Phytohormones are signal molecules that are important in plant defense to pathogens. It has been known that jasmonic acid (JA), ethylene (ET), and abscisic acid (ABA) phytohormones are elevated during a response towards a necrotrophic pathogen. These phytohormones produce and accumulate polyacetylenic lipids that are known to have antifungal activities. Falcarinol and Falcarindiol (falcarins) are the two most common polyacetylenes found in the carrots periderm. Although falcarins have been discovered decades ago, there is essentially nothing known about how pathogen-related signaling can regulate polyacetylene biosynthesis. In my current project, I am working on understanding which phytohormones either upregulate or downregulate the production of falcarins for plants defending themselves against white mold S. sclerotiorum. I am doing this by adding a certain concentration of said phytohormone and doing time points (4 hours post-inoculation (hpi), 12 hpi, and 24 hpi) with a control, mock, and phytohormone treatment. I am gathering my data by performing chemistry and molecular biology experiments such as acetylation to run them on the gas chromatography-mass spectroscopy (GC-MS) and/or flame ionization detector (FID) or cDNA synthesis to quantify the level of gene expression.

For my project, I will first Clean raw data from gas chromatography-flame ionization detector (GC-FID) utilizing OpenRefine in order to clean the data for less human errors when looking at the data and figuring out which section we need to highlight for our analysis of the quantification of the samples. The next step is already done from last class that Dr. Parchmen taught. Which is executing the clean raw data to give the statistical analysis. This was done by writing a python script in order to not plug and chug into an excel sheet that has all the data. For instance, some people put in the wrong peak area for each sample, internal standard samples, and or even the dry weight and this can cause major differences on the calculations. Not only that, but I also want to show graphs about the statistical analysis using R/ggplot. Lastly, since I talked to my professor about me leaving the laboratory – we have been discussing about all my experiments that I have performed and that I need to organize all the excel sheet or csv files from each experiment and combine into one folder. This comes to a challenge as I did not have any organization whatsoever and there were to many excel files to even count.

**Results**

Graphical user interface, text, application, email

Description automatically generated

Figure 1. OpenRefine script for cleaning GC-FID Raw data.

Table

Description automatically generatedTable

Description automatically generated

Figure 2. The differences between the clean OpenRefine version (left) versus the GC-FID raw data (right).

A picture containing text

Description automatically generated

Figure 3. Python script. Data analysis of raw GC-FID results of the area peaks of the samples and the internal standards given to calculate the accumulation of falcarins.

Text

Description automatically generated

Figure 4. R/ggplot script. Average of each sample for roots and shoots with treatment and controls within roots and shoots.

Graphical user interface, application

Description automatically generated

Figure 5. Organization of experiments. An example of how I had raw and clean data in one folder, data analysis for the python scripts, and R script and graphs in one folder for each experiment.

**Discussion**

This helped my project and my fellow research lab by cleaning all the data to eliminate the possibilities of targeting/highlighting the wrong data. OpenRefine help me avoid that option by initiating the script to the excel files and also converting it into a csv file that my python script needs to have it in. The python script allows me to avoid the possibility of adding a different area number and has all the sections in order and in view so you are able to see what number is for what sample. The second part is to add the R/ggplot script to check the averages of the samples versus its respective controls of roots and shoots. Although we do have prism – I believe this helps people understand the statistical analysis more than just plugging and chugging in prism. Lastly, since I had to go through all the data – this allowed me to organize each experiment that my peers and I have done and let me organize as such in figure 5. I would do one way anova and two way anova statistical analysis on R/ggplot to have better figures on understanding the significant differences between the controls and treatments and the differences of the roots and shoots.

organizing every excel sheet that was in there. It was confusing in that