ROUTE MAP FOR GENOTYPE CALLING AND POPULATION-GENOMIC ANALYSES

https://github.com/mennodejong1986/WGS_data_analyses

sample1_1..fq.gz + sample1_2.fq.gz
sample2_1.fq.gz + sample2_2.fq.gz

FASTQ_filter_fastp.sh

sample1.QC.1..fq.gz + sample1.QC_2.fq.gz
sample2.QC._1.fq.gz + sample2.QC_2.fq.gz

FASTQ2BAM_bwa_HIGHWAY.sh

sample1.sorted.RG.dupremoved.filtered.bam
sample2.sorted.RG.dupremoved.filtered.bam

BAM_extractregion.sh

sample1.xchrom.bam
sample2.xchrom.bam

include all sites (polymorphic AND monomorphic)
include all polymorphic sites (biallelic AND multiallelic)
include only biallelic polymorphic sites
depreciated/redundant

STARTING POINT WHEN REDOING ANALYSES

BAM2VCF_bcftools_HIGHWAY.sh

runpipeline (maskdepth=?)

combinevcf → PREFIX.masked.vcf.gz

masked_PREFIX.mybed1.txt.vcf.gz — *calcdepth* → sampleinfo.DP.txt

BAM2VCF_plotdepth.inR.txt

Retained: 94.6 %
Discarded: 5.4 %

globalfilter (min_alleles=?, snp_quality=?, removeindels=TRUE, mindepth=?,maxdepth=?)

PREFIX.globalfilter.thin0.mybed1.txt.vcf.gz — *combinevcf2* → PREFIX.globalfilter.thin0.vcf.gz

VCF_darwindow.sh
VCF_darwindow.plotinR.txt

→ mywindow.20000.he.allsites.roh.txt

thinallsites (thinallbp=??)

combinethin → PREFIX.globalfilter.thin??.vcf.gz

PREFIX.globalfilter.thin??.mybed1.txt.vcf.gz

selectvariants(biallelic=FALSE)
selectvariants(biallelic=TRUE)

CONVERT_VCF2phylip.sh
Create input fur ML tree building
→ PREFIX.mysnps.thin.phy

PREFIX.mysnps.thin??.mybed1.txt.vcf.gz — *combinevcf3* → PREFIX.mysnps.thin??.vcf.gz

VCF_annotate_SNPs.sh
For McDonald-Kreitman test
→ PREFIX.annotated.vcf.gz

missfilter (minalleles)

PREFIX.mysnps.thin??.missfilter.mybed1.txt.vcf.gz

VCF_calcdist.sh
VCF_calcdist_parallel.sh
→ allvcfdist.txt

combinevcf4

PREFIX.mysnps.thin??.missfilter.vcf.gz

thinsnps (thinsnpbp=?????)

SAMBAR.txt

add2inds2()

PREFIX.mysnps.thin??_?????.missfilter.vcf.gz — *convert2ped* → PREFIX.mysnps.ped — *importdata()* → genlight, snps, inds, inds2

VCF_easySFS.sh
SFS-based analyses

VCF_slidingwindow.sh
Genomic landscape of Dxy, pi and Fst along chromosomes

VCF_Dsuite
ABBA-BABA analyses