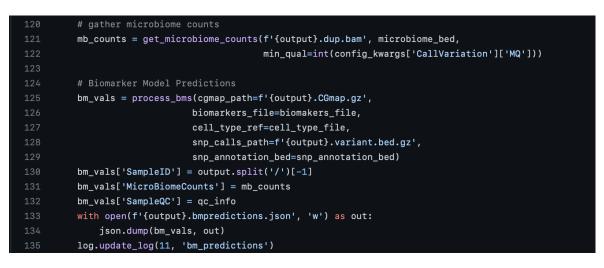
## Prosper Pipeline and Genotype Matrix on Hoffman2

- 1) Clone the prosper processing pipeline to your hoffman2 working directory https://github.com/NuttyLogic/ProsperProcessingContainer
- 2) Remove lines 120 to 135 from ProcessingScripts/ProcessingPipeline.py



(if you don't want CGmap files created you can also remove lines 98-106 as well):

108	# Variant Calling	
109		
110	call_args = config_kwargs['CallVariation']	
111	call_args['DB'] = bsb_db	
112	call_args['I'] = f'{output}.dup.bam'	
113	call_args['0'] = f'{output}.variant'	
114	<pre>if probe_target_bed is not None:</pre>	
115	call_args['BR'] = probe_target_bed	
116		
117	variant_cmd = extract_bsbolt_cmd(call_args)	
118	run_subprocess(variant_cmd, output, 10, 'CallVariation', log)	

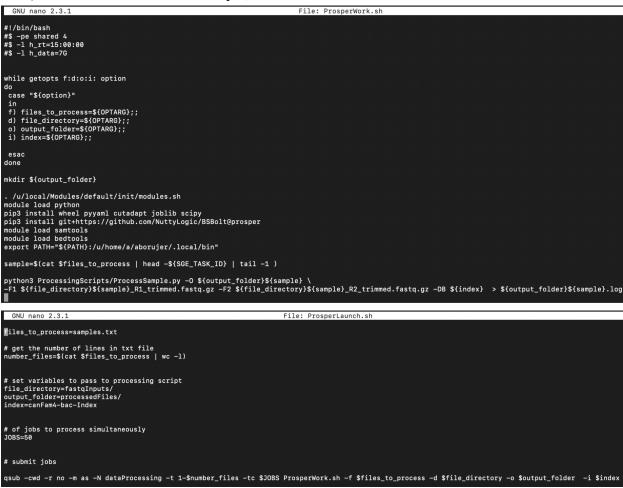
- 3) Set up environment and dependencies:
  - a) cd ProsperProcessingContainer
  - b) Module load python
  - c) pip3 install wheel pyyaml cutadapt joblib scipy git+https://github.com/NuttyLogic/BSBolt@prosper
  - d) export PATH="\${PATH}:/u/home/{first letter of username}/{username}/.local/bin"
  - e) Module load samtools
  - f) Module load bedtools
- 4) Generate Index for species
  - a) Download reference fasta file for species
  - b) Run BSBolt Index to generate index, here is the bash script I used:

GNU nano 2.3.1	File: canFam4_bac_index.sh	
#!/bin/bash #\$ -pe shared 8 #\$ -l h_rt=20:00:00 #\$ -l h_data=4G		
echo "@@@ Starting bsb Index on integrated_genomes.fna \$(date)"		
. /u/local/Modules/default/init/modules.sh module load python		
python3 -m bsbolt Index -G integrated_genomes.fna.gz -DB canFam4-ba	c-Index 1> bsbIndex.stdout 2> bsbIndex.stderr	
echo "@@@ Script reached end. \$(date)"		

- 5) Create line separated text file of sample names you will be processing
  - a) My fastq files were formatted like this:

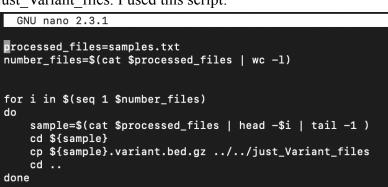
PK9-31201050607829\_R1\_trimmed.fastq.gz PK9-31201050607829\_R2\_trimmed.fastq.gz So for this specific sample I would just write PK9-31201050607829 in the samples.txt file

- 6) Place all fastq inputs into a new directory, I called mine fastqInputs
- 7) Create work and launch scripts, here are mine



8) Launch pipeline through bash scripts (./ProsperLaunch.sh)

- 9) To generate genotype matrix:
  - a) Move your .variant.bed files into a single directory, I called mine
    - just Variant files. I used this script:

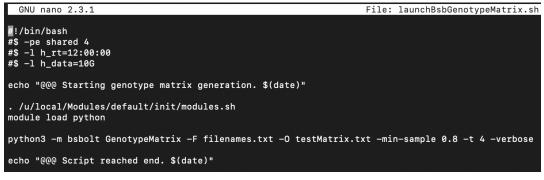


- b) You can clone my repository that has the genotype matrix method https://github.com/aborujerdpur/BSBolt-VariantMatrix-Test
- c) cd BSBolt
- d) Create filenames.txt inside BSBolt that contains the path to your variant files that you want to generate a matrix from, mine looks like this:

File: getVariants.sh

GNU nano 2.3.1 File: filenames.txt
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00108_S84.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00109_S110.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00110_S105.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00111_S89.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00112_S88.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00113_S82.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00114_S95.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00115_S56.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00116_S77.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00117_S48.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00118_S58.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00119_S76.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00120_S102.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00121_S70.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00122_S45.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00123_S59.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00124_S107.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00125_S75.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00126_S103.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00127_S114.variant.bed.gz
/u/scratch/a/aborujer/ProsperProcessingContainer/just_Variant_files/DogWolf00128_S93.variant.bed.gz

e) Create a launch script, mine looks like this:



f) Launch the command using ./launchBsbGenotypeMatrix.sh