unzip files:

[guest@centos6]$ cd project2

[guest@centos6 project2]$ gunzip gencommand_proj2_data.tar.gz

[guest@centos6 project2]$ tar xvc gencommand_proj2_data.tar

1. How many alignments does the set contain?

[guest@centos6 project2]$ samtools view athal_wu_0_A.bam | head

[guest@centos6 project2]$ samtools view athal_wu_0_A.bam | wc -l

2. How many alignments show the read's mate unmapped?

[guest@centos6 project2]$ samtools view athal_wu_0_A.bam | cut -f7 | grep "*" | wc -l

3. How many alignments contain a deletion (D)?

[guest@centos6 project2]$ samtools view athal_wu_0_A.bam | cut -f6 | grep "D" | wc -l

4. How many alignments show the read's mate mapped to the same chromosome?

[guest@centos6 project2]$ samtools view athal_wu_0_A.bam | cut -f7 | grep "=" | wc -l

5. How many alignments are spliced?

```
[guest@centos6 project2]$ samtools view athal_wu_0_A.bam | cut -f6 |
grep "N" | wc -l
```

6. How many alignments does the set contain?

```
[guest@centos6 project2]$ nohup samtools sort athal_wu_0_A.bam
athal_wu_0_A.sorted &
[guest@centos6 project2]$ samtools index athal_wu_0_A.sorted.bam
[guest@centos6 project2]$ samtools view athal_wu_0_A.sorted.bam
"Chr3:11777000-11794000" | head
```

```
[guest@centos6 project2]$ samtools view athal_wu_0_A.sorted.bam
"Chr3:11777000-11794000" | wc -l
```

7. How many alignments show the read's mate unmapped?

```
[guest@centos6 project2]$ samtools view athal_wu_0_A.sorted.bam
"Chr3:11777000-11794000" | cut -f7 | grep "*" | wc -l
```

8. How many alignments contain a deletion (D)?

```
[guest@centos6 project2]$ samtools view athal_wu_0_A.sorted.bam
"Chr3:11777000-11794000"  | cut -f6 | grep "D" | wc -l
```

9. How many alignments show the read's mate mapped to the same
chromosome?

[guest@centos6 project2]$ samtools view athal_wu_0_A.sorted.bam

"Chr3:11777000-11794000" | cut -f7 | grep "=" | wc -l

10. How many alignments are spliced?

[guest@centos6 project2]$ samtools view athal_wu_0_A.sorted.bam

"Chr3:11777000-11794000" | cut -f6 | grep "N" | wc -l

11. How many sequences are in the genome file?

12. What is the length of the first sequence in the genome file?

13. What alignment tool was used?

[guest@centos6 project2]$ samtools view -H athal_wu_0_A.bam

14. What is the read identifier (name) for the first alignment?

15. What is the start position of this read's mate on the genome? Give this

as 'chrom:pos' if the read was mapped, or '*' if unmapped.

[guest@centos6 project2]$ samtools view athal_wu_0_A.bam | head

16. How many overlaps (each overlap is reported on one line) are

reported?

[guest@centos6 project2]$ bedtools bamtobed -i athal_wu_0_A.bam >

data.bed

[guest@centos6 project2]$ head data.bed

[guest@centos6 project2]$ bedtools intersect -wo -a

athal_wu_0_A_annot.gtf -b data.bed | head

[guest@centos6 project2]$ bedtools intersect -wo -a

athal_wu_0_A_annot.gtf -b data.bed | wc -l

17. How many of these are 10 bases or longer?

[guest@centos6 project2]$ bedtools intersect -wo -a

athal_wu_0_A_annot.gtf -b data.bed | cut -f16 > count_base.txt

18. How many alignments overlap the annotations?

[guest@centos6 project2]$ bedtools intersect -wo -a

athal_wu_0_A_annot.gtf -b data.bed | wc -l

19. Conversely, how many exons have reads mapped to them?

[guest@centos6 project2]$ bedtools intersect -wo -a

athal_wu_0_A_annot.gtf -b data.bed | cut -f4 | sort -u | wc -l

[guest@centos6 project2]$ bedtools intersect -wo -a

athal_wu_0_A_annot.gtf -b data.bed | cut -f5 | sort -u | wc -l

20. If you were to convert the transcript annotations in the file

"athal_wu_0_A_annot.gtf" into BED format, how many BED records would

be generated?

[guest@centos6 project2]$ bedtools intersect -wo -a

athal_wu_0_A_annot.gtf -b data.bed | cut -f9 | cut -d " " -f4 | head

```
[guest@centos6 project2]$ bedtools intersect -wo -a
athal_wu_0_A_annot.gtf -b data.bed | cut -f9 | cut -d " " -f4 | sort -u | wc -l
```