# Host-Parasite

# SNP Calling Pipeline

## Requirements

The following modules/software are required to run the pipeline:

- \*\*FastQC\*\*

- \*\*cutadapt\*\*

- \*\*Stacks (version 2.67)\*\*

- \*\*VCFtools (version 0.1.15)\*\*

## Pipeline Steps

1. \*\*Quality Control\*\*: Remove adapter contamination and assess quality using FastQC.

2. \*\*Demultiplexing\*\*: Separate samples using barcodes.

3. \*\*SNP Calling\*\*: Generate SNPs using `denovo\_map.pl` (de novo assembly).

4. \*\*Filtering\*\*: Apply quality filters to the resulting VCF files.

# Quality Control Script

module load FastQC

fastqc \*fq

module load cutadapt

cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -m 75 --length 75 -q 25 \

-o trimmed\_S1\_R1\_001.fastq -p trimmed\_S1\_R2\_001.fastq \

Vahid\_GBS\_S1\_R1\_001.fastq.gz Vahid\_GBS\_S1\_R2\_001.fastq.gz

# Check the quality of trimmed files

fastqc trimmed\_S1\_R1\_001.fastq trimmed\_S1\_R2\_001.fastq

#!/bin/bash

# Demultiplexing Script

# Create required directories

mkdir -p raw samples

# Copy trimmed files into the raw folder

cp trimmed\_S1\_R1\_001.fastq raw/

cp trimmed\_S1\_R2\_001.fastq raw/

# Run process\_radtags

module load Stacks

process\_radtags -P -p raw/ -o samples/ -b Barcodes.txt -e pstI -r -c -q --inline-inline

# Calculate read counts

output\_file="read\_counts.txt"

for file in samples/\*.fq.gz; do

read\_count=$(zcat "$file" | wc -l)

read\_count=$((read\_count / 4))

echo "File: $file, Read Count: $read\_count" >> "$output\_file"

done

# SNP Calling Script

module load Stacks

denovo\_map.pl --samples samples/ --paired --popmap Fish.txt -o Fish.with.het.filter \

-M 2 -n 3 -p 0.8 -X "populations: --max-obs-het 0.6 --write-random-snp --vcf"

# Filter based on missing data and minor allele frequency

module load VCFtools

vcftools --vcf populations.snps.vcf --max-missing 0.5 --maf 0.04 --minDP 3 \

--recode --recode-INFO-all --out miss50

# Remove individuals with more than 80% missing data

vcftools --vcf miss50.recode.vcf --missing-indv

awk '$5 > 0.80' out.imiss | cut -f1 > lowDP-80.indv

vcftools --vcf miss50.recode.vcf --remove lowDP-80.indv --recode --recode-INFO-all --out miss50.INDV

# Apply additional filtering

vcftools --vcf miss50.INDV.recode.vcf --min-meanDP 5 --recode --recode-INFO-all --out final\_snps