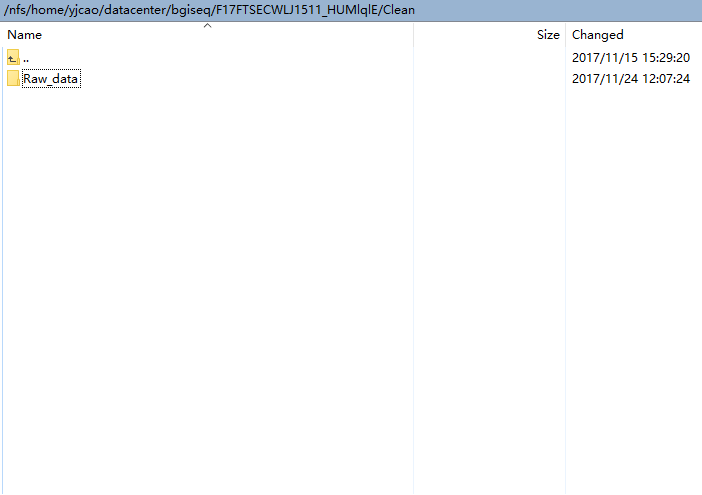
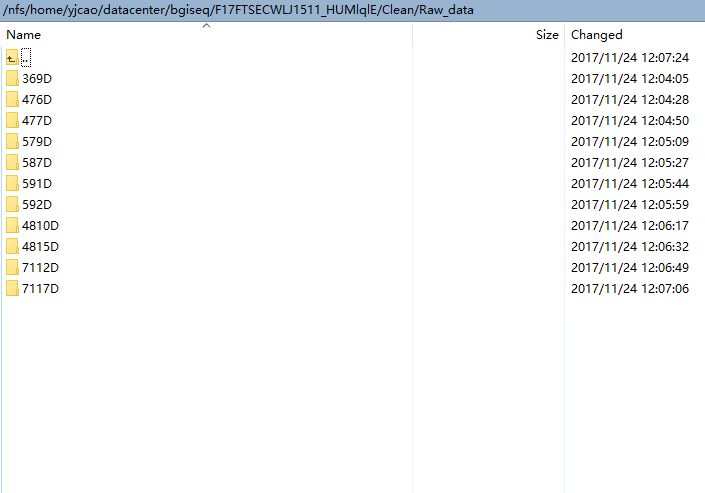
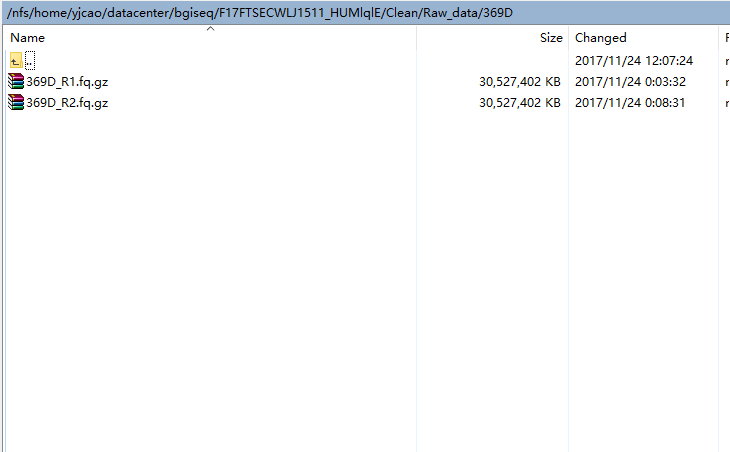
**Easy\_WES tools by Cao Yujie 2017-11-29**

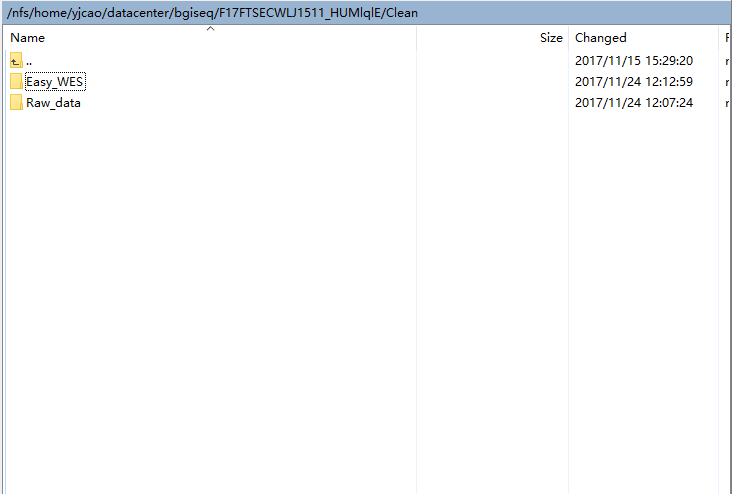
* Upload your Raw\_data to the project folder naming the folder Raw\_data as show below:



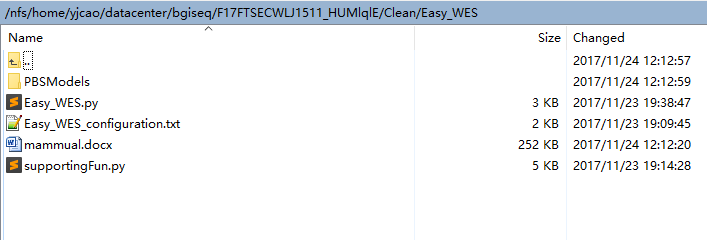




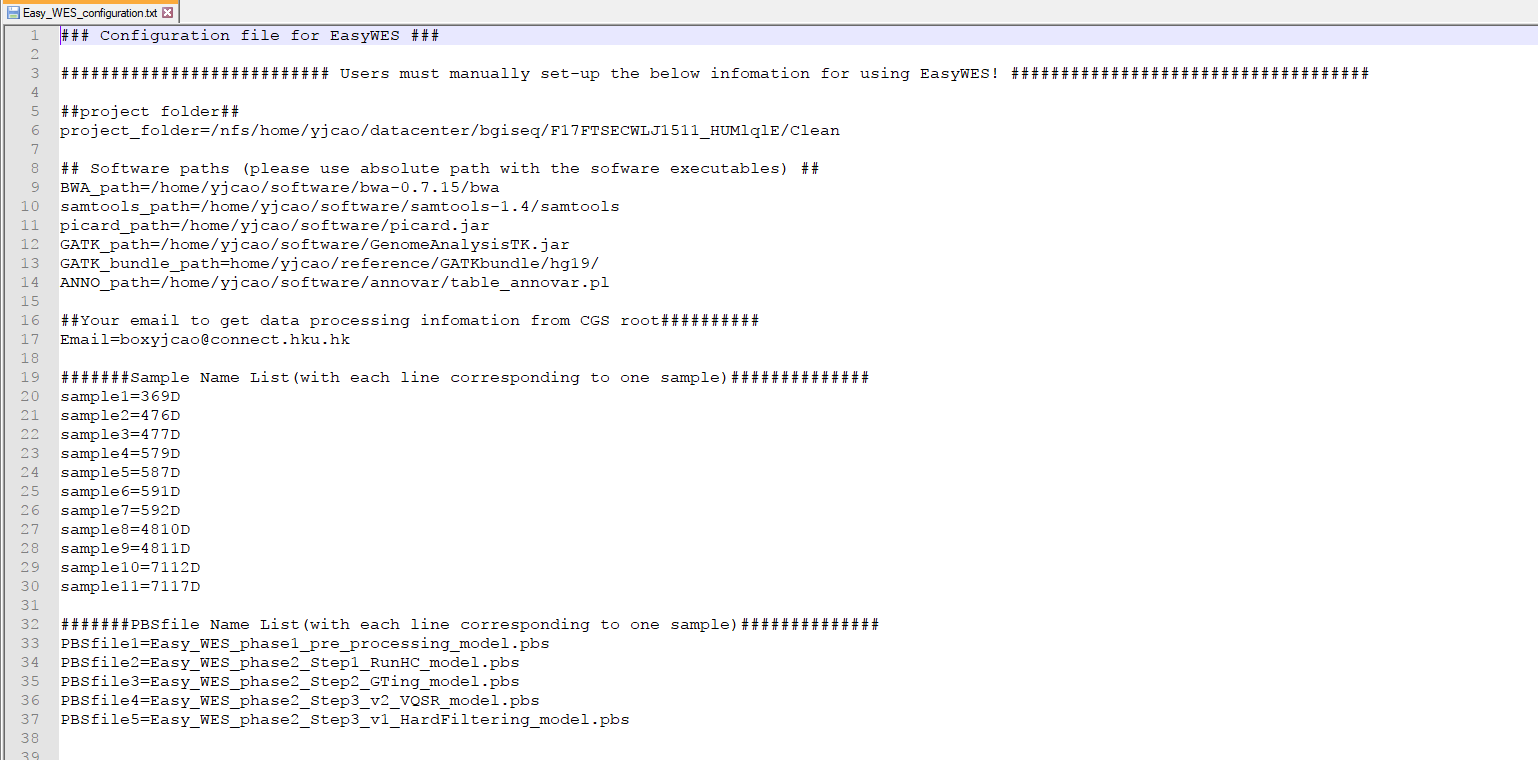
* Your file name should be in the format of SampleID+R1orR2.fq.gz, otherwise you need to rename file name using bash command mv.
* Upload Easy\_WES folder to the project folder, same folder as Raw\_data folder as show below:



* Then, cd Easy\_WES

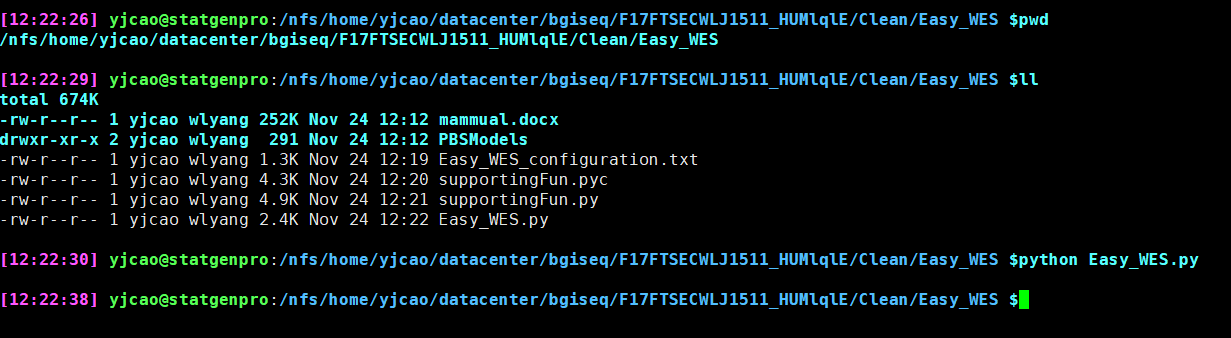


* Then, change configures in file Easy\_WES\_configuration.txt, such as working path, software path, sample name list and PBSscripts lists:

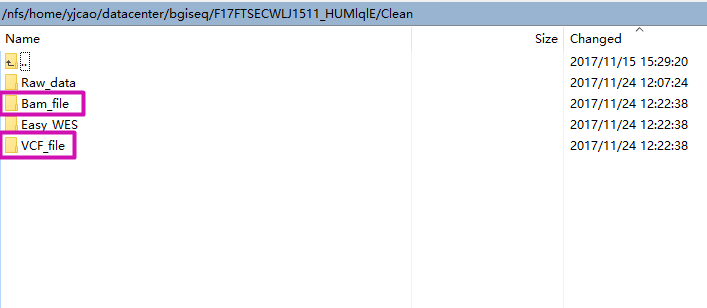


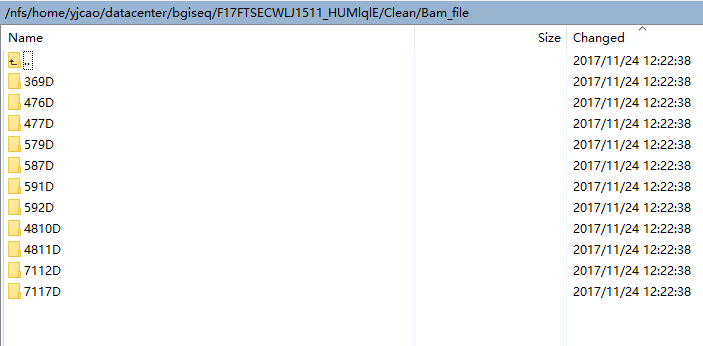
* Then, cd Easy\_WES, run the scripts:

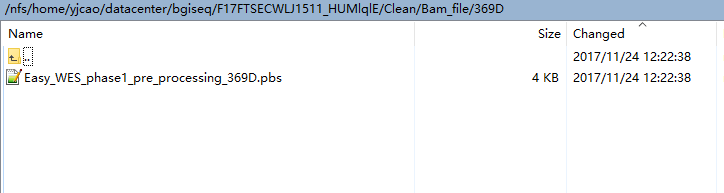
python Easy\_WES.py

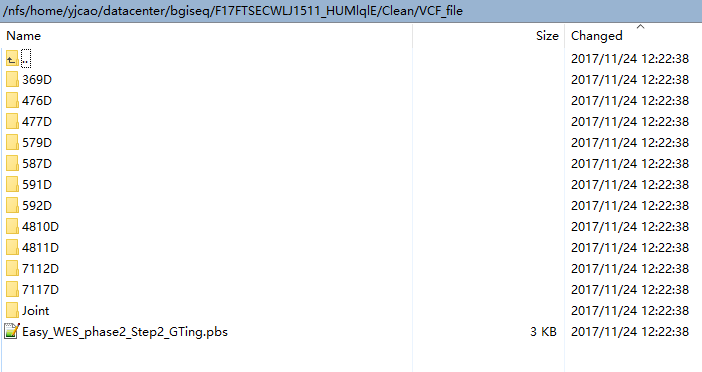


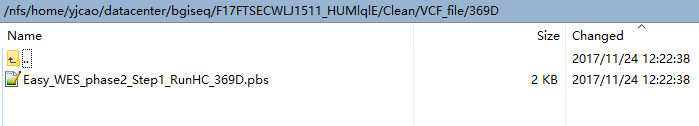
* You will get:

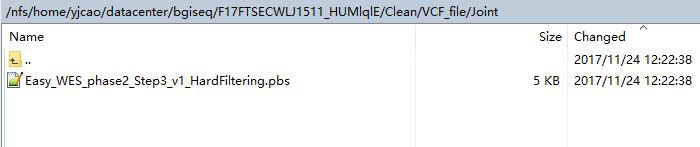




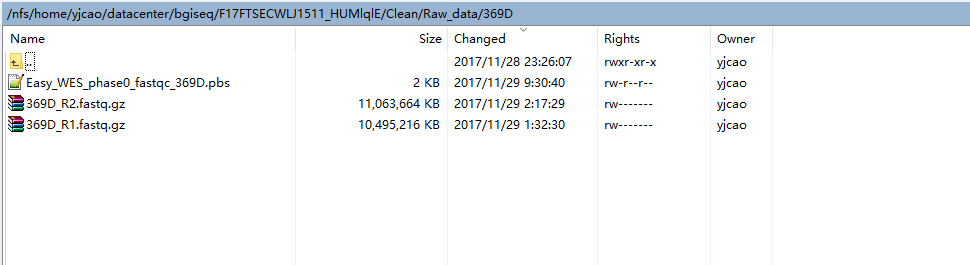
****



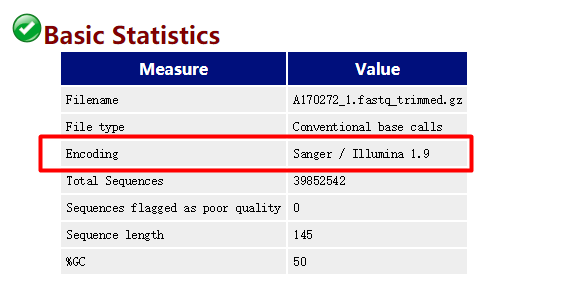


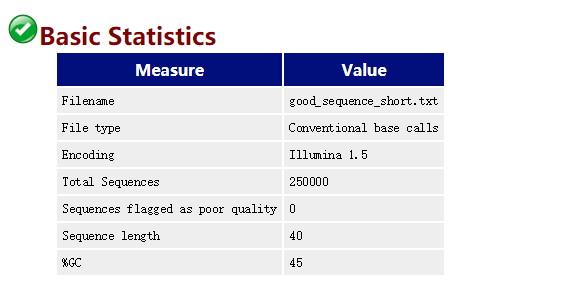


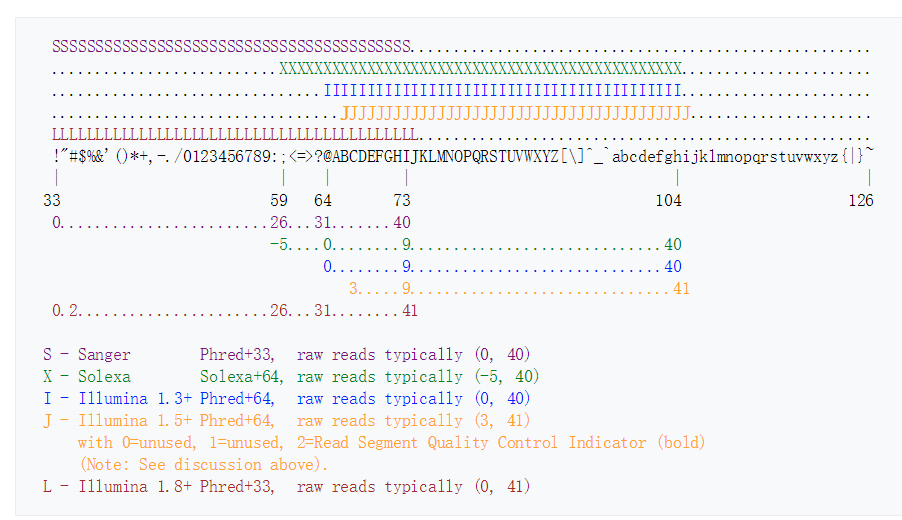
* And under each Raw\_data sample folder, you will get fastqc pbs files, such as this case, under /Raw\_data/369D/:



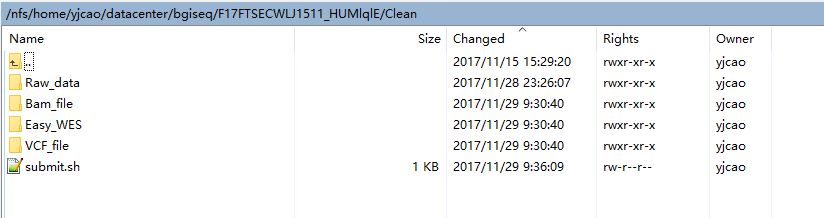
* If necessary, run fastqc for each sample. In the reporter of fastqc, you need to make sure base quality for fastq files should be Illumina 1.8+ Phred+33(https://en.wikipedia.org/wiki/FASTQ\_format). Otherwise you need to correct it use tools like seqtk.

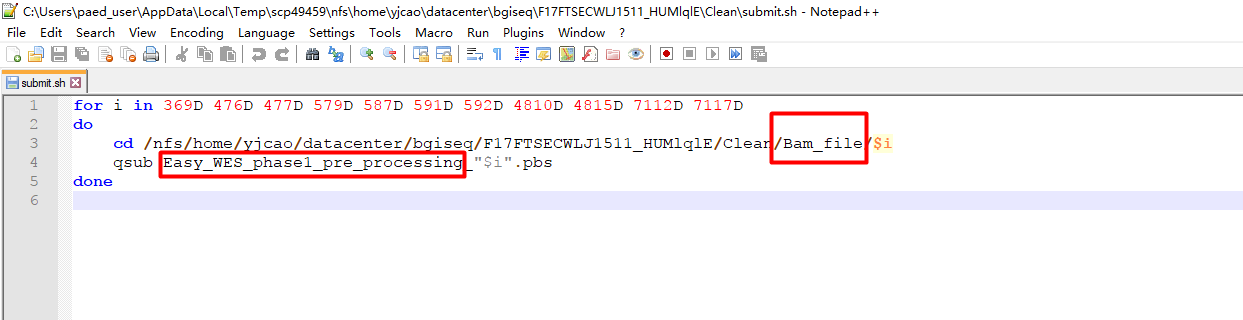






* Submit your tasks to CGS queen





* Only manually change the highlighted folders or prefix, and run the following bash commands:

sh submit.sh

in the following orders:

cd /path/to/projectfolder/Raw\_data/$i

qsub Easy\_WES\_phase0\_Fastqc\_"$i".pbs

cd /path/to/projectfolder/Bam\_file/$i

qsub Easy\_WES\_phase1\_pre\_processing\_"$i".pbs

cd /path/to/projectfolder/VCF\_file/$i

qsub Easy\_WES\_phase2\_Step1\_RunHC\_"$i".pbs

Under VCF\_file folder, qsub Easy\_WES\_phase2\_Step2\_GTing.pbs, then under /VCF\_file/Joint there will be a file named All\_Genotyping.vcf.

Then if your sample size is large than 30, qsub the Easy\_WES\_phase2\_Step3\_v2\_VQSR.pbs, otherwise qsub Easy\_WES\_phase2\_Step3\_v1\_HardFiltering.pbs

Both of them will perform filtering on the callset, and do the genotype refinement, followed by annotate the callset.

