SUPPLEMENTARY MATERIAL I

Protocol for MetaboAnalyst and Excel Analyses

Instructions for MetaboAnalyst upload were adapted from an Honors Thesis by RU student Yekaterina A. Rousakova
Instructions were updated by Elizabeth Baldo to be specific to the accompanying thesis. Excel instructions are entirely specific to this thesis.
These steps were done on an Apple computer, some steps might vary slightly for Windows

Principal Component Analysis (PCA) and Partial-Least-Squares Discriminant Analysis (PLS-DA) graphs were created using MetaboAnalyst, which is a comprehensive tool suite for metabolomics data analysis. Before using MetaboAnalyst, each replicate for each *Emiliania huxleyi* and *Oxyrrhis marina* strain must be formatted a certain way via Excel. The first part of this protocol details the steps required to format data in Excel for eventual upload onto MetaboAnalyst. The second part of this protocol details how to upload and normalize this data in MetaboAnalyst, as well as the steps used in MetaboAnalyst to perform statistical analyses and download data obtained from these tests. The third part of the protocol details how this data from MetaboAnalyst was manipulated and analyzed in Microsoft Excel. This document focuses on how to prepare *E. huxleyi* lipids for upload; the process is the same for *O. marina*, only different columns are selected from the original lipids data sheet.

In Microsoft Excel

- EACH replicate (i.e., grazed (reps 34-36) and grazed (reps 37-39)) needs to have its own CSV file, this information can be copy and pasted from the original lipids data sheet.
- Make a different Excel sheet for each replicate. (For this project, this means 24 Excel sheets)
- Each Excel sheet should contain:
    - Column A: row m/z
    - Column B: row retention time
    - Column C: peak area (this is the actual data for each lipid in a replicate aka a column such as Ehux 624_34 grazed)
- **Save each Excel sheet as a CSV file!** This is a very important step or MetaboAnalyst will not be able to read your data.
- In order to save as a CSV file:
    - In the Excel file go to file -> save as
    - Under “File Format” select CSV UTF-8 (Comma delimited) (.csv)
    - Hit save
- Do this for every treatment, and every strain. These files need to be distinguishable from one another and every treatment. An example would be “374_34”; the strain, followed by the replicate number.
- After this is done, group the treatment together in a folder.

    - Ex. Three separate trials were done for exposure to grazers, and three CSV files were labeled "374_34, 374_35, and 374_36". A new folder was created called "374 Grazed".
- Once you have done this for all the treatments of a strain, nest all the treatments together in a new folder.
    - Ex. There were two different treatments for 374, with three replicates each in a folder (374 Grazed), those two treatment folders were grouped together and a new folder called "374" was created.
- After this, convert your final grouped file into a zipped folder
    - Right click on a grouped folder -> scroll down to "compress"
    - This will zip the folder

In MetaboAnalyst

- Click on "Statistical Analysis [one factor]"
- **Upload your data under "A compressed file (.zip)"**
    - Data Type: MS peak list
    - Data File: Choose your newly created zip folder
- **Processing MS peak list data:**
    - Mass Tolerance (m/z): 0.025
    - Retention time tolerance: 30.0
- **Data Filtering: Interquartile range (IQR)**
    - *Ignore* "Filtering features if their RSDs are > …."
- **Normalization overview:**
    - Sample normalization: Normalize by median
    - Data transformation: Log transformation (base 10)
    - Data scaling: Auto scaling
    - Hit "Normalize" then "Proceed" after viewing results
- **You can now view all the different data generated.**

- ***To change color of PCA and other plots*** 
    - Click the little color palette (see image above) in the upper right above the plot you are looking at.
    - Go to "Colors & Shapes"
    - From here, you can adjust the color and shape
    - Hit "Submit"
    - On the PCA plot page, hit "Update"

- ***To Download Data***
- To transfer all the data from MetaboAnalyst go to "Download" in the side bar
- Download whatever you need or select "Download.zip" to download everything in one file.
    - For my thesis, *all* data was downloaded for every comparison
- Click on "Generate Report" to download a readable PDF summary of all the statistics you did and the graphs MetaboAnalyst created.

In Microsoft Excel

PCA and PLS-DA Loadings Analysis

Analysis of the PCA and PLS-DA plots are done in the same way using the exact same steps. The only difference is that to analyze the loadings of a PCA axis, the PC1 or PC2 loadings are used. In a PLS-DA, the weights are labeled as Component 1 and Component 2, etc. To maintain clarity, the directions below are for a PCA analysis performed on PC1 loadings. To perform an equivalent PLS-DA analysis, simply replace "PCA" with "PLS-DA" and "PC" with "Component" in the directions below. In some cases, an analysis will be done on PC2 or Component 2 loadings. In that case, simply replace "PC1 loadings" with "PC2 loadings" etc., in the directions below.

- **Download relevant information from MetaboAnalyst reports into Excel**
    - *Put each in a different Excel spreadsheet or tab*
    - Download *Fold Change* and *log2(FC)* data from MetaboAnalyst into Excel
    - Download *PCA Score* data from MetaboAnalyst into Excel
    - Download *T-Test* data from MetaboAnalyst onto Excel
    - Download *PCA Loadings* from MetaboAnalyst onto Excel
- **Move PC1 loadings to a new tab in Excel**
    - Sort the PC1 loadings from low to high
    - Highlight the lowest 30 PC1 loadings and the highest 30 PC1 loadings
        - *You can highlight any number of lowest and highest lipids you would like. 30 were used for my analyses.*
        - Duplicate the tab
        - Delete the PC1 loadings between your selected loadings
        - These 60 PC1 loadings are what the rest of the analysis will be performed on
    - For each of the 60 PC1 loadings, match the fold change, log2(FC), t stat, and p-value to the corresponding PC1 loading.
        - This ensures that the lipid identified by that PC1 loading is significant, and lets us know how much the lipid changed (quantitatively) between the control and grazed treatments
    - For each of the selected PC1 loadings, use the given m/z and retention time numbers to locate the corresponding lipid it belongs to
        - Open the original lipids data sheet, created by Dr. Poulson and colleagues
        - Go to the EhuxLip(2)_normal tab (or OxyLip(1)_normal tab, if you are analyzing *O. marina* lipids)
        - Search for the peakgroup_mz value
            - The numbers from MetaboAnalyst are rounded, so remove one or two decimals from the number in order to locate the lipid
            - Check against the peakgroup_rt value to ensure the correct lipid is identified

        - Copy the lipid ID# and Compound Name from the original lipids sheet onto the PC1 loadings tab, next to the matching PC1 loading
    - Once complete, create columns for *Lipid Class* and *Lipid Name*
        - Fill these in accordingly from the original lipids data sheet
        - Google searches and academic papers are used to identify the lipid names and lipid classes of these selected lipids
            - Lipid class is also available on the original lipids data sheet from Harvey et al. (2015)
- **Create graphics in Excel to visualize results**
    - Pie graphs were used in this experiment, but many other graphics could be used instead
        - Other tools could be used to create graphics, such as R or Prism
    - Pie graphs were created in Excel for both the lipid class distribution and the lipid subclass distribution in the highest 30 and lowest 30 PC1 scores
    - These pie graphs were used to locate unique lipids in both treatments, and understand how lipids and lipid classes changed due to treatment

SUPPLEMENTARY MATERIAL II

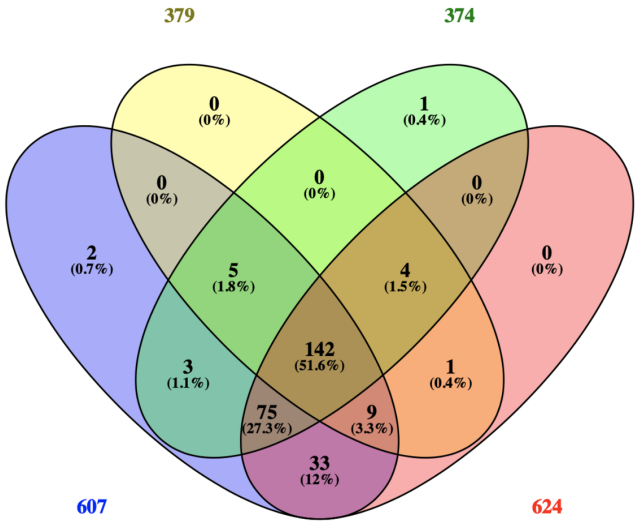

Figure S1. Venn diagram comparing the grazed lipidomes of all four *E. huxleyi* strains (created using Venny 2.1).

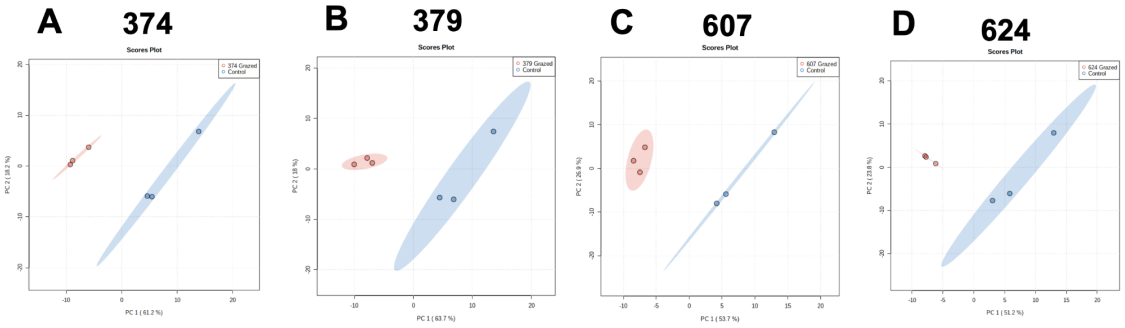

Figure S2. Results of the unfed vs. fed lipidome analysis of *Oxyrrhis marina* fed each of the four *Emiliania huxleyi* strains. PCA scores plot of the *O. marina* lipidome before and after grazing on *E. huxleyi* **A**, strain 374 **B**, strain 379 **C**, strain 607 **D**, strain 624. The negative scores on each plot correspond to *O. marina* lipids present after grazing on a specified *E. huxleyi* strain (red), while the positive scores on each plot correspond to *O. marina* lipids in an unfed state (blue).