1 **<u>M</u>ultIscale <u>M</u>ultiObjective <u>S</u>ystems <u>A</u>nalysis (MIMOSA): an advanced metabolic modeling**

2 **framework for complex systems**

3

4 Joseph J. Gardner[1], Bri-Mathias S. Hodge[1,2,3], Nanette R. Boyle[1,*]

5

6 [1]Chemical & Biological Engineering, Colorado School of Mines, 1613 Illinois St. Golden, CO

7 80403; [2]National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO

8 80401; [3]Electrical, Computer and Energy Engineering, 425 UCB, University of Colorado,

9 Boulder, CO 80309

10 *Corresponding Author

11

12 **ABSTRACT**

13 In natural environments, cells live in complex communities and experience a high degree of

14 heterogeneity internally and in the environment. Unfortunately, most of the metabolic modeling

15 approaches that are currently used assume ideal conditions and that each cell is identical, limiting

16 their application to pure cultures in well-mixed vessels. Here we describe our development of

17 MultIscale MultiObjective Systems Analysis (MIMOSA), a metabolic modeling approach that can

18 track individual cells in both space and time, track the diffusion of nutrients and light and the

19 interaction of cells with each other and the environment. As a proof-of concept study, we used

20 MIMOSA to model the growth of *Trichodesmium erythraeum*, a filamentous diazotrophic

21 cyanobacterium which has cells with two distinct metabolic modes. The use of MIMOSA

22 significantly improves our ability to predictively model metabolic changes and phenotype in more

23 complex cell cultures.

24   **KEYWORDS**

25   systems biology, agent-based modeling, metabolic flux, multi-paradigm, advanced metabolic

26   model

27   **BACKGROUND**

28   Microbes live in complex communities where they must interact with other organisms and compete

29   for resources to thrive. By leveraging the capabilities of each individual in the community,

30   consortia can achieve outcomes that are not possible by any one individual species. Metabolic

31   engineers are learning from nature and are engineering synthetic consortia to take advantage of

32   endogenous capabilities of specialists to achieve higher yields than pure cultures. One tool that has

33   been used extensively to aide in the rational design of strains are metabolic models (11). The most

34   widely used stoichiometric metabolic models are constraint-based linear programming models

35   which vary in complexity from the relatively simple flux balance analysis (FBA) to more complex

36   FBA models which integrate regulatory and/or thermodynamic constraints (12, 13) or time-

37   dependent responses (14). The wide use of these models is due to the ease of constructing them;

38   access to the genome sequence is enough to build a draft metabolic network. The simplicity of this

39   technique does come at a cost: typical model formulations are limited to modeling steady-state

40   growth of axenic cultures assuming homogenous environmental conditions; while this works for

41   traditional metabolic engineering efforts on single species, it cannot accurately predict the behavior

42   of consortia. There have been a few attempts to expand the applicability of these models to

43   communities (3, 14, 15), but these models require assumptions that oversimplify the system, such

44   as no diffusional limitations and identical or static growth rates for the different organisms. The

45   current benchmarks for constraints-based metabolic modeling of microbial consortia are OptCom

46   (16) and d-OptCom (17) (OptCom's dynamic version). These approaches use inner and outer

2

47   linear optimization problems to satisfy species and community level objectives, leveraging the

48   inner solution as a constraint for the outer problem. However, this approach still relies on *a priori*

49   determination of relative objective preference as well as predetermination of both species-level

50   and community-level objectives. Additionally, cells are treated as homogenous spatial groups (13)

51   or homogenous species groups (14), which limits the accurate simulation of cells acting

52   individually, interacting with their environment, and ultimately forming communities. These

53   approaches thus discount the complexity of individual cells forming communities and, instead of

54   acting uniformly with neighbors or species, create dynamic intercellular and inter-environmental

55   reactions (13, 18-20).

56

57   To more accurately model the complexity of community growth, a new modeling approach must

58   be developed. We have developed <u>M</u>ult<u>I</u>scale <u>M</u>ulti<u>O</u>bjective <u>S</u>ystems <u>A</u>nalysis (MIMOSA), an

59   advanced metabolic modeling framework for complex systems. This approach uses a multi-scale

60   multi-paradigm metabolic modeling approach can leverage simple, powerful stoichiometric

61   metabolic models and integrate spatio-temporal tracking of cells, nutrient diffusion, cell-cell

62   interactions and cell-environmental interactions. This approach requires the use of both continuous

63   and discrete variables as well as several different mathematical formalisms to reflect the multilevel

64   behavior in populations. Therefore, we use an agent-based modeling (ABM) framework to allow

65   direct interaction of different levels through the encapsulation of physiological, environmental,

66   and metabolic models. ABM is a bottom-up modeling approach; the model is made up of a set of

67   agents, which are allowed to act independently as long as they follow distinct rules of behavior

68   defined by the user, this allows us to simulate emergent behavior of complex communities that

69   arise from individual agent behaviors (21-24). The system behavior emerges as a result of the

3

70 many (tens, hundreds, thousands, millions) individuals, each following their own behavior rules,

71 living in a defined environment, interacting with each other and the environment (22). The

72 integration of multiple modeling formalisms to represent disparate sub-systems is a trend common

73 in engineering and science domains (25-29) and has recently seen some developments in the

74 systems biology area (5, 30). Agent-based modeling has been previously applied to both

75 intercellular (12, 31) and multi-cellular processes (32, 33) but has not previously been used to

76 model metabolic fluxes. This multi-scale multi-paradigm approach represents a novel method of

77 integrating individuals (through agents) with previously leveraged dFBA formulations (34, 35),

78 thereby discretizing and separating variables for computational efficient solutions with low *a*

79 *priori* knowledge.

80

81 As a proof-of-concept study, we chose to model *Trichodesmium erythraeum*, a filamentous

82 diazotrophic cyanobacterium. *T. erythraeum* is a major contributor to the global nitrogen cycle; it

83 is responsible for fixing an estimated 42% of all marine biological nitrogen (36) and it leaks 20-

84 50% of the nitrogen it fixes (37), providing surrounding organisms with a biologically available

85 nitrogen source. Unlike other diazotrophs, which either spatially or temporally separate the oxygen

86 sensitive nitrogenase enzyme from the water splitting reaction of photosynthesis (oxygen

87 production), *T. erythraeum* is unique because it simultaneously carries out nitrogen and carbon

88 fixation during the day in different cells along the same filament (trichome). Therefore, it is the

89 ideal model system for the development of MIMOSA: it has structurally identical cells that operate

90 in two distinct metabolic modes (photoautotrophic and diazotrophic), a published genome scale

91 model (3), transcriptome data, and a plethora of *in situ* and laboratory data to both train the model

92 and validate predictions. We use this organism as a test-case for the modeling framework and

93   illustrate how it can be used to develop a predictive model that can also be used to investigate

94   cellular physiology by elucidating rules of behavior.

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_2, N_2, 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.
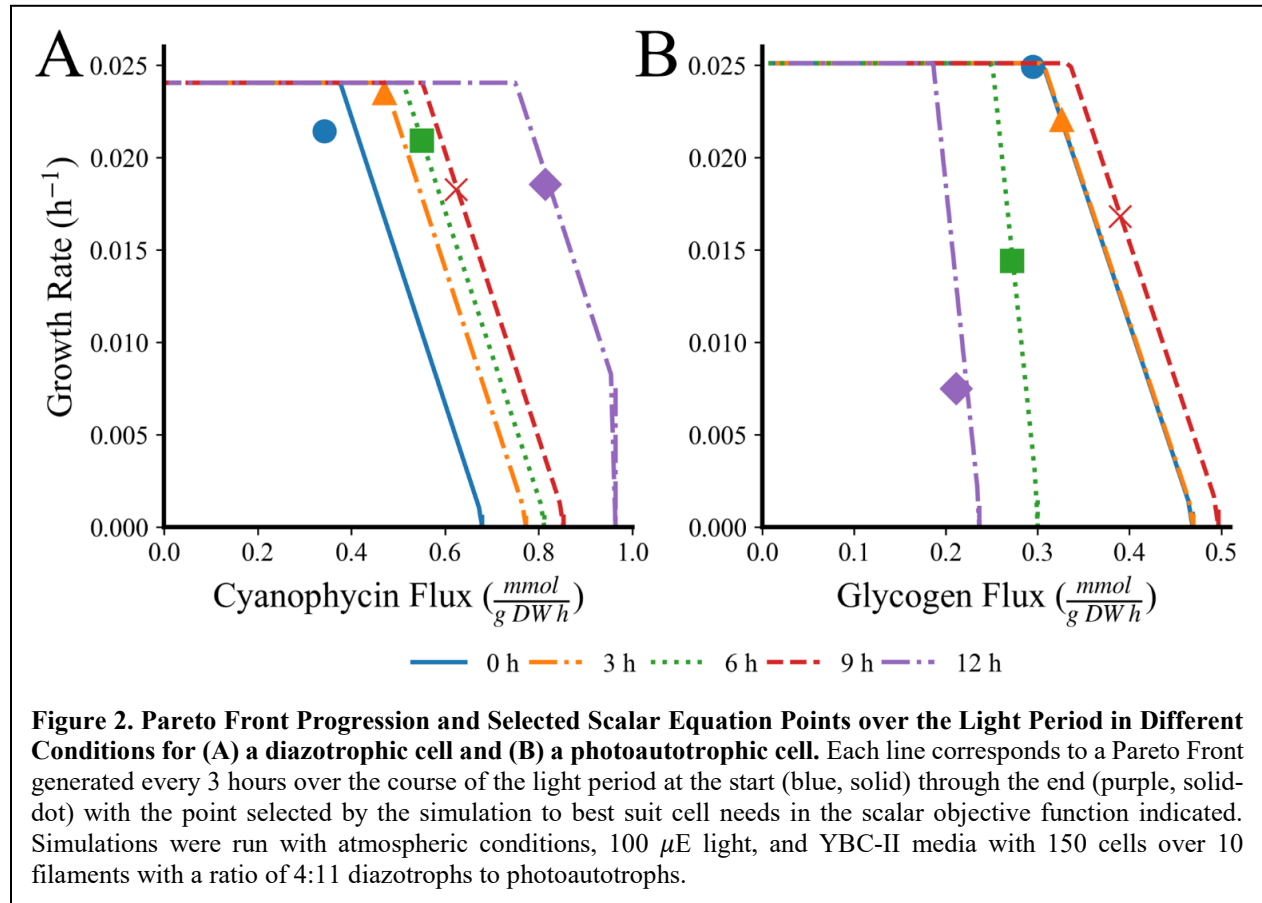
95

5

96 **RESULTS**

97 **Model Formulation.** We developed MIMOSA by integrating an updated version of the genome-

98 scale metabolic model (3) (Table S1 for updated reactions) with nutrient diffusion, light diffusion,

99 cell/cell interaction and cell/environment interactions (see Figure 1) using an agent based modeling

100 framework. We have also implemented the use of multiobjective optimization to account for the

101 dual cellular objective of producing biomass and the metabolite which is transacted between cells

102 (glycogen or β-aspartyl arginine, depending on cell type). Constraints were imposed on the model

103 as reported previously (3) with two notable exceptions. First, the ultimate product of nitrogen

104 fixation was changed from ammonium to β-aspartyl arginine, which is the monomer used to create

105 cyanophycin, a nitrogen storage polymer in *T. erythraeum* and other diazotrophic cyanobacteria

106 (38-40). Second, the two major storage polymers, glycogen (modeled as maltose, or two linked

107 glucoses) and cyanophycin (modeled as β-aspartyl arginine), were decoupled from the biomass

108 formation equation so that they could freely accumulate or be metabolized. More detail about the

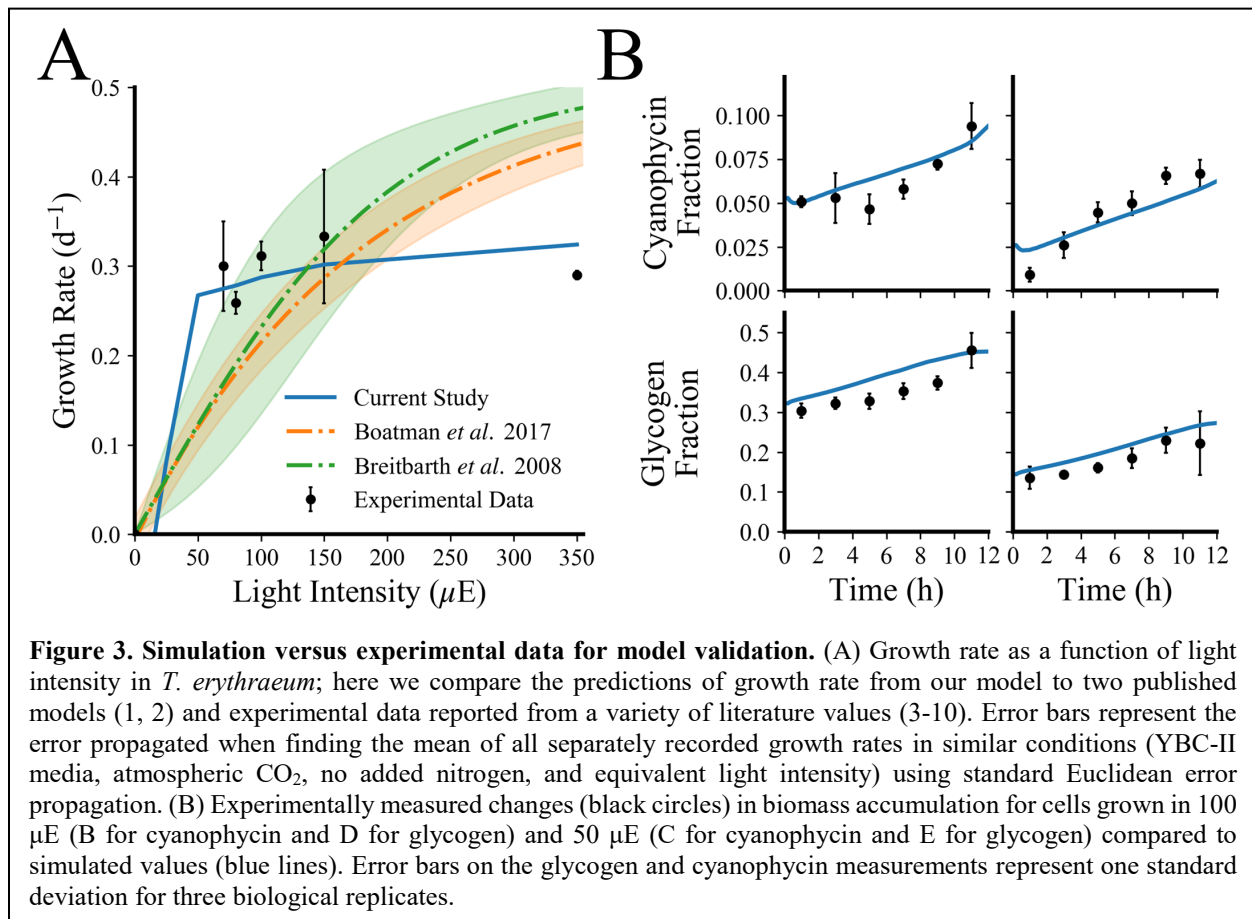109 formulation of the model is provided in Methods and Supplemental Text.

110

111 **Tracking Changing Cellular Objectives.** MIMOSA evaluates the cellular objective for each cell

112 for each time step based on the changing environmental conditions. As an example of this, we

113 have tracked how the Pareto front changes for both photoautotrophic and diazotrophic cells over

114 time (Figure 2). With increasing time, diazotrophs shift their objective away from biomass toward

115 the production of cyanophycin as carbon becomes more available (Figure 2A). In contrast,

116 photoautotrophic cells see a maximum production of glycogen at 9 hours after the onset of light

117 and then their productivity decreases (Figure 2B). It is notable that every cell in the population is

118 performing these decisions in parallel and Figure 2 is for a single representative cell of each cell

6

119   type. Cell optimization changed based on environmental conditions and agent rules and the Pareto

120   Fronts representing this behavior in these contexts is visualized in Fig. S1.
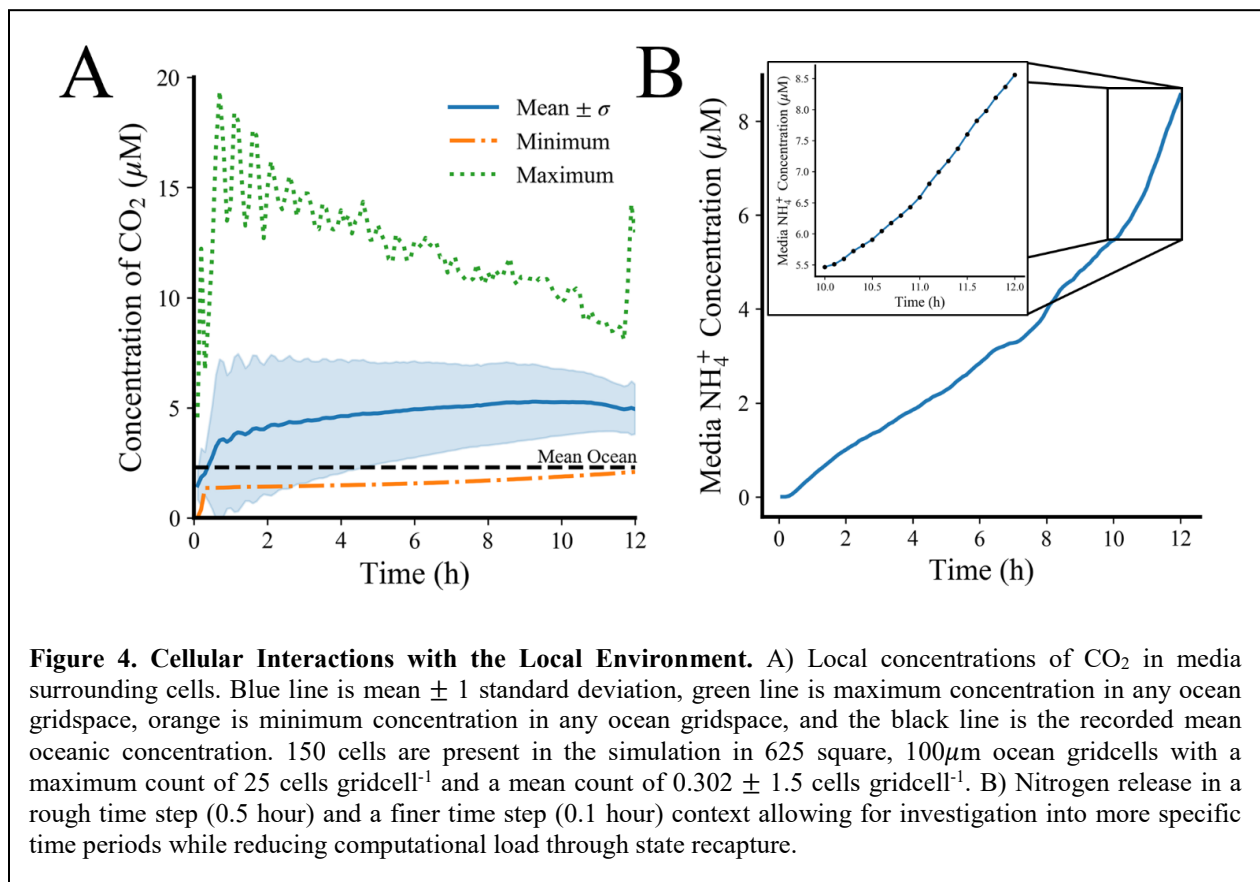
121

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. Simulations were run with atmospheric conditions, 100 μE light, and YBC-II media with 150 cells over 10 filaments with a ratio of 4:11 diazotrophs to photoautotrophs.

122

123   **Model Validation.** In order to test the predictive accuracy of the model, we predicted growth rate

124   for a variety of different light intensities (Figure 3A) and compared to other published models for

125   *T. erythraeum* (1, 2) as well as other experimentally measured growth rates (1, 3, 6, 9, 10, 41)

126   exhibiting light saturation at higher light intensities. Ultimately, our model is a metabolic model,

127   so it is important that it can also capture the metabolic changes that occur in response to changes

128   in the environment. Therefore, we compared predictions of biomass changes to data collected in

129   our laboratory for growth in different light intensities (see Figure 3B). The model was trained on

130   data collected in 100 µE light and was validated with data collect in 50 µE light over a twelve-

131   hour light period.

Figure 3. Simulation versus experimental data for 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 (1, 2) and experimental data reported from a variety of literature values (3-10). Error bars represent the error propagated when finding the mean of all separately recorded growth rates in similar conditions (YBC-II media, atmospheric 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 µE (B for cyanophycin and D for glycogen) and 50 µE (C for cyanophycin and E for glycogen) compared to simulated values (blue lines). Error bars on the glycogen and cyanophycin measurements represent one standard deviation for three biological replicates.

132

133   **Cells Alter Their Microenvironment.** An advantage of the modeling approach we have

134   developed is that we can track nutrients in the environment. Carbon dioxide (CO_2) is typically the

135   limiting substrate in aquatic photosynthetic growth due to low ambient concentrations and low

136   solubility; for ambient CO_2, Henry's law defines an equilibrium concentration of 2.3 µM in the

137   ocean. It is well known that photosynthetic microorganisms use carbon concentrating mechanisms

138   (CCM) to concentrate CO_2 near the carbon fixing enzyme, ribulose-1,5- bisphosphate

139   carboxlyase/oxygenase (RuBisCO) to overcome low selectivity (42); our simulations imply that

140   cells also increase the local concentration of CO_2 immediately surrounding the cell (Figure 4A)

8

141   and the release of nitrogen to the media including at more frequent time steps (Figure 4B). The

142   simulation covers 150 cells and 10 filaments in a model 0.625 mm^3 environment, corresponding

143   to a filament density of 16×10^6 trichomes m^{-3}, well within the *in situ* ranges of free trichome

144   density (43). This illustrates that the simulation corresponds well quantitatively to realistic local

145   environments. At the end of our simulation, the cells on average can create a microenvironment

146   that is roughly 2 fold higher in CO_2 than the surrounding ocean. By looking at flux through major

147   pathways, it appears that the CO_2 is derived from high fluxes through the oxidative PPP and TCA

148   Cycle in diazotrophic cells (Figure 5).

Figure 4. Cellular Interactions with the Local Environment. A) Local concentrations of CO_2 in media surrounding cells. Blue line is mean $\pm$ 1 standard deviation, green line is maximum concentration in any ocean gridspace, orange is minimum concentration in any ocean gridspace, and the black line is the recorded mean oceanic concentration. 150 cells are present in the simulation in 625 square, $100\mu m$ ocean gridcells with a maximum count of 25 cells gridcell^{-1} and a mean count of 0.302 ± 1.5 cells gridcell^{-1}. B) Nitrogen release in a rough time step (0.5 hour) and a finer time step (0.1 hour) context allowing for investigation into more specific time periods while reducing computational load through state recapture.

149

150   **Modeling a Heterogeneous Cell Population.** One of the main advantages of this new modeling

151   approach is that individual cells can be tracked in space and time so the heterogeneity of the

152   population can be quantified (in terms of metabolic flux distributions). As an example, we tracked

9

153   150 cells over a 12-hour time period with time steps of 6 seconds which results in a total of 18,000

154   metabolic flux maps. Since this is an overwhelming amount of data to visualize, we have chosen

155   to focus on a few representative flux maps (see Figure 5). In the left column, we track how the

156   ammonium composition of the environment surrounding the cells changes with time from the

157   initial seeding of cells at 0 hours to the middle of the daytime period (6 hours) to right before the

158   onset of night (12 hours). These panels depict the release of ammonium into the environment as

159   time progresses, and it is higher in areas where the cell density is highest. This agrees well with *in*

160   *situ* data which reports that *T. erythraeum* leaks 30-50% of the nitrogen it fixes (37); our

161   simulations predict that approximately 20% of the nitrogen fixed by the community is excreted

162   into the medium. It is also important to note that the majority of ammonium is released by the cells

163   in the second half of the day; during the first 6 hours, the cells release a total of 1.28 μmoles

164   compared to 4.61 μmoles in the last six hours of the day. Again, this agrees with previous literature

165   reports that the rate of nitrogen fixation peaks at midday (44), therefore we would expect more

166   secretion of ammonium after peak nitrogenase activity. Select flux maps of cells growing in areas

167   of low ammonium (top), medium ammonium (middle) and high ammonium (bottom) are depicted

168   in the middle column of Figure 5. At the beginning of the simulations, cells are seeded in an

169   environment that is identical to the defined marine medium YBC-II and because of this, they have

170   identical flux maps as shown by the distribution graph in the right column. At time 0, we have a

171   bimodal distribution because there are two cell types: photoautotrophic and diazotrophic.

172   Photoautotrophic cells have high flux through the Calvin Cycle and the diazotrophic cells are

173   operating in a more respiratory mode, with high flux through both the oxidative PPP and TCA

174   Cycle. As the cells grow and start to experience more heterogeneity in their environment, they

175   respond by differentiating their metabolism within the filament (Figure S2). First, this is evident

10

176   in the frequency distribution plot, where they are both diverging in terms of total metabolic flux

177   distributions and moving toward achieving optimal flux in terms of the objective function for both

178   t = 6 hours and t = 12 hours. By comparing the changes that occurs in metabolic flux between areas

179   of low, medium, and high ammonium, we can learn a few things about cellular physiology. In all

180   cases, photoautotrophic cells have high flux through the Calvin cycle and an incomplete TCA

181   Cycle, which has been widely reported in cyanobacteria grown phototrophically (45). In the case

182   of *T. erythraeum*, succinic semialdehyde is derived from the nitrogen storage compound

183   cyanophycin and is fed into the TCA Cycle to support the production of biomass precursors and

184   glycogen (through gluconeogenesis). When external ammonium is high, photoautotrophic cells

185   have less flux to glycogen, presumably because they do not need to provide as much to the

186   diazotrophic cells to obtain fixed nitrogen in return. Investigations into imbalances in both

187   metabolites and relative cell quantity display mechanisms of ammonium loss to the environment.

188   Figure S3A illustrates how a lack of glycogen flux results in a higher loss of ammonium (with the

189   exceptions of recently divided cells which metabolize glycogen with high ammonium loss) while

190   Figure S3B visualizes a clear minimum ammonium release in the recorded range of percent

191   diazotrophs per filament (between 15 and 30%). Diazotrophic cells have high flux through both

192   the oxidative PPP and the TCA cycle which still utilizing carbon fixation reactions such as

193   RuBisCO and PEP carboxylase and carbon conserving reactions like the glyoxylate shunt. Flux

194   through the glyoxylate shunt increases as the availability of ammonium increases outside the cell,

195   which is likely in response to the lower glycogen transfer from the photoautotrophs.

11

Figure 5. Examples of the type 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 μm (2.5 mm) and each grid-cell with dimensions of 100 μm (10 *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. Compound abbreviations for flux maps are provided in the "Abbreviations" section.

196

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

12

200   improvement of the model, some rules can be elucidated. One trait that is widely variable in nature

201   is filament length. It has been widely accepted that the average filament length is 100 cells (46)

202   but more recent studies have suggested that they are typically much shorter, with a geometric mean

203   of 13.2 ± 2.3 cells per filament, but with a mean range of 1.2 to 685 cells per filament *in situ* (47).

204   Conditions for *in situ* sampling are widely variable so we hypothesized that filament length plays

205   a role in maintaining growth in different environments: low light, low CO_2 and low N_2. We used

206   the model to investigate which conditions might favor shorter or longer filaments (Figure 6). For

207   each simulation, 150 total cells were seeded but in different trichome lengths (10, 30, 75, and 150

208   cells/filament) with and a ratio of diazotrophs to photoautotrophs of 3:7. In terms of growth rate,

209   across all conditions we tested the shorter filaments had faster growth. This implies that diffusional

210   limitations of nutrients into the cell and metabolites within the filament between different cell types

211   start to hamper growth rate at longer filament lengths. The relative decline in growth rate is less

212   dramatic for 25 μE when comparing across filament length, but when compared to other light

213   conditions, there is a dramatic drop in growth rate for shorter filaments at low light. This indicates

214   that longer filaments are capable of compensating for less light better than shorter filaments,

215   perhaps due to increased surface area. Next, we examined the effect of filament length on

216   cyanophycin composition for the same growth conditions as above. In every condition except low

217   nitrogen, filaments with 75 cells appear to have more cells with above average cyanophycin

218   content than other filaments lengths. Smaller nitrogen compounds (NH_4^+, amino acids, urea, etc.)

219   can theoretically be used to support growth, permitting cyanophycin to be a longer-term storage

220   compound. This is a possible explanation for the increase of cyanophycin in longer filaments. As

221   filaments are longer, diffusive limitations become more pronounced, meaning that nitrogen

222   gradients will remain in nitrogen replete cells longer and will be remade into cyanophycin as

13

223   opposed to being metabolized for growth. This makes intuitive sense: not only is there a final drop-

224   off at 150 cells, the distribution of cyanophycin content within the cells becomes larger, suggesting

225   that some cells are starved for nitrogen and some are nitrogen replete. It is probable that filaments

226   have adapted to leverage diffusion to both sequester nitrogen and to mitigate futile cycling of

227   carbon and nitrogen compounds when diatomic nitrogen is available. The pattern of cyanophycin

228   content diverges for cells in nitrogen limited environments due to overall shortages of nitrogen

229   within the filament.
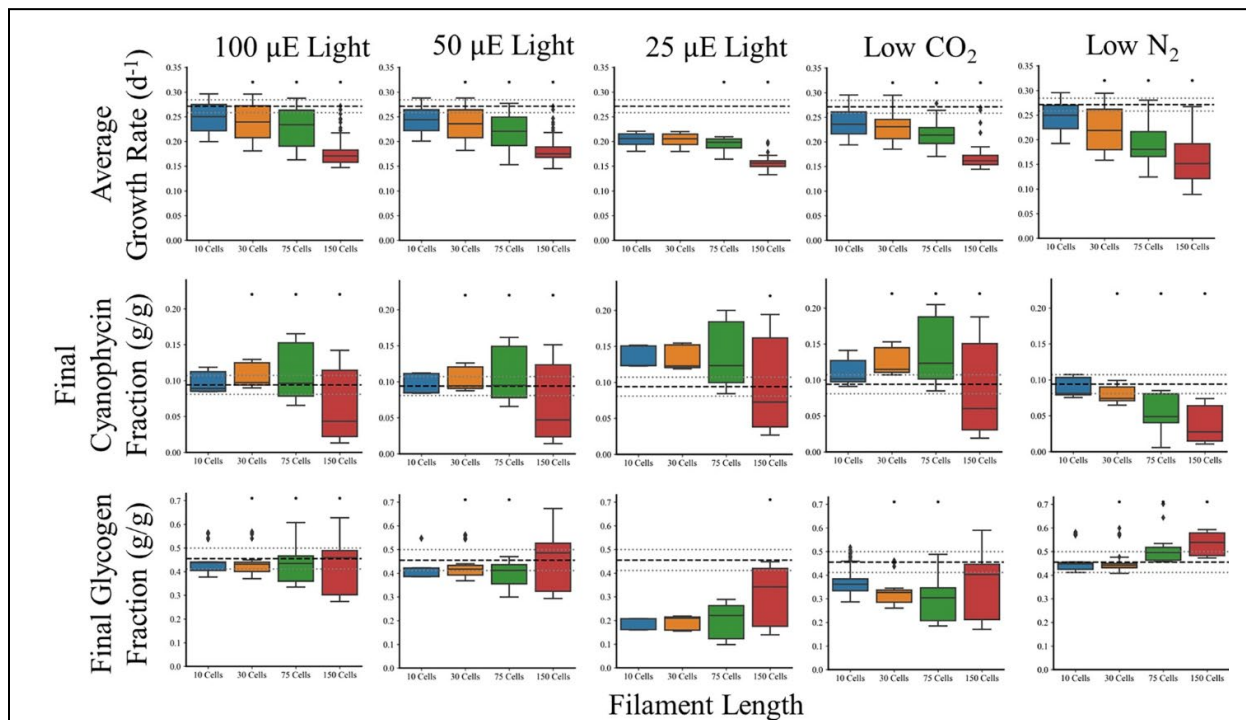
Figure 6. Trichome length affects the performance of the community. Trichomes were varied in initial length from short (10 cells filament^{-1}) to long (150 cells filament^{-1}) but with identical diazotroph: photoautotroph ratios of 3:7 (excepting 75 cells filament^{-1} which was 3.1:7) and initial 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 μE 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 d^{-1} for growth rate, 0.0939 $\pm$ 0.0131 g cyanophycin/g DW, 0.455 $\pm$ 0.0440 g glycogen/g DW). Growth conditions represent the columns progressing as follows: 100 μE light in YBC-II media and atmospheric conditions, 50 μE light in YBC-II media and atmospheric conditions, 25 μE light in YBC-II media and atmospheric conditions, 1.25 mM (versus 2.5 mM) HCO$_3^-$ and 200 ppm (versus 400 ppm) CO$_2$ in otherwise atmospheric conditions and YBC-II media with 100 μE light, and 0.395 atm (versus 0.79 atm) N$_2$ in otherwise atmospheric conditions and YBC-II media with 100 μE.

230

14

231   Finally, we investigated how glycogen content of cells changes due to filament length. The first

232   pattern to note is that as length increases, the heterogeneity of the filament in terms of glycogen

233   content also increases. This illustrates the importance of tracking individual cells because they are

234   experiencing different environments and responding in different ways. Longer filaments also

235   appear to be able to maintain glycogen content more readily than shorter filaments in all stress

236   conditions we tested. Finally, nitrogen limited growth results in increased glycogen content as seen

237   in other cyanobacteria (48). It appears that longer filaments in N limited growth can accumulate

238   more carbon, perhaps again due to higher surface area and hence more energy from light

239   harvesting. Our simulations agree well with published studies; it has been reported that growth

240   rate and light intensity are both inversely correlated to filament length (49). This data indicates

241   that filament length is largely determined by external cues rather than genetically.

242

243   **DISCUSSION**

244   MIMOSA enables the most detailed and accurate metabolic modeling of complex systems to date

245   by allowing coupling of several different mathematical formalisms describing natural phenomena,

246   behavioral rules, and metabolism into a multi-scale multi-paradigm model. In constructing

247   MIMOSA, we have added several features to enable us to more accurately predict phenotypes. A

248   key feature of MIMOSA is the use of a multi-objective optimization approach. Unlike fast growing

249   bacteria, which have successfully been modeled using a single objective function of maximum

250   biomass (50), slow growing organisms have more complex objectives. In our simulations, *T.*

251   *erythraeum* cells must achieve a delicate balance between biomass formation and the production

252   of either glycogen or cyanophycin due to the symbiotic relationship between two cell types in the

253   same filament. Photoautotrophs cannot function optimally without a biologically available form

15

254   of nitrogen from the diazotrophs and the diazotrophs cannot support their metabolism without

255   reduced carbon from the photoautotrophic cells. The use of multi-objective optimization allows us

256   to describe this trade-off more accurately and by calculating the Pareto Front *a priori* we can also

257   reduce computational effort. We have also accounted for changes in biomass composition that

258   occur in response to changes in the environment or as a result of building carbon and nitrogen

259   reserves during the day by decoupling the biomass equation. This allows the model to respond

260   more fluidly to changes in the environment, which more closely mimics what cells experience in

261   nature; for example, macro- and micro-nutrient stresses have been well known to cause changes

262   in metabolism such as lipid and carbon accumulation (51-56). As such, the inclusion of metabolite

263   and nutrient diffusion to augment metabolic optimization is a critical aspect of the model.

264

265   The influences of nutrient and energy availability in conjunction with population characteristics

266   were studied to determine community and cellular adaptations to environmental perturbations. The

267   model allows us to quantify the changes in the microenvironment around the cell compared to the

268   bulk properties of the environment (Figure 4A) as well as to see how these changes affect the

269   distribution of carbon and nitrogen inside the cell (Figure 5). These can be supplemented with

270   "zooming in" on specific time steps to enhance investigation to rapidly occurring phenomena

271   (Figure 4B). Not only did our predicted growth rates quantitatively match the experimental data,

272   it was better able to capture effect of light saturation on growth rate; light intensities above 100 μE

273   have little to no effect on growth rate (2, 49, 57-59). Our simulations agree well with the

274   experimental data, however, there are differences that can be explained by the differences between

275   our experimental conditions and our simulations. The main difference being the effect of diurnal

276   light; *T. erythraeum* will not grow without diurnal day/night patterns, therefore the experimental

16

277   data was collected from cells that were grown in 12 h: 12 h day/night cycles but the model is for a

278   single 12-hour day time period. The addition of diurnal light patterns in future iterations of this

279   model will help to improve the light dependent growth phenotype. Even so, the model is able to

280   visualize community coordination and development during the 12 hour light period, exhibiting the

281   increased release of ammonium to the media in the afternoon, consistent with the observation that

282   nitrogenase activity peaks midday (44). Moreover, the individualized resolution of metabolic

283   optimization can probe the nuances of intercellular, intracellular, and cell-environment

284   interactions. Analysis of metabolic flux reveals a spontaneous partial/linear TCA Cycle in

285   photoautotrophic cells consistent with previous reports (45). Cells also naturally coordinate to

286   provide glycogen and cyanophycin transfer between cells, yielding oxidative behavior in

287   diazotrophic cells through glycolysis with the possible side effect of oxygen consumption as a

288   mechanism to protect nitrogenase as suggested in experimentation (58). Meanwhile,

289   photoautotrophs naturally perform reductive carbon fixation coupled with utilization of the lower

290   TCA Cycle to degrade arginine. These metabolic functions are affected by extracellular forces

291   which are integrated into this model. For example, high ammonium environments result in

292   declining gluconeogenesis in photoautotrophs (12 hours in Figure 5), likely since these cells are

293   energetically limited and use cyanophycin as an energy source instead of light. Diazotrophs are

294   prone to these environmental cues as well as low ammonium environments enhance light TCA

295   Cycle to enhance recycling of amino acid byproducts from a lack of nitrogen. These observations,

296   coupled with the diversity of metabolic profiles available to a relatively small population, By

297   integrating modeling of other phenomenon with constraints based metabolic models, we were able

298   to simulate *T. erythraeum* cultures that more accurately represent both *in situ* and laboratory data.

299

300   One of the many advantages of using this multi-paradigm framework is that we can simulate

301   emergent behavior of a population. *In situ* data reports a wide mean range of trichome length from

302   1.2 to 685 cells (47); we used the model to investigate possible causes because this is a difficult

303   phenotype to investigate experimentally. Our simulations suggest that even though longer

304   filaments suffer from diffusional effects that limit growth, they are better able to handle stress

305   (Figure 6) consistent with literature. Increased surface area in longer filaments minimizes the effect

306   of lower light because the filament can harvest more light per volume. Also, the larger filaments

307   are better able to maintain the average composition of storage compounds despite low carbon or

308   low nitrogen conditions. Therefore, we would expect in areas of nutrient or light stress, the filament

309   length would be longer.

310

311   One of the other unusual phenotypes of *Trichodesmium* that we were able to investigate using

312   MIMOSA was leaking 30-50% of the nitrogen it fixes. Nitrogen fixation is an incredibly energy

313   intensive process, costing the cell 8 ATP per ammonium, so it is not clear why *T. erythraeum*

314   would excrete 30 – 50%. Despite using optimization to solve for fluxes, which should minimize

315   energy losses, our simulations predict approximately 20% of the fixed nitrogen is excreted into the

316   medium (Figure S4) which implies that this is a metabolically driven phenomenon. Further

317   investigation has led us to develop three hypotheses on why this occurs: carbon limitation in

318   diazotrophs, energy limitation in photoautotrophs, and imbalances between photoautotroph:

319   diazotroph ratios. In the first case, photoautotrophs are unable to create glycogen chains and

320   instead must start from a higher energy substrate than carbon dioxide (like succinic-semialdehyde)

321   or must perform glycolysis on arginine derivatives to achieve energetic viability (Figure S3A).

322   Second, population imbalances cause nitrogen to be produced faster than it can be anabolized into

18

323   β-aspartyl-arginine chains and is released into the media, meaning there is an optimal ratio of cell

324   types (Figure S3B). It is also possible that carbon limited diazotrophs are unable to manufacture

325   full β-aspartyl-arginine chains and proton imbalances require ammonium release to the medium

326   instead of passage to surrounding photoautotrophs.

327

328   MIMOSA enables the tracking of cellular-level environmental changes and the impact that they

329   have on a metabolic model, opening the door to more accurate modeling of multi-cellular systems

330   and the *in silico* investigation of the complex interactions between different cell types within an

331   organism, and different species in a community. This is the first report of a metabolic model that

332   integrates nutrient and light diffusion, cell/cell interactions and cell/environment interactions and

333   we have used it to accurately predict growth, cellular composition and to investigate the unique

334   physiology of *T. erythraeum*, which has filaments of both diazotrophs and photoautotrophs in close

335   proximity. It establishes that this organism can effectively adapt to different conditions at three

336   levels: the genetic level through division of labor in separate cell types, the metabolic level through

337   relatively open-ended metabolic capabilities as well as further division within types, and at the

338   population level to harness diffusional and physical interactions with the environment. MIMOSA

339   is also a readily adaptable modeling framework – the addition of additional species to the model

340   only requires the availability of a genome-scale metabolic and a few rules of behavior to be added.

341   While we focused the proof-of-concept study of *T. erythraeum*, MIMOSA is a modeling

342   framework that can be used to model a variety of more complex systems including applications in

343   ecology, human health and metabolic engineering.

344

345   **MATERIALS/METHODS**

19

Cell Culture Conditions

Cells were grown as described previously (3). *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 μE in 12h light/12h dark cycles. Cells were grown in artificial seawater YBC-II medium (60) at pH 8.15-8.20. CO_2 was maintained at atmospheric concentration. All chemicals were obtained from Sigma-Aldrich (St. Louis, MO). Growth rate was monitored by measuring chlorophyll absorbance (61) 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 μm cellulose-nitrate filter and dried overnight at 100°C.

Biomass Quantification

Carbohydrates were measured colorimetrically using the anthrone method (62) against glycogen as a standard. Cyanophycin was extracted by disrupting 740 μL of 250 mL cells concentrated to 2 mL via filtration and rinsing with TE buffer with 2.70 mg/mL lysozyme overnight at 37 ºC, centrifuging at 16,100 x 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 (63).

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:

20

369

$$U = M + G + P + L \tag{1}$$

370   Where U corresponds to uptake, M is the nitrogen or carbon required for maintenance metabolism,

371   G is the accumulation of fixed carbon or nitrogen during growth into non-biomass metabolites, P

372   is the accumulation of fixed carbon or nitrogen that is passed to the other cells, and L is the carbon

373   or nitrogen leaked into the surrounding media. This can be further detailed into carbon and nitrogen

374   energy balances (defined as above with the subscript "N" for nitrogen and "C" for carbon):

$$U_N = 2v_{N_2} + v_{NO_3^-} + v_{NO_2^-} + 2\,v_{urea} \tag{2}$$

$$M_N = -5\frac{\Delta m_{C_{ph},Night}}{N_{C_{ph}}\bar{X}_t\Delta t} \tag{3}$$

$$G_N = \mu Y_{N/X} \tag{4}$$

$$P_N = 5\frac{\Delta m_{C_{ph},Day}}{N_{C_{ph}}\bar{X}_t\Delta t} \tag{5}$$

$$L_N = -\frac{\Delta C_{NH_4^+}}{\bar{X}_t\Delta t} \tag{6}$$

375   Where v corresponds to flux of the substrate (indicated in the subscript), m is mass, N is molar

376   mass, the subscript c_{ph} is cyanophycin, $\bar{X}_t$ is average biomass over the measured time period (Δt),

377   μ is growth rate, $Y_{N/X}$ is the nitrogen stoichiometry in biomass estimated by the biomass

378   composition. The stoichiometric coefficients represent the number of nitrogen atoms in each

379   molecule; 2 per diatomic nitrogen and 5 per β-aspartyl arginine. The flux of nitrogen (v_{N_2}) is

380   measured and recorded via the acetylene assay for nitrogenase activity as recorded in the literature

381   for the same growth conditions. G_N was approximated using the model's prediction for cellular

382   composition of nitrogen using the biomass equation and balanced equations. The faction of

21

383   cyanophycin in biomass was measured analytically at 6 timepoints throughout a single 12-hour

384   daytime period (See *Biomass Quantitation* in methods). Ammonium release over a 12-hour period

385   was below detectable limits (< 1 µg/L) (64) in our laboratory experiments. If we re-arrange this

386   equation to solve for the average flux of nitrogen into a single cell, $\bar{v}_N$, we obtain the following

387   equation which can be used to solve for $\overline{v_N}$ or v_{N_2}:

$$\bar{v}_N = 2v_{N_2} + \frac{\Delta C_{NH_4^+}}{\bar{X}_t \Delta t} = \mu Y_{N/X} - 5\frac{\Delta m_{cph,Night}}{N_{cph}\bar{X}_t\Delta t} + 5\frac{\Delta m_{cph,Day}}{N_{cph}\bar{X}_t\Delta t} \tag{7}$$

388   Assuming each cell requires the same amount of nitrogen, that only diazotrophs reduce diatomic

389   nitrogen, that the average ratio is 4:1 photoautotrophs to diazotrophs (44) for estimation of training

390   data for consumption and production, and that cells do not release ammonium at optimal

391   production, maximum nitrogen flux into a photoautotrophic cell can be approximated as:

$$\bar{v}_{cph} = \frac{2}{5}v_{N_2} \tag{8}$$

392   The same approach is taken for the carbon mass balance.

$$U = M + G + P \tag{9}$$

393   Where:

$$U_C = v_{CO_2} \tag{10}$$

$$M_C = -12\frac{\Delta m_{gly,Night}}{N_{gly}\bar{X}_t\Delta t} \tag{11}$$

$$G_C = \mu Y_{C/X} \tag{12}$$

$$P_C = 12\frac{\Delta m_{gly,Day}}{N_{gly}\bar{X}_t\Delta t} \tag{13}$$

394   And:

22

$$v_{CO_2} = \mu Y_{C/X} - 12\frac{\Delta m_{gly,Night}}{N_{gly}\bar{X}_t \Delta t} + 12\frac{\Delta m_{gly,Day}}{N_{gly}\bar{X}_t \Delta t} \tag{14}$$

395   In this case, the variables are the same except for subscripts C (carbon), gly (glycogen). And CO_2.

396   G_C represents the stoichiometric predictions of elemental composition and v_{CO_2} is approximated

397   using equation 10. This allows prediction of maximal glycogen flux (assuming 12 carbon

398   molecules per glycogen, since it is modeled as disaccharide glucose or maltose) using:

$$\bar{v}_{gly} = \frac{v_{CO_2}}{12} \tag{15}$$

399   **Development of Agent Based Model.** Repast Simphony (65) in Java was used as the agent-based

400   modeling framework in which differentiated multi-objective metabolic models of *Trichodesmium*

401   *erythraeum* are contained. It contains three agent types – Ocean, Cells, and Filaments. The Cells

402   agent contains two sub-agents representing each cell type: photoautotrophs and diazotrophs and is

403   responsible for intracellular processes and decisions. The Ocean agent defines and calculates the

404   extracellular environment and the Filaments agent organizes the Cells and modulates their

405   transactions.

406   *Cells*

407   Cell agents (cells) are generated for each individual cell in the model. These contain two subtypes,

408   photoautotrophs and diazotrophs, but contain several consistent elements between the two.

409   Simulation variables are summarized in Table S4. All cells reproduce according to the same rules:

410   cells divide according to sampling from the weighting distribution described above if that sample

411   is bigger than the cell mean cell, cells only extend from the ends, and cells can only divide into

412   diazotrophs if there is a diazocyte under development (decided at the filament level if the filament

413   is nitrogen limited). When a cell is large enough, it converts to fully stationary growth, producing

414   only metabolites and creating a larger and larger metabolic gradient between cells without *de novo*

23

415   biomass synthesis. This prevents a cell from becoming excessively large in the center of the

416   filament. Cells will die if they cannot produce the requisite maintenance ATP through metabolism

417   or catabolism.

418

419   Cells allow metabolites to diffuse through the lipid bilayer using permeabilities reported in the

420   literature (Table S5). This mechanism represents a non-zero leakage scenario that was nevertheless

421   much slower than intrafilamental diffusion (Table S6). Scavenging from the environment for

422   compounds which carried no evidence of active transport followed these same rules and was

423   therefore prone to concentration gradients. Active transporters, on the other hand, allowed the cell

424   to uptake whatever concentration of compound was necessary subject to its presence in the local

425   ocean grid. Allowable exchange of metabolites between cells is illustrated in Figure 7. If several

426   cells compete in that grid space, access to the available molecule was divided equally among those

427   cells.

Figure 7. 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 β-aspartyl arginine is the foundation for cyanophycin.

428

429   *Subclasses: Photoautotrophs and Diazotrophs*

430   Both subclasses define the uptake constraints and send to a Python file that decides whether the

431   cell metabolizes or catabolizes based on those constraints using the multi-objective metabolic

432   model previously described (see *Supplemental Information: Routine Metabolic Optimizations*).

433   The cell then updates its internal metabolites based on the optimization results, diffuses

434   metabolites, divides if possible, and uptakes from its local environment. Model bounds are

435   calculated using local concentrations to calculate maximum flux bounds excepting β-aspartyl

436   arginine which is further limited to 8% of available nutrients (See Figure S5). These methods are

437   handled by three ScheduledMethods that Repast Simphony schedules in specific progression.

438   Together with the Ocean Agent's updates, the individual cell actions (as dictated by the metabolic

439   model) form the core of the simulation. A more detailed flow chart for cell decision making can

25

440   be found in Figure S6. Progression through these steps is identical for cell types, but the metabolic

441   static variables (objectives, gas uptake, etc.) are different between the two subclasses, necessitating

442   separate methods.

443

444   *Ocean*

445   The ocean agent is responsible for tracking cells and modeling external nutrients. Its main task is

446   facilitating diffusion between cells and locations as well as approximating an uptake radius for

447   cells. Each ocean represents a uniform, static, abstract area of the overall grid space with a uniform

448   dimension space of $\delta \times \delta$ where δ is a user defined parameter. This set of simulations was

449   conducted with time steps of 0.1 hours as a moderate value between diffusion phenomena (on the

450   order of seconds along the length of a filament) and doubling time (on the order of 50 hours).

451   Metabolites are assumed to freely diffuse in a dilute seawater environment between cell filaments

452   (Table S6) and assumed to be uniform within the grid, given the relatively long time step compared

453   to the rate of diffusion over such small dimensions. If the impacts of metabolic diffusion limitations

454   were of interest, the time step within the framework could be made appropriately small to more

455   accurately track metabolites, at cost of increased computational burden. Each ocean gridcell

456   diffuses molecules into its adjacent ocean gridcells assuming discretized slab diffusion in two

457   dimensions. This is done using a previously developed discrete algorithm for diffusion in a grid

458   (66):

459

$$\Delta \bar{f}(x_i) = A \sum_{j=1}^{n} (f(x_i^j) - f(x_i)) e^{-d_j^2/\eta} \qquad (16)$$

460

26

$$d_j = |x_i - x_i^j| \qquad (17)$$

461

$$A \sum_{k=1}^{n} e^{-d_j^2/\eta} = 1 \qquad (18)$$

462

463   Where $\Delta\bar{f}(x_i)$ is the change in concentration of metabolite in grid space x_i over a time step, A is

464   the normalization constant to be calculated by solving the third equation within the entire

465   neighborhood to ensure conservation of mass within the neighborhood, d_j is the distance between

466   grid space x_i and its grid neighbor x_i^j and η is the diffusivity control of the system over the time

467   step. As $\eta \to 0$ diffusion halts and as $\eta \to \infty$ diffusion becomes instantaneous. In this study, $\eta =$

468    $4\mathcal{D}\Delta t$ as in the original Fick's Law.

469

470   Diffusion is calculated using two steps, one forward and one reversing the order of gridspace

471   calculation, to mitigate the effect of order on estimating the concentration gradient (Figure S7).

472   Excess ammonium is secreted into the environment using a membrane diffusion coefficient. Cells

473   are allowed to uptake any metabolite/nutrient in YBC-II medium; the only extracellular products

474   allowed in simulations are small molecules, such as CO_2 and NH_4^+ which diffuse through the

475   membrane, as well as compounds that have experimental evidence of transporters from proteomic

476   analysis or transcriptomic analysis (estimated using membrane diffusion outwardly and free

477   diffusion for gases or active transporters for ions/molecules inwardly) in Table S7 (67, 68).

478

479   The Ocean Agents also manage diffusion of metabolites from marine sinks and through the gas-

480   liquid surface interface with the atmosphere. This is done assuming equilibrium concentrations of

27

481   dissolved gases defined by Henry's Law and mono-directional slab diffusion for CO_2, O_2, and N_2.

482   Table S6 lists the free diffusivities of compounds and Table S8 lists the Henry's Constants for

483   atmospheric compounds, and Figure S7 demonstrates the movement of diffusive molecules

484   through the simulation.

485

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

487   equation:

$$I = I_0 e^{-ky} \tag{19}$$

488   Where I is light intensity, k is the extinction coefficient of light in seawater, and y is the depth

489   below the surface of the individual cell.

490

491   *Filaments*

492   Filament Agents are responsible for organizing cells, managing movement, splitting to promote

493   diazotroph development, and defining cell type after division. Random walk movement (to

494   simulate the lack of control cells have over lateral motion) is simulated by generating a random

495   direction that has an empty grid space for every cell in the filament. Cells move within a user

496   defined interval of time or if their growth is impeded by another filament, in which case growth is

497   halted until the cells move away from each other. The filament forces splitting into two separate

498   filaments when nitrogen is limiting growth and neither filament end is undergoing diazotroph

499   development (meaning that another diazocyte is required). Filament Agents decide the next cell

500   type using this inequality:

501

28

$$\frac{\sum_i \varepsilon_i^{DZ}}{n_{DZ}} > \frac{\sum_i \varepsilon_i^{PA}}{n_{PA}} \tag{20}$$

502

503   Where ε is the Pareto Efficiency of the given cell type and n is the quantity of that cell type in the

504   filament. The Pareto Efficiency is quantified as the sum of the objective fluxes divided by their

505   Pareto Optimum (from experimental results) divided by the number of objectives.

$$\varepsilon_i^c = \sum_j \frac{v_j}{v_j^{exp}} \tag{21}$$

506   If inequality (20) is satisfied, the cell prioritizes diazotroph development, otherwise it prioritizes

507   photoautotroph development. If a diazotroph region is currently under development, the filament

508   adds another cell to that region. If there is no diazotroph under development, or if the C:N ratio

509   becomes higher than physiological bounds, the filament splits to expose a region where diazotroph

510   development may begin. A photoautotroph can be placed at any open site. Since there are two ends

511   on every filament, up to two of these decisions are being made during each simulation time step.

512   After filament splitting, if the split results in a homogenous region of either diazotrophs or

513   photoautotrophs, the missing cell type is preferred. Filaments split in the middle of the longest

514   region of homogenous cells and are prevented from splitting to result in a single cell, meaning that

515   the shortest possible resulting splits are two cells in length. Cell division completes within one

516   time step when metabolites and biomass are equally divided between parent and daughter cell and

517   the filament updates to contain the cell at its end. This decision is a memoryless process conducted

518   each time step. This means that cell division is completely metabolically motivated (which is

519   affected, in turn, by diffusion and physiological processes).

520

521 **Parameter Estimation.** As described previously, to improve the accuracy of simulations, the

522 model was fit to experimental data for cells grown in 100 μE light in YBC-II medium. Maintenance

523 energy, in the form of the ATP hydrolysis reaction, is the main parameter that is adjusted in FBA

524 formulations (3, 69-72) to match simulations with growth rate. Since maintenance energy at 100

525 μE was higher than the energetic capacity of the model for growth at 50 μE, a linear correlation

526 was interpolated from experiments at 100 μE and 80 μE with ATP maintenance flux fit to both

527 cases for each cell type:

$$v_{ATP}(I) = mI + v_0 \tag{22}$$

528 Where m and v_0 are calculated using the point-slope equation for a linear equation:

$$m = \frac{v_{ATP,I_1} - v_{ATP,I_2}}{I_1 - I_2} \tag{23}$$

$$v_0 = -I_1 m + v_{ATP,I_1} \tag{24}$$

529 Where I_1 is 100 μE and I_2 is 80 μE. The estimated values of the linear equation are recorded below

530 in Table 1. If the model is unable to satisfy its maintenance demand (through any metabolic

531 process, including catabolizing its own biomass), the cell dies. L_0 is the energy required in zero

532 light to maintain the cell without active metabolism.

533

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

Cell Type	Maintenance Energy, v_{ATP}		m $\left(\frac{mmol}{g\ DW\ h\ μE}\right)$	v_0 $\left(\frac{mmol}{g\ DW\ h}\right)$	L_0 (μE)
	80 μE $\left(\frac{mmol}{g\ DW\ h}\right)$	100 μE $\left(\frac{mmol}{g\ DW\ h}\right)$			
Photoautotroph	34.3	53.3	0.952	-16.7	16.9
Diazotroph	62.3	82.0	0.987	-41.9	44.0

30

534

535   **Multi-Objective Optimization.** Unlike typical formulations of flux balance analysis (14, 71-76),

536   which use a single objective function to predict fluxes, our model uses multi-objective

537   optimization to more accurately approximate the true objectives of the cell: to optimize biomass

538   while also producing the metabolite they exchange between cell types. Implementation of multi-

539   objective optimization is more complex and computationally intense than single objective

540   optimization therefore, to minimize computational effort, Pareto Fronts were generated *a priori* by

541   iteratively increasing ATP maintenance flux and using every permutation of objective weights to

542   fit to a dominant front (see *SI Methods* and Figure S8 for more complete details). For each point

543   along the Pareto Front, Euclidean distance was used to determine the relative weight of each

544   objective function, which was then used to generate a single, scalarized reaction. Each cell in the

545   simulation calculates its scalar objective function separately during each time step based on its

546   internal constitution and requirements.

547

548   **Implementation of Mutable Objective Functions.** Previous studies have used static objective

549   functions, where production is consistent during every phase of growth. However, organisms

550   accumulate and digest metabolites during growth and development. To reflect this, we inserted a

551   "mutable" objective function where relative preferences of storage compounds and biomass

552   production can be tailored by the agent based on cell biomass. The scalarized objective equation

553   was thus broken into two main components: storage compounds (cyanophycin modeled as β-

554   aspartyl arginine and glycogen modeled as maltose) and biomass (lipids, proteins, DNA, RNA,

555   chlorophyll, phycoerythrin, etc.). We assumed that biomass remained relatively stable throughout

556   the day while the amount of storage compound was allowed to vary. The scalar weights, or

31

557   production priorities, were manipulated assuming cells do not grow beyond twice their average

558   cell without dividing: lower biomass prioritizes growth and higher biomass prioritizes vegetative

559   storage compound production. Mathematically, this is modeled such that the scalar objective

560   equation's biomass coefficient was inversely adjusted by cumulative probability of a cell's

561   biomass in the distribution. The normal distribution was formulated assuming cubic 10 μm cells

562   with density of water (77) as the average mass and a narrow distribution with a standard deviation

563   of 0.433 times the mean size. This value was chosen to promote switch-like bistable behavior

564   between cell phenotypes: either cells are biomass driven (exponential) or they are metabolite

565   driven with combinations of probabilities in between. This is because a single sample of a cell

566   from a distribution of cells would have a probability of 99% to fall between 0 and twice the mean

567   size. The final distribution is:

$$f(X) \sim N(1.029\ ng, 0.433 \cdot 1.029\ ng) \tag{25}$$

568   Calculation of new objective coefficients was done by first finding the cumulative probability (z)

569   of another randomly selected cell's non-metabolite biomass being less than or equal to the

570   objective cell's biomass at each time point for each cell:

$$z = F(X \leq x_i) \tag{26}$$

571   This is used to adjust the average, experimentally matched objective coefficient ($\overline{w}_b$) for biomass

572   by multiplying that coefficient by the probability of the cell being larger than that size, a value that

573   represents the probabilistic expansion space (ϵ) of the cell:

$$\epsilon = 1 - z \tag{27}$$

$$\widehat{w}_b = \epsilon \cdot \overline{w}_b \tag{28}$$

32

574 Major metabolite coefficients for the scalarized objective equation were also adjusted using this

575 probability, increasing as the cell's size increased:

$$\widehat{w}_m = z \cdot \overline{w}_m \tag{29}$$

576 Finally, the coefficients are normalized such that:

$$a \sum_i \widehat{w}_i = 1 \tag{30}$$

577 Or:

$$a = \frac{1}{\sum_i \widehat{w}_i} \tag{31}$$

578 Which yields final objective coefficients of:

$$w_k = \frac{\widehat{w}_k}{\sum_i \widehat{w}_i} \ \forall \ k \in \boldsymbol{\mathcal{O}} \tag{32}$$

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

580 Performance evaluation of the mutable objective function, validation of the mutable objective

581 function versus the static version, and justification of non-metabolite biomass as the independent

582 objective are provided in *SI Methods* and Figure S9.

583

584 **ABBREVIATIONS.** 6PG: 6-phospho-D-gluconate, 6PGDL: 6-phosph-D-glucono-1,5-lactone,

585 ABM: Agent-Based Modeling, AcCoA: Acetyl-CoA, AKG/αKG: α-ketoglutarate/2-oxoglutarate,

586 ALA: L-alanine, βASP-ARG, β-aspartyl arginine, βG6P: β-glucose-6-phosphate, CBB: Calvin-

587 Benson-Bassham Cycle, CDeg: Cyanophycin Degradation to Amino Acids, cEFMA: community

588 Elementary Flux Mode Analysis, CIT: Citrate, COBRA: Constraint-Based Reconstruction and

589 Analysis Toolbox, CobraPy: Constraint-Based Reconstruction and Analysis Toolbox for Python,

590 COMETS: Computation of Microbial Ecosystems in Time and Space, CSV, Comma Separated

591   Value, cph: Cyanophycin, CSyn: Cyanophycin Synthesis from Amino Acids, dFBA: dynamic Flux

592   Balance Analysis, DHAP: Dihydroxyacetone phosphate, DON: Dissolved Organic Nitrogen,

593   E4P: Erythrose-4-phosphate, EF: Efflux, EX: Export, F6P: Fructose-6-phosphate, FBA: Flux

594   Balance Analysis, FDP: Fructose 1,6-diphosphate, FOR: Formate, FUM: Fumarate, FVA: Flux

595   Variability Analysis, G6P: Glucose-6-phosphate, GAP: Glyceraldehyde 3-phosphate, GDeg:

596   Glycogen degradation, GLX: Glyoxylate (flux maps) or Glyoxylate and dicarboxylate metabolism

597   (Figure 4), GLY: Glycine (flux maps) or Glycogen/Gluconeogenesis (Figure 4), gly: Glycogen,

598   GLYR: Glycerate, GOL: Glycerol, GP: 3-phosphoglycerate, GSyn: Glycogen synthesis,

599   ICIT: Isocitrate, IN: Influx, JSON: JavaScript Object Notation, jyCOBRA: java-python integrated

600   COBRA, KEGG: Kyoto Encyclopedia of Genes and Genomes, LIP: Lipid metabolism,

601   MAL: Malate, MSM: Multiscale Modeling, OAA: Oxaloacetate, PEP: Phosphoenolpyruvate,

602   PGOL: Phosphoglycolate, Pi: Inorganic phosphate, PPP: Oxidative Pentose Phosphate Pathway,

603   PRO: Protein synthesis, PYR: Pyruvate, R5P: Ribose-5-phosphate, REF: Reflux, Ru5P: Ribulose-

604   5-phosphate, RuBP: Ribulose 1,5-bisphosphate, S17P: Sedoheptulose 1,7-bisphosphate,

605   S7P: Sedoheptulose 7-phosphate, SBML: Systems Biology Markup Language, SUCC: Succinate,

606   SUCSAL: Succinic semialdehyde, SDA: Subsystem Distribution Analysis, TCA: Tricarboxylic

607   acid cycle, TCP/IP: Transmission Control Protocol/Internet Protocol, X5P: Xylulose 5-phosphate

608   **DECLARATIONS**

609   ***Competing Interests***

610   The authors declare that they have no competing interests.

611   ***Funding***

612   This work was supported by a grant from the Department of Energy Office of Science, Biological

613   and Environmental Research (BER) Early Career Program grant no. DE-SC0019171.

614   ***Availability of data and materials***

615   The ABM framework and datasets resulting/analyzed from simulations are available on the

616   GitHub repository at https://github.com/boylelab/iTery101-ABM with username "boylelab" and

617   password "diazo101". We will make this public upon publication.

618   ***Author contributions***

619   JJG, BMSH and NRB designed the research. JJG performed the research. JJG and NRB analyzed

620   the data. JJG, BMSH and NRB wrote the manuscript.

621   ***Acknowledgements***

622   Not applicable.

623

624   **REFERENCES**

625   1.   Boatman TG, Lawson T, Geider RJ. A key marine diazotroph in a changing ocean: The
626   interacting effects of temperature, CO2 and light on the growth of Trichodesmium erythraeum
627   IMS101. PloS one. 2017;12(1):e0168796.
628   2.   Breitbarth E, Wohlers J, Kläs J, LaRoche J, Peeken I. Nitrogen fixation and growth rates
629   of Trichodesmium IMS-101 as a function of light intensity. Marine Ecology Progress Series.
630   2008;359:25-36.
631   3.   Gardner JJ, Boyle NR. The use of genome-scale metabolic network reconstruction to
632   predict fluxes and equilibrium composition of N-fixing versus C-fixing cells in a diazotrophic
633   cyanobacterium, Trichodesmium erythraeum. BMC systems biology. 2017;11(1):4.
634   4.   Blazeck J, Alper H. Systems metabolic engineering: Genome-scale models and beyond.
635   Biotechnology Journal. 2010;5(7):647-59.
636   5.   Tenazinha N, Vinga S. A Survey on Methods for Modeling and Analyzing Integrated
637   Biological Networks. Computational Biology and Bioinformatics, IEEE/ACM Transactions on.
638   2011;8(4):943-58.
639   6.   Kranz SA, Dieter S, Richter K-U, Rost B. Carbon acquisition by Trichodesmium: the effect
640   of pCO2 and diurnal changes. Limnology and Oceanography. 2009;54(2):548-59.
641   7.   Flynn KJ. Ecological modelling in a sea of variable stoichiometry: dysfunctionality and the
642   legacy of Redfield and Monod. Progress in Oceanography. 2010;84(1-2):52-65.
643   8.   Mahadevan R, Schilling C. The effects of alternate optimal solutions in constraint-based
644   genome-scale metabolic models. Metabolic engineering. 2003;5(4):264-76.
645   9.   Hutchins D, Fu F-X, Zhang Y, Warner M, Feng Y, Portune K, et al. CO_2 control of
646   *Trichodesmium* N_2 fixation, photosynthesis, growth rates, and elemental ratios: Implications for
647   past, present, and future ocean biogeochemistry. Limnology and Oceanography. 2007;52(4):1293-
648   304.
649   10.   Spungin D, Berman-Frank I, Levitan O. Trichodesmium's strategies to alleviate
650   phosphorus limitation in the future acidified oceans. Environmental microbiology.
651   2014;16(6):1935-47.
652   11.   Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, et al. Metabolic
653   modeling of microbial strains in silico. Trends Biochem Sci. 2001;26.
654   12.   Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival Jr B, et al. A whole-
655   cell computational model predicts phenotype from genotype. Cell. 2012;150(2):389-401.
656   13.   Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, et al. Metabolic
657   resource allocation in individual microbes determines ecosystem interactions and spatial
658   dynamics. Cell reports. 2014;7(4):1104-15.
659   14.   Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in
660   *Escherichia coli*. Biophysical journal. 2002;83(3):1331-40.
661   15.   Zhuang K, Ma E, Lovley DR, Mahadevan R. The design of long-term effective uranium
662   bioremediation strategy using a community metabolic model. Biotechnology and bioengineering.
663   2012;109(10):2475-83.
664   16.   Zomorrodi AR, Maranas CD. OptCom: a multi-level optimization framework for the
665   metabolic modeling and analysis of microbial communities. PLoS Comput Biol.
666   2012;8(2):e1002363.
667   17.   Zomorrodi AR, Islam MM, Maranas CD. d-OptCom: dynamic multi-level and multi-
668   objective metabolic modeling of microbial communities. ACS synthetic biology. 2014;3(4):247-
669   57.

36

670   18.   Oberhardt MA, Zarecki R, Gronow S, Lang E, Klenk H-P, Gophna U, et al. Harnessing the
671   landscape of microbial culture media to predict new organism–media pairings. Nature
672   communications. 2015;6:8493.
673   19.   Oberhardt M, Palsson B, Papin J. Applications of genome-scale metabolic reconstructions.
674   Mol Syst Biol. 2009;5:320.
675   20.   Fiegna F, Moreno-Letelier A, Bell T, Barraclough TG. Evolution of species interactions
676   determines microbial community productivity in new environments. The ISME journal.
677   2015;9(5):1235.
678   21.   Borshchev A, Filippov A, editors. From system dynamics and discrete event to practical
679   agent based modeling: reasons, techniques, tools. Proceedings of the 22nd international conference
680   of the system dynamics society; 2004: Citeseer.
681   22.   Klann M, Lapin A, Reuss M. Agent-based simulation of reactions in the crowded and
682   structured intracellular environment: Influence of mobility and location of the reactants. BMC
683   Systems Biology. 2011;5(1):71.
684   23.   Segovia-Juarez JL, Ganguli S, Kirschner D. Identifying control mechanisms of granuloma
685   formation during *M. tuberculosis* infection using an agent-based model. Journal of Theoretical
686   Biology. 2004;231(3):357-76.
687   24.   Parunak HVD. Practical and industrial applications of agent-based systems. Environmental
688   Research Institute of Michigan (ERIM). 1998.
689   25.   Bin C, Gang G, Xiaogang Q, editors. The Application of Multi-paradigm Modeling in
690   Social Computation. Digital Manufacturing and Automation (ICDMA), 2013 Fourth International
691   Conference on; 2013 29-30 June 2013.
692   26.   Hardebolle C, Boulanger F. Exploring Multi-Paradigm Modeling Techniques.
693   SIMULATION. 2009;85(11-12):688-708.
694   27.   Hodge B-M, Shukla A, Huang S, Reklaitis G, Venkatasubramanian V, Pekny J. Multi-
695   Paradigm Modeling of the Effects of PHEV Adoption on Electric Utility Usage Levels and
696   Emissions. Industrial & Engineering Chemistry Research. 2011;50(9):5191-203.
697   28.   Hodge B-MS, Huang S, Siirola JD, Pekny JF, Reklaitis GV. A multi-paradigm modeling
698   framework for energy systems simulation and analysis. Computers & Chemical Engineering.
699   2011;35(9):1725-37.
700   29.   Mosterman PJ, Vangheluwe H. Computer Automated Multi-Paradigm Modeling: An
701   Introduction. SIMULATION. 2004;80(9):433-50.
702   30.   Machado D, Costa R, Rocha M, Ferreira E, Tidor B, Rocha I. Modeling Formalisms in
703   Systems Biology. AMB Express. 2011;1(1):45.
704   31.   Simons M, Misra A, Sriram G. Genome-Scale Models of Plant Metabolism. In: Sriram G,
705   editor. Plant Metabolism. Methods in Molecular Biology. 1083: Humana Press; 2014. p. 213-30.
706   32.   Heinken A, Sahoo S, Fleming RM, Thiele I. Systems-level characterization of a host-
707   microbe metabolic symbiosis in the mammalian gut. Gut microbes. 2013;4(1):28-40.
708   33.   Libourel IGL, Shachar-Hill Y. Metabolic Flux Analysis in Plants: From Intelligent Design
709   to Rational Engineering. Annual Review of Plant Biology. 2008;59(1):625-50.
710   34.   Mahadevan R, Edwards JS, Doyle Iii FJ. Dynamic Flux Balance Analysis of Diauxic
711   Growth in *Escherichia coli*. Biophysical Journal. 2002;83(3):1331-40.
712   35.   Zhuang K, Ma E, Lovley DR, Mahadevan R. The design of long-term effective uranium
713   bioremediation strategy using a community metabolic model. Biotechnol Bioeng. 2012;109.
714   36.   Berman-Frank I, Lundgren P, Falkowski P. Nitrogen fixation and photosynthetic oxygen
715   evolution in cyanobacteria. Research in Microbiology. 2003;154:157-64.

716 37.   Glibert PM, Bronk DA. Release of Dissolved Organic Nitrogen by Marine Diazotrophic
717 Cyanobacteria, *Trichodesmium* spp. Applied and Environmental Microbiology.
718 1994;60(11):3996-4000.
719 38.   Burnat M, Herrero A, Flores E. Compartmentalized cyanophycin metabolism in the
720 diazotrophic filaments of a heterocyst-forming cyanobacterium. Proceedings of the National
721 Academy of Sciences. 2014;111(10):3823-8.
722 39.   Sherman DM, Tucker D, Sherman LA. Heterocyst development and localization of
723 cyanophycin in N_2-fixing cultures of *Anabaena* sp. PCC 7120 (cyanobacteria). Journal of
724 Phycology. 2000;36(5):932-41.
725 40.   Simon RD. Cyanophycin granules from the blue-green alga *Anabaena cylindrica*: a reserve
726 material consisting of copolymers of aspartic acid and arginine. Proceedings of the national
727 academy of sciences. 1971;68(2):265-7.
728 41.   Levitan O, Brown CM, Sudhaus S, Campbell D, LaRoche J, Berman-Frank I. Regulation
729 of nitrogen metabolism in the marine diazotroph Trichodesmium IMS101 under varying
730 temperatures and atmospheric CO2 concentrations. Environmental microbiology.
731 2010;12(7):1899-912.
732 42.   Eichner M, Thoms S, Kranz SA, Rost B. Cellular inorganic carbon fluxes in
733 Trichodesmium: a combined approach using measurements and modelling. Journal of
734 experimental botany. 2014;66(3):749-59.
735 43.   Buitenhuis E, Vogt M, Moriarty R, Bednarsek N, Doney S, Leblanc K, et al. MAREDAT:
736 towards a world atlas of MARine Ecosystem DATa. Earth System Science Data. 2013;5:227-39.
737 44.   Berman-Frank I, Lundgren P, Chen Y-B, Küpper H, Kolber Z, Bergman B, et al.
738 Segregation of Nitrogen Fixation and Oxygenic Photosynthesis in the Marine Cyanobacterium
739 *Trichodesmium*. Science. 2001;294(5546):1534-7.
740 45.   Zhang S, Bryant DA. The Tricarboxylic Acid Cycle in Cyanobacteria. Science.
741 2011;334(6062):1551-3.
742 46.   Luo Y, Doney S, Anderson L, Benavides M, Berman-Frank I, Bode A, et al. Database of
743 Diazotrophs in Global Ocean: Abundance, Biomass, and Nitrogen Fixation Rates. Earth System
744 Science Data. 2012;4(1).
745 47.   White AE, Watkins-Brandt KS, Church MJ. Temporal Variability of Trichodesmium spp.
746 and Diatom-Diazotroph Assemblages in the North Pacific Subtropical Gyre. Frontiers in Marine
747 Science. 2018;5:27.
748 48.   Yoo S-H, Keppel C, Spalding M, Jane J-l. Effects of growth condition on the structure of
749 glycogen produced in cyanobacterium Synechocystis sp. PCC6803. International journal of
750 biological macromolecules. 2007;40(5):498-504.
751 49.   Cai X, Gao K, Fu F, Campbell DA, Beardall J, Hutchins DA. Electron transport kinetics in
752 the diazotrophic cyanobacterium Trichodesmium spp. grown across a range of light levels.
753 Photosynthesis research. 2015;124(1):45-56.
754 50.   Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting
755 intracellular fluxes in Escherichia coli. Mol Syst Biol. 2007;3:119.
756 51.   Blaby IK, Glaesener AG, Mettler T, Fitz-Gibbon ST, Gallaher SD, Liu B, et al. Systems-
757 Level Analysis of Nitrogen Starvation–Induced Modifications of Carbon Metabolism in a
758 *Chlamydomonas reinhardtii* Starchless Mutant. The Plant Cell Online. 2013;25(11):4305-23.
759 52.   Boyle NR, Page MD, Liu B, Blaby IK, Casero D, Kropat J, et al. Three Acyltransferases
760 and Nitrogen-responsive Regulator Are Implicated in Nitrogen Starvation-induced Triacylglycerol
761 Accumulation in *Chlamydomonas*. Journal of Biological Chemistry. 2012;287(19):15811-25.

762   53.   Breuer G, Lamers PP, Martens DE, Draaisma RB, Wijffels RH. The impact of nitrogen
763   starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. Bioresource
764   Technology. 2012;124:217-26.

765   54.   Hockin NL, Mock T, Mulholland F, Kopriva S, Malin G. The Response of Diatom Central
766   Carbon Metabolism to Nitrogen Starvation Is Different from That of Green Algae and Higher
767   Plants. Plant Physiology. 2012;158(1):299-312.

768   55.   Tedesco MA, Duerr EO. Light, temperature and nitrogen starvation effects on the total
769   lipid and fatty acid content and composition of *Spirulina platensis* UTEX 1928. Journal of Applied
770   Phycology. 1989;1(3):201-9.

771   56.   Kropat J, Hong-Hermesdorf A, Casero D, Ent P, Castruita M, Pellegrini M, et al. A revised
772   mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*.
773   The Plant journal : for cell and molecular biology. 2011;66(5):770-80.

774   57.   Boatman TG, Davey PA, Lawson T, Geider RJ. The physiological cost of diazotrophy for
775   Trichodesmium erythraeum IMS101. PloS one. 2018;13(4):e0195638.

776   58.   Boatman TG, Davey PA, Lawson T, Geider RJ. CO2 modulation of the rates of
777   photosynthesis and light-dependent O2 consumption in Trichodesmium. Journal of experimental
778   botany. 2018;70(2):589-97.

779   59.   Boatman TG, Mangan NM, Lawson T, Geider RJ. Inorganic carbon and pH dependency
780   of photosynthetic rates in Trichodesmium. Journal of experimental botany. 2018:ery141.

781   60.   Chen Y-B, Zehr JP, Mellon M. Growth and Nitrogen Fixation of the Diazotrophic
782   Filamentous Nonheterocytous Cyanobacterium *Trichodesmium* sp. IMS 101 in Defined Media:
783   Evidence for a Circadian Rhythm. Journal of Phycology. 1996;32(6):916-23.

784   61.   Harris EH, Stern DB, Witman G. The *Chlamydomonas* sourcebook: Cambridge Univ
785   Press; 2009.

786   62.   Yemm E, Willis A. The estimation of carbohydrates in plant extracts by anthrone.
787   Biochemical journal. 1954;57(3):508.

788   63.   Messineo L. Modification of the Sakaguchi reaction: spectrophotometric determination of
789   arginine in proteins without previous hydrolysis. Archives of Biochemistry and Biophysics.
790   1966;117(3):534-40.

791   64.   Liddicoat M, Tibhitts S, Butler E. The determination of ammonia in seawater. Limnology
792   and Oceanography. 1975;20(1):131-2.

793   65.   Collier N, North M. Parallel agent-based simulation with repast for high performance
794   computing. Simulation. 2013;89(10):1215-35.

795   66.   Grajdeanu A. Modeling Diffusion in a Discrete Environment. George Mason University
796   Technical Report Series. 2007:1-5.

797   67.   Sandh G, Ran L, Xu L, Sundqvist G, Bulone V, Bergman B. Comparative proteomic
798   profiles of the marine cyanobacterium *Trichodesmium erythraeum* IMS101 under different
799   nitrogen regimes. Proteomics. 2011;11(3):406-19.

800   68.   Pfreundt U, Kopf M, Belkin N, Berman-Frank I, Hess WR. The primary transcriptome of
801   the marine diazotroph *Trichodesmium erythraeum* IMS101. Scientific reports. 2014;4.

802   69.   Boyle NR, Shastri AA, Morgan JA. Network Stoichiometry In: Schwender J, editor. Plant
803   Metabolic Networks. New York: Springer New York 2009. p. 211-43.

804   70.   Boyle N, Morgan J. Flux balance analysis of primary metabolism in *Chlamydomonas*
805   *reinhardtii*. BMC Systems Biology. 2009;3(1):4.

806   71.   Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S, Lockau W, et al. Flux Balance
807   Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis* sp. PCC 6803.
808   PLoS Comput Biol. 2013;9(6):e1003081.
809   72.   Misra A, Conway MF, Johnnie J, Qureshi TM, Derrick AM, Agbo EC, et al. Metabolic
810   analyses elucidate nontrivial gene targets for amplifying dihydroartemisinic acid production in
811   yeast. Frontiers in Microbiology. 2013;4.
812   73.   Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotech. 2010;28(3):245-
813   8.
814   74.   Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. Current Opinion
815   in Biotechnology. 2003;14(5):491-6.
816   75.   Edwards J, Ramakrishna R, Schilling C, Palsson B. Metabolic flux balance analysis.
817   Metabolic engineering. 1999.
818   76.   Asadollahi MA, Maury J, Patil KR, Schalk M, Clark A, Nielsen J. Enhancing sesquiterpene
819   production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. Metabolic
820   Engineering. 2009;11(6):328-34.
821   77.   van Baalen C, Brown Jr RM. The ultrastructure of the marine blue green alga,
822   *Trichodesmium erythraeum*, with special reference to the cell wall, gas vacuoles, and cylindrical
823   bodies. Archiv für Mikrobiologie. 1969;69(1):79-91.

824