1. Methodology

The Garrido-Oter lab aims to develop a high-throughput phycosphere modeling system to study synthetic communities (henceforth, SynComs) or rationally designed compositions of microorganisms **[cite syncom paper]**. The proposed system makes use of the ubiquitous soil algae *C. Rheinhardtii*(Chlamy) and bacterial strains (henceforth, Strain(s)) from the Indexed Chlamydomonas-associated bacterial Library (ICL). Through the successful development of this system, the group hopes to uncover fundamental rules and processes, which determine community establishment, dynamics and stability **[garrido-oter]**. Absorbance and fluorescence spectroscopy are used to probe the system as an efficient way of quickly iterating over hypotheses and experiments. The two main instruments employed in the generation of the data used in this research are the Tecan M200 Inifinite Pro and the Photon Systems Instrument Multi-Cultivator 1000-OD (henceforth Tecan and PBR, respectively). A third instrument, the Beckman Coulter Counter Multisizer 3 (henceforth, Coulter) is employed to obtain cell count estimates of PBR experiment samples. Two main data sets were created for analysis: Screening and PBR absorbance data. Some parts of each kind were generated prior to the commencement of this Thesis, other parts were created specifically for it and are subsequently analyzed through predictive models grounded on the Beer-Lambert law, which is derived from Tecan-measured calibrations.

3.2 Growing Chlamy & Strains

Before any experiment can be conducted, Strains and Chlamy are grown to quantities allowing experimentation; this typically means stocks of optical densities between 0.1 > OD > 1.0. Chlamy and Strains are individually obtained from stored samples in the ICL or other bacterial collections housed at the MPIPZ. After thawing, the microorganism are placed in respective pre-growing vessels for 2-5 days or until a desired cell density has been achieved. They are co-inoculated with respect to individual experimental designs. Pre-growing vessels are either tryptic soy agar (50%) plates for Strains or TAP medium for Chlamy. Once SynComs have been designed and their strains pre-grown, individual members are left to grow continuously whenever possible.

3.2.1 Media

Several growing media are used during Tecan experimentation. Tris-acetate-phosphate (TAP) medium is the standard (full) growing solution for Chlamy: Acetate and ammonium serve as Carbon and Nitrogen sources, respectively, and Tris helps buffer the pH **[Harris, 1989]**. Chlamy is a heteroautotroph, henceit can grow via photosynthesis or use external organic carbon sources. In order to force Chlamy to photosynthesize, TP Medium (TAP without acetate) is used; however, this requires CO2 to be introduced into the system. TAP-N, is TAP medium without Nitrogen and TP-N has no acetate or Nitrogen. Chlamy is unable to grow in any of the -N media and are used as controls **[Figure X]**. Previous work has determined which Strains grew in TAP medium. The produced results clarified which Strains cannot grow in full media, as shown in **FigureX** as flat lines, and the list is provided in Supplementary Data. Recall that the ICL was generated from harvested Strains, which were co-inoculated alongside Chlamy during growth. It thus e stands to reason that those which cannot grow in TAP by themselves must receive some nutrition either from Chlamy or through other favorable metabolic exchanges with other community members. This notion helps guide SynCom design and experimentation.

|  |
| --- |
|  |
| **FigureX**. Pilot experiment determining which bacterial strains from the ICL grow in TAP medium. |

3.3 Experiments & Data

Two general data sets were obtained for this Thesis: screening experiments absorbance and fluorescence data, and photo bioreactor experiments absorbance data with its sample’s Tecan measurements, cell count estimations and 16S profiles. Tecan data are generated as excel files while PBR data is continuously measured and stored in CSV files. Processing of this data is discussed in another section. The screening experiments were conducted to obtain Strain-specific or SynCom-specific interaction information as well as to test parameters for phycosphere modeling. PBR experiments were conducted to study the structure and dynamics of the specific SynComs themselves across longer time spans. This data is provided in Supplementary Data.

3.3.1 Data Generators: the Tecan & Multi-Cultivator

The Tecan and PBR instruments employ the same fundamental physics properties to measure absorbance values and by doing so allow the Beer-Lambert law to be leveraged on their data. While the PBR is sold for the purpose of continuous production of microorganism under fixed parameters, in the lab PBR serves to model test and to model the phycospheres over long periods. Conversely, the Tecan serves to prototype and test inoculation mixtures and their stability, as well as to iterate over and test hypotheses without the time overhead that is required to prepare and set-up PBR experiments. The Tecan’s measurements are made over standard flat-bottomed Greiner96-well plates (henceforward, microplates) which hold up to 300 μL of sample per well. Tecan experiments last no longer than two weeks and measurements are conducted daily. In comparison, the PBR hosts eight tubular vessels, each capable of housing 80 mL samples; measurements are taken continuously over user-specified time intervals across several weeks to months: the latest experiment PBR4 is still ongoing **X MANY Days LATER**. The Tecan served to generate the screening data as well as the calibration data that allowed for the derivation of Beer-Lambert law based models. Samples collected from some of the PBR experiments are also measured with the Tecan which allows testing the model’s transferability across these two systems. While it is possible to generate models using the PBR, thiswould not only require greater quantities of growing media, Strains, and Chlamy, but also time: the Beer-Lambert law and the Tecan help avoid this waste. Fluorescence is measured during some of the Tecan experiments (the PBR is unable to measure this), but these values are discarded for reasons discussed later. [**cite manuals]**

3.3.2 Screening Experiments

S1 is conducted to obtain information on binary interactions between specific Strains and Chlamy in the four growing media previously mentioned. Strain-only wells serve as controls in this and subsequent screenings as do the -N media. Previous analysis of S1 concluded that nothing grew in -N media (as is expected) thus subsequent screens no longer made use of these media. The specific strains used in each screening experiment areprovided in the Supplementary Data. Absorbance measurements were collected at 600 and 750 nm wavelengths (A600 and A750, respectively), and fluorescence was measured at 680 nm wavelength after 480 nm excitation (henceforward F680). S2 was conducted over three conditions in TP media: Strain-only controls, Strain and Chlamy binary interactions, and Strain and Chlamy in the context of a defined SynCom (denoted as B, B+C, SC+C, respectively). Absorbance measurements were taken at A750, and fluorescence at F680. S3 was conducted to obtain information on whether the initial inoculum ratios of Chlamy-to-Strain in TP media make an impact on the overall growth. These ratios were 1:1.6, 1:16, 1:160 as measured by OD. Absorbances A600 and A750 were measured as well as F680. S4 is conducted to determine if SynCom growth necessitated Chlamy or if Chlamy supernatant was sufficient for their growth and was meant to provide further evidence for the modulating effect of Chlamy. Ratios were used to determine growth impact as in S3. Absorbances were measured at A680 and A750 and fluorescence at F680. The screening experiments 1-3 (S1-S3) were conducted prior to commencement of this Thesis while Screening 4 (S4) was conducted for this thesis. The S1-S3 well fractions had been resolved and thus could had not been interpreted and thus are given to be analyzed along with S4 for this project.

3.3.3 PBR Experiments

Given that the phycosphere modeling system is still under development, the PBR generated data does not only serve to elucidate SynCom properties, it also informs on the conditions that best keep the system from crashing. This is especially true for experiments one and two (PBR1-2). This data was generated before the commencement of this Thesis and helped guide the growing conditions for PBR3, which was conducted directly for this project. Unfortunately, PBR3 failed two weeks post-inoculation: The resulting data was nevertheless analyzed. This particular experiment was restarted as PBR4 after modulating growing conditions and has been continuously growing and generating data since. The PBR is making continuous measurements over specified time intervals and storing the information on a CSV file. The information can be accessed d live via device-specific software or downloaded to be analyzed. **[show how this data looks like]**

|  |  |
| --- | --- |
|  |  |
| **Figure X. A.** Photo Bioreactor & Accessories Schematic: 1. Fresh Media; 2. Turbidostat; 3. PBR 4. Waste Tank. **B.** PBR in the dark shining green due to Chlamy | |

3.3.4 PBR Correlating Data & Usage

As noted earlier, PBR experiments are not only measured for absorbance values; it is rather that each vessel—i.e synthetic phycosphere—is sampled and subsequently has its cell contents estimated with the Coulter and its 16S sequences profiled. This data will be compared against concentration predictions and should, at the very least, facilitate the tuning of growing conditions of the system. As the system gradually becomes more robust and consistent, the group aims to perform sequencing less frequently. The general sampling steps are as follows: First, a 1 mL sample is taken from each of the eight vessels. Second, a 250 uL aliquot from each sample is transferred on to a microplate and measured with the Tecan at A680 and A750 wavelengths and the data is stored. Another 24 uL aliquot is taken from each sample and transferred into its respective polymerase chain reaction (PCR) tube. Then 40 uL of a DNA buffer is introduced into each tube and thermocycled for 30 minutes at 94 oC **[cite 16S methods]**. After thermocylcing is complete, 40 uL of a different DNA buffer is pipetted into each tube and subsequently frozen (as are the remainder of the samples). Once either an arbitrary number of daily samples has been collected or the experiment is considered complete, the frozen samples are thawed and prepared for 16S sequencing **[provide a protocol]**.

Typical analyses of 16S sequencing results are performed over the type of bar graph shown in **FigureX.** For this Thesis, the relative abundances (RA) of the bacterial strains and the Chlamy are compared against the predictions of the models as a form of validation for the goodness of using this method. As an example, each vessel’s 16S sequencing reveals the composition of the community in terms of Class relative abundances for any given day. The resulting sum of the bacterial Class RA is then compared against the Beer-Lambert model predictions for bacterial content as are the Chlamy’s RA to their predicted fraction.

|  |
| --- |
|  |
| FigureX. Relative abundances of 16S sequencing from PBR4. A fellow group member, Dr. Pepe Flores, provides preparation for sequencing, sequencing and the visualization of this data. |

3.4 Modeling I

The Beer-Lambert law states that the absorbance of a substance is proportional to the concentration of the sample and the path-length through which the light travels. The systems, which employ this law often, standardize the length of the vessel through which light travels to 1cm, and by doing so simplifying the procurement of the extinction coefficient, which typically has units of (M-1cm-1). This standardization of the path-length transforms the Beer-Lambert law to Equation 8 and facilitates the derivation of the constant as it would be equal to the slope of the resulting calibration curve having units of molarity (M-1) instead. The Tecan, however, does not measure absorbance across 1cm sample vessels, and thus employs a path-length correction algorithm that utilizes an internal standard. The magnitude of the slope of the lines from these resulting calibration curves can hence be written to conform to standard extinction coefficient units (M-1cm-1) as needed, and facilitate the portability of the extinction coefficients to other systems employing absorbance.

|  |  |
| --- | --- |
|  | Equation 7 |
|  | Equation 8 |

3.4.1 Extension of the Beer-Lambert Law

Commonly, the Beer-Lambert law is employed to generate simple regression curves for single component solutions of chemical compounds or macromolecules (e.g. proteins) allowing researchers to quickly estimate the concentration of their sample and move onto experimentation and analysis **[measurements by OD]**. Additionally, a more elegant and expanded form of this law exists, which can be employed to resolve the concentrations of individual species in multiple component solutions and serves as the foundation of this study. The extension of the Beer-Lambert law states that in a multiple component system, its measured absorbance is equal to the sum of the individual measured absorbance of each species (Equations 9 & 10).

|  |  |
| --- | --- |
|  | Equation 9 |
|  | Equation 10 |
|  | Equation 11 |

Consequently, in order to determine the concentrations of N species, N wavelengths must be measured, and an N equation system needs to be solved for each species’ concentration. In this theoretical system, there are N2 coefficients which would first need to be determined as well. The lab’s envisioned system can theoretically be modeled in this way if every Strain is considered to be an individual component of each species’ extinction coefficient is determination **[go into details in the discussion]**. As of now, however, the goal is to ensure that the resolution between Chlamy and SynCom bacterial strains is possible. Consequently, this Thesis only considers a two component system (Strains and Chlamy) with four extinction coefficients.

3.4.2 Forced Constraints on Modeling

The Tecan is capable of taking absorbance measurements between 200 and 1000 nm, but the design of the Tecan screening experiments that were provided for analysis and the component specifications of the PBR dictate which wavelengths are to be used for modeling **[in the discussion discuss always measuring over a spectrum]**. S1 and S3 measurements were taken at A600, A750 and F680. S2 was measured at A750 and F680, and S4 at A680, A750 and F680. Moreover, the particular Multi-Cultivator model currently serving to prototype this system only measures absorbances at A680 and A720. Lastly, the Tecan measurements of PBR 1 & 2 samples were measured at A680 and A730 and these wavelengths are also modeled. In order to probe the data provided, the models must consequently be generated using those wavelengths as regressands.

3.4.3 Exclusion of Fluorescence Measurements & Models

During the initial development phase of the phycosphere modeling system, daily fluorescence measurements were found to be a good proxy for measuring the growth of Chlamy alone. This is due to Chlamy’s Chlorophyll being a fluorophore and fluoresces near 680 nm **[cite chlorophyll fluorescne and absrobance paper]**. As these cells grew, the signal became stronger than expected, however, and paradoxically, Chlorophyll also absorbings at this wavelength likely had the effect of underestimating the actual Chlamy content of those measurements. An example of these initial measurements is shown in **FigureX [as seen in pepe’s initial work, show image]**. This overlapping of absorbance and fluorescence wavelengths causes the Inner Filter Effect (IFE); in essence, as one Chlorophyll molecule becomes excited and fluoresces, the neighboring molecule is likely to absorb the emitted photon. As the system is rich in Chlamy’s Cholorphyll, the probability of emitted photons being re-absorbed is high. As noted in the Background, the derivation of fluorescent regression models is best employed when samples for these types of measurements are very dilute. While this technique is highly sensitive and allows for the determination of very dilute concentrations as low as 10-12M **[cite fluorescence tome],** it does not work well in a context where absorbance measurements are OD > 0.1. Furthermore, when creating models to deconvolve these signals, the fluorescent measurements show another behavior, which is dependent on bacterial load. **[show the graph where Fluorescent measurements decline with incraeasing bacterial concetraiton].** Note in **FigureBacterialLoadX** that the overall signal decreases as the bacterial load increases. This dependence on the bacterial concentration, as well as the IFE and the fact that research is conducted has yielded measurements of absorbances greater than what is recommended for fluorescence regression models, makes the creation of these models unreasonable. Therefore, fluorescence measurements are disregarded altogether.

|  |
| --- |
|  |
| FigureX. F Measurements show increasing growth of Chlamy increases in the initial growth phase the signal is always show as increasing |

3.4.4 Calibration Samples Preparation for Modeling

Two types of calibration data were obtained to generate the models: a single-strain calibration and SynCom (strains only) calibration (henceforth SS and SC calibrations, respectively). Both of these calibration data are expected to behave similarly and are also expected to produce similar predictions; but were made at different times during this Thesis and are discussed later. Chlamy and bacteria samples were obtained from continuously growing stocks. For the SC calibration, several bacterial stocks were mixed in equal parts **[True?]** to compose SynCom 007 stock (the used Strains are provided in Supplementary Data), whereas for the SS calibration, only one of these strains was used. From these stocks, 100uL & 20uL aliquots were taken and diluted into TP media to compose 1:10 and 1:50 ratios of 1mL samples; these were subsequently measured using A750 to estimate the total concentration in absorbance units of the original stocks (cell count measurements are also measured). From these determined original stock concentrations, serial dilutions of 1mL are prepared to range between 0.0 to 1.0 theoretical (i.e. unmeasured) absorbance units for both Chlamy and Strain(s). Theoretical absorbance concentration units are obtained using the serial dilution equation below and the estimated absorbances. Note that these values are irrelevant and are used only to ensure serially diluted and accordance with thespecified range in order to be able to successfully apply the Beer-Lambert law.

|  |  |
| --- | --- |
|  | Equation 12 |

Next, the introduction of the serial dilutions into the microplates is executed in a specific manner with the final volume in each well being 300μL. Note that microplates are labeled using numbers (1-12, left-to-right) for columns of wells and letters for the rows of wells (A-H, top-down), **FigureX.** From each of the twelve serial dilutions of Chlamy, 150μL are pipetted into row H, beginning with the lowest dilution at H12 and going up in gradient to H1. The same process is repeated for each row resulting in every column containing the same concentration of Chlamy. Then, from each of the eight serial dilutions of SS or SC samples, 150μL are introduced at column 12, beginning with the lowest dilution at H12 and going up in gradient to A12 using a multi-channel pipette. The same process is done for every column resulting in every row containing the same bacterial load. H12 contains 300uL of TP medium only and is used as a correctional means for the absorbance of the medium for all wells. **FigureX** shows how the gradient set-up looks against a black backdrop. The calibration plates are subsequently measured using the Tecan: Row H and column 12 serve to generate the calibration models; all other rows serve to test the models before they are applied to the screening and PBR data. For the SS calibration, the Tecan is set up to measure the microplate between 300 and 800nm wavelengths using a 5nm step-size. Fluorescence measurements were taken at F680. For the SC calibration, the Tecan is set up to measure the microplate between 500 and 750nm using a 10nm step-size without taking fluorescence measurements. The SC calibration data have multiple measurements for A600, A680, A720, and A750 and are averaged as used to make the required models. Single measurements are made at all other wavelengths in both SC and SS.

|  |  |
| --- | --- |
| IMG_256 |  |
|  | |

3.5 Data Processing

All Tecan generated data are measured over several days and produce an excel file per condition in any given experiment including the calibration data. Tecan data thus result in a variable number of files for any given experiment (all data are provided in Supplementary Data) and must be parsed into their own manageable CSV files for efficient use. The resulting CSV files and those of the PBR are then pre-processed. Pre-processing of the screening experiments is done extensively in order to facilitate model application and time-series analysis; details are provided in their respective Jupyter Notebook. The PBR CSV files were easily pre-processed and required only the removal of empty values and formatting for model application. All work was accomplished using Python (v. 3.7.3) libraries on Jupyter Notebooks (v. 1.0.2). The use of Pandas (v. 0.24.2) is extensive for the majority of the work conducted. Modeling relies heavily on Numpy (v. 1.16.4), Scikit-Learn (v. 0.21.2) and SciPy (v. 1.3.0) libraries. Visualization functions rely mostly on MatplotLib (v 3.1.0), but Seaborn (v. 0.9.0) was used throughout for quick visualizations as needed. See the provided Notebooks **GitHub MAKE LINK** for the coding utilized to create the models and the visualization of the data and predictions. **[cite the relevant libraries]**

3.6 Modeling II

To create predictive models, the extinction coefficients need to be extracted from the calibration data. However, at the onset of this Thesis, it was not self-evident how best to achieve this given that measurements of the concentration were ‘theoretical’ and unit-less and not in molarity (M) as expected. Nevertheless, the imposed constraints of the existing data serve as a good starting point and scatter plots at the imposed wavelengths (A680& A750 for S4) are made against the theoretical concentrations of the SS calibration’s row H and column 12, **FigureX**. A preliminary regression on these plots reveals the general idea, as did the guidance of colleagues. Ultimately, any wavelengths measured can serve as both regressor and regressand, and therefore all measured wavelengths can serve as a proxy for concentration to regress back to.

|  |  |
| --- | --- |
| A | B |
| FigureX. The four fitted scatter plots represent microplate row H and column 12. At this row and column, either Chlamy or bacterial strain or SynCom is present. The regressions on [bacter] are very close together and hard to discern. | |

3.6.1 Obtaining the Extinction Coefficients

Once the general idea was clear, it was a matter of utilizing Linear Regression to obtain the coefficients needed for the models. This statistical analysis technique is used to find the best fitting line through a set of points and takes the general form of Equation 13 **[cite statistical learning tome]**. To apply the Beer-Lambert law, however, only Simple Linear Regressions (SLR) need to be solved as each regressand can only depend on one regressor at a time. The regressors are the absorbance values at any measured wavelength, moonlighting as the concentrations of either Chlamy or Strain(s) in the x-coordinate.

|  |  |
| --- | --- |
|  | Equation 13 |
|  | Equation 14 |

Note that Equation 14 has the familiar form of the Beer-Lambert law (Equation 8) where represents , corresponds to , and represents the error in each prediction. To obtain these coefficients, Scikit-Learn’s Linear Regression class was used **[cite scikit-learn]**. Scikit-Learn’s implementation uses Ordinary Least Squares (OLS) to find the slope—and hence extinction coefficient—of the line that best fits the data. The OLS algorithm solves Equation 15 to estimate the slope .

|  |  |
| --- | --- |
|  | Equation 15 |
|  | Equation 16 |

For the calibration data sets SS and SC, the wavelengths mentioned in section 3.4.2 are modeled against all other measured wavelengths, and the obtained extinction coefficients for each of these regressions are provided in the Supplementary Data.

3.6.2 General Form of Predictive Models

The extended form of the Beer-Lambert law is used to create the predictive models, which will be applied to the collected absorbance data. To create these models extinction coefficients are first derived as described above and are put into a two-equation system, which is then solved for the respective variables representing bacteria and Chlamy concentrations. The derivation of the models is illustrated in Equation 10 for wavelengths A680, A750 which are used to analyze S4 data. Equation 10 is first rewritten for these specific wavelengths and new variable symbols are chosen to represent bacteria ) and Chlamy ) concentrations as measured at an arbitrary absorbance wavelength, .

|  |  |
| --- | --- |
|  | Equation 10 |
|  | System 1 |
|  |  |
|  |  |
|  | |
|  | |
|  | |
|  | |

|  |  |
| --- | --- |
|  |  |

The solutions to System 1 state that given absorbance measurements at A680 and A750 of a multiple component system (Chlamy + Strain(s)), predictions, represented with the hat (^) symbol, can be obtained for each component in the system. Furthermore, it follows from the Beer-Lambert law that since each component’s prediction is actually an absorbance prediction, they must also obey the law and be equal to the empirical measurements of the microplate wells at the respective absorbance wavelength . Therefore, in order to evaluate the predictive power of the models, the sum of the predictions can be compared against the empirical measurements of wells A-G:1-11 of the calibration plates. Furthermore, since we know how much of each species was introduced into the wells A-G:1-11, each prediction or at each of these wells must also equal the known concentrations. Model application was not trivial as in the application in compound or macromolecule predictions and required thorough analysis for its application in this system; thus, the models are further discussed in Results.

3.6.3 Model Application to PBR Data

In order for the Tecan-generated models to be applied to PBR data, they must be converted to fit its path length. In practice, this is done at the moment of model application after the extinction coefficient has been determined and it is only a matter of scaling these coefficients accordingly. Recall that the general form of the Beer-Lambert law is as Equation 7 states. For this equation to hold true, the path-length of the vessel must be known and the extinction coefficient must have units which cancel out with respect to the concentration and path-length units. The obtained extinction coefficients constants are unit-less since they are produced by regressions that use unitless absorbance values as concentration. However, the standardization of the values by the Tecan means that the magnitudes of the and are equal and we denote them as having cm-1 units. Thus, model application to the PBR measurements only require that we scale these extinction coefficients by the path-length which is 2.7cm **[cite manual]**. This will ensure that the returned predictions are also in unit-less absorbance values.

3.7 Absorbances to Cell-Counts

Having models that predict absorption values as concentration does not make for easy analyses. This is due to the difficulty of forming an intuition for what these values mean with respect to completely different microorganisms. The majority of soil bacterium measures under 1.2μm while Chlamy cells conservatively measure 10μm **[size distribution, chlamy Tome]**. If the cells are modeled as spheres, this would mean that Chlamy has 500 times the volume of a typical bacterial cell and will absorb much more light per unit cell than the Strains. To abstract away this layer, Coulter cell estimates are regressed against the measured spectrum of the SC calibration data (400 - 750 nm) using SLR as described above. FigureX shows this regression for A560. All other regression coefficients for these are provided in Supplementary Data. For quickly predicted absorbances-to-cell counts and cell-cell ratio estimates, Table 2 has been devised. The table is to be read as follows: predicted absorbances are in black, values in green and red correspond to cell-count estimates, Chlamy and Strain respectively, from the regression in FigureX. The blue 250 diagonal represents 1:1 absorbance ratios with roughly 250 times as many bacterial cells as Chlamy cells (B:C). Green and red-boxed values are fractions or multiples of this value. Note, however, that the regressions show a large gap in the ranges of 0 to 1; this is because absorbance measurements at these wavelengths are not very sound.

|  |
| --- |
|  |
| FigureX. Coulter Cell Count Estimates on A560. |

|  |  |
| --- | --- |
|  | d |