**Assessment Data README**

You should have three folders:

* Assessment Data
* Genome Files
* Toy Data

**Genome Files**

This folder contains the fasta file and the GFF file for the genome that was used to call SNPs. These files can be used with either Assessment Data or the Toy Data.

**Assessment Data**

This folder contains the full SNP dataset that I will use when assessing your scripts. There are three files:

* assessmentData.vcf
* assessmentData.vcf.gz
* assessmentData.vcf.tbi

The file called assessmentData.vcf is the unzipped data in vcf format. This file is plain text - you can open it and read it. This file is included so you can inspect the file to understand it.

The file called assessmentData.vcf.gz is **the same data**, but it has been zipped using the bgzip method. You won’t be able to read this file, but this is the one you should use when you run your script. I have also included a tabix index for the zipped file (assessmentData.vcf.gz.tbi)

**Toy Data**

This folder contains a toy dataset that you can use when developing and testing your code. It contains four files:

* testData.vcf
* testData.vcf.gz
* testData.vcf.gz.tbi
* test.tsv

This dataset contains only two SNPs. I have selected one that is in a positive strand CDS and one that is in a negative strand CDS. The SNP data is in the same formats as described above. For these two SNPs, I have also provided the expected output, in the file called test.tsv.