Data Wrangling and Processing for Genomics - Parte 4

Evidencia:

Nueva carpeta y descargamos el genoma de referencia de E. coliREL606:

```
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/trimmed_fastq$ cd ~/dc_workshop
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ mkdir -p data/ref_genome
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ sudo curl -L -o data/ref_genome/ecoli_rel606.fasta
.gz ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/017/985/GCA_000017985.1_ASM1798v1/GCA_00001
7985.1_ASM1798v1_genomic.fna.gz
[sudo] password for estuardo8u14:
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 1343k  100 1343k    0     0   625k      0  0:00:02  0:00:02 --:--:--  625k
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ gunzip data/ref_genome/ecoli_rel606.fasta.gz
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ head data/ref_genome/ecoli_rel606.fasta
>CP000819.1 Escherichia coli B str. REL606, complete genome
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGCTTCTGAACTG
GTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAATATAGGCATAGCGCACAGAC
AGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTACCACAGGT
AACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTCGACCAAAGGT
AACGAGGTAACAACCATGCGAGTGTTGAAGTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGGGTTGCCGA
TATTCTGGAAAGCAATGCCAGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGGTGG
CGATGATTGAAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCCGAACTTTTG
ACGGGACTCGCCGCCGCCCAGCCGGGATTCCCGCTGGCGCAATTGAAAACTTTCGTCGATCAGGAATTTGCCCAAATAAA
ACATGTCCTGCATGGCATTAGTTTGTTGGGGCAGTGCCCGGATAGCATCAACGCTGCGCTGATTTGCCGTGGCGAGAAAA
```

Nuestro primer paso es indexar el genoma de referencia para que lo utilice BWA:

```
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ sudo curl -L -o sub.tar.gz https://ndownloader.fig
share.com/files/14418248
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
100  109M  100  109M    0     0  3492k      0  0:00:32  0:00:32 --:--:-- 3599k
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ tar xvf sub.tar.gz
sub/
sub/SRR2584866_1.trim.sub.fastq
sub/SRR2589044_1.trim.sub.fastq
sub/SRR2589044_2.trim.sub.fastq
sub/SRR2584863_2.trim.sub.fastq
sub/SRR2584866_2.trim.sub.fastq
sub/SRR2584863_1.trim.sub.fastq
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ mv sub/ ~/dc_workshop/data/trimmed_fastq_small
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ mkdir -p results/sam results/bam results/bcf resul
ts/vcf
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ bwa index data/ref_genome/ecoli_rel606.fasta
[bwa_index] Pack FASTA... 0.15 sec
[bwa_index] Construct BWT for the packed sequence...
[bwa_index] 2.21 seconds elapse.
[bwa_index] Update BWT... 0.06 sec
[bwa_index] Pack forward-only FASTA... 0.02 sec
[bwa_index] Construct SA from BWT and Occ... 1.32 sec
[main] Version: 0.7.17-r1188
[main] CMD: bwa index data/ref_genome/ecoli_rel606.fasta
[main] Real time: 4.928 sec; CPU: 3.824 sec
```

Comenzaremos alineando las lecturas de solo una de las muestras en nuestro conjunto de datos (SRR2584866) y tras correr la línea bwa mem data/ref_genome/ecoli_rel606.fasta data/trimmed_fastq_small/SRR2584866_1.trim.sub.fastq data/trimmed_fastq_small/SRR2584866_2.trim.sub.fastq > results/sam/SRR2584866.aligned.sam :

```
[M::mem_pestat] analyzing insert size distribution for orientation FF...
[M::mem_pestat] (25, 50, 75) percentile: (235, 720, 1382)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 3676)
[M::mem_pestat] mean and std.dev: (770.19, 676.46)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 4823)
[M::mem_pestat] analyzing insert size distribution for orientation FR...
[M::mem_pestat] (25, 50, 75) percentile: (219, 350, 560)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1242)
[M::mem_pestat] mean and std.dev: (404.41, 221.74)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 1583)
[M::mem_pestat] skip orientation RF as there are not enough pairs
[M::mem_pestat] analyzing insert size distribution for orientation RR...
[M::mem_pestat] (25, 50, 75) percentile: (325, 557, 721)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1513)
[M::mem_pestat] mean and std.dev: (501.66, 309.62)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 1909)
[M::mem_pestat] skip orientation FF
[M::mem_pestat] skip orientation RR
[M::mem_process_seqs] Processed 45746 reads in 3.633 CPU sec, 3.443 real sec
[main] Version: 0.7.17-r1188
[main] CMD: bwa mem data/ref_genome/ecoli_rel606.fasta data/trimmed_fastq_small/SRR2584866_1.t
rim.sub.fastq data/trimmed_fastq_small/SRR2584866_2.trim.sub.fastq
[main] Real time: 28.329 sec; CPU: 28.838 sec
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ samtools view -S -b results/sam/SRR2584866.aligned
.sam > results/bam/SRR2584866.aligned.bam
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ samtools sort -o results/bam/SRR2584866.aligned.so
rted.bam results/bam/SRR2584866.aligned.bam
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ samtools flagstat results/bam/SRR2584866.aligned.s
orted.bam
351169 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
1169 + 0 supplementary
0 + 0 duplicates
351103 + 0 mapped (99.98% : N/A)
350000 + 0 paired in sequencing
175000 + 0 read1
175000 + 0 read2
346688 + 0 properly paired (99.05% : N/A)
349876 + 0 with itself and mate mapped
58 + 0 singletons (0.02% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$
```

Llamado de variantes

Paso 1: Calcule la cobertura de lectura de posiciones en el genoma

```
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ bcftools mpileup -O b -o results/bcf/SRR2584866_ra
w.bcf \
> -f data/ref_genome/ecoli_rel606.fasta results/bam/SRR2584866.aligned.sorted.bam
[mpileup] 1 samples in 1 input files
[mpileup] maximum number of reads per input file set to -d 250
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$
```

Paso 2: detectar los polimorfismos de un solo nucleótido (SNP)

Paso 3: filtrar e informar las variantes de SNP en formato de llamada de variante (VCF)

Viendo los datos:

```
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##bcftoolsVersion=1.10.2+htslib-1.10.2-3
##bcftoolsCommand=mpileup -O b -o results/bcf/SRR2584866_raw.bcf -f data/ref_genome/ecoli_rel606.fasta results/bam/SRR2584866.aligned.sorted.bam
##reference=file://data/ref_genome/ecoli_rel606.fasta
##contig=<ID=CP000819.1,length=4629812>
##ALT=<ID=*,Description="Represents allele(s) other than observed.">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=IDV,Number=1,Type=Integer,Description="Maximum number of raw reads supporting an indel">
##INFO=<ID=IMF,Number=1,Type=Float,Description="Maximum fraction of raw reads supporting an indel">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=<ID=VDB,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts in RNA-seq data (bigger is better)",Version="3">
##INFO=<ID=RPB,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
##INFO=<ID=MQB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
##INFO=<ID=BQB,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
##INFO=<ID=MQSB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
##INFO=<ID=SGB,Number=1,Type=Float,Description="Segregation based metric.">
##INFO=<ID=MQ0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=ICB,Number=1,Type=Float,Description="Inbreeding Coefficient Binomial test (bigger is better)">
##INFO=<ID=HOB,Number=1,Type=Float,Description="Bias in the number of HOMs number (smaller is better)">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="Number of high-quality ref-forward , ref-reverse, alt-forward and alt-reverse bases">
##INFO=<ID=MQ,Number=1,Type=Integer,Description="Average mapping quality">
##bcftools_callVersion=1.10.2+htslib-1.10.2-3
##bcftools_callCommand=call --ploidy 1 -m -v -o results/bcf/SRR2584866_variants.vcf results/bcf/SRR2584866_raw.bcf; Date=Tue Nov 10 12:55:30 2020
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO    FORMAT  results/bam/SRR2584866.aligned.sorted.bam
CP000819.1      1521    .       C       T       207     .       DP=9;VDB=0.993024;SGB=-0.662043;MQSB=0.974597;MQ0F=0;AC=1;AN=1;DP4=0,0,4,5;MQ=60       GT:PL   1:237,0
CP000819.1      1612    .       A       G       225     .       DP=13;VDB=0.52194;SGB=-0.676189;MQSB=0.950952;MQ0F=0;AC=1;AN=1;DP4=0,0,6,5;MQ=60       GT:PL   1:255,0
CP000819.1      9092    .       A       G       225     .       DP=14;VDB=0.717543;SGB=-0.670168;MQSB=0.916482;MQ0F=0;AC=1;AN=1;DP4=0,0,7,3;MQ=60      GT:PL   1:255,0
CP000819.1      9972    .       T       G       214     .       DP=10;VDB=0.022095;SGB=-0.670168;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,8;MQ=60      GT:PL   1:244,0
CP000819.1      10563   .       G       A       225     .       DP=11;VDB=0.958658;SGB=-0.670168;MQSB=0.952347;MQ0F=0;AC=1;AN=1;DP4=0,0,5,5;MQ=60       GT:PL   1:255,0
CP000819.1      22257   .       C       T       127     .       DP=5;VDB=0.0765947;SGB=-0.590765;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,3;MQ=60      GT:PL   1:157,0
CP000819.1      38971   .       A       G       225     .       DP=14;VDB=0.872139;SGB=-0.680642;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,4,8;MQ=60      GT:PL   1:255,0
CP000819.1      42306   .       A       G       225     .       DP=15;VDB=0.969686;SGB=-0.686358;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,5,9;MQ=60      GT:PL   1:255,0
CP000819.1      45277   .       A       G       225     .       DP=15;VDB=0.470998;SGB=-0.680642;MQSB=0.95494;MQ0F=0;AC=1;AN=1;DP4=0,0,7,5;MQ=60       GT:PL   1:255,0
CP000819.1      56613   .       C       G       183     .       DP=12;VDB=0.879703;SGB=-0.676189;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,8,3;MQ=60      GT:PL   1:213,0
CP000819.1      62118   .       A       G       225     .       DP=19;VDB=0.265944;SGB=-0.689466;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,7,9;MQ=60      GT:PL   1:255,0
CP000819.1      64042   .       G       A       225     .       DP=18;VDB=0.451328;SGB=-0.689466;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,7,9;MQ=60      GT:PL   1:255,0
CP000819.1      78808   .       C       T       225     .       DP=23;VDB=0.885435;SGB=-0.691153;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,13,5;MQ=60      GT:PL   1:255,0
CP000819.1      80113   .       A       G       165     .       DP=9;VDB=0.989691;SGB=-0.651104;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,6,2;MQ=60       GT:PL   1:195,0
CP000819.1      81158   .       A       C       225     .       DP=13;VDB=0.154969;SGB=-0.676189;MQSB=0.950952;MQ0F=0;AC=1;AN=1;DP4=0,0,6,5;MQ=60       GT:PL   1:255,0
CP000819.1      87462   .       A       G       225     .       DP=10;VDB=0.345209;SGB=-0.636426;MQSB=1.01283;MQ0F=0;AC=1;AN=1;DP4=0,0,4,3;MQ=60       GT:PL   1:255,0
:
```

```
CP000819.1      4248049 .       T       C       206     .       DP=9;VDB=0.633718;SGB=-0.662043;MQSB=0.924584;MQ0F=0;AC=1;AN=1;DP4=0,0,6,3;MQ=60       GT:PL   1:236,0
CP000819.1      4250292 .       C       T       225     .       DP=13;VDB=0.00888554;SGB=-0.676189;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,8,3;MQ=60    GT:PL   1:255,0
CP000819.1      4265633 .       G       A       225     .       DP=12;VDB=0.708194;SGB=-0.676189;MQSB=0.8618;MQ0F=0;AC=1;AN=1;DP4=0,0,6,5;MQ=59 GT:PL   1:255,0
CP000819.1      4286515 .       C       T       190     .       DP=9;VDB=0.267186;SGB=-0.662043;MQSB=0.974597;MQ0F=0;AC=1;AN=1;DP4=0,0,5,4;MQ=60       GT:PL   1:220,0
CP000819.1      4342010 .       A       G       225     .       DP=13;VDB=0.406086;SGB=-0.676189;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,10,1;MQ=60     GT:PL   1:255,0
CP000819.1      4347725 .       C       T       116     .       DP=5;VDB=0.348933;SGB=-0.556411;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,1,3;MQ=60       GT:PL   1:146,0
CP000819.1      4347803 .       A       G       143     .       DP=6;VDB=0.368038;SGB=-0.616816;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,4;MQ=60       GT:PL   1:173,0
CP000819.1      4349281 .       ATTTTTTTT       ATTTTTTT        44.2229 .       INDEL;IDV=9;IMF=0.818182;DP=11;VDB=0.95493;SGB=-0.670168;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,1,3,7;MQ=60      GT:PL   1:71,0
CP000819.1      4356652 .       G       A       217     .       DP=9;VDB=0.162453;SGB=-0.651104;MQSB=0.900802;MQ0F=0;AC=1;AN=1;DP4=0,0,3,5;MQ=60       GT:PL   1:247,0
CP000819.1      4363338 .       C       A       176     .       DP=10;VDB=0.399237;SGB=-0.670168;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,1,9;MQ=60      GT:PL   1:206,0
CP000819.1      4377265 .       A       G       225     .       DP=16;VDB=0.921692;SGB=-0.683931;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,4,9;MQ=60       GT:PL   1:255,0
CP000819.1      4380317 .       A       G       181     .       DP=7;VDB=0.228249;SGB=-0.636426;MQSB=1.01283;MQ0F=0;AC=1;AN=1;DP4=0,0,3,4;MQ=60 GT:PL   1:211,0
CP000819.1      4380609 .       C       T       225     .       DP=12;VDB=0.568173;SGB=-0.636426;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,6,1;MQ=60      GT:PL   1:255,0
CP000819.1      4382294 .       C       T       212     .       DP=9;VDB=0.598931;SGB=-0.651104;MQSB=0.900802;MQ0F=0;AC=1;AN=1;DP4=0,0,3,5;MQ=60       GT:PL   1:242,0
CP000819.1      4384556 .       C       T       225     .       DP=13;VDB=0.691465;SGB=-0.676189;MQSB=0.950952;MQ0F=0;AC=1;AN=1;DP4=0,0,5,6;MQ=60       GT:PL   1:255,0
CP000819.1      4392683 .       ACCCC   ACCCCC  155     .       INDEL;IDV=16;IMF=0.888889;DP=18;VDB=0.989366;SGB=-0.688148;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,3,8,7;MQ=60    GT:PL   1:182,0
CP000819.1      4423016 .       C       T       225     .       DP=12;VDB=0.789033;SGB=-0.680642;MQSB=0.982603;MQ0F=0;AC=1;AN=1;DP4=0,0,6,6;MQ=60       GT:PL   1:255,0
CP000819.1      4423593 .       G       T       225     .       DP=13;VDB=0.815149;SGB=-0.683931;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,10,3;MQ=60     GT:PL   1:255,0
CP000819.1      4424649 .       T       C       153     .       DP=8;VDB=0.493668;SGB=-0.636426;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,1,6;MQ=60       GT:PL   1:183,0
CP000819.1      4429654 .       G       A       149     .       DP=7;VDB=0.561812;SGB=-0.616816;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,1,5;MQ=60       GT:PL   1:179,0
CP000819.1      4431393 .       TGG     T       228     .       INDEL;IDV=11;IMF=1;DP=11;VDB=0.138406;SGB=-0.662043;MQSB=0.950952;MQ0F=0;AC=1;AN=1;DP4=0,2,6,3;MQ=60    GT:PL   1:255,0
CP000819.1      4433347 .       A       G       149     .       DP=7;VDB=0.427991;SGB=-0.616816;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,5,1;MQ=60       GT:PL   1:179,0
CP000819.1      4439519 .       A       G       162     .       DP=7;VDB=0.810254;SGB=-0.636426;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,5;MQ=60       GT:PL   1:192,0
CP000819.1      4461084 .       G       A       186     .       DP=9;VDB=0.651328;SGB=-0.616816;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,4,2;MQ=60       GT:PL   1:216,0
CP000819.1      4462040 .       C       T       103     .       DP=5;VDB=0.510154;SGB=-0.556411;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,1,3;MQ=60       GT:PL   1:133,0
CP000819.1      4472216 .       C       T       190     .       DP=9;VDB=0.295648;SGB=-0.651104;MQSB=0.900802;MQ0F=0;AC=1;AN=1;DP4=0,0,5,3;MQ=60       GT:PL   1:220,0
CP000819.1      4486277 .       C       T       212     .       DP=9;VDB=0.887079;SGB=-0.651104;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,6,2;MQ=60       GT:PL   1:242,0
CP000819.1      4491013 .       A       G       96      .       DP=5;VDB=0.0456508;SGB=-0.590765;MQ0F=0;AC=1;AN=1;DP4=0,0,0,5;MQ=60     GT:PL   1:126,0
CP000819.1      4498157 .       G       A       40.4148 .       DP=2;VDB=0.26;SGB=-0.453602;MQ0F=0;AC=1;AN=1;DP4=0,0,0,2;MQ=60  GT:PL   1:70,0
CP000819.1      4498856 .       T       C       225     .       DP=12;VDB=0.945914;SGB=-0.676189;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,8,3;MQ=60       GT:PL   1:255,0
CP000819.1      4503582 .       T       C       111     .       DP=6;VDB=0.446842;SGB=-0.556411;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,2;MQ=60       GT:PL   1:141,0
CP000819.1      4509290 .       C       T       225     .       DP=15;VDB=0.254982;SGB=-0.680642;MQSB=0.95494;MQ0F=0;AC=1;AN=1;DP4=0,0,7,5;MQ=60       GT:PL   1:255,0
CP000819.1      4588672 .       A       G       118     .       DP=5;VDB=0.459448;SGB=-0.556411;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,1,3;MQ=60       GT:PL   1:148,0
CP000819.1      4593186 .       G       A       225     .       DP=18;VDB=0.269766;SGB=-0.686358;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,9,5;MQ=60       GT:PL   1:255,0
CP000819.1      4603391 .       C       T       208     .       DP=12;VDB=0.206708;SGB=-0.670168;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,8;MQ=60       GT:PL   1:238,0
CP000819.1      4603678 .       A       G       57      .       DP=4;VDB=0.0298006;SGB=-0.556411;MQ0F=0;AC=1;AN=1;DP4=0,0,4,0;MQ=43      GT:PL   1:87,0
CP000819.1      4603681 .       C       T       58      .       DP=4;VDB=0.0298006;SGB=-0.556411;MQ0F=0;AC=1;AN=1;DP4=0,0,4,0;MQ=43      GT:PL   1:88,0
CP000819.1      4603730 .       C       T       28.4205 .       DP=3;VDB=0.121629;SGB=-0.511536;MQSB=1;MQ0F=0.333333;AC=1;AN=1;DP4=0,0,1,2;MQ=29        GT:PL   1:58,0
CP000819.1      4603737 .       T       C       82      .       DP=7;VDB=0.102049;SGB=-0.590765;MQSB=0.666667;MQ0F=0.142857;AC=1;AN=1;DP4=0,0,2,3;MQ=41 GT:PL   1:112,0
CP000819.1      4604220 .       G       A       225     .       DP=12;VDB=0.874935;SGB=-0.680642;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,8,4;MQ=60       GT:PL   1:255,0
CP000819.1      4604577 .       G       C       218     .       DP=11;VDB=0.774083;SGB=-0.662043;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,7,2;MQ=60       GT:PL   1:248,0
CP000819.1      4613668 .       C       T       196     .       DP=9;VDB=0.671615;SGB=-0.662043;MQSB=0.974597;MQ0F=0;AC=1;AN=1;DP4=0,0,4,5;MQ=60       GT:PL   1:226,0
CP000819.1      4616538 .       A       C       92      .       DP=5;VDB=0.950351;SGB=-0.556411;MQ0F=0;AC=1;AN=1;DP4=0,0,4,0;MQ=60     GT:PL   1:122,0
CP000819.1      4626487 .       C       T       225     .       DP=13;VDB=0.665828;SGB=-0.670168;MQSB=1.00775;MQ0F=0;AC=1;AN=1;DP4=0,0,6,4;MQ=60    GT:PL   1:255,0
CP000819.1      4629225 .       C       T       225     .       DP=10;VDB=0.89913;SGB=-0.670168;MQSB=1.00775;MQ0F=0;AC=1;AN=1;DP4=0,0,6,4;MQ=60 GT:PL   1:255,0
(END)
```

Use the grep and wc commands you've learned to assess how many variants are in the vcf file.

Respuesta:

```
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ bcftools call --ploidy 1 -m -v -o results/bcf/SRR2584866_variants.vcf results/bcf/SRR258486
6_raw.bcf
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ vcfutils.pl varFilter results/bcf/SRR2584866_variants.vcf  > results/vcf/SRR2584866_final_v
ariants.vcf
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ less -S results/vcf/SRR2584866_final_variants.vcf
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$ grep -v "#" results/vcf/SRR2584866_final_variants.vcf | wc -l
767
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop$
```

767