WHOLE EXOME VARIANT AND INDIVIDUAL FILTERING

0. INITIAL VARIANT FILTERS (VCFtools/0.1.14)

0.1 Exclude non-PASS variants.
0.2 Exclude Indels
0.3 Exclude X,Y chromosomes
0.4 Exclude non-biallelic sites
0.5 Exclude fixed sites.

1. VCF QUALITY CONTROL (VCFtools/0.1.14)

1.1 Depth of Coverage (per Individual / per Site)
1.2 Missingness (per Individual / per Site)
1.3 GQ (per Individual / per Site)
1.4 Ts/Tv ratio

2. VARIANT AND INDIVIDUAL FILTERING (VCFtools/0.1.14)

2.1. Exclude sites with DP < 5X, GQ < 20 and missingness > 5%.
2.2. Exclude individuals with missingness > 5%, heterozygosity > 4SD, coverage <40.
2.3. Filter SNPs in HW disequilibrium pvalue 10^{-3} (vcftools per Population vs PCAngsd)

3. RELATEDNESS ESTIMATION (PLINK/1.9b; KING)

3.1. MAF 0.01
3.2. Linkage disequilibrium pruning (50 5 0.5)
3.3. Estimate relatedness (KING vs PLINK)
3.4. Remove related individuals (the one with less missing sites)