STEPS TAKEN FOR PENNYCRESS DATA ANALYSIS

1. Make a new folder where the lipidomics will take place. This will be the 'working directory'.

a. Here I've made a folder named: pennycress\_lipidomics

b. Inside I've moved my Multiquant data, R scripts, and made two new folders: polar & neutral

2. Opening the R script for the polar lipids data analysis will open a panel at the upper half of RStudio.

a. Lines 18-19 detail what columns are included in the data table that will be prepared from the Multiquant data.

b. Note: not all of the columns are necessary for the actual data analysis (e.g "mrm", "height") but could be useful in future analyses

3. Making a data table suitable for the R script requires moving the Multiquant data within an Excel spreadsheet to have the appropriate column names, sample numbers (i.e. sample\_no), tissue identity (i.e. tissue), and the tissue weight (i.e. tissue\_weight)

a. copy over the columns of data and information desired or required

i. THE ABSOLUTELY REQUIRED COLUMNS INCLUDE: sample\_no (to be made), tissue (to be made), component\_name, component\_group, area, and tissue\_wt (to be made). AND THE NAMES OF THOSE COLUMNS MUST BE TYPED EXACTLY AS STATED HERE AND IN THE R SCRIPT.

b. Change: Sample Name --> sample\_id

Component Name --> component\_name

Mass Info --> mrm

Component Group Name --> component\_group

Area --> area

c. Add: sample\_no

tissue

tissue\_wt

d. Fill in the required information for the new columns.

e. For quality of life, Find/Replace all 'N/A' in the 'area' column and replace it with 0.

f. Next, save the data table from the spreadsheet as a tab-delimited text file. This can be done by copying the content from the spreadsheet, pasting it into a text editor (e.g. Notepad), and then saving the resulting file.

i. Here I've saved the .txt file as: seed\_polar\_lipids.txt

4. The R script requires editing so that it knows where your data is located.

a. On line 22, change the "PATH TO YOUR DATA FILE" with the file directory where your tab-delimited text file (seed\_polar\_lipids.txt) is saved, keeping the quotes ("").

i. The easiest way of doing this and to avoid typing errors is to copy from the File Explorer address and pasting the result into the R script.

\*Note: I’ve since learned R on Windows will require ‘/’ in the path name not ‘\’

b. On line 29, change the "FILENAME OF YOUR DATA FILE" to the name of your tab-delimited .txt file (seed\_polar\_lipids.txt)

i. Note: Be sure your data file is copied into your working directory (i.e. the 'polar' folder, in this case)

5. Other parts of the R script /may/ need to be changed depending on the requirments of the experiment or analysis:

a. If you would like to use a different internal standard or adjust the amount used, this is located on lines 130-181

i. An example for PC is shown here:

seed\_polars\_PC\_stds\_IS <- filter(seed\_polars\_PC\_stds, component\_name == "PC-35:1-d5 (17:0\_18:1)")

setDT(seed\_polars\_PC\_stds\_IS); setDT(seed\_polars\_PC\_species)

seed\_polars\_PC\_species[seed\_polars\_PC\_stds\_IS, nmol\_mg := (3.85 \* area) \* (1 / i.area) \* (1 / tissue\_wt), on = "sample\_no"]

These three lines tell the R script to normalize the PC molecular species by the internal standard "PC-35:1-d5 (17:0\_18:1)" which is one of the standards in the Avanti Polar Lipids UltimateSPLASH mix. The last line indicates "3.85 nmol" have been added to the samples. If you would like to normalize against a different internal standard, then you would need to change "PC-35:1-d5 (17:0\_18:1)" to your desired internal standard as it is listed in the data table (any errors in the typing of the name will cause errors) retaining the quotes (""), and you will need to change the 3.85 to the correct nmols added your samples.

ii. Note: ONLY CHANGE THESE IF YOU ARE CERTAIN TO USE A DIFFERENT INTERNAL STANDARD.

iii. This example shows changing the internal standard for PC, but the process is the same for any other lipid class.

6. The labels for the graphs to be made will need to be changed, beginning on line 430 up to line 550.

a. The easiest way to do this is to highlight what needs to be changed, in this case ("14 DAP", "17 DAP", "20 DAP", "23 DAP", "mature", "germ"), and then click on the Find/Replace button above the script editor (a magnifying glass icon). You'll notice the Find field is already filled with what you've highlighted. Next type in ("14 DAP", "mature") into the Replace field. Then press All to replace all instances of those labels with your desired labels.

7. With all of this complete, you are now ready to run the R script. Select all lines (Ctrl + A), and then press Run at the top of the R script editor (a little white box with a green arrow point right icon).

The neutral lipids analysis is done much in the same way.

8. Start a new session of RStudio, or clear your Environment panel, Console, and Plots panel (a little broom icon in each)

9. Open the R script for neutral lipids.

a. Line 9 indicates which columns are included in the data table to be prepared from the Multiquant file.

10. Prepare the data table that is suitable for the R script

a. Copy over the desired columns of data and information to a new spreadsheet.

i. THE ABSOLUTELY REQUIRED COLUMNS INCLUDE: sample\_no (to be made), tissue (to be made), component\_name, component\_group, area, and tissue\_wt (to be made). AND THE NAMES OF THOSE COLUMNS MUST BE TYPED EXACTLY AS STATED HERE AND IN THE R SCRIPT.

b. Change: Sample Name --> sample\_id

Component Name --> mrm\_component\_name

Component Group Name --> component\_group

Area --> area

c. Add: component\_name

sample\_no

tissue

tissue\_wt

d. Note: The R script makes use of regularized naming for the component names. This is why there is the mrm\_component\_name and component\_name columns. The former is what is indicated in the MS method, but the latter is what is required in the R script. The MS method could be changed to have the correct names from the component\_name column, which would negate the need for the separate mrm\_component\_name column.

e. Fill in the required information for the new columns. The component\_names for the molecular species here are in the second tab of the provided sample data.

f. For quality of life, Find/Replace all 'N/A' in the 'area' column and replace it with 0.

11. There are special considerations for TAG and their individual FA compositions. Prior to preparing the R sript or data table, you will have gone through each of the TAG-FA extracted ion chromatograms (XIC) in Analyst to determine the FA makeup of each TAG molecular species. Here you will have also determined which MRM transitions to use for the quantification in the R script. Some of those transitions may require being divided by 2 or 3 (e.g. multiple of the same FA in the same TAG molecular species, or the same FA found in multiple TAG molecular species whose peaks overlap). To adjust the divisions of the required TAG molecular species, lines 27-31 within the R script will need to be adjusted.

a. On line 29, you can add to the list already present (or remove those unnecessary) by adding the molecular species to be divided by 2. To divide by 3 you can add a new line, copying either lines 30 or 31 and adjusting accordingly.

b. Note: To indicate the determined FA compositions of TAG (as determined when reviewing the individual XIC in Analyst), you will need to type this for the specific component\_name within the prepared data table before saving it as a tab-delimited .txt file. This has already been done within the sample data, but, as an example, it would involve changing 'TAG-52:6-FA16:0' to 'TAG-52:6-FA16:0\_(16:0/18:3/18:3)' (no quotes)

c. Also note that unlike the polar lipids R script, the neutral lipids R script explicitly indicates which molecular species it is using for quantification for DAG and TAG (not for MAG). This means that for the molecular species you are interested in quantifying, the list of indicated DAG on line 45 and TAG on line 47, will need to be changed. This can be done either in the R script editor panel, a text editor, or an Excel spreadsheet.

d. Once you've indicated the FA compositions for TAG and DAG within the spreadsheet of Multiquant data, this can be saved as a tab-delimited text file as done with the polar lipids data.

i. Here I've saved the .txt file as: seed\_neutrals\_lipids.txt

12. Edit the R script as was done in step 4 (above) for the "PATH TO YOUR DATA FILE" (the neutral folder) and "FILENAME OF YOUR DATA FILE" (seed\_neutral\_lipids.txt)

13. Lines 50-63 indicate what internal standards will be used to normalize the data. Change these as indicated in step 5 if necessary, otherwise do not change them.

14. The labels for the graphs to be made will need to be changed, similarly to step 6, beginning on line 99 and up to line 160.

15. With all of this complete, you are now ready to run the R script. Select all lines (Ctrl + A), and then press Run at the top of the R script editor (a little white box with a green arrow point right icon).

Other notes for Windows: .png files may be better for images (Windows lacks any native .svg viewer as far as I can tell). R may require ‘svglite’ library to save .svg images.