# CFP Internship Diary January-March 2023

Week 1 commencing 03/01/23:

For my first week as an intern, I was working from home; getting to grips with my project titled "Completion and analysis of Defra oak GWAS panel". This project is about Acute Oak Decline (AOD) and finding genetic variants that might be associated with the disease by conducting genome-wide association studies (GWAS). This project is important as AOD is a major threat to oak trees across the UK, the underlying causes of which are still unclear. The results of this research could help point to mechanisms by which oaks might be resistant to AOD and therefore will help to combat the disease.

I spent this week doing a lot of background reading on research techniques and analysis in this field as well as the subject of my project. I learnt about techniques such as Illumina DNA sequencing, aligning short reads to a reference genome and GWAS and read the current research on AOD.
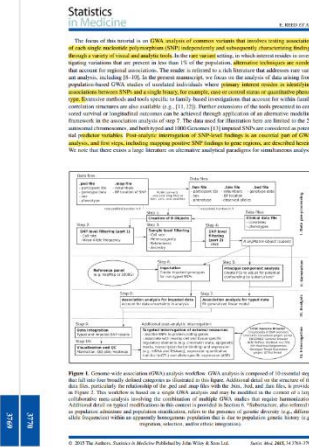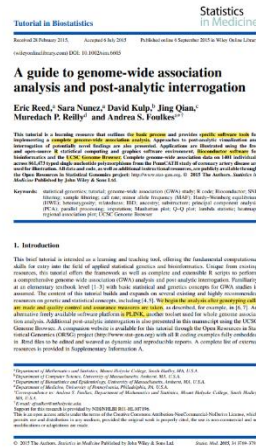
Week 2 commencing 09/01/23:
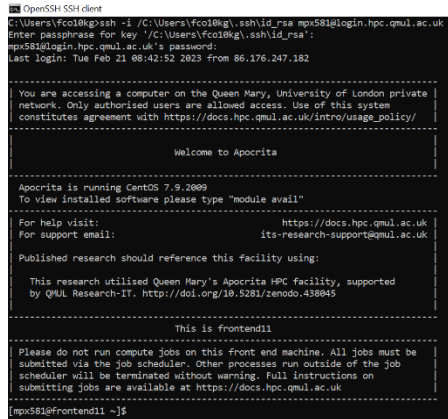
During this week, I had the opportunity to go to Kew and attend welcome sessions as well as meeting the other interns. I also learnt how to access the high-performance computing (HPC) cluster I would be using later on during my internship 'Apocrita' which is managed by ITS research at Queen Mary University of London.

This week I also attended a conference with the CFP where I was able to obtain an overview of all the different projects being carried out with the CFP.

Week 3 commencing 16/01/23:

This week I was working in the lab watching DNA extractions in order to learn the protocol. I learnt how to carry out a 3-day CTAB protocol for bad quality samples as well as how to quantify the results of the extractions using agarose gel electrophoresis, spectrophotometry and a fluorometer.

I also learnt some of the Linux programming language, Bash, through online tutorials. This was particularly useful as when using the HPC cluster, all programming and code is written using this language.

Week 4 commencing 23/01/23:

During this week, I worked through a genomics tutorial in which practiced with a bacterial genome learning how to align short DNA reads to a reference genome and call variants using the command line and Apocrita. This also involved learning to use genomics software such as FastQC, Samtools and bcftools.

I also started working on a database search to find previously published oak whole genome sequencing data which will be used in later analyses to augment the current oak dataset.
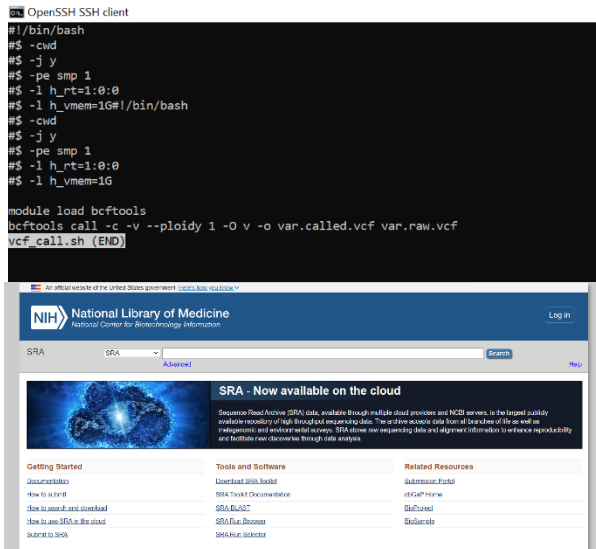
Week 5 commencing 30/01/23:

This week I was working in the lab preparing ~200 leave samples for DNA extractions. This involved cutting off 30-50mg of the greenest part of the leaf and putting it into an Eppendorf tube along with 2 steel ball bearings.

I also quantified extracted DNA using the Nanodrop spectrophotometer.
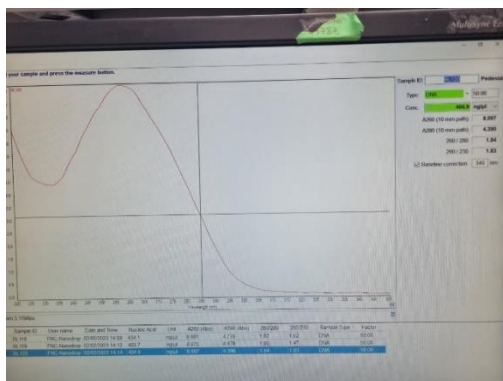
Week 6 commencing 06/02/23:

This week, I continued working in the lab, helping with the DNA extractions of over 300 samples. This helped me learn the protocol and improve my lab skills.
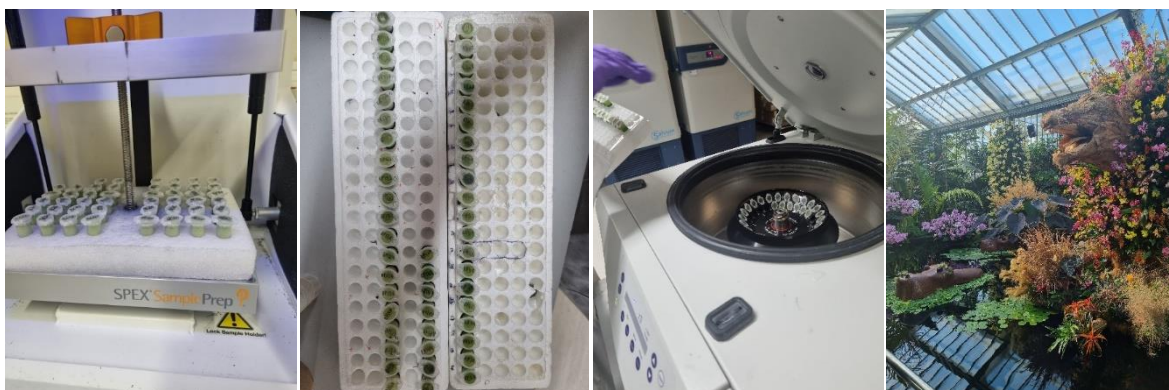
Week 7 commencing 13/02/23:

Continuing work in the lab, this week I had the opportunity to work through the entire CTAB extraction protocol from start to finish, re-extracting DNA from the samples that failed the first time.

I also quantified these results using the Nanodrop spectrophotometer and Quantus fluorometer.

The extracted samples were then sent off to an external company for library prep and sequencing.

Week 8 Commencing 20/02/23:

Since I had finished with lab work, this week I returned to my task of searching for oak genomic datasets to use later in the analysis.

I also helped out at and attended the GCBC conference tackling plant blindness which was very interesting.

Week 9 commencing 27/02/23:

During this week, since we had received DNA sequence data for the first 100 individuals, my supervisor showed me how to perform quality control for these samples as well as alligning the reads to the reference genome.
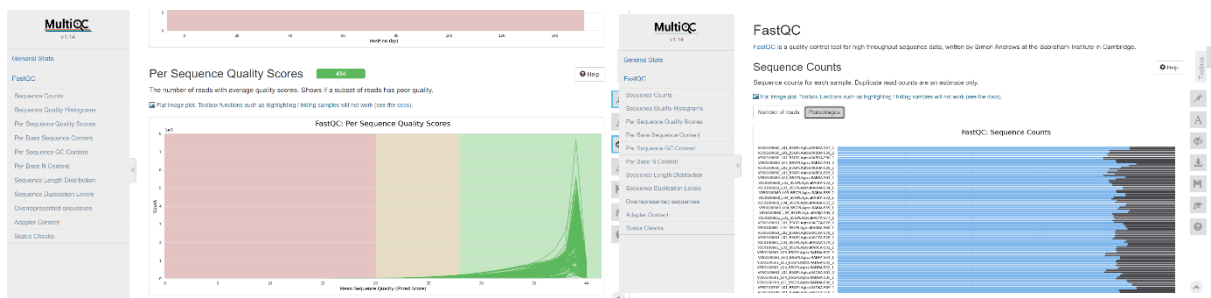
Week 10 commencing 06/03/23:

This week I conducted a literature search to find new areas of research for use later in the analysis. I was looking for fitness traits and traits of commercial value in oaks that we could test for associations with AOD and climatic variables.

Week 11 commencing 13/03/23:

Another part of the project was to look at oak resilience to climate change. To contribute to this I created a dataset of climatic variables for each of the sites in our dataset. I achieved this by downloading publicly available raster data of world climatic variables and clipped them to the UK. I then plotted the coordinates of each of our sites and extracted the associated climate data. I also did this for soil variables such as pH and nitrogen content.

As well as this, I started analysing some previous oak data in order to learn the GWAS pipeline. Using Apocrita and PLINK software, I converted hard filtered .vfc files into PLINK binary files (.bim, .bed and .fam files) and performed a principal component analysis (PCA) to obtain covariates.

Week 12 commencing 20/03/23:

During this week I ran the next steps in the GWAS pipeline- obtaining the actual GWAS results to show SNPs that are most likely to be associated with AOD. I also plotted these results on a Manhattan plot.

Week 13 commencing 28/03/23:

For my last week as an intern, I was working on the last stage of the GWAS pipeline- identifying the SNPs that are likely to be associated with AOD and working out if they are near a gene and therefore hypothesised what effect the SNP might have on the individual.

I also created a plot to show each of the sites and the proportion of trees in each site that have AOD.