To understand the allele frequency difference between the time points 12 and 18 we extracted the time points for each generation using vcftools and placed the subsets into new vcf files. In this example we extacted all generations at time point eighteen and used the recode option to place the output into another vcf file.

vcftools --gzvcf Combined.Q30.recode.vcf.gz --indv "YEE\\_0112\\_03\\_02\\_18" --indv "YEE\\_0112\\_03\\_03\\_18" --indv "YEE\\_0112\\_03\\_05\\_18" --indv "YEE\\_0112\\_03\\_07\\_18" --indv "YEE\\_0112\\_03\\_10\\_18" --recode --recode-INFO-all --out ra.g18

We later determined we needed to extract each generation and time point to a separate file. However, these vcf files were utilized in determining the depth in the below code.

vcftools --vcf ra.g18.recode.vcf --depth --out ra.g18

The vcf files were also utilized to produce stats files using vcf-stats. Prior to this, the files were compressed using bgzip and indexed using tabix.

bgzip ra.recode.vcf

tabix -p vcf ra.g18.recode.vcf.gz

vcf-stats ra.g18.recode.vcf.gz >> ra.g18.stats.txt

In order to compare the allele frequency, we utilized vcf-compare. Samples were extracted utilizing the freq option and then compared as you can see in this example.

vcftools --gzvcf Combined.Q30.recode.vcf.gz --indv "YEE\\_0112\\_03\\_02\\_18" --indv "YEE\\_0112\\_03\\_03\\_18" --indv "YEE\\_0112\\_03\\_05\\_18" --indv "YEE\\_0112\\_03\\_07\\_18" --indv "YEE\\_0112\\_03\\_10\\_18" --freq --out ra.g18

vcf-compare ra.g12.frq ra.g18.frq >> compare.freq.txt

Samples were extracted again using the vcftool option freq that gives us an output file showing the number and areas of allele frequency. Here is an example of this code option:

vcftools --gzvcf Combined.Q30.recode.vcf.gz --indv "YEE\\_0112\\_03\\_02\\_18" --freq --out r1.g18

Even though frequency was able to be determined, the format of the .frq file did not allow to add the true events of allele frequency. We were only able to count the amounts of alleles in that region. These values were the same for both generation, however, the number of allele events were not the same between the two time points. Due to this, we were given a file in order to determine these counts and produce subsequent plots.