Supplementary Material
Fiore CF, Longnecker K, Kido Soule MC, Kujawinski EB. Release of ecologically relevant metabolites by the cyanobacterium Synechococcus elongatus CCMP 1631 *Environmental Microbiology*

Methods

Culture of Synechococcus elongatus *CCMP 1631*

The filters containing DAPI-stained samples from the cultures (see main text) were checked for bacterial contamination on a Zeiss Axiostar plus microscope, with a mercury vapor short-arc lamp (HBO 50), where 25-35 fields or at least 150 cells were counted. Contamination by heterotrophic cells was determined by alternating between fluorescence excitation filters to view DAPI fluorescence and chlorophyll (autofluorescence). It can be difficult to distinguish DNA-containing debris from small cells by DAPI staining. Therefore, we counted a putative cell, which fluoresced by DAPI but was not autofluorescent, as a heterotrophic cell.

Metabolite extractions and instrument methods

The intracellular metabolites were originally extracted using one quarter of the filter; however, this was determined to be too dilute and another set of extractions were performed using the remaining three quarters of each filter. The FT-ICR mass spectrometer was externally calibrated with a calibration solution provided by the manufacturer.

Standards for the targeted method were purchased from Sigma-Aldrich (MO, USA), with the exception of dimethylsulfoniopropionate (DMSP), purchased from Research Plus Inc. (NJ, USA), and 2,3-dihydroxypropane-1-sulfonate (DHPS), which was generously donated by Dr. Mary Ann Moran at the University of Georgia.

Samples for TOC analysis were stored at 4°C until analysis with a Shimadzu TOC-V_{CSH} total organic carbon analyzer (Longnecker *et al*., 2015). Nutrient analysis was performed at the Nutrient Analytical Facility at WHOI using a SEAL AA3 four-channel segmented flow analyzer according to USEPA approved protocols to determine concentrations of ammonium (NH_4^+), nitrate plus nitrite (NO_x^-), and phosphate (PO_4^-).

Metabolomics data analysis

The packages gplots (Warnes *et al*., 2013) and ggplot2 (Wickham, 2009) in R were used to generate the heatmap and other figures using TOC- and volume-normalized concentrations or peak areas. Non-metric multidimensional scaling (nMDS) was performed with quality-checked mass feature data from the untargeted metabolomics method. Data were scaled between 0 and 1 (clusterSim package; Walesiak & Dudek, 2014) prior to generating the Bray-Curtis dissimilarity matrix for nMDS analysis. The function metaMDS in R (vegan package; Oksanen *et al*., 2013) uses multiple random starts to find a stable solution and the number of dimensions used in the function is specified by the user. The dimensionality was assessed by Monte Carlo analysis using 20 iterations with real data and 20 iterations with randomized data for each of several dimensions (n=2-7). Based on the results from the Monte Carlo Analysis, the final nMDS was performed with two dimensions. Mantel tests were performed to determine the coefficient of determination (r^2) between ordinal distance of the nMDS and the Bray-Curtis dissimilarity for both axes and each axis separately. Analysis of similarity (ANOSIM) using the vegan package was performed to test for significant differences in the composition and abundance of mass features between groups detected in the nMDS.

We used draft metabolic networks of several *Synechococcus* species (*S. elongatus* PCC 6301, *S. elongatus* PCC 7942, *Synechococcus* sp. WH8102), the heterotrophic bacterium *Ruegeria pomeroyi*, and the diatom *Thalassiosira pseudonana* to further identify mass features of interest from the thousands of features in the untargeted data set. *Synechococcus elongatus* PCC 6301 and PCC 7942 are both freshwater cyanobacteria while WH8102 is an open ocean strain (Palenik *et al.*, 2003). Any unknown feature within 3 ppm of a predicted metabolite mass was selected as a putative metabolite match. Putative matches were then manually screened for fragmentation spectra. We then compared the mass, retention time, and fragmentation spectrum of the mass feature of interest with those of the commercial standard. Lists of predicted metabolites of *Synechococcus* species were generated using free-form advanced query in BioCyc (Karp *et al.*, 2005). The metabolic networks for *R. pomeroyi* and *T. pseudonana* were generated using Pathway Tools (Karp *et al.*, 2011) according to the user manual with genome sequences and annotation files downloaded from the integrated microbial genomes database (IMG; Markowitz *et al.*, 2007). Recently, *R. pomeroyi* and *T. pseudonana* pathway genome databases (PGDB) were added to BioCyc online and are nearly identical to those generated manually in the current study; therefore we did not submit the created PGDBs to BioCyc.

Search for genes with homology to tryptophan catabolic genes in Synechococcus *spp.*

Two tryptophan oxidation products, kynurenine and indole 3-acetic acid, were detected in the *S. elongatus* extracellular metabolite fraction. Kynurenine is not a predicted metabolite of *S. elongatus*, and although IAA is a predicted metabolite of *S. elongatus*, the pathway producing IAA is not well characterized. Therefore, putative

genes involved in these two pathways were further investigated. The genome of *Synechococcus elongatus* CCMP 1631 is not yet sequenced; therefore, the genomes of *S. elongatus* PCC 6301, *S. elongatus* PCC 7942, *Synechococcus* sp. WH8102 (hereafter referred to as WH8102), and *Synechococcus* sp. CC9311 (hereafter referred to as CC9311) were used as a proxy for CCMP 1631 in a comparative genomics approach to identify putative genes involved in tryptophan catabolism. The strain CC9311 is a marine coastal water strain of *Synechococcus* (Palenik *et al.*, 2006). Gene and protein sequences (nucleotide and amino acid sequences) encoding for important enzymes involved in the kynurenine pathway (Table S4) were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/) from several bacteria (Table S5). Gene sequences encoding for important enzymes involved in tryptophan degradation to indole-3-acetic acid (Table S4) were also downloaded from NCBI (Table S5). The BLAST tool available through the IMG database was used to search the *Synechococcus* spp. genomes for putative matches using the downloaded gene sequences. Nucleotide (BLASTn) and protein (BLASTp) searches were performed and the E-value was set low ($1e^{-2}$) to allow for any potential match.

Additional genome comparisons were performed using the online program STRING v9.1 (Franceschini *et al.*, 2012). STRING allows genomes to be queried by specific genes of interest and for comparison of these genes to other genomes in the STRING database. STRING also provides information on gene neighbors, expression, and previous literature for a gene of interest. We searched the STRING database for the genes described above and used the results to query *Synechococcus* spp. genomes. Putative matches were examined in terms of homology as well as neighboring genes and any known expression or co-occurrence of the gene of interest. Multiple and pair-wise alignments were also performed on putative matches and known genes of interest derived

from organisms in the STRING database. Alignments were generated in MEGA 2.1 (Tamura *et al.*, 2011) with translated protein sequences (Clustal W with Gonnet protein weight matrix and default settings in MEGA) and using pairwise BLAST (bl2seq). Sequence alignments were manually checked for overall quality and for conserved amino acid residues specific to the enzyme of interest.

Results and Discussion

Targeted analysis of intracellular metabolites

Intracellular samples from the targeted method did not completely cluster by time point, reflecting the variability observed by nMDS with untargeted data (Figure 2). Three clusters of interest are highlighted here in the targeted intracellular data (Figure 2). Cluster A includes amino acids involved in polyamine synthesis (methionine, arginine, citrulline) and clustered together with spermidine. Cluster B includes the two osmolytes, glycine betaine and DMSP, as well as the amino acid aspartic acid, and were generally more abundant earlier in the growth curve. Lastly, cluster C includes the amino acid glutamine, the vitamin thiamin, and the pyrimidine precursor orotic acid, which generally exhibited highest concentrations in the later growth stages. The ratio of two amino acids of interest due to their role in nitrogen cycling, arginine and citrulline, increased from approximately 1 to 5 over time.

Untargeted analysis of extracellular DOM

Mainly nitrogen- and carbon-rich compounds were detected in the media, but there may be other modified compounds (e.g., phosphorylated) released by the cells that

were either below detection levels of the instruments or were not retained by the SPE method used for extracellular metabolite extractions. Baran *et al*. (2011), however, noted a similar lack of phosphorylated compounds in their experiments using a centrifugation method to extract metabolites from the media of *Synechococcus* cultures.

Several metabolites were identified in the extracellular fraction using untargeted metabolomics analysis and confirmed with authentic standards (Table S3). The mass feature putatively identified as the purine nucleoside xanthosine, increased over time in the extracellular fraction but decreased at the last time point in both replicates (Figure S5). Both xanthosine replicates appeared to increase only slightly, with one noticeably high data point, over time when normalized to TOC. This likely indicates that xanthosine is constitutively released during the growth of *S. elongatus*. The mass feature identified as another purine nucleoside, guanosine, showed a similar trend to xanthosine, with one noticeably high data point (Figure S5). Both guanosine replicates had a higher concentration in the later time point samples relative to the early time points when normalized to TOC. The amino acid tyrosine and the monosaccharide *N*-acetylmuramic acid were also observed to accumulate over time (Figures S6, S7). The compounds described here show a distinct increase in concentration from days 10 to 15 when normalized to TOC and may indicate higher metabolic overflow during exponential growth.

Other features derived from untargeted analysis of the dissolved fraction were putatively identified as metabolites based on two criteria. First these features matched by exact mass to a metabolite from one of the predicted metabolomes and second, the associated fragmentation spectrum matched to the same metabolite using *in silico* analysis (MetFrag). These features (Table S3) did not show a clear increase or decrease

in concentration over time but represents targets that may be of interest for future studies on the physiology and ecology of *S. elongatus* and other cyanobacteria.

Identification of genes in Synechococcus that are homologous to genes in the kynurenine pathway and IAA production pathways

Kynurenine is not predicted to be produced by *S. elongatus* based on available *Synechococcus* genomes; therefore, we used comparative genomics analysis to investigate whether sequenced *Synechococcus* spp. contain the enzymes responsible for tryptophan oxidation to kynurenine. Enzymes involved in tryptophan oxidation leading to kynurenine and other enzymes involved in the kynurenine pathway (Figure S8) were queried against the *Synechococcus* genomes. There were some hits in the *Synechococcus* genomes, but none of these proved to be genes that would likely function in the kynurenine pathway based on pairwise alignments with characterized genes (Table S4).

Indole-3-acetic acid (IAA) was observed in the extracellular metabolite fraction from *S. elongatus;* however, the biochemical pathway generating IAA is not well characterized in cyanobacteria. Tryptophan catabolism to IAA has been suggested to occur via the indole-3-pyruvate pathway in cyanobacteria (Sergeeva *et al.*, 2002) (Figure S8). Sergeeva *et al*., (2002) characterized the *idp*C gene (encoding for indolepyruvate decarboxylase) in several cyanobacteria although no *Synechococcus* were tested. There are at least three genes involved in IAA formation that have been previously identified in different strains of *Synechococcus* (*idp*C gene in CC931; nitrilase encoding genes in WH8102 and *S. elongatus* PCC 7942; amidase encoding genes in *S. elongatus* PCC 6301 (Figure S8)), although the full biochemical pathway is not characterized in any of these organisms. The *idp*C gene of CC9311 as well as of *Bradyrhizobium* sp. BTAi1, were

used to query *Synechococcus* spp. genomes for putative homologs. No putative orthologs were detected in the *S. elongatus* genomes, but there was a hit to a putative *ilv*B gene (acetolactate synthase; SYNW1746) in WH8102 (Table S4). Only one of the other enzymes involved in tryptophan degradation to IAA investigated in the present study, tryptophan decarboxylase, yielded a potential match to a gene in the *Synechococcus* genomes: a pyridoxal-dependent decarboxylase family protein (Table S4). Most amino acid decarboxylases are pyridoxal 5'-phosphate-dependent enzymes (Capitani *et al*., 2003) and it is possible that this enzyme could act on tryptophan, converting it to tryptamine in the first step of tryptophan catabolism to IAA (Mitoma *et al*., 1960; Carreño-Lopez *et al*., 2000). The pyridoxal-dependent decarboxylase genes identified here could be targets for future genetic and physiological experiments to determine the substrate and role of the enzyme. This also provides an example of the potential of metabolomics studies for improving genome annotation for marine microbes.

Literature Cited

Carreño-Lopez R, Campos-Reales N, Elmerich C, Baca BE. (2000). Physiological evidence for differently regulated tryptophan-dependent pathways for indole-3-acetic acid synthesis in *Azospirillum brasilense*. *Mol Gen Genet* **264**:521-530.

Capitani G, Tramonti A, Bossa F, Grütter MG, De Biase D. (2003). The critical structural role of a highly conserved histidine residue in group II amino acid decarboxylases. *FEBS Lett* **554**:41-44.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, *et al.* (2012). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research* **41**:D808–D815.

Ishii S, Mizuguchi H, Nishino J, Hayashi H, Kagamiyama H. (1996). Functionally important residues of aromatic L-amino acid decarboxylase probed by sequence alignment and site-directed mutagenesis. *The Journal of Biochemistry* **120**:369–376.

Karp PD, Latendresse M, Caspi R. (2011). The Pathway Tools Pathway Prediction Algorithm. *Stand Genomic Sci* **5**:424–429.

Karp PD, Ouzounis CA, Moore-Kohlacs C, Goldovsky L, Kaipa P, Ahrén D, *et al.* (2005). Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research* **33**:6083–6089.

Koga J. (1995). Structure and function of indolepyruvate decarboxylase, a key enzyme in indole-3-acetic acid biosynthesis. *Biochimica et Biophysica Acta* **1249**:1–13.

Lima WC, Varani Am, Menck CFM. (2009). NAD biosynthesis evolution in bacteria: lateral gene transfer of kynurenine pathway in Xanthomonadales and Flavobacteriales. *Mol Biol Evol* **26**:399-406.

Longnecker K, Kido Soule MC, Kujawinski EB. (2015). Dissolved organic matter produced by *Thalassiosira pseudonana*. *Marine Chemistry* **168**:114-123.

Markowitz VM, Szeto E, Palaniappan K, Grechkin Y, Chu K, Chen I-MA, *et al.* (2007). The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Research* **36**:D528–D533.

Mitoma C, Udenfriend S. (1960). Bacterial tryptophan decaroboxylase. *Biochim Biophys Acta* **37:**356-357.

Momany C, Levdikov V, Blagova L, Lima S, Phillips RS. (2004). Three-Dimensional Structure of Kynureninase from Pseudomonas fluorescens†,‡. *Biochemistry* **43**:1193–1203.

Oksanen J, Guillaume Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, *et al*. (2013). vegan: community ecology package. R package version 2.0-10. http://CRAN.R-project.org/package=vegan

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain PSG, *et al.* (2003). The genome of a motile marine *Synechococcus*. *Nature* **424**:1037–1042.

Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, *et al.* (2006). Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proceedings of the National Academy of Sciences* **36**:13555–13559.

Sergeeva E, Liaimer A, Bergman B. (2002). Evidence for production of the phytohormone indole-3-acetic acid by cyanobacteria. *Planta* **215**:229–238.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**:2731–2739.

Walesiak M, Dudek A. (2014). clusterSim: searching for optimal clustering procedure for a data set. R package version 0.43-4. http://CRAN.R-project.org/package=clusterSim

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, *et al*. (2013). gplots: various R programming tools for plotting data. R package version 2.12.1. http://CRAN.R-project.org/package=gplots

Wickham H. (2009). ggplot2: elegant graphics for data analysis. Springer New York.

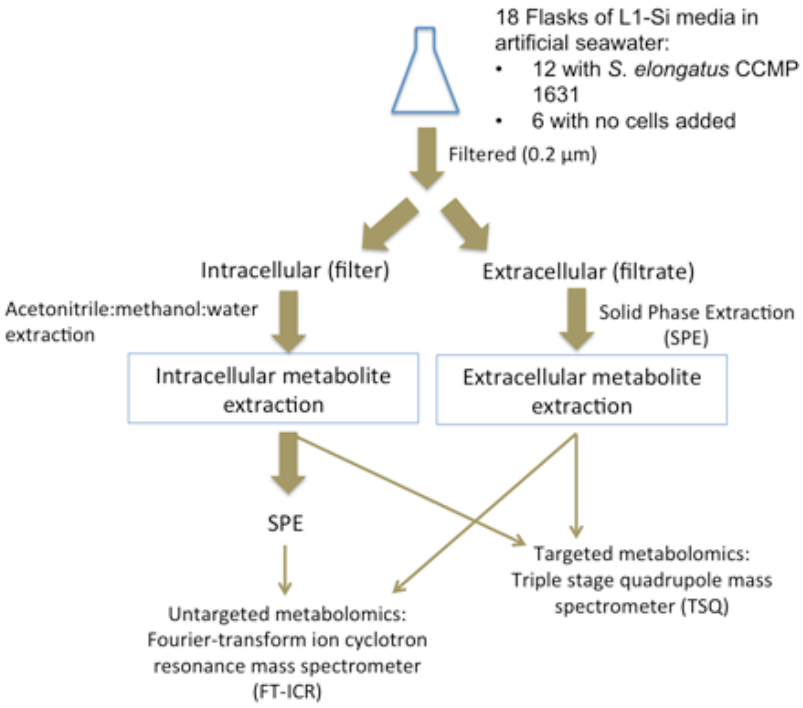

Figure S1. Overview of the methods used for metabolite extraction and analysis by targeted and untargeted metabolomics. Thin arrows represent the final step where metabolite extractions were used for analysis.

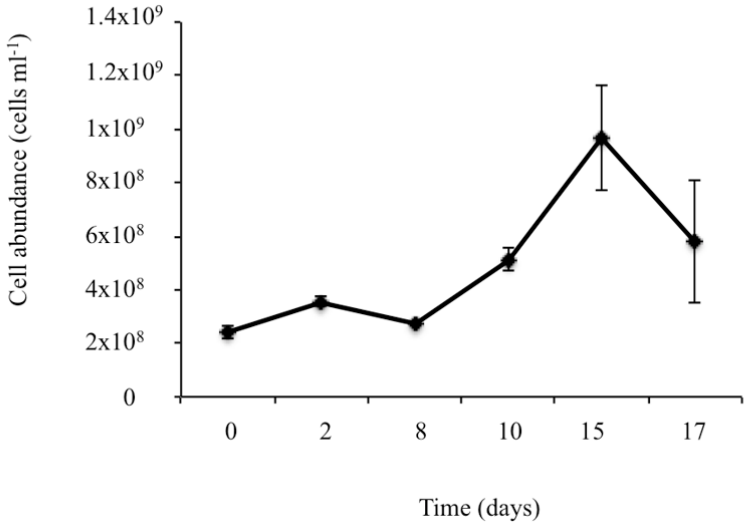

Figure S2. Growth curve of *Synechococcus elongatus* CCMP 1631 during the three-week experiment. Cell abundance was calculated using cell counts from DAPI staining and replicate samples were averaged with the exception of day 8, for which one sample was considered an outlier. Standard deviation between replicates is shown. Metabolomics samples were collected at each marker (days 0, 2, 8, 10, 15, 17).

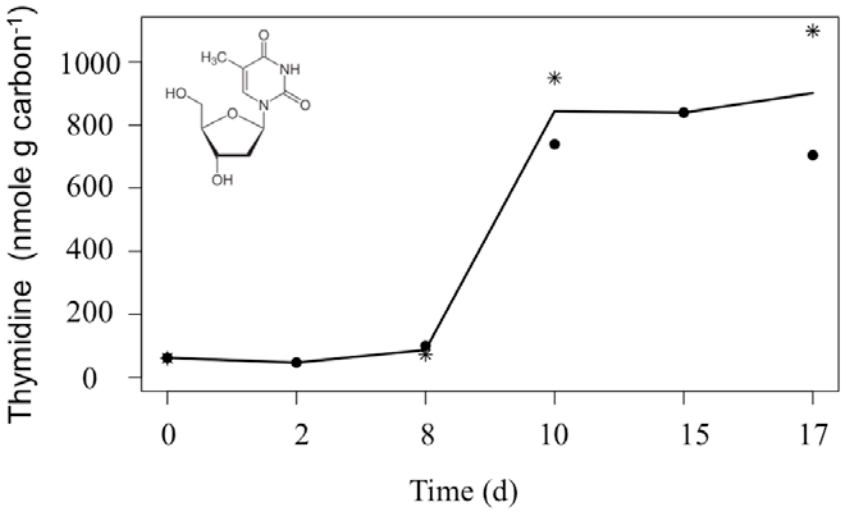

Figure S3. TOC-normalized concentrations of thymidine in both the extracellular and intracellular fractions over time. The circles and asterisks represent replicate cultures, except for days 2 and 15, where only one replicate was available. The black line represents the average of the replicates, where available. The structure of thymidine is shown in the inset.

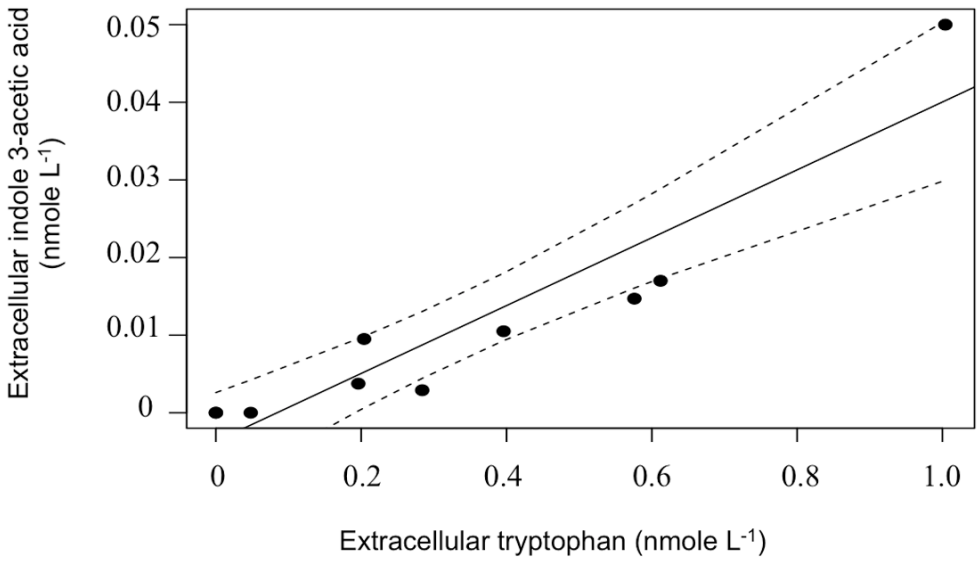

Figure S4. Volume-normalized tryptophan concentrations and indole 3-acetic acid concentrations from the extracellular fractions, showing a significant correlation (Pearson's $r = 0.94$, $p < 0.01$).

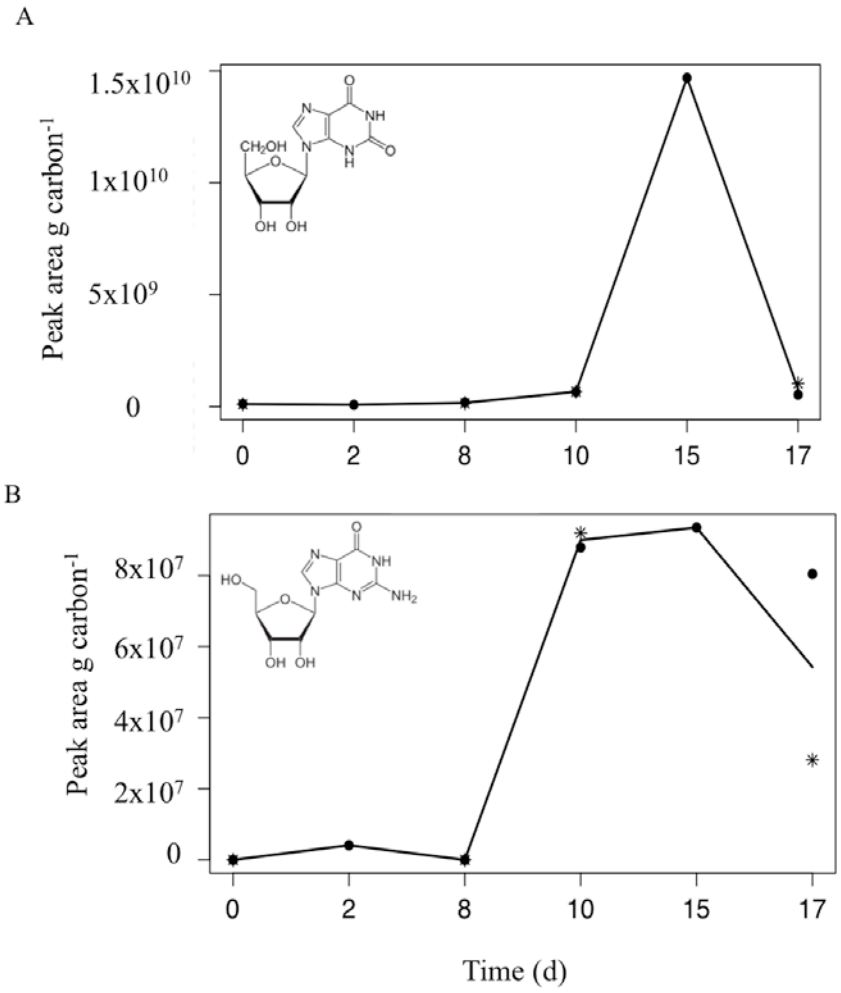

Figure S5. TOC-normalized peak area of the mass features identified as xanthosine (A) and guanosine (B) (extracellular) over time (positive ion mode). The structure of each is shown in the inset. Black dots and asterisks represent replicate cultures at each time point. The black lines represent the average of replicate cultures. For days 2 and 15, data from one only replicate is available.

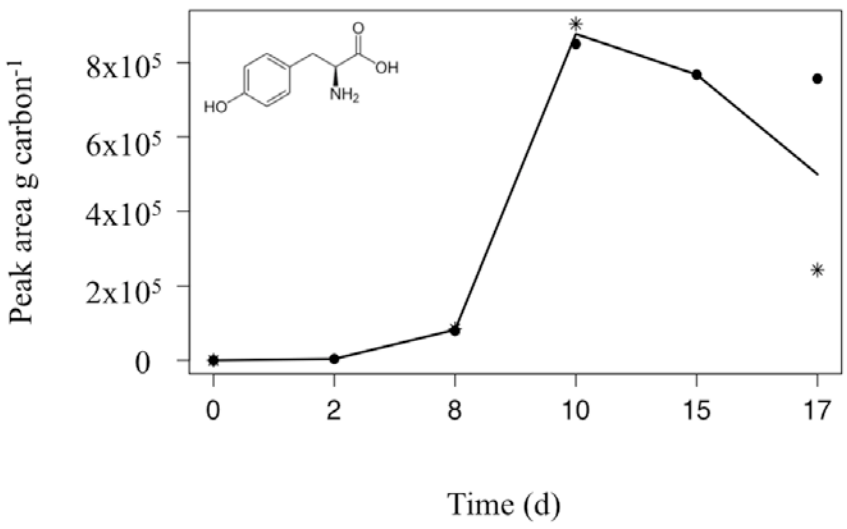

Figure S6. TOC-normalized peak area of the mass feature identified as tyrosine in the extracellular fraction over time (positive ion mode). The structure of tyrosine is shown in the inset. Black dots and asterisks represent replicate cultures at each time point. The black lines represent the average of replicate cultures. For days 2 and 15, data from one only replicate is available.

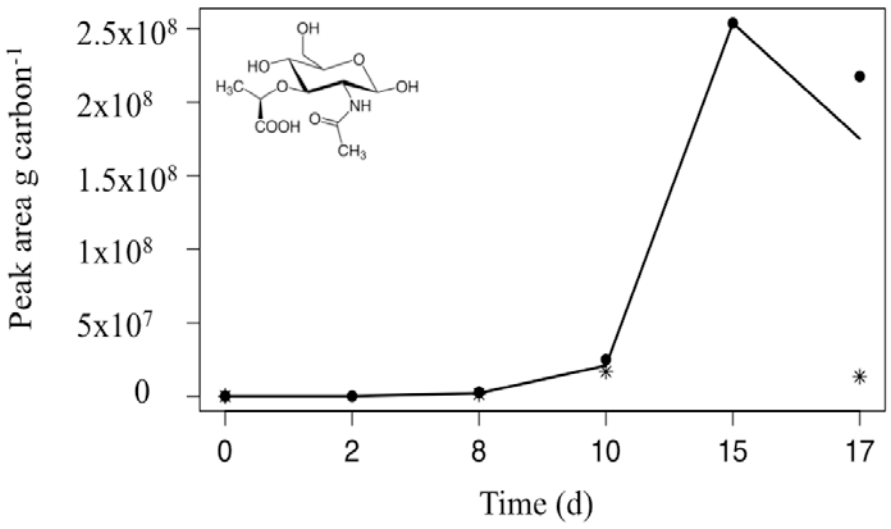

Figure S7. TOC-normalized peak area of the mass feature identified as *N*-acetylmuramic acid (extracellular) over time (positive ion mode). The structure of *N*-acetylmuramic acid is shown in the inset. Black dots and asterisks represent replicate cultures at each time point. The black lines represent the average of replicate cultures. For days 2 and 15, data from one only replicate is available.

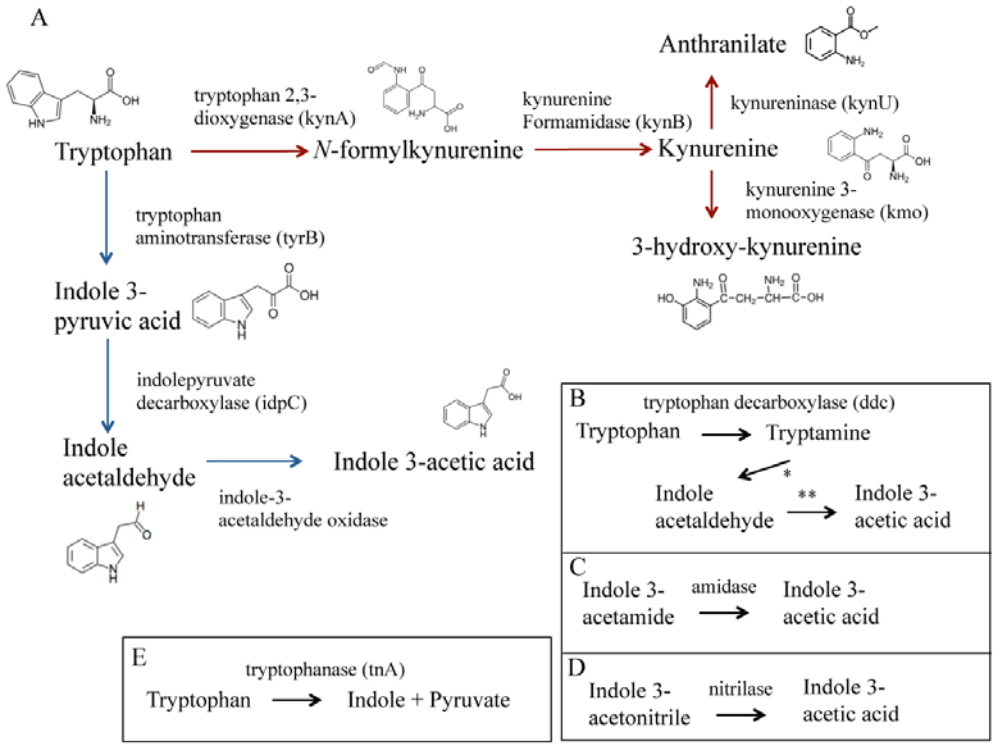

Figure S8. Overview of kynurenine pathway (red arrows) and indole 3-pyruvic acid pathway (blue arrows) (A). Other potential pathways involved in degradation of tryptophan and generation of indole 3-acetic acid are also shown (B-E). Pathways shown in C and D are present in PCC 6301 and PCC 7942 respectively. Pathways shown in B and E are other potential pathways for tryptophan catabolism to indole compounds. Gene names are listed in parentheses. Genes encoding for enzymes listed here were used to query *Synechococcus* spp. genomes for putative gene matches. Two enzymes marked as asterisks (B) are monoamine oxidase A (*) and (indole-3-y) acetaldehyde:oxygen oxidoreductase (**).

Table S1. Overview of metabolites in the targeted metabolomics method.

Metabolite	Intracellular	Extracellular
2,3-dihydroxybenzoic acid	ND	D
3-mercaptopropionate	D	D
3-phosphoglyceric acid	D	NE
4-aminobenzoic acid	D	D
4-hydroxybenzoic aicd	D	D
5'-adenosine monophosphate (AMP)	D	NE
5'-methylthioadenosine (MTA)	D	D
6-phosphogluconic acid	D	NE
Adenosine	D	D
Arginine	D	NE
Aspartic acid	D	NE
Biotin	D	D
Chitobiose	D	NE
Choline	D	NE
Citrulline	D	NE
Cyanocobalamin	D	D
Cytosine	D	NE
Deoxy-D-xylulose-5-phosphate	D	NE
Desthiobiotin	D	D
dihydroxyacetone phosphate	D	NE
Dimethylsulfoniopropionate (DMSP)	D	NE
Folic acid	D	NE
Glucosamine-6-phosphate	D	NE
Glucose-6-phosphate	D	NE
Glutamic acid	D	NE
Glutamine	D	NE
Glutathione (reduced)	D	NE
Glycerol-3-phosphate	D	NE
Glycine betaine (Betaine)	D	NE
Homoserine	D	NE
Indole 3-acetic acid (IAA)	ND	D
Inosine	D	D
Inosine monophosphate (IMP)	D	NE

D = Detected, ND = Not Detected, NE = low extraction efficiency (Not Extracted)

Table S1 continued.

Metabolite	Intracellular	Extracellular
Leucine	D	NE
Malic acid	D	NE
Methionine	D	NE
Muramic acid	D	NE
N-acetylglucosamine	D	NE
N-acetylglutamic acid	D	D
N-acetyltaurine	D	NE
Nicotinamide adenine dinucleotide (NAD)	D	NE
Nicotinamide adenine dinucleotide phosphate (NADP)	D	NE
Ornithine	D	NE
Orotic acid	D	NE
Pantothenic acid	D	NE
Phenylalanine	D	D
Phospho*enol*pyruvate (PEP)	D	NE
Proline	D	NE
Putrescine	D	NE
Pyroxidine	D	NE
Riboflavin	D	D
Ribose-5-phosphate	D	NE
Serine	D	NE
Spermidine	D	NE
Succinic acid	D	D
Taurine	D	NE
Taurocholate	D	D
Threonine	D	NE
Thymidine	D	D
Tryptophan	D	D
Uracil	D	NE
Uridine monophosphate (UMP)	D	NE

Table S2. Extraction efficiency (total carbon) for a subset of filtrate samples processed by solid phase extraction (SPE) and inorganic nutrients of all culture samples and controls.

Sample (day of sampling)	Total carbon extraction efficiency for SPE (%)	NH_4^+ (µM)	NO_x^- (µM)	PO_4^- (µM)
0	2.8	21	386	14
0	ND	37	829	34
2	ND	26	835	34
2	3.4	22	740	28
8	ND	31	885	36
8	5.2	28	862	33
10	3.6	25	819	34
10	ND	21	658	29
15	ND	34	908	35
15	12.8	25	510	23
17	7.9	14	484	20
17	ND	29	930	35
0 control	ND	21	526	22
2 control	4.0	18	406	19
8 control	ND	25	229	14
10 control	6.9	24	717	26
15 control	ND	27	928	34
17 control	7.6	27	738	29

ND = not determined. NO_x^- is the sum of nitrate (NO_3^-) and nitrite (NO_2^-).

Table S3. Mass features identified in the untargeted method and confirmed with authentic standards (top half of table) and features putatively identified based on exact mass and *in silico* analysis of the fragmentation spectrum using MetFrag.

Mass feature (*m/z*)	**Adduct**	**Metabolite name**
268.10387	$[M+H]^+$	Adenosine
284.09878	$[M+H]^+$	Guanosine
209.09152	$[M+H]^+$	Kynurenine
294.11829	$[M+H]^+$	Muramic acid
166.08522	$[M+H]^+$	Phenylalanine
243.09759	$[M+H]^+$	Thymidine
205.09702	$[M+H]^+$	Tryptophan
182.08107	$[M+H]^+$	Tyrosine
285.08293	$[M+H]^+$	Xanthosine
		Putative metabolite name
199.02536	[M-H]-	2-hydroxy-5-carboxymethylmuconate semialdehyde
219.06463	[M+H]+	2-Hydroxy-6-oxo-6-phenylhexa-2,4-dienoate
239.12723	[M+H]+	3-[(3aS,4S,7aS)-7a-Methyl-1,5-dioxo-octahydro-1H-inden-4-yl]propanoate
175.06174	[M-H]-	3-isopropylmalic
315.11946	[M+H]+	7,8-dihydropteroate
173.00971	[M-H]-	Aconitic acid
305.24695	[M+H]+	Eicosatetraenoate
271.16871	[M+H]+	Estrone
193.05118	[M-H]-	Ferulate
319.22622	[M+H]+	Leukotriene A_4
197.04609	[M-H]-	Vanillylmandelic acid
206.04423	[M+H]+	Xanthurenate

Table S4. Enzymes and corresponding genes involved in tryptophan catabolism that were used to query *Synechococcus* spp. genomes.

Enzyme	Gene name	Putative match in *Synechococcus* genomes	Homology (%)	Coverage (%)	Conserved residues	Reference for conserved residues
Kynurenine pathway:						
tryptophan 2,3-dioxygenase (TDO)	kynA	None				
kynurenine formamidase (KFA)	kynB	None				
kynureninase	kynU	putative cysteine desulfurase (WH8102)	28	21	1 of 4	Momany *et al.*, 2004
		serine:pyruvate:glyoxylate aminotransferase (WH8102)	29	29	none	Momany *et al.*, 2004
kynurenine 3-monooxygenase (KMO)	kmo	geranylgeranlyl hydrogenase (WH8102)	25	37	none	Lima *et al*., 2009; Amaral *et al*. 2013
		FAD-dependent monooxygenase	24	93	none	Lima *et al*., 2009; Amaral *et al*. 2013
		hypothetical protein	24	41	none	Lima *et al*., 2009; Amaral *et al*. 2013
Indole 3-acetic acid pathway:						
tryptophanse	tnaA	none				
tryptophan/aromatic aminotransferase	tyrB	none				
tryptophan decarboxylase (TDC) (a.k.a aromatic L-amino acid decarboxylase)	ddc	L-2-diaminobutyrate decarboxylase (PCC 6301)	26	95	7 of 7	Ishii *et al*., 1996; Capitani *et al*., 2003
		L-2-diaminobutyrate decarboxylase (PCC 7942)	26	85	7 of 7	Ishii *et al*., 1996; Capitani *et al*., 2003
		pyridoxal-dependent decarboxylase family protein (WH8102)	24	75	6 of 7	Ishii *et al*., 1996; Capitani *et al*., 2003
		pyridoxal-dependent decarboxylase family protein (CC9311)	24	75	6 of 7	Ishii *et al*., 1996; Capitani *et al*., 2003
indolepyruvate decarboxylase	idpC	acetolactate synthase (WH8102)	28	98	1 of 4	Koga, 1995

Results of the queries are provided as the name of the enzyme encoded by the putative gene match, the strain containing the match, and the homology and coverage of the putative match to the characterized gene of interest based on pairwise alignment. Conserved residues detected in the query based on published information on the enzyme of interest are also shown.

Table S5. Organisms used for generation of draft metabolic network and in comparative genomics analysis. Pathway genome databases for *Synechococcus elongatus* PCC 6301, *S. elongatus* PCC 7942, and *Synechococcus* sp. WH8102 were available through BioCyc.

Species	Database	Identification number	Analysis
Ruegeria pomeroyi	NCBI	NC_006569	Draft predicted metabolome
Thalassiosira pseudonana	NCBI	NC_012064-012069	Draft predicted metabolome
Pseudomonas aeruginosa	NCBI	NC_002516	kynurenine pathway
Nostoc sp. PCC 7107	NCBI	NC_0197676	kynurenine pathway
Synechococcus sp. CC9311	NCBI	NC_008319	indole 3-acetic acid production pathway
Escheriachia coli	NCBI	NC_000913	indole 3-acetic acid production pathway
Oceanithermus profundus	NCBI	NC_014761	indole 3-acetic acid production pathway
Vibrio parahaemolyticus	NCBI	NC_004605	indole 3-acetic acid production pathway
Saccharomyces cerevisiae	NCBI	NC_003075	indole 3-acetic acid production pathway
Bradyrhizobium sp. BTAil	NCBI	NC_009475	indole 3-acetic acid production pathway
Homo sapiense	STRING	ENSP00000350616	indole 3-acetic acid production pathway