Software Needed:
Picard tools
GATK
vcftools

1. Merge & index bam files

First, we merge ALL bam files for a species, even across multiple lanes. Make a new directory for all the bam files. Move bam files from all lanes into new directory. Make sure names don't overlap here!

```
mkdir bam
mv Plate2/bam/*.bam bam/
```

Use picard tools MergeSamFiles tool to merge all files into one big file

```
java -jar picard.jar \
     MergeSamFiles $(printf 'I=%s ' bam/*.bam)
     OUTPUT=SNPCalling/YWAR_merged.bam SORT_ORDER=coordinate
```

Index merged bam file

```
samtools index YWAR_merged.bam
```

2. Index Reference

```
samtools faidx YWAR_min1000_sm.fasta

java -jar picard.jar CreateSequenceDictionary
R=YWAR_min1000_sm.fasta O=YWAR_min1000_sm.dict
```

3. Call SNPs

```
java -Xmx8G -jar GenomeAnalysisTK.jar \
-T HaplotypeCaller \
-R References/YWARv0/YWAR_min1000_sm.fasta \
-I SNPCalling/YWAR_merged.bam \
-stand_call_conf 20.0 \
-stand_emit_conf 20.0 \
-o SNPCalling/YWAR.vcf \
--genotyping_mode DISCOVERY \
        -nct 16
```

Note: This takes a long time. The max time limit on Hoffman2 highp queue is 336 hours (2 weeks). If this is not enough or your are in a hurry you can split the bam file.

4. Filter SNPs

```
vcftools --vcf YWAR.vcf --remove-indels --min-alleles 2 --max-alleles 2 --minGQ 20 --minDP 10 --max-missing 0.5 --recode --out YWAR
```

Parameters used:

—remove-indels	keep only SNP sites
—min-alleles	minimum number of alleles at site
—max-alleles	maximum number of alleles at site
—minGQ	minimum genotype quality score (Phred scale)
—minDP	minimum read depth
—max-missing	maximum proportion missing data at a site
—recode	output new vcf

vcftools can also be used to output into 012 format (easy for reading in R or other stats programs)

```
vcftools --vcf YWAR_premature.recode.vcf --012 --out YWAR
```