Problem Set 1

Matt Grobelny

Q1.

```
cut -f 2 batch_2.fst_2-3.tsv | grep -E "[0-9]+" | sort -h |
wc -l
```
number of SNP: 29588

Q2.

```
cut -f 2 batch_2.fst_2-3.tsv | grep -E "[0-9]+" | sort -h |
uniq -c | wc -l
```
number of unique loci: 9006

Q3.

```
cut -f 5 batch_2.fst_2-3.tsv | grep -vE "Chr" | sort -h |
uniq -c | sort -h
```

Group #	Number of SNPs Per group
groupIII	4912
groupIX	4952
groupII	5491
groupI	6857
groupIV	7376

Q4.

```
$ cut -f 2,5,6 batch_2.fst_2-3.tsv | grep -w "groupII" | sort -k 3 -n | tail -n 1
```

Highest base-pair position SNP:

Loci	Group#	base-pair position
2834	groupII	23,291,075

Q5.

```
cut -f 9 batch_2.fst_2-3.tsv | sort -h | head -n 1
```

Lowest value Fst:

-0.2085994040

Number of loci with lowest fst value (-0.2085994040):

```
grep -c -- "-0.2085994040" batch_2.fst_2-3.tsv
```

Number of loci: 1

Q6.

```
cut -f 9 batch_2.fst_2-3.tsv | sort -h | uniq -c | sort -h | tail -n 1
```

Fst value with the highest number of occurrences:

-0.1359223301

```
$ grep -n -- "-0.1359223301" batch_2.fst_2-3.tsv | cut -f 2 | sort -h | uniq -c | sort -h | wc -l
```

Number of loci which have that value:

1110