

## Practice 2 : Get Connecting on a linux server by `ssh`

```
MacBook-Pro-de-Christine:~ tranchan$ ssh tranchant@10.2.14.18
tranchant@10.2.14.18's password:
Last login: Thu Oct 4 12:12:25 2018 from 10.2.14.149
[tranchant@master ~]$
```

## Practice 3 : First steps : prompt & `pwd`

- What is the current/working directory just by looking the prompt?

Voici le prompt / invite de commande : `[tranchant@master ~]$`

Le répertoire de travail est : `~` c'est à dire `/home/formation1` (si le login est `formation1`) ou `/home/tranchant` (login = `tranchant`)

- Check the name of your working directory with `pwd` command?

```
[tranchant@master ~]$ pwd
/home/tranchant
```

- On the console, type your 2 first linux commands to get data necessary for the next (we will explain the two commands latter):

# on copie le répertoire compressé `LINUX-TP.tar.gz` présent dans le répertoire `/data/FORMATION/2018/user/` dans le répertoire où on est `.`, ici `/home/tranchant`

```
[tranchant@master ~]$ cp /data/FORMATION/2018/user/LINUX-TP.tar.gz .
[tranchant@master ~]$ ls
LINUX-TP.tar.gz
```

# on décompresse le répertoire

```
[tranchant@master ~]$ tar zxvf LINUX-TP.tar.gz
LINUX-TP/
LINUX-TP/Bank/
LINUX-TP/Bank/referenceRnaseqGFF.gff3
LINUX-TP/Bank/referencePindelChr1.fasta
LINUX-TP/Bank/referencePindelChr1.fasta.fai
LINUX-TP/Bank/referencelrigin.dict
LINUX-TP/Bank/referenceArcad.fasta
LINUX-TP/Bank/referenceRnaseq.fa
LINUX-TP/Bank/referencelrigin.fasta
LINUX-TP/Bank/referencelrigin.fasta.fai
LINUX-TP/Fasta/
LINUX-TP/Fasta/enterobacteries.fasta
LINUX-TP/Data/T-coffee/sample_aln4.aln
..
LINUX-TP/Script/hash.pl
LINUX-TP/Script/transliterate.pl
LINUX-TP/transcriptsAssembly.fasta
[tranchant@master ~]$
```

- Check through filezilla the content of your home directory on the server now (cf. filetree just below)
- Delete through filezilla the file `LINUX-TP.tar.gz` on the server

## Practice 4 : List the files using `ls` command

- List the content of your home directory

```
[tranchant@master ~]$ ls
LINUX-TP
```

- List the content of the directory `Fasta` by using its absolute path in first then its relative path - `ls` command

# chemin absolu

```
[tranchant@master ~]$ ls /home/tranchant/LINUX-TP/Fasta/
C_AllContigs.fasta enterobacteries.fasta uniprot_sprot.fasta
contig_tgicl.fasta sequence.fasta
```

# chemin relatif

```
[tranchant@master ~]$ pwd
/home/tranchant
```

```
[tranchant@master ~]$ ls LINUX-TP/Fasta/
C_AllContigs.fasta enterobacteries.fasta uniprot_sprot.fasta
contig_tgicl.fasta sequence.fasta
```

- List the content of the directory `Data` with the `ls` command and the option `-R`

# chemin absolu

```
[tranchant@master ~]$ ls /home/tranchant/LINUX-TP/Data/
fastq phylip rnaseq samBam T-coffee vcf
[tranchant@master ~]$
```

# chemin relatif

```
[tranchant@master ~]$ pwd
/home/tranchant
```

```
[tranchant@master ~]$ ls LINUX-TP/Data/
fastq phylip rnaseq samBam T-coffee vcf
[tranchant@master ~]$
```

# avec l'option `-R`, on affiche le contenu également des sous dossiers

```
[tranchant@master ~]$ ls LINUX-TP/Data/ -R
LINUX-TP/Data/:
fastq phylip rnaseq samBam T-coffee vcf
```

```
LINUX-TP/Data/fastq:
assembly      pairedTwoIndividusGzippedIrigin pairingRepairing
pairedOneIndividuArcad pairedTwoIndividusIrigin   singleTwoIndividuIrigin
```

```
LINUX-TP/Data/fastq/assembly:
ebolaAssembly pairedOneIndividuPacaya
```

...

- List the content of the directory `Bank` with the `ls` command and the option `-al` or `-a -l`

```
[tranchant@master ~]$ ls LINUX-TP/Bank/ -R
LINUX-TP/Bank/:
referenceArcad.fasta referenceIrigin.fasta.fai referenceRnaseq.fasta
referenceIrigin.dict referencePindelChr1.fasta referenceRnaseqGFF.gff3
referenceIrigin.fasta referencePindelChr1.fasta.fai
```

## Practice 5 : List the files using `ls` command and metacaracter \*

- List the content of the directory T-coffee. Are there only fasta files ? - `ls` command

```
[tranchant@master ~]$ ls LINUX-TP/Data/T-coffee/  
1D4V-1.pdb.gz      sample_profile1.aln  
1D4V-2.pdb.gz      sample_profile2.aln  
1D4V-3.pdb.gz      sample_rnaseq1.fasta  
1EXT-1.pdb.gz      sample_rnaseq2.alifold  
1EXT-2.pdb.gz      sample_rnaseq2.aln.gz  
...  
sample_lib4.tc_lib02  tlara_tRNA_aln50.output.gz  
sample_lib5.tc_lib   tlara_tRNA_aln51.output.gz  
sample_list1.lib_list two_profiles.template_file  
sample_param_file.param x.gz
```

- List only the files starting by sample (in the directory T-coffee ) - `ls` command & \*

```
[tranchant@master ~]$ cd LINUX-TP/Data/T-coffee/  
[tranchant@master T-coffee]$ pwd  
/home/tranchant/LINUX-TP/Data/T-coffee  
[tranchant@master T-coffee]$ ls sample*  
sample_aln1_1.aln  sample_lib1.mocca_lib.gz  sample_rnaseq2.fasta  
sample_aln1_2.aln  sample_lib1.tc_lib.gz     sample_rnaseq2.stk.gz  
sample_aln1.aln    sample_lib2.tc_lib.gz     sample_rnaseq2.tcoffee.stk.gz  
sample_aln1.dnd    sample_lib3.tc_lib.gz     sample_seq1.aln  
sample_aln1.fasta  sample_lib4.tc_lib02     sample_seq1.fasta  
sample_aln1.tc_lib.gz  sample_lib5.tc_lib     sample_seq2.fasta  
sample_aln2.aln    sample_list1.lib_list     sample_seq3.fasta  
sample_aln2.fasta  sample_param_file.param  sample_seq4.fasta  
sample_aln2.msf    sample_profile1.aln      sample_seq5.fasta  
sample_aln3.aln    sample_profile2.aln      sample_seq6.fasta  
sample_aln4.aln    sample_rnaseq1.fasta     sample_seq7.fasta  
sample_aln5.aln.gz  sample_rnaseq2.alifold   sample_seq8.seq  
sample_aln6.aln    sample_rnaseq2.aln.gz    sample_tree2.dnd  
sample_aln6.cache  sample_rnaseq2.cons.stk  sample_tree3.dnd  
sample_aln7.aln    sample_rnaseq2.cw_aln.gz  sample_tree.dnd  
sample_aln7.cache_aln  sample_rnaseq2.cw.stk.gz  
sample_dnaseq1.fasta  sample_rnaseq2.dnd
```

- List only the files with the fasta extension (in the directory T-coffee ) - `ls` command & \*

```
[tranchant@master T-coffee]$ ls *fasta  
3d.fasta      sample_rnaseq2.fasta  sproteases2_small.fasta  
3d_sample2.fasta  sample_seq1.fasta    sproteases_oneseq.fasta  
3d_sample3.fasta  sample_seq2.fasta    sproteases_pair.fasta  
profile1_pdb1.fasta  sample_seq3.fasta    sv.fasta  
profile2_pdb2.fasta  sample_seq4.fasta    three_cdna.fasta  
sample_aln1.fasta  sample_seq5.fasta    three_pdb.fasta  
sample_aln2.fasta  sample_seq6.fasta    three_pdb_two_seq.fasta  
sample_dnaseq1.fasta  sample_seq7.fasta  
sample_rnaseq1.fasta  sproteases1_small.fasta  
[tranchant@master T-coffee]$ ls sample*fasta  
sample_aln1.fasta  sample_rnaseq1.fasta  sample_seq2.fasta  sample_seq5.fasta  
sample_aln2.fasta  sample_rnaseq2.fasta  sample_seq3.fasta  sample_seq6.fasta  
sample_dnaseq1.fasta  sample_seq1.fasta    sample_seq4.fasta  sample_seq7.fasta
```

## Practice 6 : Moving into file system using `cd` and `ls` command

- Go to the directory `Script` and check in the prompt you have changed correctly your working directory (`pwd`).

```
[tranchant@master ~]$ pwd
/home/tranchant
[tranchant@master ~]$ cd LINUX-TP/Script/
[tranchant@master Script]$
```

- List the dir content with `ls`.

```
[tranchant@master Script]$ ls -l
total 44
-rwxr-xr-x 1 tranchant tranchant 411 3 mars 2018 array.pl
-rwxr-xr-x 1 tranchant tranchant 726 3 mars 2018 codon_usage.pl
-rwxr-xr-x 1 tranchant tranchant 371 3 mars 2018 hash.pl
-rwxr-xr-x 1 tranchant tranchant 67 3 mars 2018 helloWorld.pl
-rwxr-xr-x 1 tranchant tranchant 120 3 mars 2018 loops-for.pl
-rwxr-xr-x 1 tranchant tranchant 205 3 mars 2018 matching.pl
-rwxr-xr-x 1 tranchant tranchant 260 3 mars 2018 readFasta.pl
-rwxr-xr-x 1 tranchant tranchant 282 3 mars 2018 retrieve-accession.pl
-rwxr-xr-x 1 tranchant tranchant 467 3 mars 2018 sorting-array.pl
-rwxr-xr-x 1 tranchant tranchant 194 3 mars 2018 string-array.pl
-rwxr-xr-x 1 tranchant tranchant 461 3 mars 2018 transliterate.pl
```

- Go to the `Fasta` directory using `../`

```
[tranchant@master Script]$ cd ../Fasta/
[tranchant@master Fasta]$
```

- Go to the `Fastq` directory . From this directory, and without any change in your working dir, list what's in `samBam` directory

```
[tranchant@master Fasta]$ cd ../Data/fastq/
[tranchant@master fastq]$ ls ../samBam/
oneBam oneBamUnsorted oneSam samBamSV twoBamsIorigin
```

- List `vcf` directory using `-R` option. What is there in this dir ?

```
[tranchant@master fastq]$ ls ../vcf/ -R
../vcf/:
duplicVCF singleVCF testsnmf.geno vcfForRecalibration vcfForSNIPlay
```

```
../vcf/duplicVCF:
smallDuplic-filtered.vcf smallDuplic.vcf
```

```
../vcf/singleVCF:
GATKVARIANTFILTRATION.vcf
```

```
../vcf/vcfForRecalibration:
control.vcf
```

```
../vcf/vcfForSNIPlay:
testsnmf.vcf
```

```
[tranchant@master fastq]$
```

- Come back to the home directory.

```
[tranchant@master fastq]$ pwd
/home/tranchant/LINUX-TP/Data/fastq
```

```
[tranchant@master fastq]$ cd
```

```
[tranchant@master ~]$ pwd
/home/tranchant
```

## Practice 7 : Manipulating Files and Folders

We will prepare our blast analysis performed after by creating directory and moving files as showing in the image just below :

- Create a directory called BlastAnalysis with mkdir command.

```
[tranchant@master ~]$ cd LINUX-TP/
[tranchant@master LINUX-TP]$ mkdir BlastAnalysis
```

- List the content of LINUX-TP and BlastAnalysis with ls command.

```
[tranchant@master LINUX-TP]$ pwd
/home/tranchant/LINUX-TP
```

```
[tranchant@master LINUX-TP]$ ls
AllEst.fasta BlastAnalysis Fasta transcritsAssembly.fasta
Bank      Data      Script
```

```
[tranchant@master LINUX-TP]$ ls BlastAnalysis/
[tranchant@master LINUX-TP]$
```

- Copy AllEst.fasta in Bank directory with cp command.

```
[tranchant@master LINUX-TP]$ pwd
/home/tranchant/LINUX-TP
```

```
[tranchant@master LINUX-TP]$ ls
AllEst.fasta BlastAnalysis Fasta transcritsAssembly.fasta
Bank      Data      Script
```

```
[tranchant@master LINUX-TP]$ cp AllEst.fasta Bank/
```

- List the content of LINUX-TP and Bank directories.

```
[tranchant@master LINUX-TP]$ pwd
/home/tranchant/LINUX-TP
```

```
[tranchant@master LINUX-TP]$ ls
AllEst.fasta BlastAnalysis Fasta transcritsAssembly.fasta
Bank      Data      Script
```

```
[tranchant@master LINUX-TP]$ ls Bank/
AllEst.fasta      referencePindelChr1.fasta
referenceArcad.fasta  referencePindelChr1.fasta.fai
referenceIrgin.dict  referenceRnaseq.fa
```

```
referenceIrgin.fasta referenceRnaseqGFF.gff3
referenceIrgin.fasta.fai
```

- Remove the file AllEst.fasta in LINUX-TP directory with rm command.

```
[tranchant@master LINUX-TP]$ pwd
/home/tranchant/LINUX-TP
```

```
[tranchant@master LINUX-TP]$ ls
AllEst.fasta BlastAnalysis Fasta transcritsAssembly.fasta
Bank Data Script
```

```
[tranchant@master LINUX-TP]$ rm AllEst.fasta
[tranchant@master LINUX-TP]$ ls
Bank BlastAnalysis Data Fasta Script transcritsAssembly.fasta
```

- Copy the whole directory T-coffee with the name T-coffee-copy into LINUX-TP directory.

```
[tranchant@master Data]$ pwd
/home/tranchant/LINUX-TP/Data
```

```
[tranchant@master Data]$ ls
fastq phylip rnaseq samBam T-coffee vcf
```

```
[tranchant@master Data]$ cp T-coffee/ ../T-coffee-copy/
cp: omission du répertoire « T-coffee/ »
```

```
[tranchant@master Data]$ cp T-coffee/ ../T-coffee-copy/ -R
```

```
[tranchant@master Data]$ ls
fastq phylip rnaseq samBam T-coffee vcf
```

```
[tranchant@master Data]$ ls ../
Bank BlastAnalysis Data Fasta Script T-coffee-copy transcritsAssembly.fasta
[tranchant@master Data]$
```

- After checking the content of LINUX-TP directory, remove the directory T-coffee-copy. How to remove a directory ?

```
[tranchant@master LINUX-TP]$ pwd
/home/tranchant/LINUX-TP
```

```
[tranchant@master LINUX-TP]$ ls
Bank BlastAnalysis Data Fasta Script T-coffee-copy transcritsAssembly.fasta
```

```
[tranchant@master LINUX-TP]$ rm T-coffee-copy/
rm: impossible de supprimer « T-coffee-copy/ »: est un dossier
```

```
[tranchant@master LINUX-TP]$ rm T-coffee-copy/ -r
```

---

## Practice 8 : Dealing with text Files

- Copy the vcf /data/FORMATION/2018/user/200000-CHR9.GATKVARIANTFILTRATION.vcf

```
[tranchant@master vcf]$ pwd
/home/tranchant/LINUX-TP/Data/vcf
```

```
[tranchant@master vcf]$ cp /data/FORMATION/2018/user/
200000-CHR9.GATKVARIANTFILTRATION.vcf all.gff3          LINUX-TP.tar.gz
```

```
[tranchant@master vcf]$ cp /data/FORMATION/2018/user/200000-CHR9.GATKVARIANTFILTRATION.vcf .
[tranchant@master vcf]$
```

- List the content of the directory /home/tranchant/LINUX-TP/Data/vcf with ls -l

```
[tranchant@master vcf]$ ls -l
total 1049908
-rw-r--r-- 1 tranchant tranchant 1074328449 5 oct. 14:38 200000-CHR9.GATKVARIANTFILTRATION.vcf
drwxr-xr-x 2 tranchant tranchant   61 3 mars 2018 duplicVCF
drwxr-xr-x 2 tranchant tranchant   39 3 mars 2018 singleVCF
-rw-r--r-- 1 tranchant tranchant  772338 3 mars 2018 testsnmf.geno
drwxr-xr-x 2 tranchant tranchant   25 3 mars 2018 vcfForRecalibration
drwxr-xr-x 2 tranchant tranchant   26 3 mars 2018 vcfForSniPlay
[tranchant@master vcf]$
```

- Displays the first lines of the vcf file - head command

```
[tranchant@master vcf]$ head 200000-CHR9.GATKVARIANTFILTRATION.vcf
##fileformat=VCFv4.1
##FILTER=<ID=HIGH-DP,Description="DP>20000">
##FILTER=<ID=LOW-DP,Description="DP<10">
##FILTER=<ID=LOW-MQ0,Description="MQ0>=4 && ((MQ0/(1.0*DP)) > 0.1)">
##FILTER=<ID=LOW-QUAL,Description="QUAL<200">
##FILTER=<ID=LowQual,Description="Low quality">
##FILTER=<ID=SnpcCluster,Description="SNPs found in clusters">
##FORMAT=<ID=AD,Number=.,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth (reads with MQ=255 or with bad mates are filtered)">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
```

- Displays the last lines of the vcf file - tail command

```
[tranchant@master vcf]$ tail 200000-CHR9.GATKVARIANTFILTRATION.vcf -n1
Chr9      22974997      .          G          C          1317.27  LOW-MQ0
AC=66;AF=1.00;AN=66;DP=590;Dels=0.00;FS=0.000;HaplotypeScore=0.0573;InbreedingCoeff=-0.1688;MLEAC=63;MLEAF=0.955;MQ=7.21;MQ0=
515;QD=7.20;SOR=8.635      GT:AD:DP:GQ:PL      ./      ./      ./      1/1:0,4:4:3:30,3,0      1/1:0,5:6:6:59,6,0      ./      ./
1/1:0,10:10:6:54,6,0./      ./      ./      1/1:0,3:3:3:30,3,0      ./      ./      ./      ./      ./      1/1:0,4:4:9:83,9,0      ./      ./
./      ./      1/1:0,5:5:3:30,3,0      ./      ./      ./      ./      ./      ./      1/1:0,4:4:6:53,6,0      ./      ./
./      ./      1/1:0,6:6:9:89,9,0      ./      ./      ./      1/1:0,4:4:6:53,6,0      ./      ./      1/1:0,3:3:3:30,3,0      ./      ./
./      ./      ./      ./      ./      ./      ./      1/1:0,5:5:3:25,3,0      ./      ./
./      ./      ./      ./      ./      1/1:0,7:7:3:30,3,0      ./      1/1:0,8:8:3:30,3,0      ./      ./
./      ./      ./      ./      ./      1/1:0,7:7:6:59,6,0      ./      1/1:0,8:8:6:63,6,0      ./      ./      1/1:0,4:4:3:39,3,0      ./
./      ./      ./      ./      ./      ./      ./      ./      ./      1/1:0,1:1:3:30,3,0
1/1:0,7:7:3:30,3,0      ./      1/1:0,6:6:3:24,3,0      ./      ./      ./      ./      ./      1/1:0,8:8:3:30,3,0      ./      ./
1/1:0,5:5:3:25,3,0      ./      ./      ./      ./      1/1:0,8:8:3:30,3,0      ./      ./      1/1:0,1:1:3:30,3,0
./      ./      ./      ./      ./      ./      ./      ./      ./      1/1:0,1:1:3:30,3,0
./      ./      ./      ./      1/1:0,6:6:3:33,3,0      ./      ./      1/1:0,9:9:3:24,3,0      ./      1/1:0,7:7:3:30,3,0
1/1:0,6:6:3:30,3,0      ./      ./      1/1:0,3:3:3:30,3,0      ./      ./      1/1:0,9:9:3:24,3,0      ./      1/1:0,7:7:3:30,3,0
1/1:0,6:6:3:30,3,0      1/1:0,7:7:3:30,3,0      ./      ./      1/1:0,8:8:3:30,3,0      1/1:0,7:7:3:30,3,0      1/1:0,8:8:3:30,3,0      1/1:0,8:8:3:30,3,0      1/1:0,8:8:3:30,3,0
./      ./      1/1:0,3:3:3:30,3,0      1/1:0,3:3:3:30,3,0      ./      ./      1/1:0,8:8:3:30,3,0      1/1:0,8:8:3:30,3,0      1/1:0,8:8:3:30,3,0
```

- Count the lines of the vcf files - wc -l command

```
tranchant@master vcf]$ wc -l 200000-CHR9.GATK VARIANTFILTRATION.vcf
200053 200000-CHR9.GATK VARIANTFILTRATION.vcf
```

## Practice 9 : Searching with grep

- Displays the first and last line of the gff file - head, tail

```
[tranchant@node0 Bank]$ tail all.gff3
ChrSy MSU_osa1r7 gene 585817 586479 . - .
ID=ChrSy.fgenes.h.gene.88;Name=ChrSy.fgenes.h.gene.88;Note=hypothetical%20protein
ChrSy MSU_osa1r7 mRNA 585817 586479 . - .
ID=ChrSy.fgenes.h.mRNA.88;Parent=ChrSy.fgenes.h.gene.88;Name=ChrSy.fgenes.h.mRNA.88
ChrSy MSU_osa1r7 CDS 586260 586479 -2.93 - 0
ID=ChrSy.fgenes.h.CDS.325;Parent=ChrSy.fgenes.h.mRNA.88;score=-2.93
ChrSy MSU_osa1r7 exon 586260 586479 -2.93 - .
ID=ChrSy.fgenes.h.exon.325;Parent=ChrSy.fgenes.h.mRNA.88;score=-2.93
ChrSy MSU_osa1r7 CDS 585817 586166 17.88 - 1
ID=ChrSy.fgenes.h.CDS.326;Parent=ChrSy.fgenes.h.mRNA.88;score=17.88
ChrSy MSU_osa1r7 exon 585817 586166 17.88 - .
ID=ChrSy.fgenes.h.exon.326;Parent=ChrSy.fgenes.h.mRNA.88;score=17.88
ChrSy MSU_osa1r7 gene 589676 589999 . + .
ID=ChrSy.fgenes.h.gene.89;Name=ChrSy.fgenes.h.gene.89;Note=expressed%20protein
ChrSy MSU_osa1r7 mRNA 589676 589999 . + .
ID=ChrSy.fgenes.h.mRNA.89;Parent=ChrSy.fgenes.h.gene.89;Name=ChrSy.fgenes.h.mRNA.89
ChrSy MSU_osa1r7 CDS 589676 589999 11.35 + 0
ID=ChrSy.fgenes.h.CDS.327;Parent=ChrSy.fgenes.h.mRNA.89;score=11.35
ChrSy MSU_osa1r7 exon 589676 589999 11.35 + .
ID=ChrSy.fgenes.h.exon.327;Parent=ChrSy.fgenes.h.mRNA.89;score=11.35

[tranchant@node0 Bank]$ head all.gff3
##gff-version 3
Chr1 MSU_osa1r7 gene 2903 10817 . + .
ID=LOC_Os01g01010;Name=LOC_Os01g01010;Note=TBC%20domain%20containing%20protein%2C%20expressed
Chr1 MSU_osa1r7 mRNA 2903 10817 . + .
ID=LOC_Os01g01010.1;Name=LOC_Os01g01010.1;Parent=LOC_Os01g01010
Chr1 MSU_osa1r7 exon 2903 3268 . + .
ID=LOC_Os01g01010.1:exon_1;Parent=LOC_Os01g01010.1
Chr1 MSU_osa1r7 exon 3354 3616 . + .
ID=LOC_Os01g01010.1:exon_2;Parent=LOC_Os01g01010.1
Chr1 MSU_osa1r7 exon 4357 4455 . + .
ID=LOC_Os01g01010.1:exon_3;Parent=LOC_Os01g01010.1
Chr1 MSU_osa1r7 exon 5457 5560 . + .
ID=LOC_Os01g01010.1:exon_4;Parent=LOC_Os01g01010.1
Chr1 MSU_osa1r7 exon 7136 7944 . + .
ID=LOC_Os01g01010.1:exon_5;Parent=LOC_Os01g01010.1
Chr1 MSU_osa1r7 exon 8028 8150 . + .
ID=LOC_Os01g01010.1:exon_6;Parent=LOC_Os01g01010.1
Chr1 MSU_osa1r7 exon 8232 8320 . + .
ID=LOC_Os01g01010.1:exon_7;Parent=LOC_Os01g01010.1
```

- Counts the number of genes - grep -c

```
[tranchant@node0 Bank]$ grep "gene" all.gff3 | head
Chr1 MSU_osa1r7 gene 2903 10817 . + .
ID=LOC_Os01g01010;Name=LOC_Os01g01010;Note=TBC%20domain%20containing%20protein%2C%20expressed
Chr1 MSU_osa1r7 gene 11218 12435 . + .
ID=LOC_Os01g01019;Name=LOC_Os01g01019;Note=expressed%20protein
Chr1 MSU_osa1r7 gene 12648 15915 . + .
ID=LOC_Os01g01030;Name=LOC_Os01g01030;Note=monocopper%20oxidase%2C%20putative%2C%20expressed
```



```

Chr1 MSU_osa1r7 gene 16292 20323 . + .
ID=LOC_Os01g01040;Name=LOC_Os01g01040;Note=expressed%20protein
Chr1 MSU_osa1r7 gene 22841 26971 . + .
ID=LOC_Os01g01050;Name=LOC_Os01g01050;Note=R3H%20domain%20containing%20protein%2C%20expressed
Chr1 MSU_osa1r7 gene 27136 28651 . + .
ID=LOC_Os01g01060;Name=LOC_Os01g01060;Note=40S%20ribosomal%20protein%20S5%2C%20putative%2C%20expressed
Chr1 MSU_osa1r7 gene 29818 34493 . + .
ID=LOC_Os01g01070;Name=LOC_Os01g01070;Note=expressed%20protein
Chr1 MSU_osa1r7 gene 35581 41180 . + .
ID=LOC_Os01g01080;Name=LOC_Os01g01080;Note=decarboxylase%2C%20putative%2C%20expressed
Chr1 MSU_osa1r7 gene 44595 47526 . + .
ID=LOC_Os01g01090;Name=LOC_Os01g01090;Note=retrotransposon%20protein%2C%20putative%2C%20unclassified%2C%20expressed
Chr1 MSU_osa1r7 gene 47856 53412 . - .
ID=LOC_Os01g01100;Name=LOC_Os01g01100;Note=retrotransposon%20protein%2C%20putative%2C%20Ty3-gypsy%20subclass%2C%20ex
pressed

```

```
[tranchant@node0 Bank]$ grep "gene" all.gff3 | tail
```

```

ChrSy MSU_osa1r7 gene 585817 586479 . - .
ID=ChrSy.fgenesh.gene.88;Name=ChrSy.fgenesh.gene.88;Note=hypothetical%20protein
ChrSy MSU_osa1r7 mRNA 585817 586479 . - .
ID=ChrSy.fgenesh.mRNA.88;Parent=ChrSy.fgenesh.gene.88;Name=ChrSy.fgenesh.mRNA.88
ChrSy MSU_osa1r7 CDS 586260 586479 -2.93 - 0
ID=ChrSy.fgenesh.CDS.325;Parent=ChrSy.fgenesh.mRNA.88;score=-2.93
ChrSy MSU_osa1r7 exon 586260 586479 -2.93 - .
ID=ChrSy.fgenesh.exon.325;Parent=ChrSy.fgenesh.mRNA.88;score=-2.93
ChrSy MSU_osa1r7 CDS 585817 586166 17.88 - 1
ID=ChrSy.fgenesh.CDS.326;Parent=ChrSy.fgenesh.mRNA.88;score=17.88
ChrSy MSU_osa1r7 exon 585817 586166 17.88 - .
ID=ChrSy.fgenesh.exon.326;Parent=ChrSy.fgenesh.mRNA.88;score=17.88
ChrSy MSU_osa1r7 gene 589676 589999 . + .
ID=ChrSy.fgenesh.gene.89;Name=ChrSy.fgenesh.gene.89;Note=expressed%20protein
ChrSy MSU_osa1r7 mRNA 589676 589999 . + .
ID=ChrSy.fgenesh.mRNA.89;Parent=ChrSy.fgenesh.gene.89;Name=ChrSy.fgenesh.mRNA.89
ChrSy MSU_osa1r7 CDS 589676 589999 11.35 + 0
ID=ChrSy.fgenesh.CDS.327;Parent=ChrSy.fgenesh.mRNA.89;score=11.35
ChrSy MSU_osa1r7 exon 589676 589999 11.35 + .
ID=ChrSy.fgenesh.exon.327;Parent=ChrSy.fgenesh.mRNA.89;score=11.35

```

```
[tranchant@node0 Bank]$ grep "gene" all.gff3 -c
57585
```

```
[tranchant@node0 Bank]$ grep -E "gene\s" all.gff3 | tail
```

```

ChrSy MSU_osa1r7 gene 550000 551936 . + .
ID=ChrSy.fgenesh.gene.80;Name=ChrSy.fgenesh.gene.80;Note=expressed%20protein
ChrSy MSU_osa1r7 gene 554038 554861 . - .
ID=ChrSy.fgenesh.gene.81;Name=ChrSy.fgenesh.gene.81;Note=expressed%20protein
ChrSy MSU_osa1r7 gene 558985 561654 . + .
ID=ChrSy.fgenesh.gene.82;Name=ChrSy.fgenesh.gene.82;Note=expressed%20protein
ChrSy MSU_osa1r7 gene 564501 565650 . + .
ID=ChrSy.fgenesh.gene.83;Name=ChrSy.fgenesh.gene.83;Note=expressed%20protein
ChrSy MSU_osa1r7 gene 565783 569113 . - .
ID=ChrSy.fgenesh.gene.84;Name=ChrSy.fgenesh.gene.84;Note=expressed%20protein
ChrSy MSU_osa1r7 gene 572904 573515 . + .
ID=ChrSy.fgenesh.gene.85;Name=ChrSy.fgenesh.gene.85;Note=expressed%20protein
ChrSy MSU_osa1r7 gene 573737 574348 . - .
ID=ChrSy.fgenesh.gene.86;Name=ChrSy.fgenesh.gene.86;Note=expressed%20protein
ChrSy MSU_osa1r7 gene 576385 580092 . + .
ID=ChrSy.fgenesh.gene.87;Name=ChrSy.fgenesh.gene.87;Note=expressed%20protein
ChrSy MSU_osa1r7 gene 585817 586479 . - .
ID=ChrSy.fgenesh.gene.88;Name=ChrSy.fgenesh.gene.88;Note=hypothetical%20protein
ChrSy MSU_osa1r7 gene 589676 589999 . + .
ID=ChrSy.fgenesh.gene.89;Name=ChrSy.fgenesh.gene.89;Note=expressed%20protein

```

```
[tranchant@node0 Bank]$ grep -E "gene\s" all.gff3 -c
55986
```

- Search for the nbs-irr genes - grep -i

```
[tranchant@node0 Bank]$ grep "nbs-irr" all.gff3 -c
```

0

```
[tranchant@node0 Bank]$ grep "nbs-lrr" all.gff3 -c -i  
124
```

```
[tranchant@node0 Bank]$ grep "nbs-lrr" all.gff3 | head -n2
```

```
[tranchant@node0 Bank]$ grep "nbs-lrr" all.gff3 -i | head -n2
```

```
Chr1 MSU_osa1r7 gene 2669952 2672924 . - .  
ID=LOC_Os01g05600;Name=LOC_Os01g05600;Note=NBS-LRR%20disease%20resistance%20protein%2C%20putative%2C%20expressed  
Chr1 MSU_osa1r7 gene 2682019 2684988 . - .  
ID=LOC_Os01g05620;Name=LOC_Os01g05620;Note=NBS-LRR%20disease%20resistance%20protein%2C%20putative%2C%20expressed
```

- Remove the lines with putative word - `grep -v`

```
[tranchant@node0 Bank]$ grep "putative" all.gff3 | head
```

```
Chr1 MSU_osa1r7 gene 12648 15915 . + .  
ID=LOC_Os01g01030;Name=LOC_Os01g01030;Note=monocopper%20oxidase%2C%20putative%2C%20expressed  
Chr1 MSU_osa1r7 gene 27136 28651 . + .  
ID=LOC_Os01g01060;Name=LOC_Os01g01060;Note=40S%20ribosomal%20protein%20S5%2C%20putative%2C%20expressed  
Chr1 MSU_osa1r7 gene 35581 41180 . + .  
ID=LOC_Os01g01080;Name=LOC_Os01g01080;Note=decarboxylase%2C%20putative%2C%20expressed  
Chr1 MSU_osa1r7 gene 44595 47526 . + .  
ID=LOC_Os01g01090;Name=LOC_Os01g01090;Note=retrotransposon%20protein%2C%20putative%2C%20unclassified%2C%20expressed  
Chr1 MSU_osa1r7 gene 47856 53412 . - .  
ID=LOC_Os01g01100;Name=LOC_Os01g01100;Note=retrotransposon%20protein%2C%20putative%2C%20Ty3-gypsy%20subclass%2C%20expressed  
Chr1 MSU_osa1r7 gene 58658 61090 . + .  
ID=LOC_Os01g01115;Name=LOC_Os01g01115;Note=pentatricopeptide%2C%20putative%2C%20expressed  
Chr1 MSU_osa1r7 gene 62059 63576 . + .  
ID=LOC_Os01g01120;Name=LOC_Os01g01120;Note=enolase-phosphatase%20E1%2C%20putative%2C%20expressed  
Chr1 MSU_osa1r7 gene 63244 66302 . - .  
ID=LOC_Os01g01130;Name=LOC_Os01g01130;Note=snurportin-1%2C%20putative%2C%20expressed  
Chr1 MSU_osa1r7 gene 72775 79938 . + .  
ID=LOC_Os01g01150;Name=LOC_Os01g01150;Note=RNA%20recognition%20motif%2C%20putative%2C%20expressed  
Chr1 MSU_osa1r7 gene 82428 84302 . + .  
ID=LOC_Os01g01160;Name=LOC_Os01g01160;Note=heat%20shock%20protein%20DnaJ%2C%20putative%2C%20expressed
```

```
[tranchant@node0 Bank]$ grep "putative" all.gff3 -v | head
```

```
##gff-version 3
```

```
Chr1 MSU_osa1r7 gene 2903 10817 . + .  
ID=LOC_Os01g01010;Name=LOC_Os01g01010;Note=TBC%20domain%20containing%20protein%2C%20expressed  
Chr1 MSU_osa1r7 mRNA 2903 10817 . + .  
ID=LOC_Os01g01010.1;Name=LOC_Os01g01010.1;Parent=LOC_Os01g01010  
Chr1 MSU_osa1r7 exon 2903 3268 . + .  
ID=LOC_Os01g01010.1:exon_1;Parent=LOC_Os01g01010.1  
Chr1 MSU_osa1r7 exon 3354 3616 . + .  
ID=LOC_Os01g01010.1:exon_2;Parent=LOC_Os01g01010.1  
Chr1 MSU_osa1r7 exon 4357 4455 . + .  
ID=LOC_Os01g01010.1:exon_3;Parent=LOC_Os01g01010.1  
Chr1 MSU_osa1r7 exon 5457 5560 . + .  
ID=LOC_Os01g01010.1:exon_4;Parent=LOC_Os01g01010.1  
Chr1 MSU_osa1r7 exon 7136 7944 . + .  
ID=LOC_Os01g01010.1:exon_5;Parent=LOC_Os01g01010.1  
Chr1 MSU_osa1r7 exon 8028 8150 . + .  
ID=LOC_Os01g01010.1:exon_6;Parent=LOC_Os01g01010.1  
Chr1 MSU_osa1r7 exon 8232 8320 . + .  
ID=LOC_Os01g01010.1:exon_7;Parent=LOC_Os01g01010.1
```

## Practice 10 : Blast analysis

Creating a custom database with `makeblastdb`

As we use a custom database for the first time, If we have a fasta format file of these sequences we have to create a database from our fasta format file `AllEst.fasta` with the `makeblastdb` command.

- Go into the bank directory and create a nucleotide database by typing:

```
makeblastdb -in AllEst.fasta -dbtype nucl -parse_seqids
```

- List the content of the directory to check if the database has been indexed

## BLASTing against our remote database

- Go into the blastAnalysis directory
- Run the blast by typing the following command with the outfmt equals to 6 :

```
blastn -query fastaFile -db databaseFile -outfmt [0-11] -out resultFile
```

## Parsing the results file

- Display the first 10 lines of the file - head

```
[tranchant@node0 Blast]$ pwd
/scratch/Blast
[tranchant@node0 Blast]$ ls NMT-EST.outfmt*
NMT-EST.outfmt0.blastn NMT-EST.outfmt6.blastn
```

- Display the first 15 lines of the file - head

```
[tranchant@node0 Blast]$ head NMT-EST.outfmt6.blastn
gi|33391745|gb|AY273814.1|      GW481759.1      94.922  709    14     2     188    896    1     687    0.0
1090
gi|33391745|gb|AY273814.1|      GW490598.1      93.624  643    19     2     498    1140   18     638    0.0
941
gi|33391745|gb|AY273814.1|      DV692795.1      89.523  754    50     8     1     731   14     761    0.0
928
gi|33391745|gb|AY273814.1|      GT690306.1      94.965  576    28     1     216    790   120    695    0.0
902
gi|33391745|gb|AY273814.1|      GW467935.1      97.852  512    11     0     220    731   1     512    0.0
885
gi|33391745|gb|AY273814.1|      GW476422.1      90.694  677    31     4     1     655   16     682    0.0
872
gi|33391745|gb|AY273814.1|      GW443231.1      97.760  491    11     0     205    695   110    600    0.0
846
gi|33391745|gb|AY273814.1|      GW443231.1      96.000  75     3     0     1     75    35     109    2.03e-26
122
gi|33391745|gb|AY273814.1|      GW397393.1      93.286  566    38     0     205    770   762    197    0.0
835
gi|33391745|gb|AY273814.1|      GW397393.1      89.535  172    17     1     859    1030  193    23     9.05e-55
217
```

- Display it last 15 lines - tail

```
[tranchant@node0 Blast]$ tail NMT-EST.outfmt6.blastn
gi|59799615|gb|AY918126.1|      GT646427.1      91.667  108    9     0     1017   1124   1     108    9.55e-35
150
```

gi 59799615 gb AY918126.1 139	GT678771.1	94.444	90	5	0	1621	1710	76	165	2.07e-31
gi 59799615 gb AY918126.1 139	GW450715.1	94.444	90	5	0	1621	1710	60	149	2.07e-31
gi 59799615 gb AY918126.1 137	GR987216.1	89.720	107	11	0	1018	1124	1	107	7.43e-31
gi 59799615 gb AY918126.1 132	GT729771.1	94.186	86	5	0	188	273	112	197	3.46e-29
gi 59799615 gb AY918126.1 97.1	DV704242.1	100.000	52	0	0	1073	1124	1	52	1.26e-18
gi 59799615 gb AY918126.1 97.1	DV706212.1	100.000	52	0	0	1073	1124	1	52	1.26e-18
gi 59799615 gb AY918126.1 97.1	DV708992.1	100.000	52	0	0	1073	1124	1	52	1.26e-18
gi 59799615 gb AY918126.1 97.1	DV713293.1	100.000	52	0	0	1073	1124	1	52	1.26e-18
gi 59799615 gb AY918126.1 87.9	GT732789.1	96.226	53	2	0	1072	1124	1	53	7.59e-16

- Count the number of line - wc

```
[tranchant@node0 Blast]$ wc -l NMT-EST.outfmt6.blastn
5537 NMT-EST.outfmt6.blastn
```

- Sort the lines using the second field (subject id) by alphabetical order, ascending then descending - csort

```
[tranchant@node0 Blast]$ head NMT-EST.outfmt6.blastn
```

gi 33391745 gb AY273814.1 1090	GW481759.1	94.922	709	14	2	188	896	1	687	0.0
gi 33391745 gb AY273814.1 941	GW490598.1	93.624	643	19	2	498	1140	18	638	0.0
gi 33391745 gb AY273814.1 928	DV692795.1	89.523	754	50	8	1	731	14	761	0.0
gi 33391745 gb AY273814.1 902	GT690306.1	94.965	576	28	1	216	790	120	695	0.0
gi 33391745 gb AY273814.1 885	GW467935.1	97.852	512	11	0	220	731	1	512	0.0
gi 33391745 gb AY273814.1 872	GW476422.1	90.694	677	31	4	1	655	16	682	0.0
gi 33391745 gb AY273814.1 846	GW443231.1	97.760	491	11	0	205	695	110	600	0.0
gi 33391745 gb AY273814.1 122	GW443231.1	96.000	75	3	0	1	75	35	109	2.03e-26
gi 33391745 gb AY273814.1 835	GW397393.1	93.286	566	38	0	205	770	762	197	0.0
gi 33391745 gb AY273814.1 217	GW397393.1	89.535	172	17	1	859	1030	193	23	9.05e-55

```
[tranchant@node0 Blast]$ csort -k2 NMT-EST.outfmt6.blastn | head
```

gi 59799612 gb AY918124.1	DV663631.1	100.000	207	0	0	435	641	1	207	9.85e-105	383
gi 59799614 gb AY918125.1	DV663631.1	100.000	207	0	0	439	645	1	207	1.02e-104	383
gi 20271027 gb AF494416.1	DV663631.1	92.565	269	20	0	288	556	1	269	4.34e-106	387
gi 33355460 gb AY273813.1	DV663631.1	92.565	269	20	0	359	627	1	269	5.14e-106	387
gi 37596302 gb AY362826.1	DV663631.1	94.007	267	16	0	342	608	3	269	1.39e-111	405
gi 59799617 gb AY918127.1	DV663631.1	95.631	206	9	0	428	633	2	207	3.21e-89	331
gi 85700272 gb DQ348078.1	DV663631.1	96.875	64	2	0	1601	1664	206	269	9.06e-22	108
gi 66774631 gb DQ010011.1	DV663631.1	96.875	64	2	0	1602	1665	206	269	8.96e-22	108
gi 85700270 gb DQ348077.1	DV663631.1	96.875	64	2	0	1608	1671	206	269	8.38e-22	108
gi 59799615 gb AY918126.1	DV663631.1	96.875	64	2	0	860	923	206	269	5.83e-22	108

```
[tranchant@node0 Blast]$ csort -k2r NMT-EST.outfmt6.blastn | head
```

gi 85700272 gb DQ348078.1	GW492135.1	98.684	76	1	0	2547	2622	1	76	4.16e-30	135
gi 66774631 gb DQ010011.1	GW492135.1	98.684	76	1	0	2518	2593	1	76	4.11e-30	135
gi 85700270 gb DQ348077.1	GW492135.1	98.684	76	1	0	2349	2424	1	76	3.84e-30	135
gi 59799614 gb AY918125.1	GW492135.1	98.684	76	1	0	1900	1975	1	76	3.14e-30	135
gi 59799612 gb AY918124.1	GW492135.1	98.684	76	1	0	1840	1915	1	76	3.04e-30	135
gi 33391745 gb AY273814.1	GW492135.1	98.000	50	1	0	1619	1668	1	50	7.40e-16	87.9
gi 20271023 gb AF494414.1	GW492135.1	98.000	50	1	0	1088	1137	1	50	5.01e-16	87.9
gi 20271027 gb AF494416.1	GW492135.1	98.000	50	1	0	1070	1119	1	50	4.93e-16	87.9
gi 20271025 gb AF494415.1	GW492135.1	98.000	50	1	0	1067	1116	1	50	4.91e-16	87.9

```

gj|37596302|gb|AY362826.1| GW492135.1 96.104 77 3 0 1161 1237 1 77 1.21e-27 126
[tranchant@node0 Blast]$

```

- Sort lines by e-value (ascending) and by "alignment length" (descending) - csort

```

[tranchant@node0 Blast]$ head NMT-EST.outfmt6.blastn
gj|33391745|gb|AY273814.1| GW481759.1 94.922 709 14 2 188 896 1 687 0.0 1090
gj|33391745|gb|AY273814.1| GW490598.1 93.624 643 19 2 498 1140 18 638 0.0 941
gj|33391745|gb|AY273814.1| DV692795.1 89.523 754 50 8 1 731 14 761 0.0 928
gj|33391745|gb|AY273814.1| GT690306.1 94.965 576 28 1 216 790 120 695 0.0 902
gj|33391745|gb|AY273814.1| GW467935.1 97.852 512 11 0 220 731 1 512 0.0 885
gj|33391745|gb|AY273814.1| GW476422.1 90.694 677 31 4 1 655 16 682 0.0 872
gj|33391745|gb|AY273814.1| GW443231.1 97.760 491 11 0 205 695 110 600 0.0 846
gj|33391745|gb|AY273814.1| GW443231.1 96.000 75 3 0 1 75 35 109 2.03e-26 122
gj|33391745|gb|AY273814.1| GW397393.1 93.286 566 38 0 205 770 762 197 0.0 835
gj|33391745|gb|AY273814.1| GW397393.1 89.535 172 17 1 859 1030 193 23 9.05e-55 217

```

```

[tranchant@node0 Blast]$ csort -k11g NMT-EST.outfmt6.blastn | head
gj|20271023|gb|AF494414.1| DV663736.1 94.565 460 22 1 2 458 1 460 0.0 708
gj|20271023|gb|AF494414.1| DV664413.1 99.790 476 0 1 1 475 35 510 0.0 872
gj|20271023|gb|AF494414.1| DV664589.1 93.558 652 41 1 160 811 1 651 0.0 970
gj|20271023|gb|AF494414.1| DV664642.1 99.398 498 3 0 640 1137 1 498 0.0 904
gj|20271023|gb|AF494414.1| DV665482.1 99.506 405 2 0 733 1137 1 405 0.0 737
gj|20271023|gb|AF494414.1| DV667027.1 93.703 667 42 0 1 667 29 695 0.0 1000
gj|20271023|gb|AF494414.1| DV667098.1 99.748 397 1 0 741 1137 1 397 0.0 728
gj|20271023|gb|AF494414.1| DV667592.1 94.937 474 21 1 1 471 45 518 0.0 739
gj|20271023|gb|AF494414.1| DV667823.1 99.219 512 4 0 202 713 8 519 0.0 924
gj|20271023|gb|AF494414.1| DV671853.1 94.208 518 27 1 1 515 48 565 0.0 787

```

```

[tranchant@node0 Blast]$ csort -k11g -k4gr NMT-EST.outfmt6.blastn | head
gj|33355460|gb|AY273813.1| DV706163.1 88.035 911 74 12 414 1302 1 898 0.0 1048
gj|33355460|gb|AY273813.1| DV712406.1 88.145 911 73 12 414 1302 1 898 0.0 1051
gj|37596302|gb|AY362826.1| GT000064.1 88.815 903 81 9 312 1213 11 894 0.0 1090
gj|37596302|gb|AY362826.1| DV703202.1 91.232 901 73 4 15 911 1 899 0.0 1221
gj|33355460|gb|AY273813.1| DV703202.1 93.207 898 60 1 33 930 3 899 0.0 1319
gj|33355460|gb|AY273813.1| GT000064.1 90.498 884 63 4 331 1193 11 894 0.0 1147
gj|20271023|gb|AF494414.1| DV000064.1 99.659 881 3 0 257 1137 11 891 0.0 1611
gj|20271025|gb|AF494415.1| GT000064.1 93.076 881 40 4 257 1116 11 891 0.0 1269
gj|20271027|gb|AF494416.1| GT000064.1 90.579 881 62 4 260 1119 11 891 0.0 1147
gj|33355460|gb|AY273813.1| GR993866.1 92.700 863 59 2 13 875 11 869 0.0 1242

```

- Extract the first 4 fields - cut

```

[tranchant@node0 Blast]$ cut -f 1-4 NMT-EST.outfmt6.blastn | head
gj|33391745|gb|AY273814.1| GW481759.1 94.922 709
gj|33391745|gb|AY273814.1| GW490598.1 93.624 643
gj|33391745|gb|AY273814.1| DV692795.1 89.523 754
gj|33391745|gb|AY273814.1| GT690306.1 94.965 576
gj|33391745|gb|AY273814.1| GW467935.1 97.852 512
gj|33391745|gb|AY273814.1| GW476422.1 90.694 677
gj|33391745|gb|AY273814.1| GW443231.1 97.760 491
gj|33391745|gb|AY273814.1| GW443231.1 96.000 75
gj|33391745|gb|AY273814.1| GW397393.1 93.286 566
gj|33391745|gb|AY273814.1| GW397393.1 89.535 172

```

- Extract query id, subject id, evalue, alignment length cut

```

[tranchant@node0 Blast]$ cut -f 1,2,11,4 NMT-EST.outfmt6.blastn | head
gj|33391745|gb|AY273814.1| GW481759.1 709 0.0
gj|33391745|gb|AY273814.1| GW490598.1 643 0.0
gj|33391745|gb|AY273814.1| DV692795.1 754 0.0
gj|33391745|gb|AY273814.1| GT690306.1 576 0.0
gj|33391745|gb|AY273814.1| GW467935.1 512 0.0
gj|33391745|gb|AY273814.1| GW476422.1 677 0.0
gj|33391745|gb|AY273814.1| GW443231.1 491 0.0
gj|33391745|gb|AY273814.1| GW443231.1 75 2.03e-26
gj|33391745|gb|AY273814.1| GW397393.1 566 0.0
gj|33391745|gb|AY273814.1| GW397393.1 172 9.05e-55

```

## Practice 11 : Redirecting a command output to a File with >

- Extract all ESTs identifiers and print them in the file ESTs\_accession.list - cut >

```
[tranchant@node0 Blast]$ cut -f 2 NMT-EST.outfmt6.blastn | wc -l
5537
```

```
[tranchant@node0 Blast]$ cut -f 2 NMT-EST.outfmt6.blastn > EST_accession.list
```

## Practice 12 : Sending data from one command to another (piping) with |

- How many sequences have a homology with EST sequences ? (TIPs: cut command with sort -u (uniq) or uniq command))

```
[tranchant@node0 Blast]$ cut -f 2 NMT-EST.outfmt6.blastn | sort -u | wc -l
293
```

## Practice 13 : Running many commands with | - grep

To get some basics stats of the output VCF file

/scratch/LINUX-TP/VCF/OgOb-all-MSU7-CHR6.GATK VARIANTFILTRATION.vcf, let's use linux command!

- How many raw polymorphisms were detected (Displaying all the lines which does not start with # / header lines)?

```
[tranchant@node0 vcf]$ pwd
/home/tranchant/LINUX-TP/Data/vcf
```

```
[tranchant@node0 vcf]$ ls
200000-CHR9.GATK VARIANTFILTRATION.vcf  duplicVCF  singleVCF  testsnmf.geno  vcfForRecalibration  vcfForSNIPlay
```

- How many polymorphisms were considered good after filtering steps by GATK VARIANTFILTRATION (ie marked PASS)?

```
[tranchant@node0 vcf]$ grep "#" 200000-CHR9.GATK VARIANTFILTRATION.vcf | head -n1
##fileformat=VCFv4.1
```

```
[tranchant@node0 vcf]$ grep "#" 200000-CHR9.GATK VARIANTFILTRATION.vcf -v | head -n1
Chr9      37898      .          G          A          780.81     SnpCluster
AC=31;AF=0.170;AN=182;BaseQRankSum=4.013;DP=223;Dels=0.00;FS=0.000;HaplotypeScore=2.3114;InbreedingCoeff=-0.1444;MLEAC=38;MLEAF=0.209;MQ=26.52;
MQ0=0;MQRankSum=2.119;QD=7.66;ReadPosRankSum=5.319;SOR=0.941  GT:AD:DP:GQ:PL  ./  0/0:1,0:1:3:0,3,22  ./  0/0:1,0:1:3:0,3,33
./  ./  ./  ./  0/0:2,0:2:6:0,6,55./  ./  ./  ./  0/0:3,0:3:6:0,6,55  ./  ./  ./
./  0/0:1,0:1:3:0,3,32  ./  ./  ./  ./  ./  0/0:1,0:1:3:0,3,33  ./  ./  ./
./  0/0:1,0:1:3:0,3,33  ./  ./  ./  ./  ./  0/0:1,0:1:3:0,3,33  0/0:1,0:1:3:0,3,22  ./
./  ./  0/0:1,0:1:3:0,3,33  ./  ./  ./  ./  ./  0/0:2,0:2:6:0,6,55  ./
./  ./  0/0:1,0:1:3:0,3,33  ./  ./  ./  ./  0/0:1,0:1:3:0,3,22  ./  ./  ./
./  ./  0/0:1,0:1:3:0,3,22  ./  0/0:1,0:1:3:0,3,33  ./  0/0:1,0:1:3:0,3,31  ./  ./  ./
0/0:1,0:1:3:0,3,22  0/0:1,0:1:3:0,3,32  0/0:2,0:2:3:0,3,22  ./  0/0:1,0:1:3:0,3,32  0/0:1,0:1:3:0,3,29  ./  ./  ./
./  ./  ./  ./  ./  0/0:1,0:1:3:0,3,33  0/0:1,0:1:3:0,3,33  0/0:1,0:1:3:0,3,33  0/0:1,0:1:3:0,3,33  ./
./  ./  ./  ./  0/0:1,0:1:3:0,3,33  0/0:1,0:1:3:0,3,25  0/0:1,0:1:3:0,3,33  0/0:1,0:1:3:0,3,33  ./
0/0:1,0:1:3:0,3,33  ./  ./  ./  ./  0/0:1,0:1:3:0,3,28  ./  0/0:2,0:2:6:0,6,55  0/0:1,0:1:3:0,3,22  0/0:1,0:1:3:0,3,33./  ./
./  0/0:1,0:1:3:0,3,33  ./  ./  ./  ./  0/0:1,0:1:3:0,3,33  0/0:2,0:2:6:0,6,67  ./  0/0:1,0:1:3:0,3,33
./  ./  ./  0/0:1,0:1:3:0,3,22./  ./  ./  ./  ./  0/0:1,0:1:3:0,3,33  0/0:1,0:1:3:0,3,33
0/0:1,0:1:3:0,3,33  ./  ./  ./  ./  ./  0/0:1,0:1:3:0,3,25  ./  ./  ./
0/0:1,0:1:3:0,3,33  ./  0/0:1,0:1:3:0,3,33  0/0:1,0:1:3:0,3,25  ./  ./  ./  ./
./  ./  ./  ./  ./  0/0:1,0:1:3:0,3,33./  0/0:3,0:3:9:0,9,78  ./  0/1:5,3:8:76:76,0,131
0/1:3,1:4:21:21,0,88  0/0:5,0:5:15:0,15,155  0/1:1,1:2:27:27,0,27  0/1:3,2:5:50:55,0,50  0/1:2,4:6:37:116,0,37  0/1:4,3:7:68:68,0,101  0/1:1,1:2:16:27,0,16
1/1:0,1:1:3:22,3,0  0/0:1,0:1:3:0,3,33  0/0:4,0:4:12:0,12,122  0/0:1,0:1:3:0,3,33  ./  1/1:0,1:1:3:22,3,0  ./  0/0:3,0:3:9:0,9,89
0/1:5,1:6:15:15,0,144  0/1:4,1:5:18:18,0,102  0/1:1,1:2:16:16,0,27  0/0:5,0:5:15:0,15,144  0/0:1,0:1:3:0,3,33  0/1:4,1:5:18:18,0,107  0/1:2,1:3:13:13,0,58  ./
```

```
1/1:0,2:2:6:67,6,0 0/0:1,0:1:3:0,3,33 0/0:2,0:2:6:0,6,67 ./ ./ 0/0:5,0:5:15:0,15,163 0/1:5,4:9:99:107,0,140 0/0:9,0:9:27:0,27,254
0/0:3,0:3:9:0,9,89 0/1:1,1:2:27:27,0,27 0/1:4,1:5:18:18,0,119 0/1:4,2:7:37:37,0,116 0/0:1,0:1:3:0,3,33 0/0:2,0:2:6:0,6,55 0/0:1,0:1:3:0,3,33
0/1:1,2:3:24:46,0,24 0/1:2,1:3:24:24,0,58 0/1:1,1:2:27:27,0,27 0/0:3,0:3:9:0,9,100 1/1:0,2:2:6:55,6,00/0:1,1:2:3:0,3,30 0/0:1,0:1:3:0,3,22
0/1:4,2:6:49:49,0,104 0/0:1,0:1:3:0,3,33 0/1:2,1:3:13:13,0,58 1/1:0,1:1:3:22,3,0 0/0:4,0:4:12:0,12,122 ./ ./ ./ ./
./ ./ ./ ./
```

```
[tranchant@node0 vcf]$ grep "#" 200000-CHR9.GATKVARIANTFILTRATION.vcf -v -c
200000
```

How many polymorphism were considered bad and filtered out (Displaying all the lines without the PASS tag )?

```
[tranchant@node0 vcf]$ grep "#" 200000-CHR9.GATKVARIANTFILTRATION.vcf -v | grep "PASS" | head -n2
```

```
Chr9 38076 . TG T 783.24 PASS
AC=26;AF=1.00;AN=26;DP=19;FS=0.000;InbreedingCoeff=-0.1710;MLEAC=26;MLEAF=1.00;MQ=27.46;MQ0=0;QD=32.36;SOR=5.892 GT:AD:DP:GQ:PL ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ 1/1:0,1:1:3:45,3,0 ./ 1/1:0,1:1:3:45,3,0 ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ 1/1:0,1:1:3:45,3,0 1/1:0,1:1:3:45,3,0 ./ ./ ./ ./ ./ ./ ./ ./
1/1:0,1:1:3:45,3,0 ./ ./ ./ 1/1:0,2:2:6:91,6,0 1/1:0,4:4:12:182,12,0 1/1:0,1:1:3:45,3,0/1:0,2:2:6:91,6,0 1/1:0,1:1:3:45,3,0
./ ./ 1/1:0,1:1:3:45,3,0 ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
Chr9 38518 . A T 45816.78 PASS
AC=483;AF=0.970;AN=498;BaseQRankSum=2.263;DP=1504;Dels=0.00;FS=0.000;HaplotypeScore=0.1188;InbreedingCoeff=-0.0550;MLEAC=489;MLEAF=0.982;MQ=28
.02;MQ0=0;MQRankSum=0.204;QD=30.46;ReadPosRankSum=7.023;SOR=8.486 GT:AD:DP:GQ:PL 1/1:0,7:7:15:165,15,0 0/1:1,2:3:24:58,0,24
1/1:0,4:4:9:100,9,0 1/1:0,2:2:6:62,6,0 1/1:0,5:5:15:167,15,0 1/1:0,8:8:24:256,24,0 1/1:0,4:4:12:120,12,0 1/1:0,5:5:15:155,15,0 1/1:0,5:5:12:134,12,0
1/1:0,4:4:12:134,12,0 ./ 1/1:0,4:4:9:89,9,0 1/1:0,11:11:33:354,33,0 1/1:0,6:6:18:177,18,0 1/1:0,3:3:9:100,9,0 1/1:0,10:10:30:270,30,0
1/1:0,13:13:39:423,39,0 1/1:0,11:11:33:365,33,0 1/1:0,5:5:15:165,15,0 1/1:0,6:6:18:201,18,0 1/1:0,2:2:6:67,6,0 1/1:0,8:8:21:214,21,0
1/1:0,3:3:9:77,9,0 1/1:0,10:10:27:277,27,0 1/1:0,3:3:9:100,9,0 1/1:0,9:9:27:301,27,0 1/1:0,8:8:24:267,24,0 1/1:0,6:6:18:201,18,0
1/1:0,12:12:36:399,36,0 1/1:0,2:2:6:55,6,0 1/1:0,4:4:12:122,12,0 1/1:0,3:3:9:100,9,0 1/1:0,6:6:18:201,18,0 1/1:0,6:6:18:193,18,0 1/1:0,6:6:18:195,18,0
1/1:0,2:2:6:67,6,0 1/1:0,3:3:9:100,9,0 1/1:0,8:8:21:234,21,0 1/1:0,8:8:24:232,24,0 1/1:0,3:3:9:100,9,0 1/1:0,5:5:12:134,12,0 1/1:0,8:8:24:243,24,0
0/1:1,4:5:17:107,0,17 1/1:0,1:1:3:33,3,0 1/1:0,6:6:15:155,15,0 1/1:0,5:5:15:155,15,0 1/1:0,4:4:12:134,12,0 1/1:0,6:6:18:189,18,0 1/1:0,4:4:12:122,12,0
1/1:0,2:2:6:67,6,0 1/1:0,7:7:21:222,21,0 1/1:0,7:7:21:234,21,0 1/1:0,4:4:12:122,12,0 1/1:0,15:15:45:485,45,0 1/1:0,6:6:18:189,18,0
1/1:0,4:4:9:100,9,1/1:0,8:8:24:251,24,0 1/1:0,7:7:21:222,21,0 1/1:0,6:6:15:159,15,0 1/1:0,3:3:9:95,9,0 1/1:0,10:10:30:334,30,0 1/1:0,5:5:15:167,15,0
1/1:0,10:10:30:334,30,0 1/1:0,7:7:21:210,21,0 1/1:0,8:8:24:256,24,0 1/1:0,2:2:6:67,6,0 1/1:0,5:5:15:167,15,0 1/1:0,7:7:21:222,21,0
1/1:0,1:1:3:33,3,0 1/1:0,6:6:18:189,18,0 1/1:0,11:11:33:344,33,0 1/1:0,5:5:15:155,15,0 1/1:0,12:12:36:389,36,0 1/1:0,7:7:18:201,18,0
1/1:0,4:4:12:110,12,0 1/1:0,14:14:42:456,42,0 1/1:0,3:3:9:89,9,0 1/1:0,1:1:3:28,3,0 1/1:0,10:10:30:299,30,0 1/1:0,11:11:33:342,33,0
1/1:0,8:8:24:256,24,0 1/1:0,1:1:3:33,3,0 1/1:0,5:5:15:165,15,0 1/1:0,7:7:18:177,18,0 1/1:0,10:10:30:318,30,0 1/1:0,5:5:15:166,15,0
1/1:0,9:9:27:289,27,0 1/1:0,5:5:15:167,15,0 1/1:0,10:10:30:323,30,0 1/1:0,11:11:33:332,33,0 1/1:0,8:8:24:267,24,0 1/1:0,5:5:15:164,15,0
1/1:0,5:5:15:155,15,0 1/1:0,9:9:24:265,24,0 1/1:0,5:5:15:161,15,0 1/1:0,9:9:27:301,27,0 1/1:0,8:8:24:267,24,0 1/1:0,5:5:15:155,15,0 1/1:0,7:7:21:222,21,0
1/1:0,6:6:18:177,18,0 1/1:0,4:4:12:134,12,0 1/1:0,14:14:42:445,42,0 1/1:0,2:2:6:67,6,0 1/1:0,8:8:24:244,24,0 1/1:0,6:6:18:189,18,0 1/1:0,8:8:24:244,24,0
1/1:0,4:4:12:134,12,0 1/1:0,9:9:27:301,27,0 1/1:0,4:4:12:133,12,0 1/1:0,4:4:12:120,12,0 1/1:0,4:4:12:134,12,0 1/1:0,7:7:21:187,21,0 1/1:0,8:8:24:256,24,0
1/1:0,2:2:6:43,6,0 1/1:0,2:2:6:55,6,0/1:0,4:4:12:134,12,0 1/1:0,11:11:30:302,30,0 1/1:0,5:5:15:167,15,0 1/1:0,7:7:21:222,21,0
1/1:0,5:5:12:122,12,0 1/1:0,7:7:15:165,15,0 1/1:0,10:10:30:323,30,0 1/1:0,8:8:24:256,24,0 1/1:0,4:4:12:122,12,0 1/1:0,7:7:18:201,18,0 1/1:0,6:6:18:201,18,0
1/1:0,4:4:12:134,12,0 1/1:0,5:5:12:134,12,0 1/1:0,6:6:18:189,18,0 1/1:0,8:8:24:267,24,0 1/1:0,6:6:18:185,18,0 1/1:0,7:7:21:210,21,0 1/1:0,7:7:21:210,21,0
1/1:0,8:8:24:267,24,0 1/1:0,4:4:9:100,9,0 1/1:0,3:3:9:100,9,0 1/1:0,8:8:24:256,24,0 1/1:0,3:3:9:103,9,0 1/1:0,8:8:24:267,24,0 1/1:0,4:4:12:134,12,0
1/1:0,4:4:12:122,12,0 1/1:0,17:17:51:554,51,0 1/1:0,4:4:12:134,12,0 1/1:0,4:4:12:134,12,0 1/1:0,5:5:15:166,15,0 1/1:0,8:8:24:256,24,0
1/1:0,9:9:24:267,24,0 1/1:0,8:8:24:250,24,0 1/1:0,4:4:12:128,12,0 1/1:0,4:4:12:134,12,0 0/1:1,6:7:12:168,0,12 1/1:0,9:9:27:277,27,0 1/1:0,8:8:24:267,24,0
1/1:0,6:6:18:189,18,0 1/1:0,3:3:6:55,6,0 1/1:0,9:9:27:277,27,0 1/1:0,6:6:18:165,18,0 1/1:0,6:6:18:189,18,0 1/1:0,4:4:12:131,12,0 1/1:0,4:4:12:122,12,0
1/1:0,5:5:12:134,12,0 1/1:0,10:10:30:323,30,0 1/1:0,8:8:24:220,24,0 0/1:1,6:7:12:162,0,1 1/1:0,1:1:3:33,3,0 1/1:0,9:9:27:301,27,0 1/1:0,8:8:24:256,24,0
1/1:0,9:9:27:289,27,0 1/1:0,3:3:9:89,9,0 1/1:0,4:4:12:122,12,0 1/1:0,5:5:15:155,15,0 1/1:0,5:5:15:155,15,0 1/1:0,7:7:21:216,21,0 1/1:0,6:6:18:199,18,0
1/1:0,3:3:9:100,9,0 1/1:0,6:6:18:201,18,0 1/1:0,6:6:18:177,18,0 1/1:0,4:4:12:134,12,0 1/1:0,2:2:6:67,6,0 1/1:0,5:5:15:161,15,0
1/1:0,12:12:36:395,36,0 1/1:0,7:7:21:234,21,0 1/1:0,3:3:9:99,9,0 1/1:0,6:6:18:201,18,0 1/1:0,7:7:15:155,15,0 1/1:0,8:8:24:244,24,0 1/1:0,7:7:21:228,21,0
1/1:0,6:6:18:183,18,0 1/1:0,9:9:27:277,27,0 1/1:0,1:1:3:33,3,0 1/1:0,1:1:3:33,3,0 1/1:0,6:6:18:201,18,0 0/1:1,6:7:12:149,0,15
1/1:1,11:12:14:334,14,0 1/1:0,4:4:12:134,12,0 1/1:0,4:4:12:134,12,0 1/1:0,6:6:18:201,18,0 1/1:0,7:7:21:210,21,0 1/1:0,2:2:6:67,6,0/1:0,6:6:18:189,18,0
0/1:1,10:11:3:271,0,3 1/1:0,9:9:27:277,27,0 1/1:0,1:1:3:33,3,0 1/1:0,5:5:15:155,15,0 1/1:0,5:5:15:154,15,0 1/1:0,7:7:18:201,18,0 0/1:1,4:5:18:117,0,18
0/1:1,9:10:6:229,0,6 0/1:1,4:5:18:119,0,18 1/1:0,4:4:12:132,12,0 1/1:0,8:8:24:256,24,0 1/1:0,5:5:15:167,15,0 1/1:0,6:6:18:182,18,0 1/1:0,3:3:9:104,9,0
1/1:0,1:1:3:29,3,0 1/1:0,3:3:9:100,9,1/1:0,5:5:15:167,15,0 1/1:0,7:7:21:222,21,0 1/1:0,5:5:15:167,15,0 1/1:0,8:8:24:265,24,0 1/1:0,3:3:9:100,9,0
1/1:0,9:9:24:267,24,0 1/1:0,5:5:15:167,15,0 1/1:1,4:5:12:134,12,0 0/1:1,3:4:21:88,0,21 0/1:1,2:3:27:27,0,27 1/1:0,3:3:9:100,9,0 1/1:0,3:3:9:89,9,0
0/1:1,4:5:18:107,0,18 1/1:0,6:6:18:189,18,0 1/1:0,3:3:9:89,9,0 0/1:1,4:5:18:111,0,18 1/1:0,3:3:9:100,9,1/1:0,6:6:18:189,18,0 1/1:0,8:8:24:275,24,0
1/1:0,5:5:15:144,15,0 0/1:1,5:6:15:149,0,15 1/1:0,4:4:12:134,12,0 0/1:1,7:8:9:207,0,9 1/1:0,4:4:12:140,12,0 1/1:0,3:3:9:100,9,0 1/1:0,4:4:9:100,9,0
1/1:0,9:9:27:266,27,0 1/1:0,12:12:36:389,36,0 1/1:0,12:12:36:389,36,0 1/1:0,7:7:18:177,18,0 1/1:0,8:8:21:210,21,0 1/1:0,7:7:18:189,18,0
1/1:0,7:7:18:189,18,0 1/1:0,2:2:6:55,6,0
```

```
[tranchant@node0 vcf]$ grep "#" 200000-CHR9.GATKVARIANTFILTRATION.vcf -v | grep "PASS" -v | head -n2
```

```
Chr9 37898 . G A 780.81 SnpCluster
AC=31;AF=0.170;AN=182;BaseQRankSum=4.013;DP=223;Dels=0.00;FS=0.000;HaplotypeScore=2.3114;InbreedingCoeff=-0.1444;MLEAC=38;MLEAF=0.209;MQ=26.52;
```



```

MQ0=0;MQRankSum=2.119;QD=7.66;ReadPosRankSum=5.319;SOR=0.941 GT:AD:DP:GQ:PL ./ 0/0:1,0:1:3:0,3,22 ./ 0/0:1,0:1:3:0,3,33
./ ./ ./ ./ 0/0:2,0:2:6:0,6,55./ ./ ./ ./ ./ 0/0:3,0:3:6:0,6,55 ./ ./ ./
./ 0/0:1,0:1:3:0,3,32 ./ ./ ./ ./ ./ ./ 0/0:1,0:1:3:0,3,33 ./ ./ ./
./ 0/0:1,0:1:3:0,3,33 ./ ./ ./ ./ ./ ./ ./ 0/0:1,0:1:3:0,3,22
./ ./ 0/0:1,0:1:3:0,3,33 ./ ./ ./ ./ ./ ./ 0/0:2,0:2:6:0,6,55 ./
./ 0/0:1,0:1:3:0,3,33 ./ 0/0:1,0:1:3:0,3,33 ./ 0/0:1,0:1:3:0,3,22 0/0:1,0:1:3:0,3,22 ./
0/0:1,0:1:3:0,3,22 0/0:1,0:1:3:0,3,32 0/0:2,0:2:3:0,3,22 ./ 0/0:1,0:1:3:0,3,32 0/0:1,0:1:3:0,3,29 ./ ./ ./
./ ./ ./ ./ 0/0:1,0:1:3:0,3,33 0/0:1,0:1:3:0,3,33 0/0:1,0:1:3:0,3,33 0/0:1,0:1:3:0,3,33 ./ ./
0/0:1,0:1:3:0,3,33 ./ ./ 0/0:1,0:1:3:0,3,33 0/0:1,0:1:3:0,3,33 0/0:1,0:1:3:0,3,33 0/0:1,0:1:3:0,3,33 ./
0/0:1,0:1:3:0,3,28 ./ ./ 0/0:1,0:1:3:0,3,33 ./ 0/0:2,0:2:6:0,6,55 0/0:1,0:1:3:0,3,22 0/0:1,0:1:3:0,3,33./
./ 0/0:1,0:1:3:0,3,33 ./ ./ 0/0:1,0:1:3:0,3,33 0/0:1,0:1:3:0,3,33 0/0:2,0:2:6:0,6,67 ./ 0/0:1,0:1:3:0,3,33
./ ./ ./ 0/0:1,0:1:3:0,3,22./ ./ ./ ./ ./ ./ 0/0:1,0:1:3:0,3,33
0/0:1,0:1:3:0,3,33 ./ ./ ./ ./ ./ ./ ./ ./ ./ 0/0:1,0:1:3:0,3,33
0/0:1,0:1:3:0,3,33 ./ 0/0:1,0:1:3:0,3,33 ./ ./ 0/0:1,0:1:3:0,3,25 ./ ./ 0/0:1,0:1:3:0,3,25
./ ./ ./ 0/0:1,0:1:3:0,3,33./ 0/0:3,0:3:9:0,9,78 ./ 0/1,5:3:8:76:76,0,131
0/1:3,1:4:21:21,0,88 0/0:5,0:5:15:0,15,155 0/1:1,1:2:27:27,0,27 0/1:3,2:5:50:55,0,50 0/1:2,4:6:37:116,0,37 0/1:4,3:7:68:68,0,101 0/1:1,1:2:16:27,0,16
1/1:0,1:1:3:22,3,0 0/0:1,0:1:3:0,3,33 0/0:4,0:4:12:0,12,122 0/0:1,0:1:3:0,3,33 ./ 1/1:0,1:1:3:22,3,0 ./ 0/0:3,0:3:9:0,9,89
0/1:5,1:6:15:15,0,144 0/1:4,1:5:18:18,0,102 0/1:1,1:2:16:16,0,27 0/0:5,0:5:15:0,15,144 0/0:1,0:1:3:0,3,33 0/1:4,1:5:18:18,0,107 0/1:2,1:3:13:13,0,58 ./
1/1:0,2:2:6:67,6,0 0/0:1,0:1:3:0,3,33 0/0:2,0:2:6:0,6,67 ./ ./ 0/0:5,0:5:15:0,15,163 0/1:5,4:9:99:107,0,140 0/0:9,0:9:27:0,27,254
0/0:3,0:3:9:0,9,89 0/1:1,1:2:27:27,0,27 0/1:4,1:5:18:18,0,119 0/1:4,2:7:37:37,0,116 0/0:1,0:1:3:0,3,33 0/0:2,0:2:6:0,6,55 0/0:1,0:1:3:0,3,33
0/1:1,2:3:24:46,0,24 0/1:2,1:3:24:24,0,58 0/1:1,1:2:27:27,0,27 0/0:3,0:3:9:0,9,100 1/1:0,2:2:6:55,6,00/0:1,1:2:3:0,3,30 0/0:1,0:1:3:0,3,22
0/1:4,2:6:49:49,0,104 0/0:1,0:1:3:0,3,33 0/1:2,1:3:13:13,0,58 1/1:0,1:1:3:22,3,0 0/0:4,0:4:12:0,12,122 ./ ./ ./
./ ./ ./ ./
Chr9 37976 . A T 570.76 SnpCluster
AC=37;AF=0.712;AN=52;BaseQRankSum=3.050;DP=31;Dels=0.00;FS=0.000;HaplotypeScore=0.0282;InbreedingCoeff=0.2376;MLEAC=37;MLEAF=0.712;MQ=27.42;M
Q0=0;MQRankSum=0.659;QD=25.94;ReadPosRankSum=1.147;SOR=1.839 GT:AD:DP:GQ:PL ./ ./ ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
0/0:1,0:1:3:0,3,33 ./ ./ ./ ./ ./ ./ ./ ./ ./ ./ ./
./ ./ ./ ./ 1/1:0,1:1:3:33,3,0 ./ ./ ./
./ ./ ./ 1/1:0,1:1:3:33,3,0 ./ ./ ./
./ ./ ./ ./ ./ ./ ./ ./ ./ ./
0/0:1,0:1:3:0,3,23 ./ ./ ./ ./ ./ 1/1:0,1:1:3:33,3,0 ./ ./
./ ./ ./ ./ ./ ./ 1/1:0,1:1:3:31,3,0 0/0:1,0:1:3:0,3,33 ./
./ ./ ./ 0/1:1,1:2:24:27,0,24 ./ ./
./ ./ 0/0:1,0:1:3:0,3,22 ./ ./ ./ ./ ./ 1/1:0,1:1:3:33,3,0 ./
./ ./ ./ ./ ./ 1/1:0,1:1:3:33,3,0 ./ ./
./ ./ ./ 1/1:0,1:1:3:22,3,0 0/0:1,0:1:3:0,3,33 1/1:0,1:1:3:22,3,0 ./
./ ./ ./ 1/1:0,1:1:3:22,3,0 ./ ./ 1/1:0,1:1:3:33,3,0
1/1:0,2:2:6:55,6,0 ./ 1/1:0,1:1:3:33,3,0 ./ 1/1:0,2:2:6:67,6,0 ./
1/1:0,1:1:3:33,3,0/1:0,1:1:3:33,3,0 ./ 0/0:1,0:1:3:0,3,22 1/1:0,1:1:3:33,3,0 ./ 1/1:0,1:1:3:33,3,0
./ ./ ./ ./ ./

```

```
[tranchant@node0 vcf]$ grep "#" 200000-CHR9.GATKVARIANTFILTRATION.vcf -v | grep "PASS" -c
65621
```

```
[tranchant@node0 vcf]$ grep "#" 200000-CHR9.GATKVARIANTFILTRATION.vcf -v | grep "PASS" -c -v
134379
```

## BONUS

- Renommer les fichiers fastq avec commande rename dans le répertoire (on retire le mot Mapped du nom du fichier) - /home/tranchant/LINUX-TP/Data/fastq/assembly/pairedOneIndividuPacaya

```
[tranchant@node0 pairedOneIndividuPacaya]$ pwd
/home/tranchant/LINUX-TP/Data/fastq/assembly/pairedOneIndividuPacaya
```

```
[tranchant@node0 pairedOneIndividuPacaya]$ ls
g02L5Mapped_R1.fq g02L5Mapped_R2.fq
```

```
[tranchant@node0 pairedOneIndividuPacaya]$ rename "Mapped" "" *fq
```

```
[tranchant@node0 pairedOneIndividuPacaya]$ ls
g02L5_R1.fq g02L5_R2.fq
```



- Substituer un motif par un autre, ici ">id" par ">chr\_id" dans le fichier /home/tranchant/LINUX-TP/Bank/referencelrigin.fasta

```
[tranchant@node0 Bank]$ pwd
/home/tranchant/LINUX-TP/Bank
```

```
[tranchant@node0 Bank]$ head referencelrigin.fasta
>21054 3211 65455
CGCCGCGGAGCGGGCGCCGCCGCTGTACATGTCGCCGTGCATCGTGTGCACGTGCGG
GCCCCGTACGTGCCACCCCCACACTGTGCGTGCATCGTGCCTCTGCCGGTGGGAG
TTGCCACGCGCTGGCGACCCATACTTCTTTATTTCTTTGGTCCGTGTACGTGCAGGTT
AATCGATCGCCTCGTCATTTTTACTCGGAACTCGCTGGTGAACCCGCCACATTTCAAG
TAGTAGTGGTAGCACGTACGAGACATAAATACATCACGTTTCACATAATCACATGAAACC
GAAGGAAAAAAAAAAGGGAAAGTGAAGAGCTGATCTGTAGTACTTCTCGAACGCTCTGT
TGGATTTGGACATGACACGAACTACGACAGGGTCTGAGAGGCAAGCTAATTACGGCTGTG
TTTATATTCAACTTTGAATCTAACTTCCAACCTTTTTTTCATCACATTAACATGTCATA
CACATAACTTTTTAGTCACATCGTACCAATTTCAACCCAACTTTCAACTACACAGCCTA
```

```
[tranchant@node0 Bank]$ sed 's/>/>chr_/ig' referencelrigin.fasta | head
>chr_21054 3211 65455
CGCCGCGGAGCGGGCGCCGCCGCTGTACATGTCGCCGTGCATCGTGTGCACGTGCGG
GCCCCGTACGTGCCACCCCCACACTGTGCGTGCATCGTGCCTCTGCCGGTGGGAG
TTGCCACGCGCTGGCGACCCATACTTCTTTATTTCTTTGGTCCGTGTACGTGCAGGTT
AATCGATCGCCTCGTCATTTTTACTCGGAACTCGCTGGTGAACCCGCCACATTTCAAG
TAGTAGTGGTAGCACGTACGAGACATAAATACATCACGTTTCACATAATCACATGAAACC
GAAGGAAAAAAAAAAGGGAAAGTGAAGAGCTGATCTGTAGTACTTCTCGAACGCTCTGT
TGGATTTGGACATGACACGAACTACGACAGGGTCTGAGAGGCAAGCTAATTACGGCTGTG
TTTATATTCAACTTTGAATCTAACTTCCAACCTTTTTTTCATCACATTAACATGTCATA
CACATAACTTTTTAGTCACATCGTACCAATTTCAACCCAACTTTCAACTACACAGCCTA
```

```
[tranchant@node0 Bank]$ sed 's/>/>chr_/ig' referencelrigin.fasta > referenceRename.fasta
```

- Calculer le nombre de paire de base dans la banque de séquences AllEst.fasta

```
[tranchant@node0 Bank]$ module load bioinfo/EMBOSS/6.6.0
[tranchant@node0 Bank]$ module list
Currently Loaded Modulefiles:
1) bioinfo/EMBOSS/6.6.0
```

```
[tranchant@node0 Bank]$ [tranchant@node0 Bank]$ infoseq -help
Display basic information about sequences
Version: EMBOSS:6.6.0.0
...
```

```
[tranchant@node0 Bank]$ infoseq -sequence AllEst.fasta | head
Display basic information about sequences
```

USA	Database Name	Accession	Type	Length	%GC	Organism	Description
fasta::AllEst.fasta:GT649553 -	GT649553	GT649553	N	539	51.21	Coffea canephora cDNA clone CC00-XX-SH3-053-F02-EM, mRNA sequence	CC00-XX-SH3-053-F02-EM.F Coffea canephora SH3
fasta::AllEst.fasta:GT649552 -	GT649552	GT649552	N	467	46.04	Coffea canephora cDNA clone CC00-XX-SH3-077-D04-EM, mRNA sequence	CC00-XX-SH3-077-D04-EM.F Coffea canephora SH3
fasta::AllEst.fasta:GT649551 -	GT649551	GT649551	N	604	33.11	Coffea canephora cDNA clone CC00-XX-SH3-073-D05-EM, mRNA sequence	CC00-XX-SH3-073-D05-EM.F Coffea canephora SH3
fasta::AllEst.fasta:GT649550 -	GT649550	GT649550	N	672	41.82	Coffea canephora cDNA clone CC00-XX-SH3-025-A08-EM, mRNA sequence	CC00-XX-SH3-025-A08-EM.F Coffea canephora SH3
fasta::AllEst.fasta:GT649549 -	GT649549	GT649549	N	514	42.22	Coffea canephora cDNA clone CC00-XX-SH3-035-C12-EM, mRNA sequence	CC00-XX-SH3-035-C12-EM.F Coffea canephora SH3
fasta::AllEst.fasta:GT649548 -	GT649548	GT649548	N	494	41.80	Coffea canephora cDNA clone CC00-XX-SH3-003-C05-EM, mRNA sequence	CC00-XX-SH3-003-C05-EM.F Coffea canephora SH3
fasta::AllEst.fasta:GT649547 -	GT649547	GT649547	N	762	44.16	Coffea canephora cDNA clone CC00-XX-SH3-037-E11-EM, mRNA sequence	CC00-XX-SH3-037-E11-EM.F Coffea canephora SH3
fasta::AllEst.fasta:GT649546 -	GT649546	GT649546	N	639	42.25	Coffea canephora cDNA clone CC00-XX-SH3-054-C03-EM, mRNA sequence	CC00-XX-SH3-054-C03-EM.F Coffea canephora SH3
fasta::AllEst.fasta:GT649545 -	GT649545	GT649545	N	365	35.07	Coffea canephora cDNA clone CC00-XX-SH3-062-H01-EM, mRNA sequence	CC00-XX-SH3-062-H01-EM.F Coffea canephora SH3

```
[tranchant@node0 Bank]$ infoseq -sequence AllEst.fasta -only -accession -length -pgc | head
Display basic information about sequences
```

```

Accession  Length %GC
GT649553  539  51.21
GT649552  467  46.04
GT649551  604  33.11
GT649550  672  41.82
GT649549  514  42.22
GT649548  494  41.80
GT649547  762  44.16
GT649546  639  42.25
GT649545  365  35.07
...

```

```

[tranchant@node0 Bank]$ infoseq -sequence AllEst.fasta -only -accession -length -pgc | head
Display basic information about sequences

```

```

Accession  Length %GC
GT649553  539  51.21
GT649552  467  46.04
GT649551  604  33.11
GT649550  672  41.82
GT649549  514  42.22
GT649548  494  41.80
GT649547  762  44.16
GT649546  639  42.25
GT649545  365  35.07

```

```

[tranchant@node0 Bank]$ infoseq AllEst.fasta -only -accession -length -pgc | awk -F " " '{sum += $2;} END {print "Total pb : ",sum, " pb, ", NR, " sequences" }'

```

```

Display basic information about sequences
Total pb : 41342276 pb, 68567 sequences
[tranchant@node0 Bank]$

```

- Lancer le programme fastq-stats (ea-utils) sur 1 fichier fastq ? sur plusieurs fastq ?

```

[tranchant@node0 fastqDir]$ module load bioinfo/ea-utils/
bioinfo/ea-utils/1.1.2 bioinfo/ea-utils/2.7
[tranchant@node0 fastqDir]$ module load bioinfo/ea-utils/2.7

```

```

[tranchant@node0 fastqDir]$ pwd
/scratch/tranchant/TPsnpSV/fastqDir

```

```

[tranchant@node0 fastqDir]$ ls
Clone10_1.fastq Clone11_2.fastq Clone12_2.fastq Clone13_2.fastq Clone15_1.fastq Clone2_2.fastq Clone4_1.fastq Clone5_2.fastq
Clone7_1.fastq Clone8_2.fastq fa
Clone10_2.fastq Clone1_1.fastq Clone1_2.fastq Clone14_1.fastq Clone15_2.fastq Clone3_1.fastq Clone4_2.fastq Clone6_1.fastq
Clone7_2.fastq Clone9_1.fastq
Clone11_1.fastq Clone12_1.fastq Clone13_1.fastq Clone14_2.fastq Clone2_1.fastq Clone3_2.fastq Clone5_1.fastq Clone6_2.fastq
Clone8_1.fastq Clone9_2.fastq

```

#Je lance fastq-stats sur 1 fichier

```

[tranchant@node0 fastqDir]$ fastq-stats -D Clone10_1.fastq
reads 125633
len 150
len mean 150.0000
len stdev 0.0000
len min 150
phred 33
window-size 125633
cycle-max 35
qual min 3
qual max 41
qual mean 37.8633
qual stdev 4.5322
%A 28.1582
%C 21.8583
%G 21.8560
%T 28.1275
%N 0.0000

```

```
total bases      18844950
```

```
#Je lance fastq-stats sur 1 fichier et je sauvegarde le résultat dans un fichier
```

```
[tranchant@node0 fastqDir]$ fastq-stats -D Clone10_1.fastq > Clone10_1.fastq-stats
```

```
[tranchant@node0 fastqDir]$ ls
```

```
Clone10_1.fastq  Clone11_1.fastq Clone12_1.fastq Clone13_1.fastq Clone14_2.fastq Clone2_1.fastq Clone3_2.fastq Clone5_1.fastq  
Clone6_2.fastq Clone8_1.fastq Clone9_2.fastq  
Clone10_1.fastq-stats Clone11_2.fastq Clone12_2.fastq Clone13_2.fastq Clone15_1.fastq Clone2_2.fastq Clone4_1.fastq Clone5_2.fastq  
Clone7_1.fastq Clone8_2.fastq fa  
Clone10_2.fastq Clone1_1.fastq Clone1_2.fastq Clone14_1.fastq Clone15_2.fastq Clone3_1.fastq Clone4_2.fastq Clone6_1.fastq  
Clone7_2.fastq Clone9_1.fastq
```

```
#Je liste le contenu du fichier de résultat
```

```
[tranchant@node0 fastqDir]$ cat *stats
```

```
reads      125633  
len        150  
len mean 150.0000  
len stdev 0.0000  
len min   150  
phred     33  
window-size      125633  
cycle-max        35  
qual min  3  
qual max  41  
qual mean      37.8633  
qual stdev     4.5322  
%A         28.1582  
%C         21.8583  
%G         21.8560  
%T         28.1275  
%N         0.0000  
total bases      18844950  
[tranchant@node0 fastqDir]$
```

```
# J'utilise une boucle for pour lister un à un les fichiers fastq (on va ensuite utiliser cette boucle pour lancer fastq-stats sur chaque fichier fastq)
```

```
[tranchant@node0 fastqDir]$ for file in *fastq; do echo $file; done;
```

```
Clone10_1.fastq  
Clone10_2.fastq  
Clone11_1.fastq  
Clone11_2.fastq  
Clone1_1.fastq  
Clone12_1.fastq  
Clone12_2.fastq  
Clone1_2.fastq  
Clone13_1.fastq  
Clone13_2.fastq  
Clone14_1.fastq  
Clone14_2.fastq  
Clone15_1.fastq  
Clone15_2.fastq  
Clone2_1.fastq  
Clone2_2.fastq  
Clone3_1.fastq  
Clone3_2.fastq  
Clone4_1.fastq  
Clone4_2.fastq  
Clone5_1.fastq  
Clone5_2.fastq  
Clone6_1.fastq  
Clone6_2.fastq  
Clone7_1.fastq  
Clone7_2.fastq  
Clone8_1.fastq  
Clone8_2.fastq  
Clone9_1.fastq  
Clone9_2.fastq  
[tranchant@node0 fastqDir]$
```

```
# je lance fastq-stats sur chaque fichier fastq à l aide d'une boucle for
[tranchant@node0 fastqDir]$ for file in *fastq; do fastq-stats -D $file > $file-stats; done;
```

```
# je vérifie que j ai bien ma liste de fichiers générés par fastq-stats
```

```
[tranchant@node0 fastqDir]$ ls
```

```
Clone10_1.fastq  Clone11_2.fastq-stats Clone1_2.fastq  Clone14_1.fastq-stats Clone2_1.fastq  Clone3_2.fastq-stats Clone5_2.fastq
Clone7_1.fastq-stats Clone9_1.fastq
Clone10_1.fastq-stats Clone1_1.fastq  Clone1_2.fastq-stats Clone14_2.fastq  Clone2_1.fastq-stats Clone4_1.fastq
Clone5_2.fastq-stats Clone7_2.fastq  Clone9_1.fastq-stats
Clone10_2.fastq  Clone1_1.fastq-stats Clone13_1.fastq  Clone14_2.fastq-stats Clone2_2.fastq  Clone4_1.fastq-stats Clone6_1.fastq
Clone7_2.fastq-stats Clone9_2.fastq
Clone10_2.fastq-stats Clone12_1.fastq  Clone13_1.fastq-stats Clone15_1.fastq  Clone2_2.fastq-stats Clone4_2.fastq
Clone6_1.fastq-stats Clone8_1.fastq  Clone9_2.fastq-stats
Clone11_1.fastq  Clone12_1.fastq-stats Clone13_2.fastq  Clone15_1.fastq-stats Clone3_1.fastq  Clone4_2.fastq-stats
Clone6_2.fastq  Clone8_1.fastq-stats fa
Clone11_1.fastq-stats Clone12_2.fastq  Clone13_2.fastq-stats Clone15_2.fastq  Clone3_1.fastq-stats Clone5_1.fastq
Clone6_2.fastq-stats Clone8_2.fastq
Clone11_2.fastq  Clone12_2.fastq-stats Clone14_1.fastq  Clone15_2.fastq-stats Clone3_2.fastq  Clone5_1.fastq-stats
Clone7_1.fastq  Clone8_2.fastq-stats
[tranchant@node0 fastqDir]$
```