GoSeqIt

Ex. 5 Recap

Q1: Locate the read name “SRR4114395.10”. How many lines include this read name? Why are there more than one line with this read name?

Answer: 2 - it is the forward and reverse reads in a pair.

Q2. What is the flag for the alignment with the first “SRR4114395.10” and what does it mean?

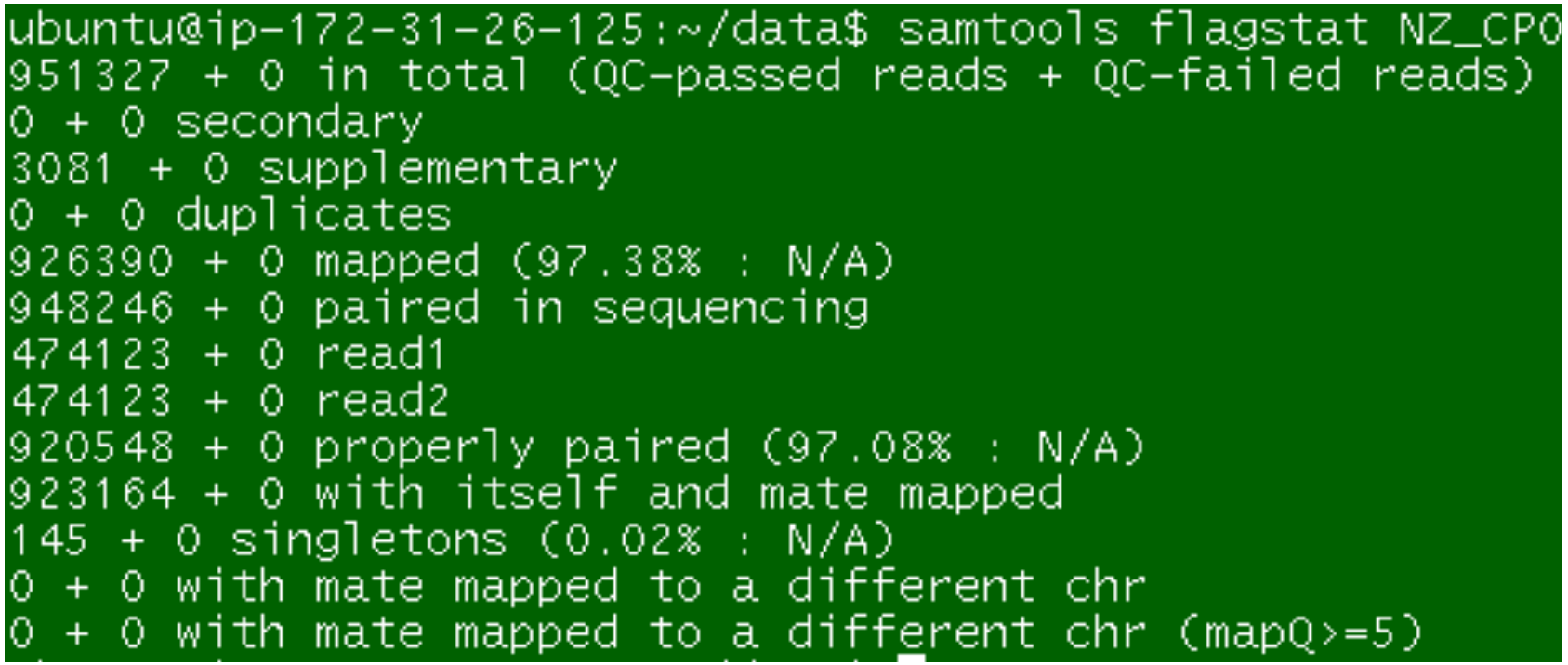
Answer: The flag is 83 and it means:

Read paired

Read mapped in proper pair

read reverse strand

first in pair



Q3. From Ex. 3 we know that there are 474123 reads in each of the files SRR4114395_1_5end_qual_trimmed.fastq.gz and SRR4114395_2_5end_qual_trimmed.fastq.gz (as also seen from the “read1” and “read2” values). Looking at the samtools flagstat output, the number of alignments listed in the first line (QC-passed reads + QC-failed reads) minus the supplementary alignments should equal the sum of reads in the two fastq files. Does it?

Answer: Yes: 951327 - 3081 = 948246 = 474123 2*

Q4. What is the percentage of unmapped reads (contamination).

*Answer: This can be calculated as 100% - the percentage of properly paired reads: 100% - 97,08% = 2,92 (and can also be calculated as ((474123*2 - 920548) / (2*474123))*100*

Q5: Use “samtools flagstat” to see how many reads were removed this way.

Answer: 474123 - 469785 = 4338 reads per input file (so 8676 in total)

Q6: How many variants were detected in the raw VFC file?

Answer: 147

Q7. How many SNPs did you find now?

Answer: 0

Ex5-extra:

I have manually mutated the reference strain by substituting an "A" at position 1708109 with a "T". Repeat the above exercise using the mutated reference strain, NZ_CP009361_mut.fasta, instead of the original one to confirm that you can find the SNP.

```
ubuntu@ip-172-31-41-5:~/data$ zcat NZ_CP009361_SRR4114395.var.mut.raw.vcf.gz | vcf-annotate --filter d=30/Q=200/c=2,10 |  grep -v "#" |
grep "PASS" | grep "1/1" | grep -v "INDEL"
NZ_CP009361.1   1708109 .       A       T       225     PASS    DP=70;VDB=0.74051;SGB=-0.693145;MQSB=1;MQOF=0;AC=2;AN=2;DP4=0,0,13,27;MQ
=60     GT:PL   1/1:255,120,0
```

```
NZ_CP009361.1    1708109  .        A        T
```