**Variant Calling with GATK**

1. Prepare HPC environment:
   1. rclone, BWA, samtools, gatk4 installed in your bubble
2. Move into your scratch directory (/scratch/alpine/$USER), activate your bio bubble
   1. cd /scratch/alpine/$USER
   2. conda activate bio
3. Index the reference genome for mapping (will take ~10-30 minutes!):
   1. gatk CreateSequenceDictionary -R Pant.hap.fasta
4. Edit your job submission script:
   1. ploidy level (see spreadsheet for ploidy levels)
   2. MarkDuplicates command
   3. AddOrReplaceReadGroups command
   4. samtools index command
   5. HaplotypeCaller command
   6. change email address
   7. check resource request
5. Submit job
   1. sbatch gatk.sh
   2. Check status of job with squeue command
      1. squeue -u $USER
6. Once job completes, check output, make sure job completed successfully
   1. samtools flagstat bamFile
   2. ls -l \*.vcf
      1. Make sure the vcf file is not small
7. Upload vcf file and duplicatesRemoved bamfile to Google Drive:
   1. rclone copy vcfFile bioinfo:2023/VCFs/ -P
   2. rclone copy duplicatesRemovedBamfile bioinfo:2023/Bams/ -P