

OPEN

# Multiscale Multiobjective Systems Analysis (MiMoSA): an advanced metabolic modeling framework for complex systems

Joseph J. Gardner<sup>1</sup>, Bri-Mathias S. Hodge<sup>1,2,3</sup> & Nanette R. Boyle<sup>1\*</sup>

In natural environments, cells live in complex communities and experience a high degree of heterogeneity internally and in the environment. Even in 'ideal' laboratory environments, cells can experience a high degree of heterogeneity in their environments. Unfortunately, most of the metabolic modeling approaches that are currently used assume ideal conditions and that each cell is identical, limiting their application to pure cultures in well-mixed vessels. Here we describe our development of Multiscale Multiobjective Systems Analysis (MiMoSA), a metabolic modeling approach that can track individual cells in both space and time, track the diffusion of nutrients and light and the interaction of cells with each other and the environment. As a proof-of concept study, we used MiMoSA to model the growth of *Trichodesmium erythraeum*, a filamentous diazotrophic cyanobacterium which has cells with two distinct metabolic modes. The use of MiMoSA significantly improves our ability to predictively model metabolic changes and phenotype in more complex cell cultures.

Biological heterogeneity is a challenge even in the most ideal growth conditions in a laboratory. Most of the commonly used methods to describe or model metabolism assume that the average cell in the population is an adequate representation of the culture if the medium is well-mixed. While these assumptions might be valid for fast growing heterotrophic bacteria and yeast, they are not as easily justified for more complex organisms or growth conditions such as filamentous bacteria or biofilms. In these cases, environments are highly variable from cell to cell; there can be wide ranges of nutrient or light availability, temperature, pH, and other important growth parameters<sup>1–6</sup>. These small differences in growth conditions can lead to large differences in phenotypes even amongst cells grown in the same culture. This presents a challenge both in experimental design as well as computational modeling of biological processes and organisms to ensure that what we are measuring or predicting is representative of the entire population.

One computational tool that has been used widely in investigating the metabolism of microorganisms are stoichiometric metabolic models<sup>7</sup>. The most widely used stoichiometric metabolic models are constraint-based linear programming models which vary in complexity from the relatively simple flux balance analysis (FBA) to more complex FBA models which integrate regulatory and/or thermodynamic constraints<sup>8–12</sup> or time-dependent responses<sup>13,14</sup>. In cases where growth occurs over a longer period, dynamic FBA (dFBA) can be used to visualize changing constraints and media concentrations<sup>13</sup>. Unfortunately, however, the way these models are formulated limits their use to modeling a single cell or an average cell in a population. These models have proved to be incredibly useful for strain design of heterotrophic bacteria in well mixed batch reactors<sup>11,13,15</sup>. Stoichiometric metabolic models have also been extended to model more heterogeneous populations, such as binary or tertiary bacterial consortia<sup>16–20</sup> by adding additional compartments for each species and allowing instantaneous metabolite exchange. This is not representative of what is actually occurring in these environments though as diffusion of metabolites is an incredibly important limitation to cell growth and interactions. The current benchmarks for modeling more heterogeneous populations are OptCom<sup>21</sup> and d-OptCom<sup>22</sup> (the dynamic version). This modeling approach uses multiple objective optimization technique which allows the model to capture any type of interaction (synergistic, antagonistic and neutral) for any number of cells or distinct organisms. While each of

<sup>1</sup>Chemical & Biological Engineering, Colorado School of Mines, 1613 Illinois St., Golden, CO, 80403, USA. <sup>2</sup>National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO, 80401, USA. <sup>3</sup>Electrical, Computer and Energy Engineering, 425 UCB, University of Colorado, Boulder, CO, 80309, USA. \*email: [nboyle@mnes.edu](mailto:nboyle@mnes.edu)

the metabolic modeling approaches above do move toward capturing more heterogeneity in the population at the cell level, they leave out important phenomena occurring outside the cell, such as diffusion of metabolites and nutrients or movement of cells.

We have developed **M**ultiscale **M**ultiobjective **S**ystems **A**nalysis (MiMoSA), an advanced metabolic modeling framework in order to more accurately model the metabolism and cellular interactions of complex systems. This approach uses a multi-scale multi-paradigm metabolic modeling approach to leverage the ease of implementation of stoichiometric metabolic models while integrating the spatiotemporal tracking of cells, nutrient diffusion, cell-cell interactions and cell-environmental interactions. This approach requires the use of both continuous and discrete variables as well as several different mathematical formalisms to reflect the multilevel behavior in populations. Therefore, we use an agent-based modeling (ABM) framework to allow direct interaction of different levels through the encapsulation of physiological, environmental, and metabolic models. ABM is a bottom-up modeling approach; the model is made up of a set of agents, which are allowed to act independently as long as they follow distinct rules of behavior defined by the user, this allows us to simulate emergent behavior of complex communities that arise from individual agent behaviors<sup>23–26</sup>. The system behavior emerges as a result of the many (tens, hundreds, thousands, millions) individuals, each following their own behavior rules, living in a defined environment, interacting with each other and the environment<sup>24</sup>. The integration of multiple modeling formalisms to represent disparate sub-systems is a trend common in engineering and science domains<sup>27–31</sup> and has recently seen some developments in the systems biology area<sup>10,12,14,32–35</sup>. Agent-based modeling has been previously applied to both intercellular<sup>8,36</sup> and multi-cellular processes<sup>37,38</sup> but has not previously been used to model metabolic fluxes. This multi-scale multi-paradigm approach represents a novel method of integrating individuals (through agents) with previously leveraged dFBA formulations<sup>13,39</sup>, thereby discretizing and separating variables for computational efficient solutions with low *a priori* knowledge.

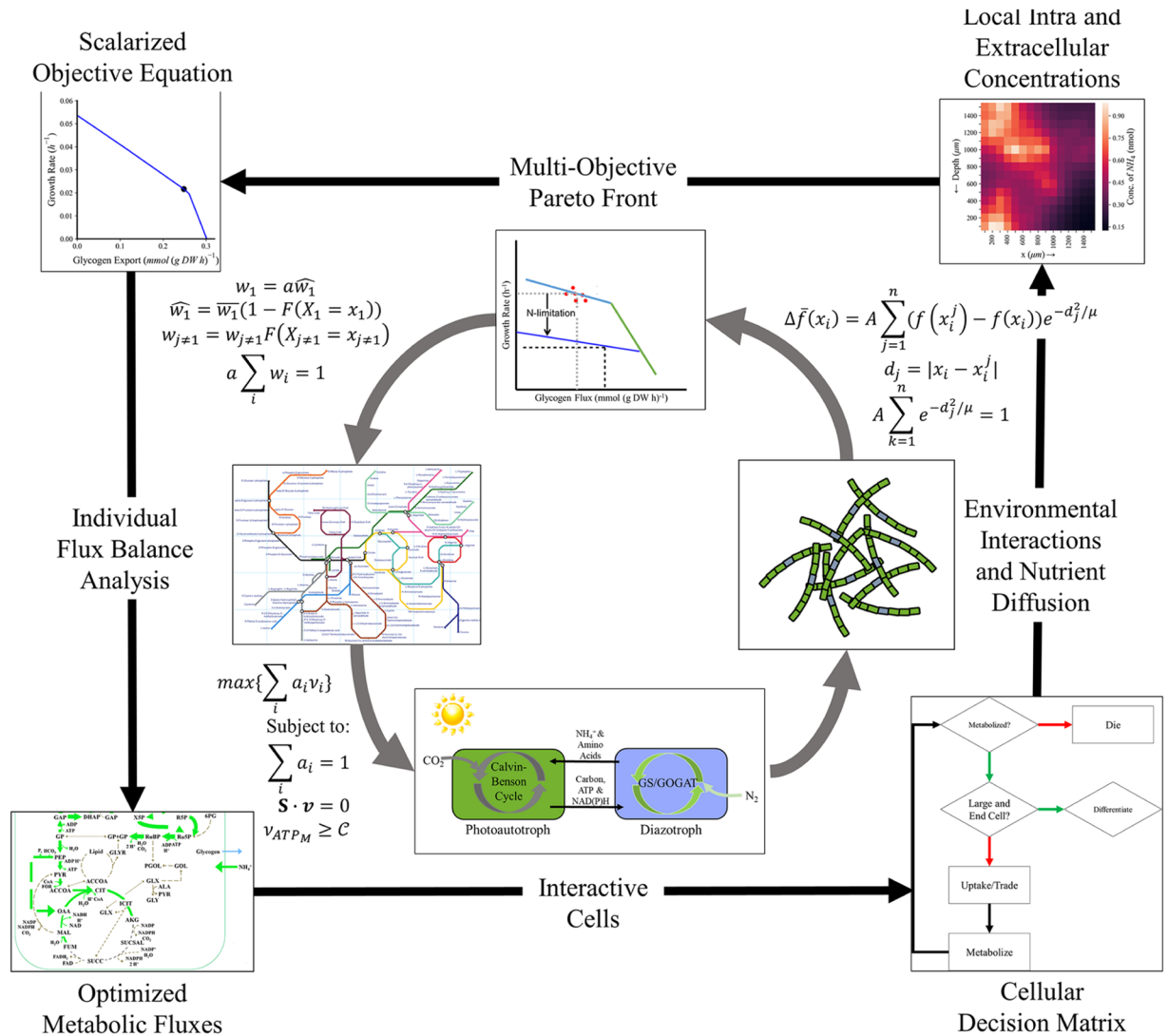
As a proof-of-concept study, we chose to model *Trichodesmium erythraeum*, a filamentous diazotrophic cyanobacterium. *T. erythraeum* is a major contributor to the global nitrogen cycle; it is responsible for fixing an estimated 42% of all marine biological nitrogen<sup>40</sup> and it leaks 20–50% of the nitrogen it fixes<sup>41</sup>, providing surrounding organisms with a biologically available nitrogen source. Unlike other diazotrophs, which either spatially or temporally separate the oxygen sensitive nitrogenase enzyme from the water splitting reaction of photosynthesis (oxygen production), *T. erythraeum* is unique because it simultaneously carries out nitrogen and carbon fixation during the day in different cells along the same filament (trichome) with metabolic as opposed to physiological control. We also have previously studied major metabolic differences between the two cell types<sup>42</sup>. Therefore, it is the ideal model system for the development of MiMoSA: it has structurally identical cells that are prone to two subsets of metabolic constraints yielding two major metabolic subsets (photoautotrophic and diazotrophic), a published genome scale model<sup>42</sup>, transcriptome data, and a plethora of *in situ* and laboratory data to both train the model and validate predictions. We use this organism to highlight the advanced capabilities of the MiMoSA framework to predict emergent behaviors of the cell and to investigate rules of cellular physiology.

## Results

**Model formulation.** We developed MiMoSA by integrating an updated version of the genome-scale metabolic model<sup>42</sup> (Table S1 for updated reactions) with nutrient diffusion, light diffusion, cell/cell interaction and cell/environment interactions (see Fig. 1) using an agent based modeling framework. We have also implemented the use of multiobjective optimization to account for the dual cellular objective of producing biomass and the metabolite which is transacted between cells (glycogen or  $\beta$ -aspartyl arginine, depending on cell type) with the capability of a full range of exchangeable metabolites that are not part of the objective function. Constraints were imposed on the model as reported previously<sup>42</sup> with two notable exceptions. First, the ultimate product of nitrogen fixation was changed from ammonium to  $\beta$ -aspartyl arginine, which is the monomer used to create cyanophycin, a nitrogen storage polymer in *T. erythraeum* and other diazotrophic cyanobacteria<sup>43–45</sup>. Second, the two major storage polymers, glycogen (modeled as maltose, or two linked glucoses) and cyanophycin (modeled as  $\beta$ -aspartyl arginine), were decoupled from the biomass formation equation so that they could freely accumulate or be metabolized. More detail about the formulation of the model is provided in Methods and Supplemental Text.

**Tracking changing cellular objectives.** MiMoSA evaluates the cellular objective for each cell for each time step based on the changing environmental conditions. As an example of this, we have tracked how the Pareto front changes for both photoautotrophic and diazotrophic cells over time (Fig. 2). The front is selected according to *Methods: Shifting Scalar Objectives* with each point corresponding to a specific scalar arrangement of variables. With increasing time, diazotrophs shift their objective away from biomass toward the production of cyanophycin as carbon becomes more available (Fig. 2A). In contrast, photoautotrophic cells see a maximum production of glycogen at 9 hours after the onset of light and then their productivity decreases (Fig. 2B). It is notable that every cell in the population is performing these decisions in parallel and Fig. 2 is for a single representative cell of each cell type. Cell optimization changed based on environmental conditions and agent rules and the Pareto Fronts representing this behavior in these contexts is visualized in Fig. S1.

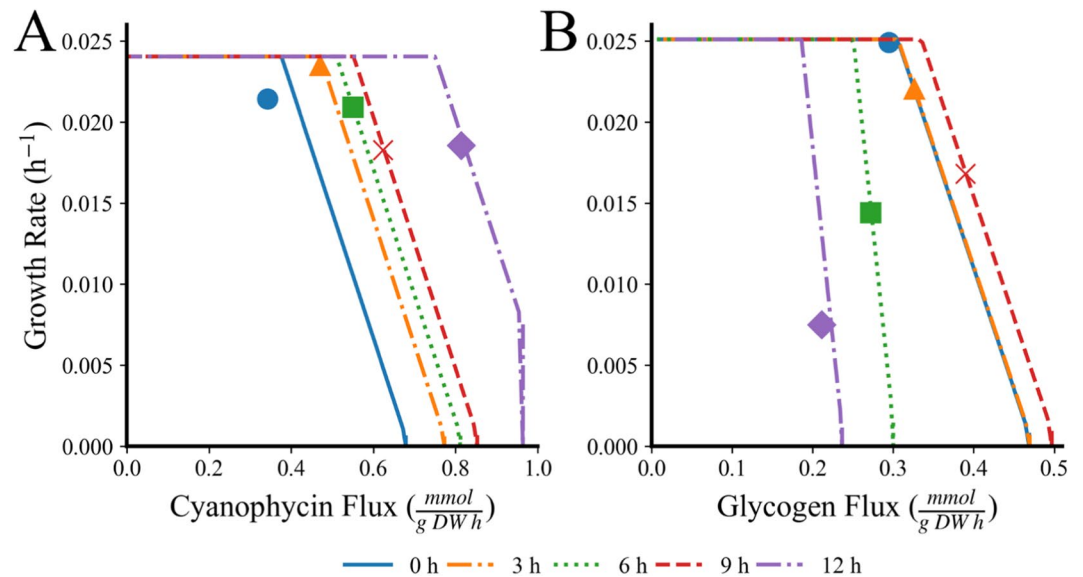
**Model validation.** In order to test the predictive accuracy of the model, we predicted growth rate for a variety of different light intensities (Fig. 3A) and compared to other published models for *T. erythraeum*<sup>46,47</sup> as well as other experimentally measured growth rates<sup>1,2,4,42,46,48</sup> exhibiting light saturation at higher light intensities. Ultimately, our model is a metabolic model, so it is important that it can also capture the metabolic changes that occur in response to changes in the environment. Therefore, we compared predictions of metabolite accumulation, representing part of the major changes to data collected in our laboratory for growth in different light



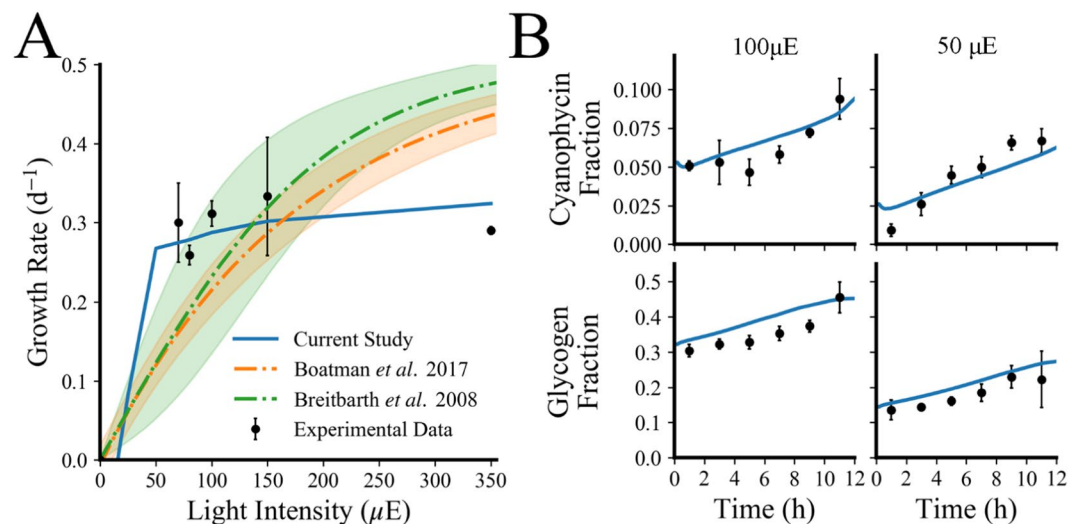
**Figure 1.** Multi-Scale Multi-Paradigm Model Generation. Before this process, the model generates an average scalar equation by fitting the organism’s Pareto Front to experimental data using the ATP hydrolysis maintenance reaction as further elucidated in Methods. Then, starting from the top and progressing with the arrows (clockwise): The multi-objective Pareto Front is corrected for environmental variables and cellular preferences using a weighting algorithm and assuming a normally distributed cell biomass (more detail in Methods). The corrected biomass equation is solved, individually, for each cell subject to existing constraints, a steady state over each time step, an appropriate maintenance ATP flux, and a scalar objective function for which all coefficients add to one. This is interpreted using the agent-based model to make individual cell and physiological decisions including (1) whether the cell should die, (2) whether the cell should reproduce (and if it does, what type of cell does it differentiate into), and (3) how it should interact with the environment and other cells. These interactions inform the status of the other cells (using an intrafilamental diffusion mechanism) and the environment (modeled with the same diffusion mechanism for CO<sub>2</sub>, N<sub>2</sub>, organic, and fixed nitrogen products, and assuming excesses of other media components). The iteration restarts with the objective equation updating each living cell (whether newly reproduced or previously established) based on the cell’s current metabolic state.

intensities (see Fig. 3B). The model was trained on data collected in 100 μE light and was validated with data collect in 50 μE light over a twelve-hour light period.

**Cells alter their microenvironment.** An advantage of the modeling approach we have developed is that we can track nutrients in the environment. Carbon dioxide (CO<sub>2</sub>) is typically the limiting substrate in aquatic photosynthetic growth due to low ambient concentrations and low solubility; for ambient CO<sub>2</sub>, Henry’s law defines an equilibrium concentration of 2.3 μM in the ocean. It is well known that photosynthetic microorganisms use carbon concentrating mechanisms (CCM) to concentrate CO<sub>2</sub> near the carbon fixing enzyme, ribulose-1,5- biphosphate carboxylase/oxygenase (RuBisCO) to overcome low selectivity<sup>49</sup>; our simulations imply that cells also increase the local concentration of CO<sub>2</sub> immediately surrounding the cell (Fig. S2A) and the release of nitrogen



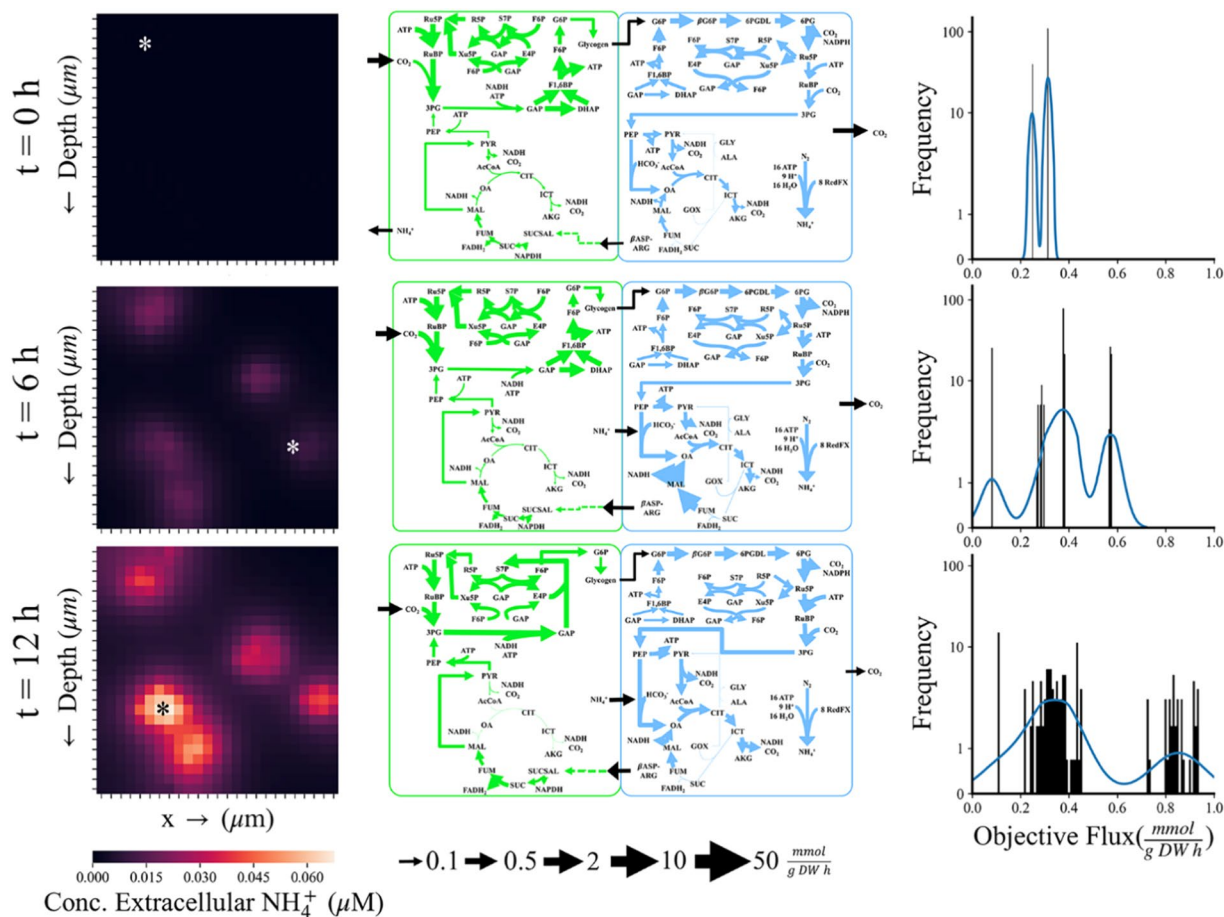
**Figure 2.** Pareto Front Progression and Selected Scalar Equation Points over the Light Period in Different Conditions for (A) a diazotrophic cell and (B) a photoautotrophic cell. Each line corresponds to a Pareto Front generated every 3 hours over the course of the light period at the start (blue, solid) through the end (purple, solid-dot) with the point selected by the simulation to best suit cell needs in the scalar objective function indicated as an indication of the permutation of objective weights found in that step and cell's objective equation (see Methods: Multiobjective Optimization). Simulations were run with atmospheric conditions, 100  $\mu\text{E}$  light, and YBC-II media with 150 cells over 10 filaments with a ratio of 4:11 diazotrophs to photoautotrophs.



**Figure 3.** MiMoSA model validation (A) Growth rate as a function of light intensity in *T. erythraeum*; here we compare the predictions of growth rate from our model to two published models<sup>46,47</sup> and experimental data reported from a variety of literature values<sup>1,2,32,42,48,82–84</sup>. Error bars represent the error propagated when finding the mean of all separately recorded growth rates in similar conditions (YBC-II media, atmospheric  $\text{CO}_2$ , no added nitrogen, and equivalent light intensity) using standard Euclidean error propagation. (B) Experimentally measured changes (black circles) in biomass accumulation for cells grown in 100  $\mu\text{E}$  and 50  $\mu\text{E}$  compared to simulated values (blue lines). Error bars on the glycogen and cyanophycin measurements represent one standard deviation for three biological replicates.

to the media including at more frequent time steps (Fig. S2B). The simulation covers 150 cells and 10 filaments in a model 0.625  $\text{mm}^3$  environment, corresponding to a filament density of  $16 \times 10^6$  trichomes  $\text{m}^{-3}$ , well within the *in situ* ranges of free trichome density<sup>50</sup>. This illustrates that the simulation corresponds well quantitatively to realistic local environments. At the end of our simulation, the cells on average can create a microenvironment that





**Figure 4.** Examples of the types of data the multi-scale multi-paradigm metabolic model can track in time and space. Heat map legend and flux map legends are at the bottom of the figure.  $x \rightarrow$  indicates horizontal space in 2-D simulations. Time progresses from top to bottom with the top row indicating the beginning of the 12 hour light cycle cycle, the middle row the midpoint at 6 hours, and the bottom row the end of the light cycle. The left column contains the distribution of ammonium in the local media over these time points. The asterisk indicates the cells selected for the flux maps. The simulation can also be depicted as a 2-D grid with an arbitrary third dimension of one grid-cell deep. The whole space is a square grid with dimensions of  $2500 \mu\text{m}$  ( $2.5 \text{ mm}$ ) and each grid-cell with dimensions of  $100 \mu\text{m}$  ( $10 \text{ T. erythraeum}$  cells). The middle column visualizes flux maps of photoautotrophs (left, green) and diazotrophs (right, blue) from the cells selected from the locations indicated by the asterisk. Flux maps were identical at 0 hours, selected from cells with the lowest extracellular ammonium in their gridspace at 6 hours, and selected from the highest extracellular ammonium in their gridspace at 12 hours. The right column is the frequency of unique metabolic profiles ranked by objective flux (a measure of cell productivity accounting for both biomass and metabolite production). Blue line is the kernel density estimate (kde) which estimates the probability of a given metabolic profile. ABBREVIATIONS. 6PG: 6-phospho-D-gluconate, 6PGDL: 6-phospho-D-glucono-1,5-lactone, AcCoA: Acetyl-CoA, AKG/ $\alpha$ KG:  $\alpha$ -ketoglutarate/2-oxoglutarate,  $\beta$ ASP-ARG,  $\beta$ -aspartyl arginine, CIT: Citrate, DHAP: Dihydroxyacetone phosphate, DON: Dissolved Organic Nitrogen, E4P: Erythrose-4-phosphate, F6P: Fructose-6-phosphate, FOR: Formate, FUM: Fumarate, G6P: Glucose-6-phosphate, GAP: Glyceraldehyde 3-phosphate, GLX: Glyoxylate, 3PG: 3-phosphoglycerate, ICIT: Isocitrate, MAL: Malate, OA: Oxaloacetate, PEP: Phosphoenolpyruvate, PYR: Pyruvate, R5P: Ribose-5-phosphate, Ru5P: Ribulose-5-phosphate, RuBP: Ribulose 1,5-bisphosphate, S17P: Sedoheptulose 1,7-bisphosphate, S7P: Sedoheptulose 7-phosphate, SUCC: Succinate, SUCSAL: Succinic semialdehyde, X5P: Xylulose 5-phosphate.

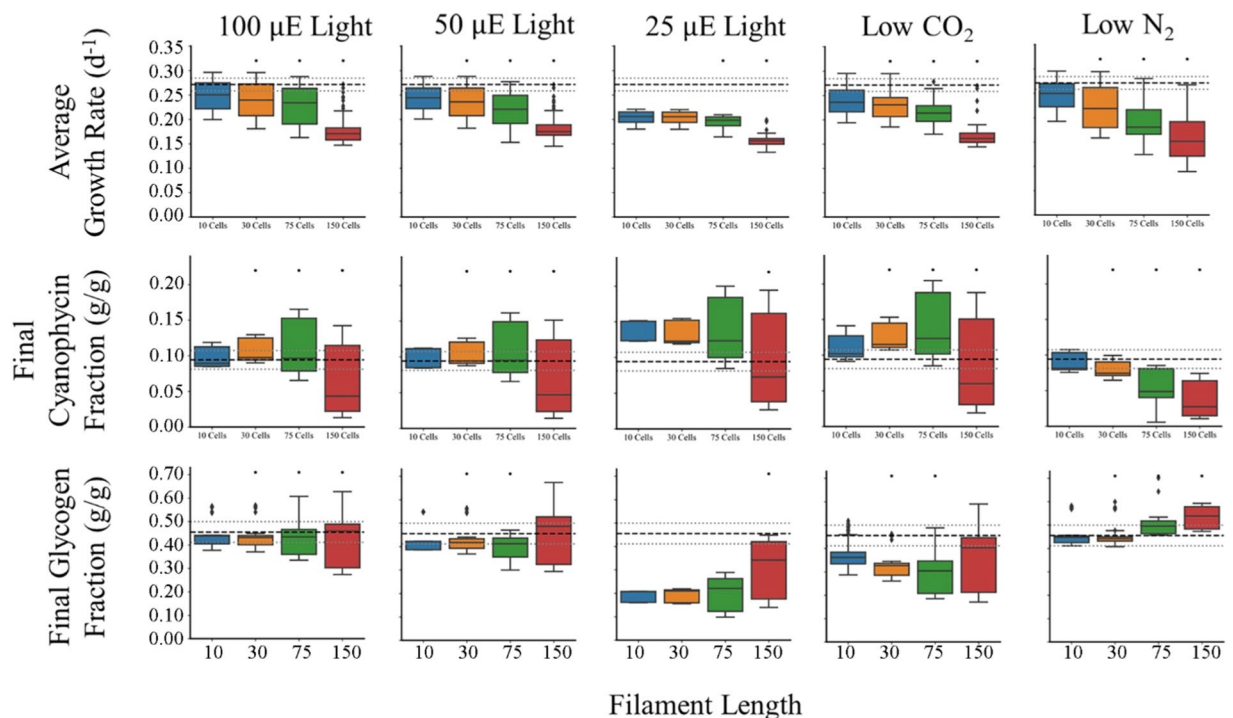
is roughly 2 fold higher in  $\text{CO}_2$  than the surrounding ocean. By looking at flux through major pathways, it appears that the  $\text{CO}_2$  is derived from high fluxes through the oxidative PPP and TCA Cycle in diazotrophic cells (Fig. 4).

**Modeling a heterogeneous cell population.** One of the main advantages of this new modeling approach is that individual cells can be tracked in space and time so the heterogeneity of the population can be quantified (in terms of metabolic flux distributions). As an example, we tracked 150 cells over a 12-hour time period with time steps of 6 seconds which results in a total of 18,000 metabolic flux maps. Since these are an overwhelming amount of data to visualize, we have chosen to focus on a few representative flux maps (see Fig. 4). In the left column, we track how the ammonium composition of the environment surrounding the cells changes with time

from the initial seeding of cells at 0 hours to the middle of the daytime period (6 hours) to right before the onset of night (12 hours). These panels depict the release of ammonium into the environment as time progresses, and it is higher in areas where the cell density is highest. This agrees well with *in situ* data which report that *T. erythraeum* leaks 30–50% of the nitrogen it fixes<sup>41</sup>; our simulations predict that approximately 20% of the nitrogen fixed by the community is excreted into the medium. It is also important to note that the majority of ammonium is released by the cells in the second half of the day; during the first 6 hours, the cells release ammonium to a total of 2.56  $\mu\text{M}$  midday compared to 9.22  $\mu\text{M}$  concentration of ammonium at the end of the day. Again, this agrees with previous literature reports that the rate of nitrogen fixation peaks at midday<sup>51</sup>, therefore we would expect more secretion of ammonium after peak nitrogenase activity. Select flux maps of cells growing in areas of low ammonium (top), medium ammonium (middle) and high ammonium (bottom) are depicted in the middle column of Fig. 4. At the beginning of the simulations, cells are seeded in an environment that is identical to the defined marine medium YBC-II and because of this, they have identical flux maps as shown by the distribution graph in the right column. At time 0, we have a bimodal distribution because there are two cell types: photoautotrophic and diazotrophic. Photoautotrophic cells have high flux through the Calvin Cycle and the diazotrophic cells are operating in a more respiratory mode, with high flux through both the oxidative pentose phosphate pathway (PPP) and tricarboxylic acid (TCA) Cycle. As the cells grow and start to experience more heterogeneity in their environment, they respond by differentiating their metabolism within the filament (Fig. S3). First, this is evident in the frequency distribution plot, where they are both diverging in terms of total metabolic flux distributions and moving toward achieving optimal flux in terms of the objective function for both  $t = 6$  hours and  $t = 12$  hours. By comparing the changes that occurs in metabolic flux between areas of low, medium, and high ammonium, we can learn a few things about cellular physiology. In all cases, photoautotrophic cells have high flux through the Calvin cycle and an incomplete TCA Cycle, which has been widely reported in cyanobacteria grown phototrophically<sup>52</sup>. In the case of *T. erythraeum*, succinic semialdehyde is derived from the nitrogen storage compound cyanophycin and is fed into the TCA Cycle to support the production of biomass precursors and glycogen (through gluconeogenesis). When external ammonium is high, photoautotrophic cells have less flux to glycogen, presumably because they do not need to provide as much to the diazotrophic cells to obtain fixed nitrogen in return. Investigations into imbalances in both metabolites and relative cell quantity display mechanisms of ammonium loss to the environment. Figure S4A illustrates how a lack of glycogen flux results in a higher loss of ammonium (with the exceptions of recently divided cells which metabolize glycogen with high ammonium loss) while Fig. S4B visualizes a clear minimum ammonium release in the recorded range of percent diazotrophs per filament (between 15 and 30%). Diazotrophic cells have high flux through both the oxidative PPP and the TCA cycle which still utilizing carbon fixation reactions such as Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and phosphoenolpyruvate (PEP) carboxylase and carbon conserving reactions like the glyoxylate shunt. Flux through the glyoxylate shunt increases as the availability of ammonium increases outside the cell, which is likely in response to the lower glycogen transfer from the photoautotrophs.

**Elucidating rules of cell physiology.** A key feature of agent-based modeling is the ability to model emergent behaviors in populations. We do not know all the rules of behaviors that define *T. erythraeum a priori* but by comparing simulations to observed *in situ* data and iterative improvement of the model, some rules can be elucidated. One trait that is widely variable in nature is filament length. It has been widely accepted that the average filament length is 100 cells<sup>53</sup> but more recent studies have suggested that they are typically much shorter, with a geometric mean of  $13.2 \pm 2.3$  cells per filament, but with a mean range of 1.2 to 685 cells per filament *in situ*<sup>54</sup>. Conditions for *in situ* sampling are widely variable so we hypothesized that filament length plays a role in maintaining growth in different environments: low light, low  $\text{CO}_2$  and low  $\text{N}_2$ . We used the model to investigate which conditions might favor shorter or longer filaments (Fig. 5). For each simulation, 150 total cells were seeded but in different trichome lengths (10, 30, 75, and 150 cells/filament) with and a ratio of diazotrophs to photoautotrophs of 3:7. In terms of growth rate, across all conditions we tested the shorter filaments had faster growth. This implies that diffusional limitations of nutrients into the cell and metabolites within the filament between different cell types start to hamper growth rate at longer filament lengths. The relative decline in growth rate is less dramatic for 25  $\mu\text{E}$  when comparing across filament length, but when compared to other light conditions, there is a dramatic drop in growth rate for shorter filaments at low light. This indicates that longer filaments are capable of compensating for less light better than shorter filaments, perhaps due to increased surface area. Next, we examined the effect of filament length on cyanophycin composition for the same growth conditions as above. In every condition except low nitrogen, filaments with 75 cells appear to have more cells with above average cyanophycin content than other filament lengths. Smaller nitrogen compounds ( $\text{NH}_4^+$ , amino acids, urea, etc.) can theoretically be used to support growth, permitting cyanophycin to be a longer-term storage compound. This is a possible explanation for the increase of cyanophycin in longer filaments. As filaments are longer, diffusive limitations become more pronounced, meaning that nitrogen gradients will remain in nitrogen replete cells longer and will be remade into cyanophycin as opposed to being metabolized for growth. This makes intuitive sense: not only is there a final drop-off at 150 cells, the distribution of cyanophycin content within the cells becomes larger, suggesting that some cells are starved for nitrogen and some are nitrogen replete. It is probable that filaments have adapted to leverage diffusion to both sequester nitrogen and to mitigate futile cycling of carbon and nitrogen compounds when diatomic nitrogen is available. The pattern of cyanophycin content diverges for cells in nitrogen limited environments due to overall shortages of nitrogen within the filament.

Finally, we investigated how glycogen content of cells changes due to filament length. The first pattern to note is that as length increases, the heterogeneity of the filament in terms of glycogen content also increases. This illustrates the importance of tracking individual cells because they are experiencing different environments and responding in different ways. Longer filaments also appear to be able to maintain glycogen content more readily than shorter filaments in all stress conditions we tested. Finally, nitrogen limited growth results in increased



**Figure 5.** Trichome length affects the performance of the community. Trichomes were varied in initial length from short (10 cells filament<sup>-1</sup>) to long (150 cells filament<sup>-1</sup>) but with identical diazotroph: photoautotroph ratios of 3:7 (excepting 75 cells filament<sup>-1</sup> which was 3.1:7) and initial cell counts (150 total) in the total population. Black asterisks are Student's T-tests P-Values < 0.05 when comparing that cellular population to the 100  $\mu\text{E}$  case in every group. Dashed black lines are experimentally measured values and the gray dashed lines are their standard errors measured in the Boyle Laboratory ( $0.271 \pm 0.0129 \text{ d}^{-1}$  for growth rate,  $0.0939 \pm 0.0131 \text{ g cyanophycin/g DW}$ ,  $0.455 \pm 0.0440 \text{ g glycogen/g DW}$ ). Growth conditions represent the columns progressing as follows: 100  $\mu\text{E}$  light in YBC-II media and atmospheric conditions, 50  $\mu\text{E}$  light in YBC-II media and atmospheric conditions, 25  $\mu\text{E}$  light in YBC-II media and atmospheric conditions, 1.25 mM (versus 2.5 mM)  $\text{HCO}_3^-$  and 200 ppm (versus 400 ppm)  $\text{CO}_2$  in otherwise atmospheric conditions and YBC-II media with 100  $\mu\text{E}$  light, and 0.395 atm (versus 0.79 atm)  $\text{N}_2$  in otherwise atmospheric conditions and YBC-II media with 100  $\mu\text{E}$ .

glycogen content as seen in other cyanobacteria<sup>55</sup>. It appears that longer filaments in N limited growth can accumulate more carbon, perhaps again due to higher surface area and hence more energy from light harvesting. Our simulations agree well with published studies; it has been reported that growth rate and light intensity are both inversely correlated to filament length<sup>56</sup>. These data indicate that filament length is largely determined by external cues rather than genetically.

## Discussion

MiMoSA enables the most detailed and accurate metabolic modeling of complex systems to date by allowing coupling of several different mathematical formalisms describing natural phenomena, behavioral rules, and metabolism into a multi-scale multi-paradigm model. In constructing MiMoSA, we have added several features to enable us to more accurately predict phenotypes. A key feature of MiMoSA is the use of a multi-objective optimization approach. Unlike fast growing bacteria, which have successfully been modeled using a single objective function of maximum biomass<sup>15</sup>, slow growing organisms have more complex objectives. In our simulations, *T. erythraeum* cells must achieve a delicate balance between biomass formation and the production of either glycogen or cyanophycin due to the symbiotic relationship between two cell types in the same filament. Photoautotrophs cannot function optimally without a biologically available form of nitrogen from the diazotrophs and the diazotrophs cannot support their metabolism without reduced carbon from the photoautotrophic cells. The use of multi-objective optimization allows us to describe this trade-off more accurately and by calculating the Pareto Front *a priori* we can also reduce computational effort. We have also accounted for changes in biomass composition that occur in response to changes in the environment or as a result of building carbon and nitrogen reserves during the day by decoupling the biomass equation. This allows the model to respond more fluidly to changes in the environment, which more closely mimics what cells experience in nature; for example, macro- and micro-nutrient stresses have been well known to cause changes in metabolism such as lipid and carbon accumulation<sup>57-62</sup>. As such, the inclusion of metabolite and nutrient diffusion to augment metabolic optimization is a critical aspect of the model.

The influences of nutrient and energy availability in conjunction with population characteristics were studied to determine community and cellular adaptations to environmental perturbations. The model allows us to

quantify the changes in the microenvironment around the cell compared to the bulk properties of the environment (Fig. S2A) as well as to see how these changes affect the distribution of carbon and nitrogen inside the cell (Fig. 4). These can be supplemented with “zooming in” on specific time steps to enhance investigation to rapidly occurring phenomena (Fig. S2B). Not only did our predicted growth rates quantitatively match the experimental data, it was better able to capture effect of light saturation on growth rate; light intensities above 100  $\mu\text{E}$  have little to no effect on growth rate<sup>47,56,63–65</sup>. Our simulations agree well with the experimental data, however, there are differences that can be explained by the differences between our experimental conditions and our simulations. The main difference being the effect of diurnal light; *T. erythraeum* will not grow without diurnal day/night patterns, therefore the experimental data were collected from cells that were grown in 12 h:12 h day/night cycles but the model is for a single 12-hour day time period. The addition of diurnal light patterns in future iterations of this model will help to improve the light dependent growth phenotype. Even so, the model is able to visualize community coordination and development during the 12 hour light period, exhibiting the increased release of ammonium to the media in the afternoon, consistent with the observation that nitrogenase activity peaks mid-day<sup>51</sup>. Moreover, the individualized resolution of metabolic optimization can probe the nuances of intercellular, intracellular, and cell-environment interactions. Analysis of metabolic flux reveals a spontaneous partial/linear TCA Cycle in photoautotrophic cells consistent with previous reports<sup>52</sup>. Cells also naturally coordinate to provide glycogen and cyanophycin transfer between cells, yielding oxidative behavior in diazotrophic cells through glycolysis with the possible side effect of oxygen consumption as a mechanism to protect nitrogenase as suggested in experimentation<sup>64</sup>. Meanwhile, photoautotrophs naturally perform reductive carbon fixation coupled with utilization of the lower TCA Cycle to degrade arginine. These metabolic functions are affected by extracellular forces which are integrated into this model. For example, high ammonium environments result in declining gluconeogenesis in photoautotrophs (12 hours in Fig. 4), likely since these cells are energetically limited and use cyanophycin as an energy source instead of light. Diazotrophs are prone to these environmental cues as well as low ammonium environments enhance light TCA cycle to enhance recycling of amino acid byproducts from a lack of nitrogen. By integrating modeling of other phenomenon with constraints based metabolic models, we were able to simulate *T. erythraeum* cultures that more accurately represent both *in situ* and laboratory data.

One of the many advantages of using this multi-paradigm framework is that we can simulate emergent behavior of a population. *In situ* data report a wide mean range of trichome length from 1.2 to 685 cells<sup>54</sup>; we used the model to investigate possible causes because this is a difficult phenotype to investigate experimentally. Our simulations suggest that even though longer filaments suffer from diffusional effects that limit growth, they are better able to handle stress (Fig. 5) consistent with literature. Increased surface area in longer filaments minimizes the effect of lower light because the filament can harvest more light per volume. Also, the larger filaments are better able to maintain the average composition of storage compounds despite low carbon or low nitrogen conditions. Therefore, we would expect in areas of nutrient or light stress, the filament length would be longer.

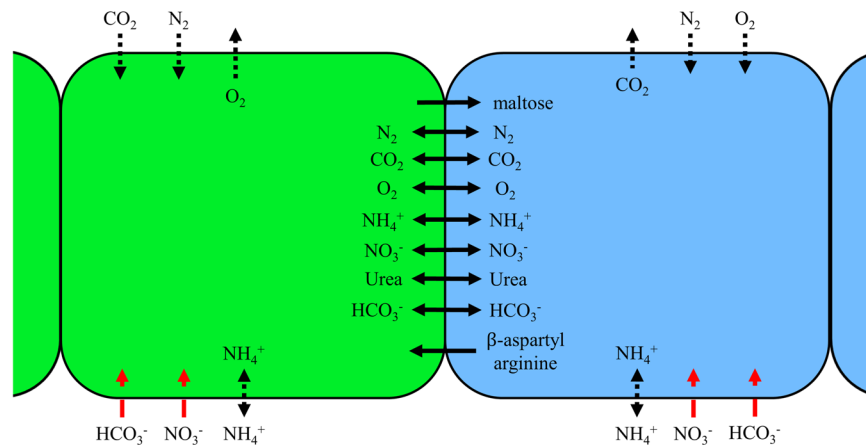
One of the other unusual phenotypes of *Trichodesmium* that we were able to investigate using MiMoSA was leaking 30–50% of the nitrogen it fixes. Nitrogen fixation is an incredibly energy intensive process, costing the cell 8 ATP per ammonium, so it is not clear why *T. erythraeum* would excrete 30–50%. Despite using optimization to solve for fluxes, which should minimize energy losses, our simulations predict approximately 20% of the fixed nitrogen is excreted into the medium (Fig. S5) which implies that this is a metabolically driven phenomenon. Further investigation has led us to develop three hypotheses on why this occurs: carbon limitation in diazotrophs, energy limitation in photoautotrophs, and imbalances between photoautotroph: diazotroph ratios. In the first case, photoautotrophs are unable to create glycogen chains and instead must start from a higher energy substrate than carbon dioxide (like succinic-semialdehyde) or must perform glycolysis on arginine derivatives to achieve energetic viability (Fig. S3A). Second, population imbalances cause nitrogen to be produced faster than it can be anabolized into  $\beta$ -aspartyl-arginine chains and is released into the media, meaning there is an optimal ratio of cell types (Fig. S3B). It is also possible that carbon limited diazotrophs are unable to manufacture full  $\beta$ -aspartyl-arginine chains and proton imbalances require ammonium release to the medium instead of passage to surrounding photoautotrophs.

MiMoSA enables the tracking of cellular-level environmental changes and the impact that they have on a metabolic model, opening the door to more accurate modeling of multi-cellular systems and the *in silico* investigation of the complex interactions between different cell types within an organism, and different species in a community. This is the first report of a metabolic model that integrates nutrient and light diffusion, cell/cell interactions and cell/environment interactions and we have used it to accurately predict growth, cellular composition and to investigate the unique physiology of *T. erythraeum*, which has filaments of both diazotrophs and photoautotrophs in close proximity. It establishes that this organism can effectively adapt to different conditions at three levels: the genetic level through division of labor in separate cell types, the metabolic level through relatively open-ended metabolic capabilities as well as further division within types, and at the population level to harness diffusional and physical interactions with the environment. MiMoSA is also a readily adaptable modeling framework – the addition of additional species to the model only requires the availability of a genome-scale metabolic and a few rules of behavior to be added. While we focused the proof-of-concept study of *T. erythraeum*, MiMoSA is a modeling framework that can be used to model a variety of more complex systems including applications in ecology, human health and metabolic engineering.

## Materials/Methods

**Cell culture conditions.** Cells were grown as described previously<sup>42</sup>. *Trichodesmium erythraeum* IMS101 cells were acquired from the Bigelow Laboratory for Ocean Sciences (East Boothbay, ME, USA). Cells were grown in a New Brunswick (Hamburg, Germany) with 100 and 50  $\mu\text{E}$  in 12 h light/12 h dark cycles. Cells were grown in artificial seawater YBC-II medium<sup>66</sup> at pH 8.15–8.20.  $\text{CO}_2$  was maintained at atmospheric concentration. All chemicals were obtained from Sigma-Aldrich (St. Louis, MO). Growth rate was monitored by measuring





**Figure 6.** Allowed Metabolite Diffusion and Exchange between and into Cells. Green cells represent photoautotrophic cells and blue represent diazotrophic cells. Solid black lines indicate free diffusion between cells, dotted black lines indicate lipid bilayer diffusion, and red dashed lines indicate coupled ion transport into the cell. Not pictured is ATP synthase. Maltose is modeled as the 12-carbon molecule that forms the foundation for glycogen while  $\beta$ -aspartyl arginine is the foundation for cyanophycin.

chlorophyll absorbance<sup>67</sup> from 50 mL of culture every two days. Cyanophycin and glycogen were measured every four hours from the beginning of the light cycle (9 AM) to its end (9 PM). Total biomass mass was determined by dry weight analysis, cells were filtered with a Whatman 0.22  $\mu\text{m}$  cellulose-nitrate filter and dried overnight at 100 °C.

**Biomass quantification.** Carbohydrates were measured colorimetrically using the anthrone method<sup>68</sup> against glycogen as a standard. Cyanophycin was extracted by disrupting 740  $\mu\text{L}$  of 250 mL cells concentrated to 2 mL via filtration and rinsing with TE buffer with 2.70 mg/mL lysozyme overnight at 37 °C, centrifuging at  $16,100 \times G$  for 5 min, and resuspending the pellet in 1 mL of 0.1 M HCl (in which cyanophycin is soluble) for 2 h. The extraction was repeated on the pellet, the supernatant fractions were combined, and cyanophycin was quantified colorimetrically using the Sakaguchi reaction<sup>69</sup>.

**Mass balance constraints.** Constraints based metabolic models are based on mass balances, therefore it is imperative that we develop accurate accounting of each element. Therefore, we used training data (Table S2) to estimate normal cellular consumption (Table S3). Average objective fluxes were estimated using mass balances around biomass and metabolite production with the formulation given in SI: I.C Estimating Mass Balance Constraints. Note that these formed training objectives: uptakes and consumption scale with nutrient molar contents around and within cells.

**Development of agent based model.** Repast Simphony<sup>70</sup> in Java was used as the agent-based modeling framework in which differentiated multi-objective metabolic models of *Trichodesmium erythraeum* are contained. It contains three agent types – Ocean, Cells, and Filaments. The Cells agent contains two sub-agents representing each cell type: photoautotrophs and diazotrophs and is responsible for intracellular processes and decisions. The Ocean agent defines and calculates the extracellular environment and the Filaments agent organizes the Cells and modulates their transactions.

**Cells.** Cell agents (cells) are generated for each individual cell in the model. These contain two subtypes, photoautotrophs and diazotrophs, but contain several consistent elements between the two. Simulation variables are summarized in Table S4. All cells reproduce according to the same rules: cells divide according to sampling from the weighting distribution described above if that sample is bigger than the cell mean cell, cells only extend from the ends, and cells can only divide into diazotrophs if there is a diazocyte under development (decided at the filament level if the filament is nitrogen limited). When a cell is large enough, it converts to fully stationary growth, producing only metabolites and creating a larger and larger metabolic gradient between cells without *de novo* biomass synthesis. This prevents a cell from becoming excessively large in the center of the filament. Cells will die if they cannot produce the requisite maintenance ATP through metabolism or catabolism.

Cells allow metabolites to diffuse through the lipid bilayer using permeabilities reported in the literature (Table S5). This mechanism represents a non-zero leakage scenario that was nevertheless much slower than intra-filamental diffusion (Table S6). Scavenging from the environment for compounds which carried no evidence of active transport followed these same rules and was therefore prone to concentration gradients. Active transporters, on the other hand, allowed the cell to uptake whatever concentration of compound was necessary subject to its presence in the local ocean grid. Allowable exchange of metabolites between cells is illustrated in Fig. 6. If several cells compete in that grid space, access to the available molecule was divided equally among those cells.

**Subclasses: photoautotrophs and diazotrophs.** Both subclasses define the uptake constraints and send to a Python file that decides whether the cell metabolizes or catabolizes based on those constraints using the multi-objective

metabolic model previously described (see *Supplemental Information: Routine Metabolic Optimizations*). The cell then updates its internal metabolites based on the optimization results, diffuses metabolites, divides if possible, and uptakes from its local environment. Model bounds are calculated using local concentrations to calculate maximum flux bounds excepting  $\beta$ -aspartyl arginine which is further limited to 8% of available nutrients (See Fig. S6). These methods are handled by three ScheduledMethods that Repast Symphony schedules in specific progression. Together with the Ocean Agent's updates, the individual cell actions (as dictated by the metabolic model) form the core of the simulation. A more detailed flow chart for cell decision making can be found in Fig. S7. Progression through these steps is identical for cell types, but the metabolic static variables (objectives, gas uptake, etc.) are different between the two subclasses, necessitating separate methods.

**Ocean.** The ocean agent is responsible for tracking cells and modeling external nutrients. Its main task is facilitating diffusion between cells and locations as well as approximating an uptake radius for cells. Each ocean represents a uniform, static, abstract area of the overall grid space with a uniform dimension space of  $\delta \times \delta$  where  $\delta$  is a user defined parameter. This set of simulations was conducted with time steps of 0.1 hours as a moderate value between diffusion phenomena (on the order of seconds along the length of a filament) and doubling time (on the order of 50 hours). Metabolites are assumed to freely diffuse in a dilute seawater environment between cell filaments (Table S6) and assumed to be uniform within the grid, given the relatively long time step compared to the rate of diffusion over such small dimensions. If the impacts of metabolic diffusion limitations were of interest, the time step within the framework could be made appropriately small to more accurately track metabolites, at cost of increased computational burden. Each ocean gridcell diffuses molecules into its adjacent ocean gridcells assuming discretized slab diffusion in two dimensions. This is done using a previously developed discrete algorithm for diffusion in a grid<sup>71</sup>:

$$\Delta \bar{f}(x_i) = A \sum_{j=1}^n (f(x_j^i) - f(x_i)) e^{-d_j^2/\eta} \quad (1)$$

$$d_j = |x_i - x_j^i| \quad (2)$$

$$A \sum_{k=1}^n e^{-d_k^2/\eta} = 1 \quad (3)$$

where  $\Delta \bar{f}(x_i)$  is the change in concentration of metabolite in grid space  $x_i$  over a time step,  $A$  is the normalization constant to be calculated by solving the third equation within the entire neighborhood to ensure conservation of mass within the neighborhood,  $d_j$  is the distance between grid space  $x_i$  and its grid neighbor  $x_j^i$  and  $\eta$  is the diffusivity control of the system over the time step. As  $\eta \rightarrow 0$  diffusion halts and as  $\eta \rightarrow \infty$  diffusion becomes instantaneous. In this study,  $\eta = 4D\Delta t$  as in the original Fick's Law.

Diffusion is calculated using two steps, one forward and one reversing the order of gridspace calculation, to mitigate the effect of order on estimating the concentration gradient (Fig. S8). Excess ammonium is secreted into the environment using a membrane diffusion coefficient. Cells are allowed to uptake any metabolite/nutrient in YBC-II medium; the only extracellular products allowed in simulations are small molecules, such as  $\text{CO}_2$  and  $\text{NH}_4^+$  which diffuse through the membrane, as well as compounds that have experimental evidence of transporters from proteomic analysis or transcriptomic analysis (estimated using membrane diffusion outwardly and free diffusion for gases or active transporters for ions/molecules inwardly) in Table S7<sup>72,73</sup>.

The Ocean Agents also manage diffusion of metabolites from marine sinks and through the gas-liquid surface interface with the atmosphere. This is done assuming equilibrium concentrations of dissolved gases defined by Henry's Law and mono-directional slab diffusion for  $\text{CO}_2$ ,  $\text{O}_2$ , and  $\text{N}_2$ . Table S6 lists the free diffusivities of compounds and Table S8 lists the Henry's Constants for atmospheric compounds, and Fig. S8 demonstrates the movement of diffusive molecules through the simulation.

Furthermore, light diffusion to cells is defined as a function of their y coordinate according to the equation:

$$I = I_0 e^{-ky} \quad (4)$$

where  $I$  is light intensity,  $k$  is the extinction coefficient of light in seawater, and  $y$  is the depth below the surface of the individual cell.

**Filaments.** Filament Agents are responsible for organizing cells, managing movement, splitting to promote diazotroph development, and defining cell type after division. Random walk movement (to simulate the lack of control cells have over lateral motion) is simulated by generating a random direction that has an empty grid space for every cell in the filament. Cells move within a user defined interval of time or if their growth is impeded by another filament, in which case growth is halted until the cells move away from each other. The filament forces splitting into two separate filaments when nitrogen is limiting growth and neither filament end is undergoing diazotroph development (meaning that another diazocyte is required). Filament Agents decide the next cell type using this inequality:

$$\frac{\sum_i \epsilon_i^{DZ}}{n_{DZ}} > \frac{\sum_i \epsilon_i^{PA}}{n_{PA}} \quad (5)$$

Cell Type	Maintenance Energy, $\nu_{ATP}$		$m \left( \frac{\text{mmol}}{\text{g DW h } \mu\text{E}} \right)$	$\nu_0 \left( \frac{\text{mmol}}{\text{g DW h}} \right)$	$L_0 \text{ (}\mu\text{E)}$
	80 $\mu\text{E} \left( \frac{\text{mmol}}{\text{g DW h}} \right)$	100 $\mu\text{E} \left( \frac{\text{mmol}}{\text{g DW h}} \right)$			
Photoautotroph	34.3	53.3	0.952	-16.7	16.9
Diazotroph	62.3	82.0	0.987	-41.9	44.0

**Table 1.** ATP maintenance flux requirements estimated as a function of light intensity for Pareto Fitting.

where  $\varepsilon$  is the Pareto Efficiency of the given cell type and  $n$  is the quantity of that cell type in the filament. The Pareto Efficiency is quantified as the sum of the objective fluxes divided by their Pareto Optimum (from experimental results) divided by the number of objectives.

$$\varepsilon_i^c = \sum_j \frac{\nu_j}{\nu_j^{exp}} \quad (6)$$

If inequality (20) is satisfied, the cell prioritizes diazotroph development, otherwise it prioritizes photoautotroph development. If a diazotroph region is currently under development, the filament adds another cell to that region. If there is no diazotroph under development, or if the C:N ratio becomes higher than physiological bounds, the filament splits to expose a region where diazotroph development may begin. A photoautotroph can be placed at any open site. Since there are two ends on every filament, up to two of these decisions are being made during each simulation time step. After filament splitting, if the split results in a homogenous region of either diazotrophs or photoautotrophs, the missing cell type is preferred. Filaments split in the middle of the longest region of homogenous cells and are prevented from splitting to result in a single cell, meaning that the shortest possible resulting splits are two cells in length. Cell division completes within one time step when metabolites and biomass are equally divided between parent and daughter cell and the filament updates to contain the cell at its end. This decision is a memoryless process conducted each time step. This means that cell division is completely metabolically motivated (which is affected, in turn, by diffusion and physiological processes).

**Parameter estimation.** As described previously, to improve the accuracy of simulations, the model was fit to experimental data for cells grown in 100  $\mu\text{E}$  light in YBC-II medium. Maintenance energy, in the form of the ATP hydrolysis reaction, is the main parameter that is adjusted in FBA formulations<sup>18,42,74–76</sup> to match simulations with growth rate. Since maintenance energy at 100  $\mu\text{E}$  was higher than the energetic capacity of the model for growth at 50  $\mu\text{E}$ , a linear correlation was interpolated from experiments at 100  $\mu\text{E}$  and 80  $\mu\text{E}$  with ATP maintenance flux fit to multiple light intensities without specifically training on 50  $\mu\text{E}$ :

$$\nu_{ATP}(I) = mI + \nu_0 \quad (7)$$

where  $m$  and  $\nu_0$  are calculated using the point-slope equation for a linear equation:

$$m = \frac{\nu_{ATP,I_1} - \nu_{ATP,I_2}}{I_1 - I_2} \quad (8)$$

$$\nu_0 = -I_1 m + \nu_{ATP,I_1} \quad (9)$$

where  $I_1$  is 100  $\mu\text{E}$  and  $I_2$  is 80  $\mu\text{E}$ . The estimated values of the linear equation are recorded below in Table 1. If the model is unable to satisfy its maintenance demand (through any metabolic process, including catabolizing its own biomass), the cell dies.  $L_0$  is the energy required in zero light to maintain the cell without active metabolism.

**Multiobjective optimization.** Unlike typical formulations of flux balance analysis<sup>13,75–80</sup>, which use a single objective function to predict fluxes, our model uses multi-objective optimization to more accurately approximate the true objectives of the cell: to optimize biomass while also producing the metabolite they exchange between cell types. The multiobjective scalar function is of the form:

$$r_{obj} = aX + (1 - a)m \quad (10)$$

where  $a$  is a scaling parameter between 0 and 1,  $X$  is non-metabolite biomass (proteins, lipids, carbohydrates, nucleic acids, and pigments) and  $m$  is the transactional metabolite ( $\beta$ -aspartyl arginine for diazotrophs and maltose for photoautotrophs). The value for  $a$  was varied between 0 and 1 over 1000 steps to generate a Pareto Front, optimizing the new objective function subject to the training constraints and ATP maintenance flux. Each point on that front corresponds to an arrangement of scalarized variables. Flux through  $r_{obj}$  can be then reinterpreted into direct biomass and metabolite fluxes via multiplication of their coefficient by the objective solution:

$$\mu = \nu_X = a\nu_{obj} \quad (11)$$

$$\nu_m = (1 - a)\nu_{obj} \quad (12)$$

Visualization of this approach can be seen in Figs 2 and S9. For each point along the Pareto Front, Euclidean distance was used to determine the relative weight of each objective function, which was then used to generate

a single, scalarized reaction. Each cell in the simulation calculates its scalar objective function separately during each time step based on its internal constitution and requirements.

**Shifting scalar objectives.** To investigate the mutability of the scalar objective equation, and the effects of intra- and inter-cellular conditions on cell objectives, an algorithm to shift the objective function along the Pareto Curve is implemented based on the biomass of the cells (assuming that structural biomass is relatively intransigent in these conditions). That means that, as biomass accumulates, a cell will behave more stationarily, i.e., it will prioritize secondary metabolites. This is adjusted assuming a normal distribution in cell sizes:

$$f(X) \sim N(\mu, \sigma) \quad (13)$$

where  $\mu$  is the average biomass (assuming cubic shape and density near water) or metabolite concentration (from experimental data) and  $\sigma$  is the standard deviation. As the standard deviation decreases, the model is less tolerant of deviations from mean values and will more readily adjust the scalar weights. As it increases, it will be more tolerant of deviations from mean values and will cause smaller perturbations to scalar weights. For this study, the standard deviation was assumed to be 0.433 so that 95% of cells would fall within a factor of seven from minimum to maximum and a mean of  $1.029 \times 10^{-9}$  g calculated using an assumed cubic shape with dimensions of 10  $\mu\text{m}$  per side and a density of seawater<sup>81</sup>.

The correction occurs by using the cumulative distribution function for the normal distribution. Given a Pareto matching algorithm resulting set of weights, the weights are shifted (inversely) as:

$$w_1 = a\hat{w}_1 \quad (14)$$

$$\hat{w}_1 = \bar{w}_1[1 - F(X_1 = x_1)] \quad (15)$$

$$w_{j \neq 1} = w_{j \neq 1}F(X_{j \neq 1} = x_{j \neq 1}) \quad (16)$$

$$a \sum_i w_i = 1 \quad (17)$$

where  $w_1$  is the new, shifted weight,  $\hat{w}_1$  is the non-normalized and shifted weight, and  $\bar{w}_1$  is the Pareto matched weight.  $X$  is the variable and  $x$  is the cell's quantity. Meaning, in each case, that as the objective increases, it responds by producing the corresponding objective: as biomass increases, it expends more resources on biomass.

**Implementation of mutable objective functions.** Previous studies have used static objective functions, where production is consistent during every phase of growth. However, organisms accumulate and digest metabolites during growth and development. To reflect this, we inserted a “mutable” objective function where relative preferences of storage compounds and biomass production can be tailored by the agent based on cell biomass. The scalarized objective equation was thus broken into two main components: storage compounds (cyanophycin modeled as  $\beta$ -aspartyl arginine and glycogen modeled as maltose) and biomass (lipids, proteins, DNA, RNA, chlorophyll, phycoerythrin, etc.). We assumed that biomass remained relatively stable throughout the day while the amount of storage compound was allowed to vary. The scalar weights, or production priorities, were manipulated assuming cells do not grow beyond twice their average cell without dividing: lower biomass prioritizes growth and higher biomass prioritizes vegetative storage compound production. Mathematically, this is modeled such that the scalar objective equation's biomass coefficient was inversely adjusted by cumulative probability of a cell's biomass in the distribution. The normal distribution was formulated assuming cubic 10  $\mu\text{m}$  cells with density of water<sup>81</sup> as the average mass and a narrow distribution with a standard deviation of 0.433 times the mean size. This value was chosen to promote switch-like bistable behavior between cell phenotypes: either cells are biomass driven (exponential) or they are metabolite driven with combinations of probabilities in between. This is because a single sample of a cell from a distribution of cells would have a probability of 99% to fall between 0 and twice the mean size. The final distribution is:

$$f(X) \sim N(1.029 \text{ ng}, 0.433 \cdot 1.029 \text{ ng}) \quad (18)$$

Calculation of new objective coefficients was done by first finding the cumulative probability ( $z$ ) of another randomly selected cell's non-metabolite biomass being less than or equal to the objective cell's biomass at each time point for each cell:

$$z = F(X \leq x_i) \quad (19)$$

This is used to adjust the average, experimentally matched objective coefficient ( $\bar{w}_b$ ) for biomass by multiplying that coefficient by the probability of the cell being larger than that size, a value that represents the probabilistic expansion space ( $\epsilon$ ) of the cell:

$$\epsilon = 1 - z \quad (20)$$

$$\hat{w}_b = \epsilon \cdot \bar{w}_b \quad (21)$$



Major metabolite coefficients for the scalarized objective equation were also adjusted using this probability, increasing as the cell's size increased:

$$\hat{w}_m = z \cdot \bar{w}_m \quad (22)$$

Finally, the coefficients are normalized such that:

$$a \sum_i \hat{w}_i = 1 \quad (23)$$

Or:

$$a = \frac{1}{\sum_i \hat{w}_i} \quad (24)$$

Which yields final objective coefficients of:

$$w_k = \frac{\hat{w}_k}{\sum_i \hat{w}_i} \quad \forall k \in \mathcal{O} \quad (25)$$

where  $\mathcal{O}$  is the set of all objective metabolites in the original scalar equation.

Performance evaluation of the mutable objective function, validation of O the mutable objective function versus the static version, and justification of non-metabolite biomass as the independent objective are provided in *SI Methods* and Fig. S10.

## Data availability

We have provided the ABM framework as supplemental file 1 and the datasets resulting from simulations are available at the Boyle Laboratory website, <https://nboylelab.com>.

Received: 29 April 2019; Accepted: 29 October 2019;

Published online: 18 November 2019

## References

- Hutchins, D. *et al.* CO<sub>2</sub> control of *Trichodesmium* N<sub>2</sub> fixation, photosynthesis, growth rates, and elemental ratios: Implications for past, present, and future ocean biogeochemistry. *Limnology and Oceanography* **52**, 1293–1304 (2007).
- Spungin, D., Berman-Frank, I. & Levitan, O. *Trichodesmium's* strategies to alleviate phosphorus limitation in the future acidified oceans. *Environmental microbiology* **16**, 1935–1947 (2014).
- Fiegna, F., Moreno-Letelier, A., Bell, T. & Barraclough, T. G. Evolution of species interactions determines microbial community productivity in new environments. *The ISME journal* **9**, 1235 (2015).
- Levitan, O. *et al.* Regulation of nitrogen metabolism in the marine diazotroph *Trichodesmium* IMS101 under varying temperatures and atmospheric CO<sub>2</sub> concentrations. *Environmental microbiology* **12**, 1899–1912 (2010).
- Westberry, T. K. & Siegel, D. A. Spatial and temporal distribution of *Trichodesmium* blooms in the world's oceans. *Global Biogeochemical Cycles* **20** (2006).
- Eichner, M. *et al.* N<sub>2</sub> fixation in free floating filaments of *Trichodesmium* is higher than in transiently suboxic colony microenvironments. *New Phytologist* (2018).
- Covert, M. W. *et al.* Metabolic modeling of microbial strains in silico. *Trends Biochem Sci* **26**, [https://doi.org/10.1016/s0968-0004\(00\)01754-0](https://doi.org/10.1016/s0968-0004(00)01754-0) (2001).
- Karr, J. R. *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401 (2012).
- Harcombe, W. R. *et al.* Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell reports* **7**, 1104–1115 (2014).
- Du, B., Zielinski, D. C., Monk, J. M. & Palsson, B. O. Thermodynamic favorability and pathway yield as evolutionary tradeoffs in biosynthetic pathway choice. *Proceedings of the National Academy of Sciences* **115**, 11339–11344 (2018).
- Fang, X. *et al.* Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. *Proceedings of the National Academy of Sciences* **114**, 10286–10291 (2017).
- Levering, J., Dupont, C. L., Allen, A. E., Palsson, B. O. & Zengler, K. Integrated regulatory and metabolic networks of the marine diatom *Phaeodactylum tricornutum* predict the response to rising CO<sub>2</sub> levels. *MSystems* **2**, e00142–00116 (2017).
- Mahadevan, R., Edwards, J. S. & Doyle, F. J. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical journal* **83**, 1331–1340, [https://doi.org/10.1016/S0006-3495\(02\)73903-9](https://doi.org/10.1016/S0006-3495(02)73903-9) (2002).
- Zuniga, C. *et al.* Predicting dynamic metabolic demands in the photosynthetic eukaryote *Chlorella vulgaris*. *Plant physiology* **176**, 450–462 (2018).
- Schuetz, R., Kuepfer, L. & Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol* **3**, 119 (2007).
- Stolyar, S. *et al.* Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology* **3**, 92–92, <https://doi.org/10.1038/msb4100131> (2007).
- Förster, J., Famili, I., Fu, P., Palsson, B. O. & Nielsen, J. Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. *Genome Research* **13**, 244–253, <https://doi.org/10.1101/gr.234503> (2003).
- Boyle, N. & Morgan, J. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Systems Biology* **3**, 4 (2009).
- Chang, R. L. *et al.* Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol Syst Biol* **7**, [http://www.nature.com/msb/journal/v7/n1/supinfo/msb201152\\_S1.html](http://www.nature.com/msb/journal/v7/n1/supinfo/msb201152_S1.html) (2011).
- Taffs, R. *et al.* In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Systems Biology* **3**, 1–16, <https://doi.org/10.1186/1752-0509-3-114> (2009).
- Zomorodi, A. R. & Maranas, C. D. OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. *PLoS Comput Biol* **8**, e1002363, <https://doi.org/10.1371/journal.pcbi.1002363> (2012).
- Zomorodi, A. R., Islam, M. M. & Maranas, C. D. d-OptCom: Dynamic Multi-level and Multi-objective Metabolic Modeling of Microbial Communities. *ACS Synthetic Biology* **3**, 247–257, <https://doi.org/10.1021/sb4001307> (2014).
- Borshchev, A. & Filippov, A. In *Proceedings of the 22nd international conference of the system dynamics society*. (Citeseer).

24. Klann, M., Lapin, A. & Reuss, M. Agent-based simulation of reactions in the crowded and structured intracellular environment: Segovia-Infuere of mobility and location of the reactants. *BMC Systems Biology* **5**, 71 (2011).
25. Segovia-Juarez, J. L., Ganguli, S. & Kirschner, D. Identifying control mechanisms of granuloma formation during *M. tuberculosis* infection using an agent-based model. *Journal of Theoretical Biology* **231**, 357–376, <https://doi.org/10.1016/j.jtbi.2004.06.031> (2004).
26. Parunak, H. V. D. Practical and industrial applications of agent-based systems. *Environmental Research Institute of Michigan (ERIM)* (1998).
27. Bin, C., Gang, G. & Xiaogang, Q. In *Digital Manufacturing and Automation (ICDMA)*, 2013 Fourth International Conference on. 1396–1400.
28. Hardebolle, C. & Boulanger, F. Exploring Multi-Paradigm Modeling Techniques. *Simulation* **85**, 688–708, <https://doi.org/10.1177/0037549709105240> (2009).
29. Hodge, B.-M. *et al.* Multi-Paradigm Modeling of the Effects of PHEV Adoption on Electric Utility Usage Levels and Emissions. *Industrial & Engineering Chemistry Research* **50**, 5191–5203, <https://doi.org/10.1021/ie101837w> (2011).
30. Hodge, B.-M. S., Huang, S., Sirola, J. D., Pekny, J. F. & Reklaitis, G. V. A multi-paradigm modeling framework for energy systems simulation and analysis. *Computers & Chemical Engineering* **35**, 1725–1737, <https://doi.org/10.1016/j.compchemeng.2011.05.005> (2011).
31. Mosterman, P. J. & Vangheluwe, H. Computer Automated Multi-Paradigm Modeling: An Introduction. *SIMULATION* **80**, 433–450, <https://doi.org/10.1177/0037549704050532> (2004).
32. Tenazinha, N. & Vinga, S. A Survey on Methods for Modeling and Analyzing Integrated Biological Networks. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* **8**, 943–958, <https://doi.org/10.1109/TCBB.2010.117> (2011).
33. Machado, D. *et al.* Modeling Formalisms in Systems Biology. *AMB Express* **1**, 45 (2011).
34. Zuñiga, C., Zaramela, L. & Zengler, K. Elucidation of complexity and prediction of interactions in microbial communities. *Microbial biotechnology* **10**, 1500–1522 (2017).
35. Nagarajan, H. *et al.* Characterization and modelling of interspecies electron transfer mechanisms and microbial community dynamics of a syntrophic association. *Nature communications* **4**, 2809 (2013).
36. Simons, M., Misra, A. & Sriram, G. In *Plant Metabolism* Vol. 1083 *Methods in Molecular Biology* (ed. Ganesh Sriram) Ch. 13, 213–230 (Humana Press, 2014).
37. Heinken, A., Sahoo, S., Fleming, R. M. & Thiele, I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut microbes* **4**, 28–40 (2013).
38. Libourel, I. G. L. & Shachar-Hill, Y. Metabolic Flux Analysis in Plants: From Intelligent Design to Rational Engineering. *Annual Review of Plant Biology* **59**, 625–650, <https://doi.org/10.1146/annurev.arplant.58.032806.103822> (2008).
39. Zhuang, K., Ma, E., Lovley, D. R. & Mahadevan, R. The design of long term effective uranium bioremediation strategy using a community metabolic model. *Biotechnol Bioeng* **109**, <https://doi.org/10.1002/bit.24528> (2012).
40. Berman-Frank, I., Lundgren, P. & Falkowski, P. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Research in Microbiology* **154**, 157–164 (2003).
41. Glibert, P. M. & Bronk, D. A. Release of Dissolved Organic Nitrogen by Marine Diazotrophic Cyanobacteria, *Trichodesmium* spp. *Applied and Environmental Microbiology* **60**, 3996–4000 (1994).
42. Gardner, J. J. & Boyle, N. R. The use of genome-scale metabolic network reconstruction to predict fluxes and equilibrium composition of N-fixing versus C-fixing cells in a diazotrophic cyanobacterium, *Trichodesmium erythraeum*. *BMC Systems Biology* **11**, 4, <https://doi.org/10.1186/s12918-016-0383-z> (2017).
43. Burnat, M., Herrero, A. & Flores, E. Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. *Proceedings of the National Academy of Sciences* **111**, 3823–3828 (2014).
44. Sherman, D. M., Tucker, D. & Sherman, L. A. Heterocyst development and localization of cyanophycin in N<sub>2</sub>-fixing cultures of *Anabaena* sp. PCC 7120 (cyanobacteria). *Journal of Phycology* **36**, 932–941 (2000).
45. Simon, R. D. Cyanophycin granules from the blue-green alga *Anabaena cylindrica*: a reserve material consisting of copolymers of aspartic acid and arginine. *Proceedings of the national academy of sciences* **68**, 265–267 (1971).
46. Boatman, T. G., Lawson, T. & Geider, R. J. A key marine diazotroph in a changing ocean: The interacting effects of temperature, CO<sub>2</sub> and light on the growth of *Trichodesmium erythraeum* IMS101. *PLoS one* **12**, e0168796 (2017).
47. Breitbart, E., Wohlers, J., Kläs, J., LaRoche, J. & Peeken, I. Nitrogen fixation and growth rates of *Trichodesmium* IMS-101 as a function of light intensity. *Marine Ecology Progress Series* **359**, 25–36 (2008).
48. Kranz, S. A., Dieter, S., Richter, K.-U. & Rost, B. Carbon acquisition by *Trichodesmium*: the effect of pCO<sub>2</sub> and diurnal changes. *Limnology and Oceanography* **54**, 548–559 (2009).
49. Eichner, M., Thoms, S., Kranz, S. A. & Rost, B. Cellular inorganic carbon fluxes in *Trichodesmium*: a combined approach using measurements and modelling. *Journal of experimental botany* **66**, 749–759 (2014).
50. Buitenhuis, E. *et al.* MAREDAT: towards a world atlas of MARine Ecosystem DATA. *Earth System Science Data* **5**, 227–239 (2013).
51. Berman-Frank, I. *et al.* Segregation of Nitrogen Fixation and Oxygenic Photosynthesis in the Marine Cyanobacterium *Trichodesmium*. *Science* **294**, 1534–1537, <https://doi.org/10.1126/science.1064082> (2001).
52. Zhang, S. & Bryant, D. A. The Tricarboxylic Acid Cycle in Cyanobacteria. *Science* **334**, 1551–1553, <https://doi.org/10.1126/science.1210858> (2011).
53. Luo, Y. *et al.* Database of Diazotrophs in Global Ocean: Abundance, Biomass, and Nitrogen Fixation Rates. *Earth System Science Data* **4** (2012).
54. White, A. E., Watkins-Brandt, K. S. & Church, M. J. Temporal Variability of *Trichodesmium* spp. and Diatom-Diazotroph Assemblages in the North Pacific Subtropical Gyre. *Frontiers in Marine Science* **5**, 27 (2018).
55. Yoo, S.-H., Keppel, C., Spalding, M. & Jane, J.-I. Effects of growth condition on the structure of glycogen produced in cyanobacterium *Synechocystis* sp. PCC6803. *International journal of biological macromolecules* **40**, 498–504 (2007).
56. Cai, X. *et al.* Electron transport kinetics in the diazotrophic cyanobacterium *Trichodesmium* spp. grown across a range of light levels. *Photosynthesis research* **124**, 45–56 (2015).
57. Blaby, I. K. *et al.* Systems-Level Analysis of Nitrogen Starvation-Induced Modifications of Carbon Metabolism in a *Chlamydomonas reinhardtii* Starchless Mutant. *The Plant Cell Online* **25**, 4305–4323, <https://doi.org/10.1105/tpc.113.117580> (2013).
58. Boyle, N. R. *et al.* Three Acyltransferases and Nitrogen-responsive Regulator Are Implicated in Nitrogen Starvation-induced Triacylglycerol Accumulation in *Chlamydomonas*. *Journal of Biological Chemistry* **287**, 15811–15825, <https://doi.org/10.1074/jbc.M111.334052> (2012).
59. Breuer, G., Lamers, P. P., Martens, D. E., Draaisma, R. B. & Wijffels, R. H. The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. *Bioresource Technology* **124**, 217–226, <https://doi.org/10.1016/j.biortech.2012.08.003> (2012).
60. Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S. & Malin, G. The Response of Diatom Central Carbon Metabolism to Nitrogen Starvation Is Different from That of Green Algae and Higher Plants. *Plant Physiology* **158**, 299–312, <https://doi.org/10.1104/pp.111.184333> (2012).
61. Tedesco, M. A. & Duerr, E. O. Light, temperature and nitrogen starvation effects on the total lipid and fatty acid content and composition of *Spirulina platensis* UTEX 1928. *Journal of Applied Phycology* **1**, 201–209, <https://doi.org/10.1007/bf00003646> (1989).
62. Kropat, J. *et al.* A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *The Plant journal: for cell and molecular biology* **66**, 770–780, <https://doi.org/10.1111/j.1365-313X.2011.04537.x> (2011).

63. Boatman, T. G., Davey, P. A., Lawson, T. & Geider, R. J. The physiological cost of diazotrophy for *Trichodesmium erythraeum* IMS101. *PLoS one* **13**, e0195638 (2018).
64. Boatman, T. G., Davey, P. A., Lawson, T. & Geider, R. J. CO<sub>2</sub> modulation of the rates of photosynthesis and light-dependent O<sub>2</sub> consumption in *Trichodesmium*. *Journal of experimental botany* **70**, 589–597 (2018).
65. Boatman, T. G., Mangan, N. M., Lawson, T. & Geider, R. J. Inorganic carbon and pH dependency of photosynthetic rates in *Trichodesmium*. *Journal of experimental botany*, ery141 (2018).
66. Chen, Y.-B., Zehr, J. P. & Mellon, M. Growth and Nitrogen Fixation of the Diazotrophic Filamentous Nonheterocytous Cyanobacterium *Trichodesmium* sp. IMS 101 in Defined Media: Evidence for a Circadian Rhythm. *Journal of Phycology* **32**, 916–923, <https://doi.org/10.1111/j.0022-3646.1996.00916.x> (1996).
67. Harris, E. H., Stern, D. B. & Witman, G. *The Chlamydomonas sourcebook*. Vol. 1 (Cambridge Univ Press, 2009).
68. Yemm, E. & Willis, A. The estimation of carbohydrates in plant extracts by anthrone. *Biochemical journal* **57**, 508 (1954).
69. Messineo, L. Modification of the Sakaguchi reaction: spectrophotometric determination of arginine in proteins without previous hydrolysis. *Archives of Biochemistry and Biophysics* **117**, 534–540 (1966).
70. Collier, N. & North, M. Parallel agent-based simulation with repast for high performance computing. *Simulation* **89**, 1215–1235 (2013).
71. Grajdeanu, A. Modeling Diffusion in a Discrete Environment. *George Mason University Technical Report Series*, 1–5 (2007).
72. Sandh, G. *et al.* Comparative proteomic profiles of the marine cyanobacterium *Trichodesmium erythraeum* IMS101 under different nitrogen regimes. *Proteomics* **11**, 406–419 (2011).
73. Pfreundt, U., Kopf, M., Belkin, N., Berman-Frank, I. & Hess, W. R. The primary transcriptome of the marine diazotroph *Trichodesmium erythraeum* IMS101. *Scientific reports* **4** (2014).
74. Boyle, N. R., Shastri, A. A. & Morgan, J. A. In *Plant Metabolic Networks* (ed Jörg Schwender) Ch. 8, 211–243 (Springer New York 2009).
75. Knoop, H. *et al.* Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis* sp. PCC 6803. *PLoS Comput Biol* **9**, e1003081, <https://doi.org/10.1371/journal.pcbi.1003081> (2013).
76. Misra, A. *et al.* Metabolic analyses elucidate nontrivial gene targets for amplifying dihydroartemisinic acid production in yeast. *Frontiers in Microbiology* **4**, <https://doi.org/10.3389/fmicb.2013.00200> (2013).
77. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat Biotech* **28**, 245–248, <http://www.nature.com/nbt/journal/v28/n3/abs/nbt.1614.html#supplementary-information> (2010).
78. Kauffman, K. J., Prakash, P. & Edwards, J. S. Advances in flux balance analysis. *Current Opinion in Biotechnology* **14**, 491–496, <https://doi.org/10.1016/j.copbio.2003.08.001> (2003).
79. Edwards, J., Ramakrishna, R., Schilling, C. & Palsson, B. Metabolic flux balance analysis. *Metabolic engineering* (1999).
80. Asadollahi, M. A. *et al.* Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. *Metabolic Engineering* **11**, 328–334, <https://doi.org/10.1016/j.ymben.2009.07.001> (2009).
81. van Baalen, C. & Brown, R. M. Jr. The ultrastructure of the marine blue green alga, *Trichodesmium erythraeum*, with special reference to the cell wall, gas vacuoles, and cylindrical bodies. *Archiv für Mikrobiologie* **69**, 79–91 (1969).
82. Blazeck, J. & Alper, H. Systems metabolic engineering: Genome-scale models and beyond. *Biotechnology Journal* **5**, 647–659, <https://doi.org/10.1002/biot.200900247> (2010).
83. Flynn, K. J. Ecological modelling in a sea of variable stoichiometry: dysfunctionality and the legacy of Redfield and Monod. *Progress in Oceanography* **84**, 52–65 (2010).
84. Mahadevan, R. & Schilling, C. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering* **5**, 264–276 (2003).

## Acknowledgements

This work was supported by a grant from the Department of Energy Office of Science, Biological and Environmental Research (BER) Early Career Program grant no. DE-SC0019171.

## Author contributions

J.J.G., B.M.S.H. and N.R.B. designed the research. J.J.G. performed the research. J.J.G. and N.R.B. analyzed the data. J.J.G., B.M.S.H. and N.R.B. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-53188-0>.

**Correspondence** and requests for materials should be addressed to N.R.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019