Bioinformatics pipeline for G3\_2018\_200625 manuscript

Title: The role of standing variation in the evolution of weediness traits in South Asian weedy rice (*Oryza Spp.*)

Authors: Zhongyun Huang, Shannon Kelly, Rika Matsuo, Lin-Feng Li, Yaling Li, Kenneth M Olsen, Yulin Jia, Ana Caicedo

1. FastQC plot 🡪 trim fastq [bwa] 🡪 sai [samtools] 🡪 sam 🡪 filter uniq (XT:A:U and XT:A:M) [samtools] 🡪 sort\_bam [merge\_bam.pl] 🡪 merge bam [Multi\_sam2.pl sam2.pl] 🡪 VCF
2. Filter VCF by depth and quality (vcf.pl) 🡪 filter.vcf [merge\_snp.pl] 🡪 merge\_snp.txt [Splice.pl] 🡪 temporary individuals [multi\_replace.pl and replace.pl] 🡪 temporary individuals [merge\_replace.pl] 🡪 merge\_unmap.txt (all samples SNP matrix)
3. Merge\_unmap.txt + group4.txt [chname.pl] 🡪 merge\_group\_unmap\_snp.txt

### group4.txt equal to sureselect\_group\_2016.xlsx ###

1. Merge\_group\_unmap\_snp.txt + select\_cover\_region\_for\_select\_snp.txt [select\_gene\_snp.pl] 🡪 region\_snp.txt [split\_group.pl] 🡪 \*.group [count\_new3.pl] 🡪 \*.new\_group
2. region\_snp.txt [group\_sample\_new.pl] 🡪 group\_sample\_new.txt
3. Input: msu6.fa, cds\_sort.gff, cds\_over\_fragment.txt <130 region>

Script: trans\_non\_length.pl

Output: cds\_syn.txt

1. Input: \*.new.group, cds\_syn.txt, gene\_region.\*.txt

Script: trans\_group\_p.pl

Output: group\_p.txt

1. Input: \*.new.group, msu6.fa, cds.sort.gff

Script: trans\_non\_new.pl

Output: \*.new.syn/nonsyn/silent.group

1. \*.new.group + \*.new.syn/nonsyn/silent.group + header.txt [watterson3\_count.pl] 🡪 \*watterson3.count [merge\_count.pl] 🡪 merge\_count.txt
2. Theta

Input: group\_p.txt, header.txt, \*.syn/nonsyn/silent.group, \*.new.group, new\_cover\_region.txt

Script: watterson3\_gene\_new.pl and merge\_watterson.pl

Output: \*.new.total/syn/silent/nonsyn.theta.txt 🡪 merge.fra.theta

1. Pi

Input: group\_p.txt, header.txt, \*.syn/nonsyn/silent.group, \*.new.group, new\_cover\_region.txt

Script: pi\_gene\_nei\_1987\_new\_win.pl and merge\_pi.pl

Output: \*.new.total/syn/silent/nonsyn.nei1987\_pi.txt 🡪 merge.fra.pi

1. FST

Input: group\_p.txt, header.txt, new\_cover\_region.txt

Script: multi\_fst.pl and fst\_gene\_hudson1992.p and fst\_total\_second\_new.pl

Output: Hudson\_fst and second\_fst

1. Tajima’s D

Input: group\_sample\_new.txt, merge\_count.txt, group\_p.txt, merge\_fra.pi

Script: d\_final.pl

Output: Tajima’s D