Erika Maldonado-Rosado

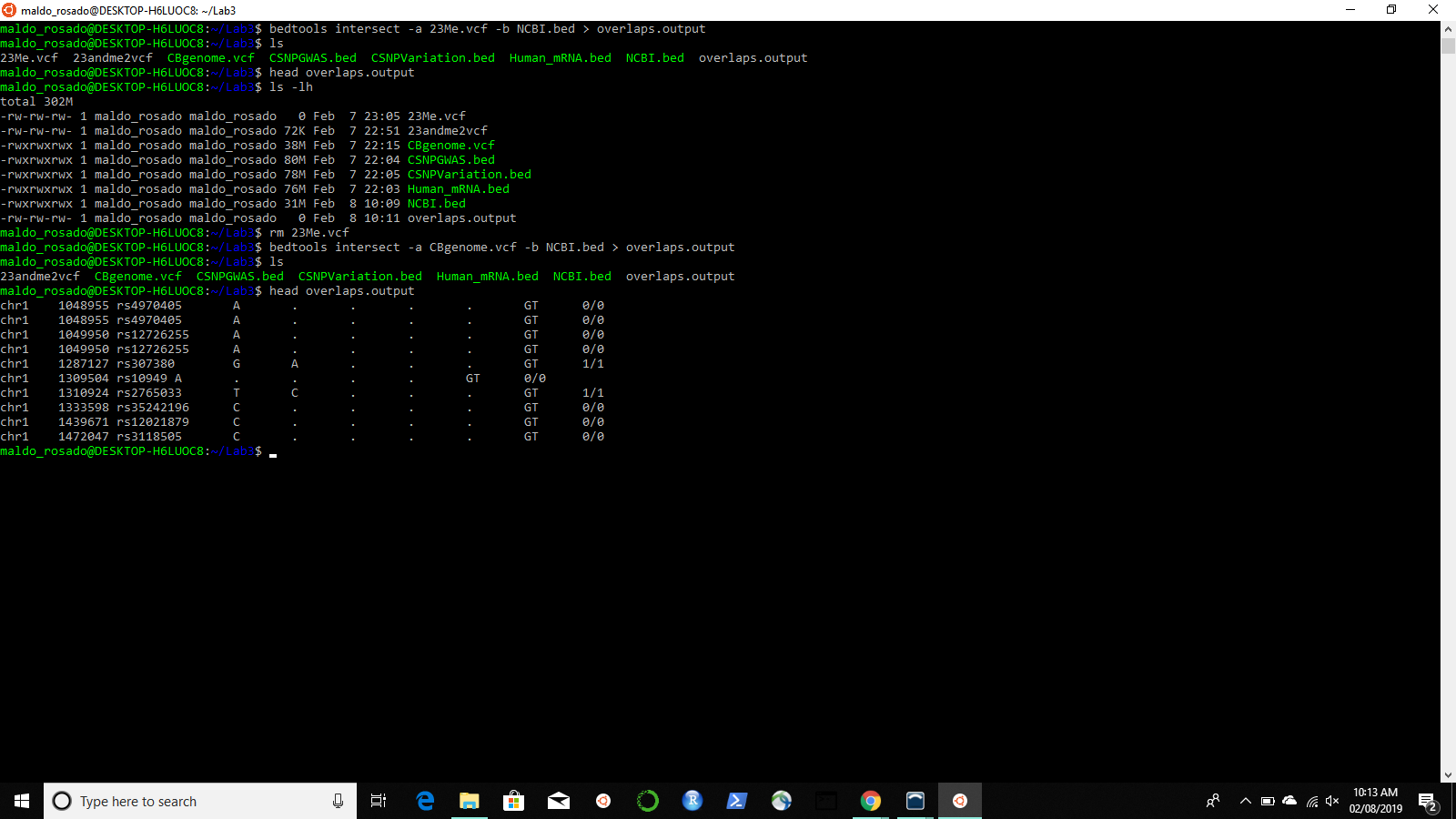
7 February 2019

Lab 3: Genome Track Math

BINF 6203

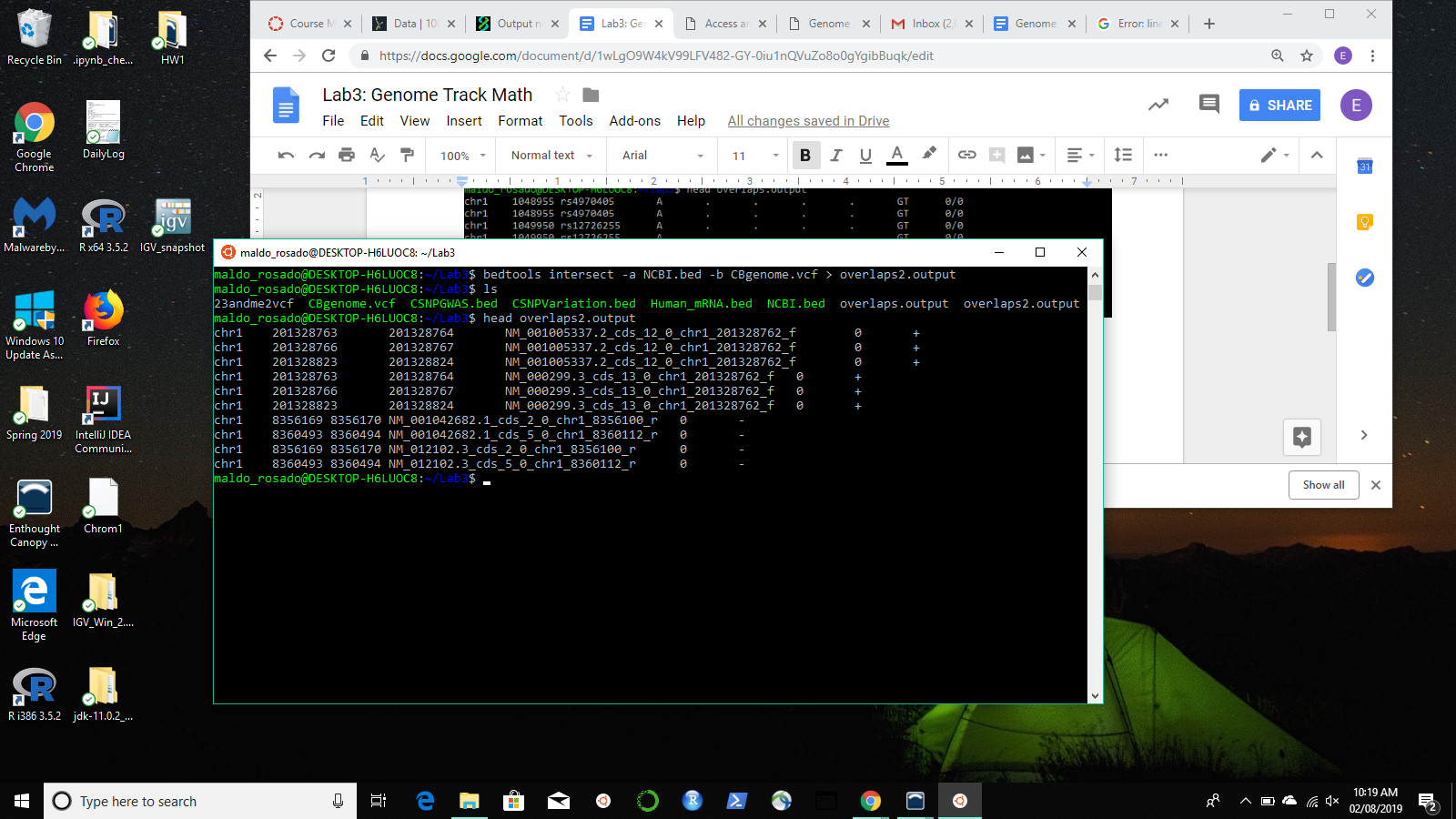
**1) What format is the output that you get?**

I got a VCF file format.



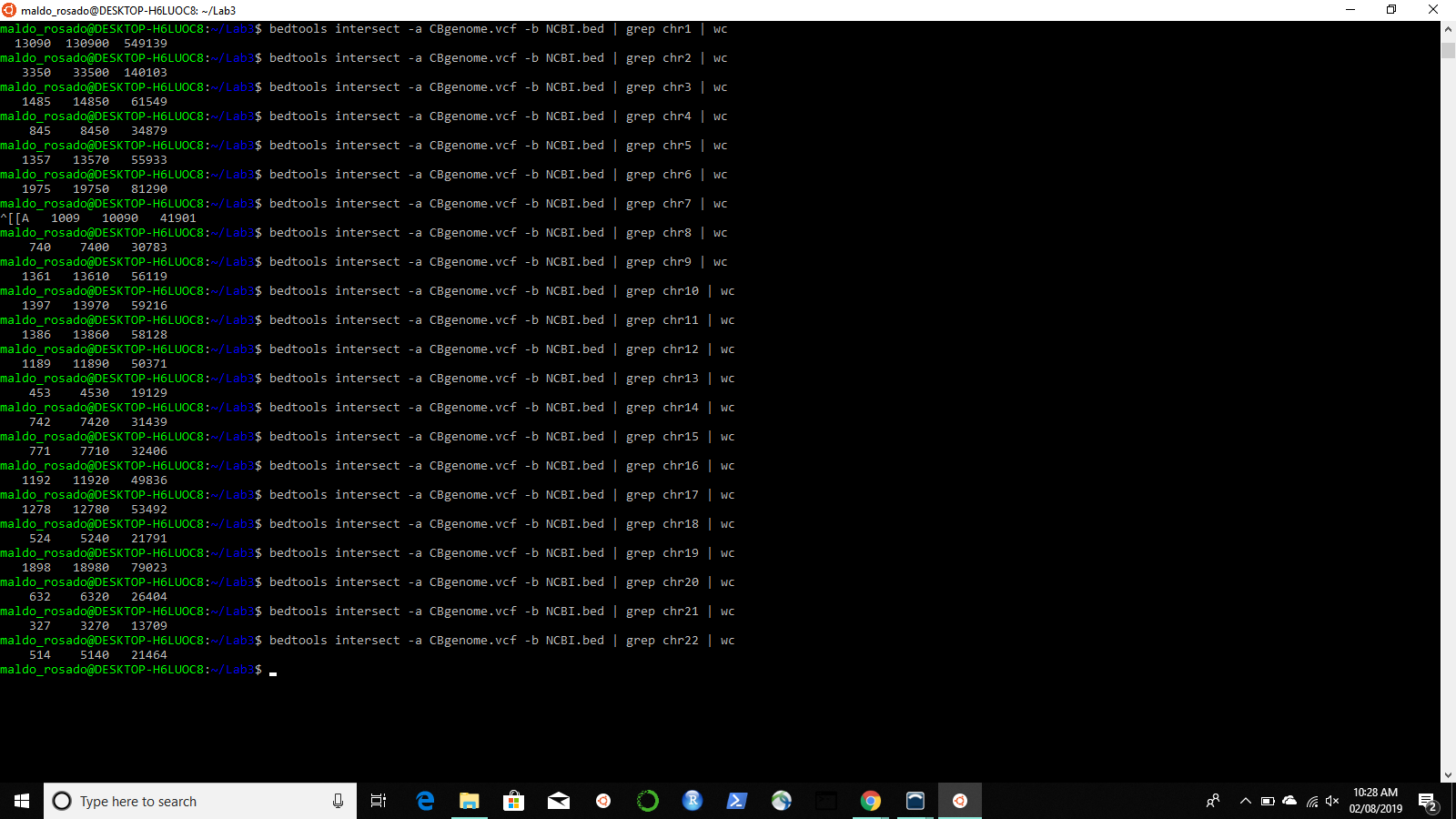
**2) What happens if you switch things up so that the -a file is the BED file and the -b file is the VCF?**

I got a BED file format.

****

**3) Which human autosome has the highest per-base-pair representation of SNPs in the 23andMe assay?**

Chromosome eight had the highest autosome per base pair.

****

**4) Which SNPs in the 23andMe assay intersect with SNPs in the GWAS (Genome Wide Association Study) Catalog (in the Phenotype and Literature track category)?**

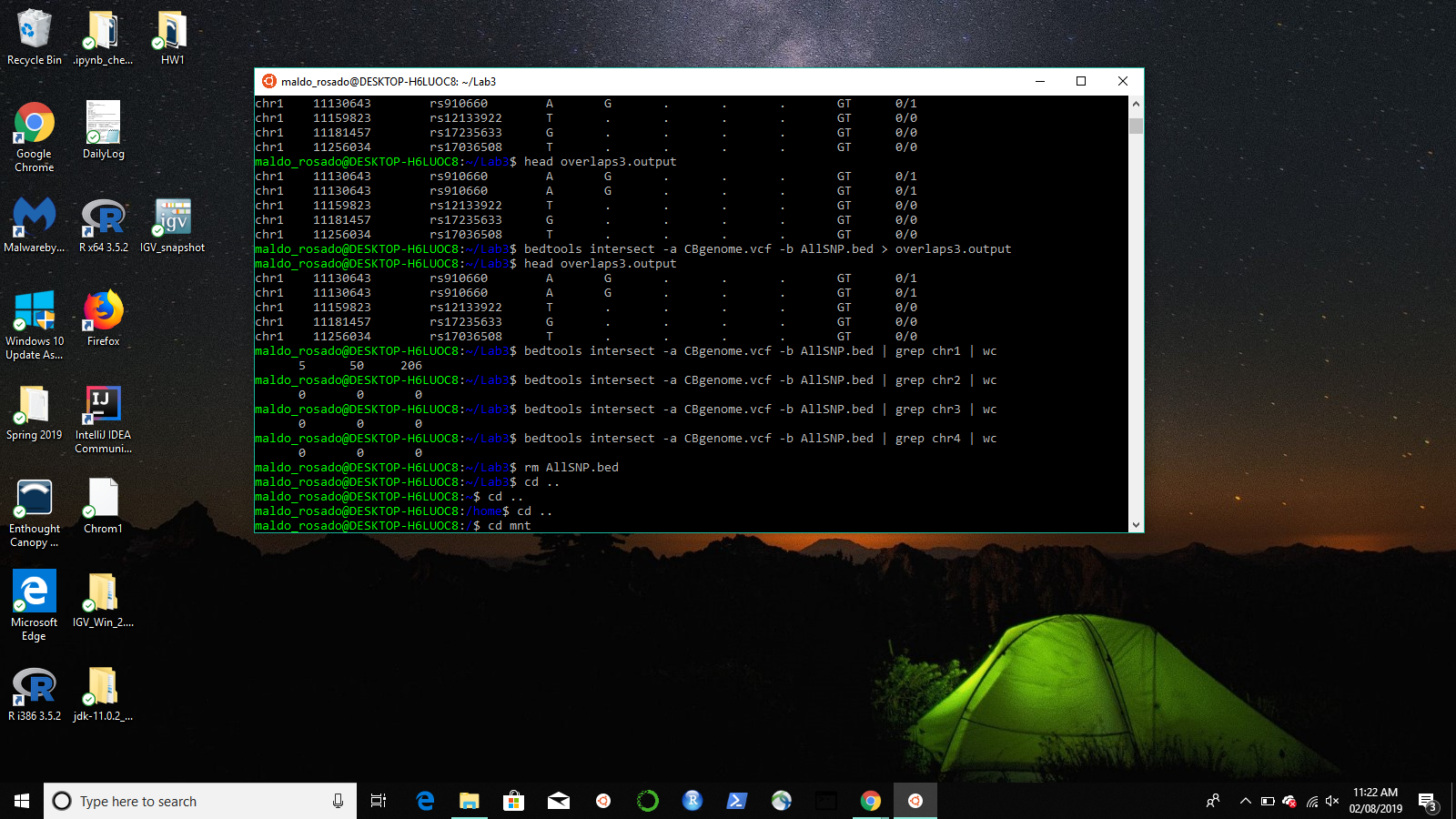
**Do you think the number that you get from a simple intersection is correct?**

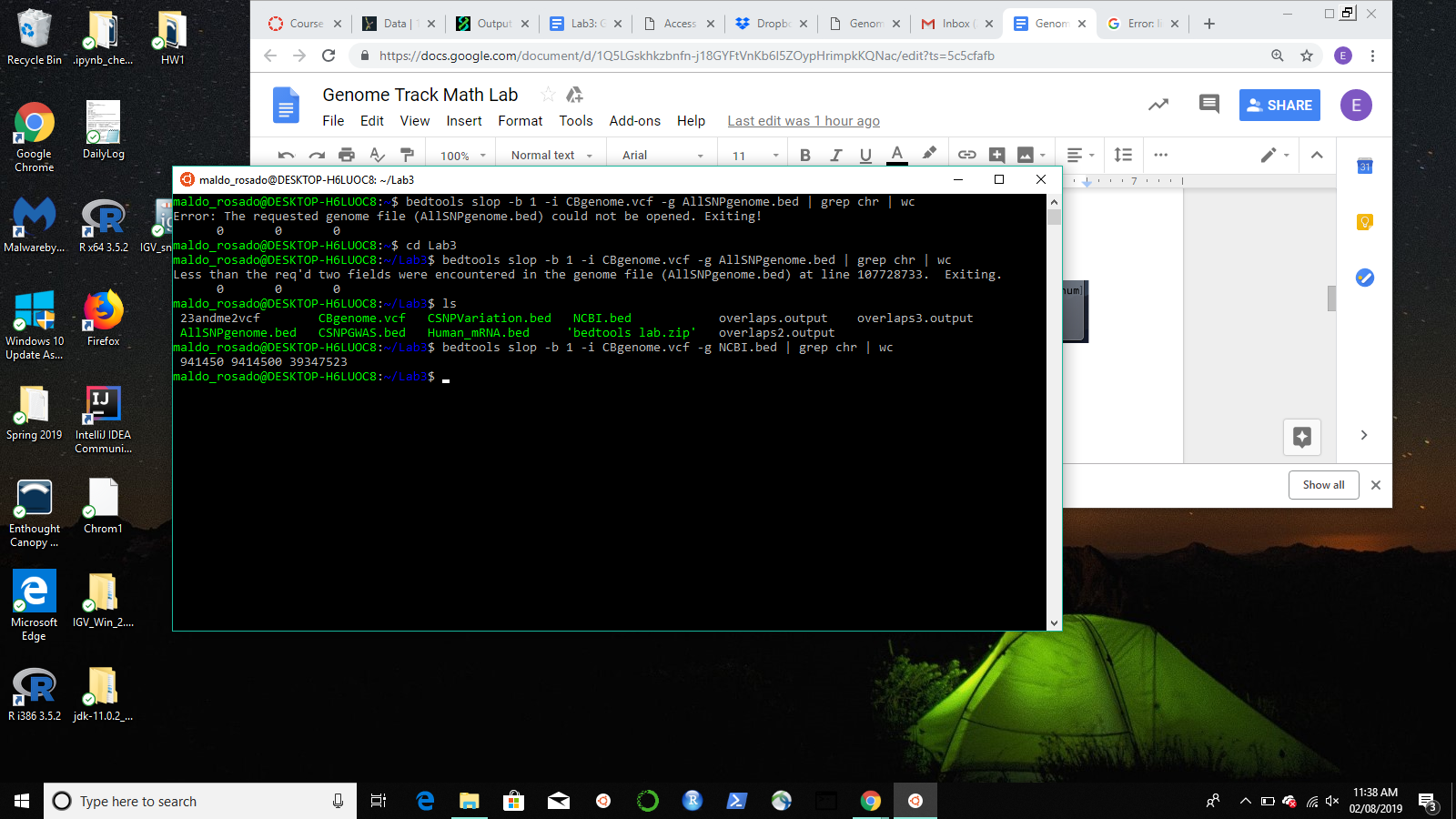
**What happens when you include flanking sequence, either via bedtools window or by selecting different options at track download time?**

There are SNPs in chromosome 1.

I do not think that the number that I got from the simple intersection is correct because it only takes into account one base pair.

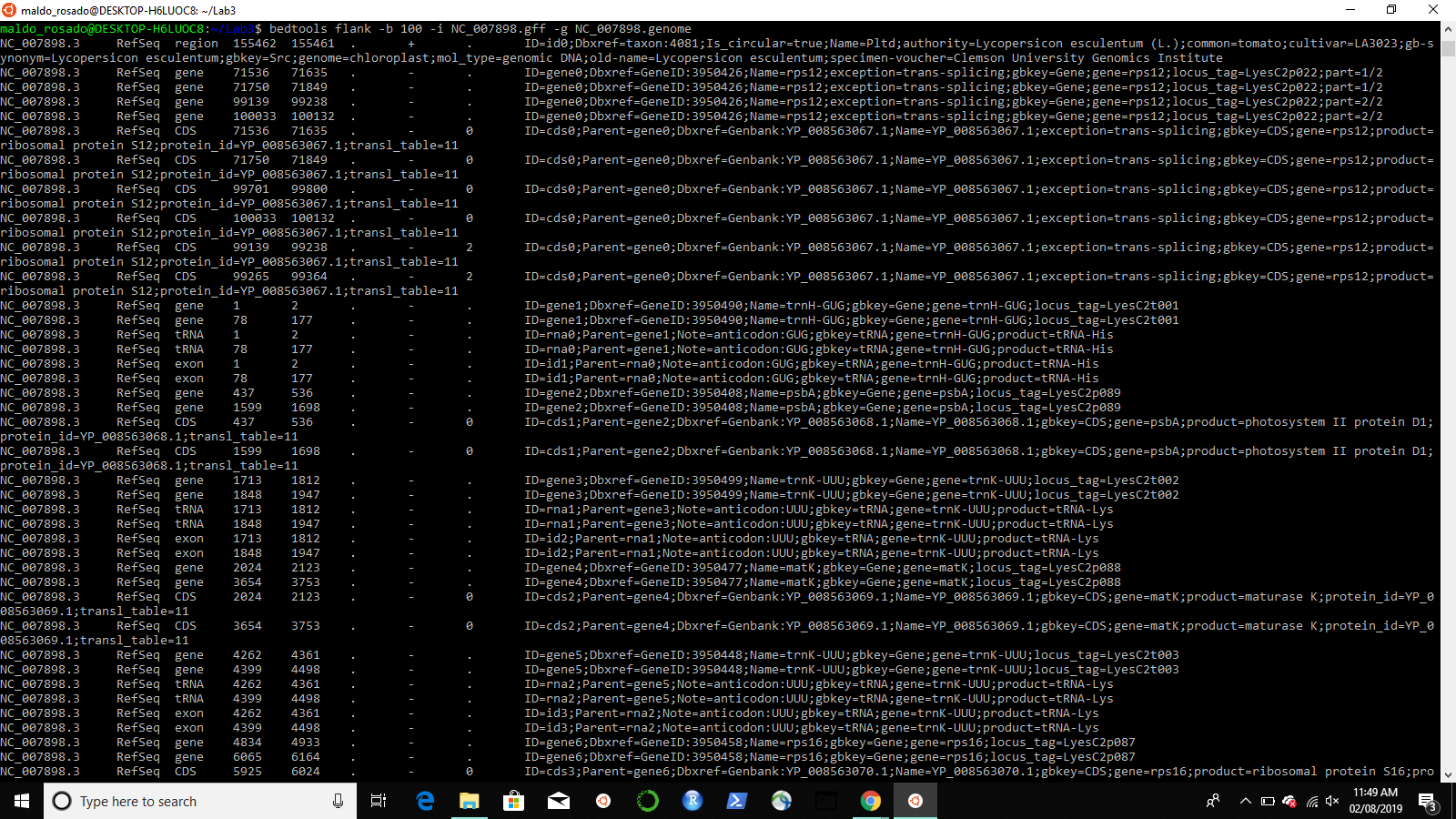
Whenever I used the flanking sequence via bedtools the number of SNPs increases.





**Problem 1: get the coordinate intervals for 100 bases flanking each feature in the GFF.**

bedtools flank -b 100 -i NC\_007898.gff -g NC\_007898.genome

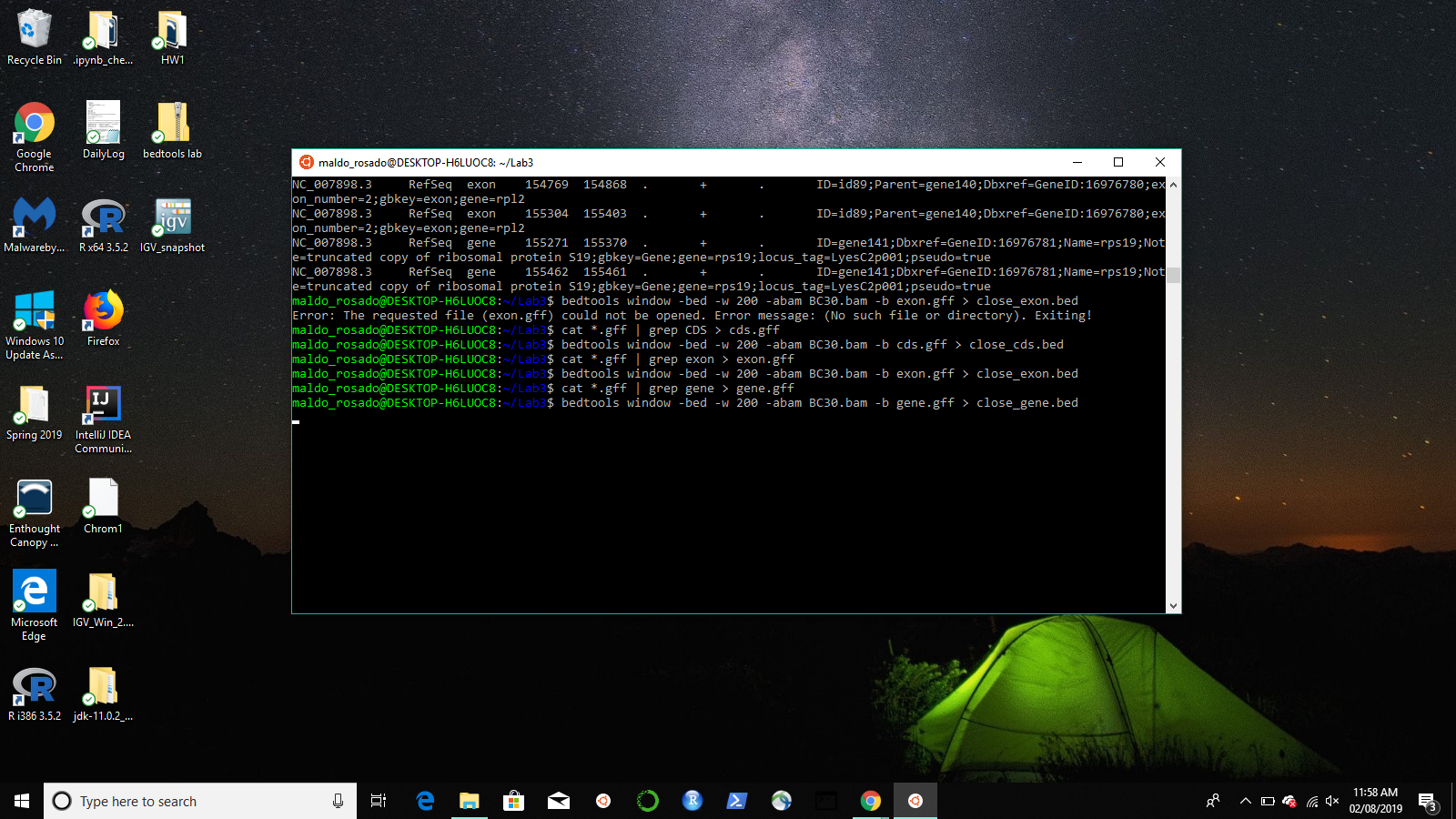
****

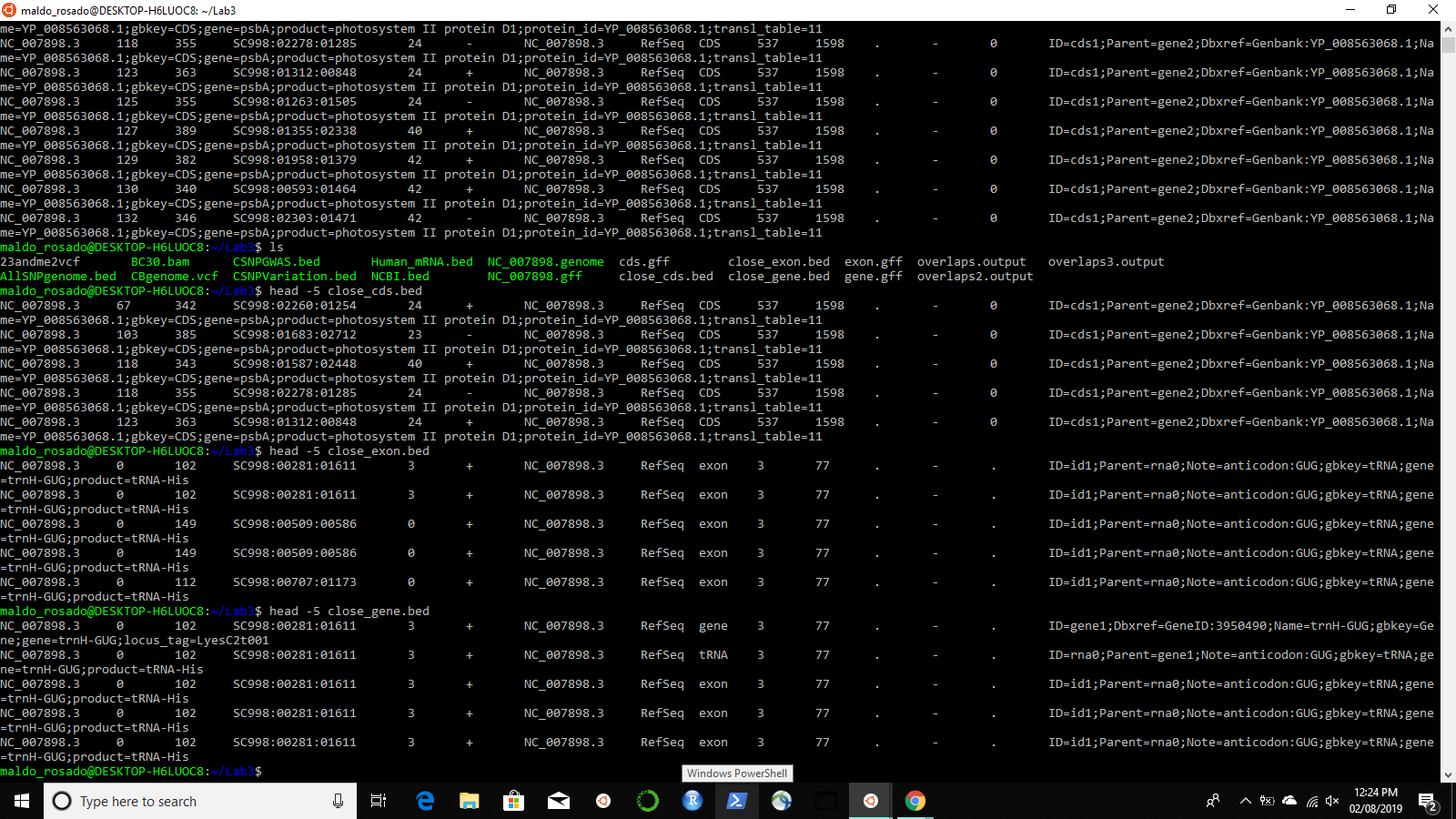
**Problem 2: for each feature in a file, find its nearest physical neighbors on the same strand (say, within 200 bp) in the chloroplast**

bedtools window -bed -w 200 -abam BC30.bam -b cds.gff > close\_cds.bed

bedtools window -bed -w 200 -abam BC30.bam -b exon.gff > close\_exon.bed

bedtools window -bed -w 200 -abam BC30.bam -b gene.gff > close\_gene.bed



****

**Problem 3. For each feature in the GFF, report only the features that have NO reads**

**overlapping then in the BAM file.**

bedtools window -bed -abam BC30.bam -b NC\_007898.gff

