

### Getting a STRUCTURE Plot

Step 1. Sometimes it is useful to generate a pruned subset of SNPs that are in approximate linkage equilibrium with each other. This can be achieved via two commands: –indep which prunes based on the variance inflation factor (VIF), which recursively removes SNPs within a sliding window; second, –indep-pairwise which is similar, except it is based only on pairwise genotypic correlation.

Step 2. Convert VCF files to .bed .bim .fam files using PLINK

/programs/plink-1.9-x86\_64-beta3.46/plink --vcf joint\_genotype\_filtered\_SNPs\_only.vcf --allow-extra-chr --out joint\_filtered\_SNPs\_only

Step 2. Choose your model complexity – Run python script (provided by FastStructure) to figure out range of possible K values.

python chooseK.py --input=join\_filtered\_SNPs\_only

###Calculating inbreeding coefficient in Plink

1. IBC analyses are compared to expected allele frequencies in a panmictic non-inbred population, therefore we need to read in a reference set of allele frequences that we can compare our observed individuals to. I made the reference alleles with the combination of all the chimps together (including Pardos paper).

#make bfiles for a combined set of chimps  
/programs/plink-1.9-x86\_64-beta3.46/plink --vcf joint\_genotype\_filtered\_SNPs\_only\_merge\_PradoMartinez.vcf --allow-extra-chr --out all\_snps\_combined  
  
/programs/plink-1.9-x86\_64-beta3.46/plink --bfile SouthDemo.b.gt.22Oct17 --allow-no-sex --nonfounders --dog --freqx --mind 0.05 --out SouthDemo\_b\_ref\_allele\_freq\_22Oct17