

1) Take the human genera (Homo) and one other genera in your vcf file (vcf: <https://www.dropbox.com/s/f0ik4a9bi9atsb2/greatapes.fixedchr21.vcf.gz> tabix: <https://www.dropbox.com/s/tlbyaqo3w89li2o/greatapes.fixedchr21.vcf.gz.tbi> bam: https://www.dropbox.com/s/8bv5tme8c2jlgr6/Intersect_filtered_cov8_chr21_rand1000.bed) and determine the genetic diversity within each genera (Use Watterson's estimator (S) and Nei and Li's estimator (π)). Which of your two genera is more genetically diverse? Is this surprising? Why or why not?
SOME CAVEATS:

- a) Our individuals are heterozygote, meaning some of them have two alleles at a site. b) To make the analysis simpler, we will randomly choose the reference or the alternate allele at each heterozygote site, effectively treating each individual as haploid.
- c) ./ usually means the genotype is missing, but since this vcf file is merged from separate vcf files for each primate genera, ./ indicates that this site matched the reference allele for all individuals in that genera.
- d) We assume that all sites are biallelic. Thus, we skip any sites that are not biallelic. e) How would you figure this out?
- f) To not worry about effect of sample size, randomly choose the same set of samples from each genera (at least five).