Pacific Northwest
NATIONAL LABORATORY

PNNL-34278

Generating Co-expression Networks for Three Cyanobacteria:

Synechococcus sp. PCC 7942, Synechococcus sp. PCC 7002, Synechocystis sp. PCC 6803

May 2023

David A. Anderson
Pavlo Bohutskyi

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-34278

Generating Co-expression Networks for Three Cyanobacteria:

Synechococcus sp. PCC 7942, Synechococcus sp. PCC 7002, Synechocystis sp. PCC 6803

May 2023

David A. Anderson
Pavlo Bohutskyi

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Generating Co-expression Networks for Three Cyanobacteria: Synechococcus sp. PCC 7942, Synechococcus sp. PCC 7002, Synechocystis sp. PCC 6803

David Anderson (2023)

Abstract

Cyanobacteria are photosynthetic organisms capable of high growth rate and represent a promising bioplatform for harnessing the sun's energy to make biofuel. Additionally, the process of photosynthesis absorbs CO2 from the environment. Understanding the metabolic processes involved in photosynthesis could lead to solutions to the recent rise of CO2 concentration in Earth's atmosphere and the associated climate change. More research on the transcriptional regulation of these cells is needed to learn how to harness the untapped potential of cyanobacteria for these applications. Transcriptional analysis via RNA-seq provides an understanding of how gene expression changes at the mRNA level under diverse growing conditions. I systematically collected and analyzed RNA-Seq data obtained under a variety of conditions and available on the NCBI database for three cyanobacteria model organisms: *Synechococcus elongatus sp.* PCC 7942, *Synechococcus sp.* PCC 7002, and *Synechocystis sp.* PCC 6803. For each organism, the data was mapped to a reference genome to characterize the RNA expression profile. Samples were checked for quality based on the number of reads and the correlation of the expression profile between labeled replicates. All samples were transformed into transcripts per million reads, followed by a log transformation to account for the wide range of sample sizes. Gene co-expression networks were generated and analyzed for each species using Cytoscape. These networks provide a base level of gene expression for each species. The network topology and high-betweeness nodes of these networks need to be analyzed further to provide insight on potential ways to harness cyanobacteria genetics. Additionally, these datasets can be used together to form a core genome network analysis- one that includes only the genes that are homologous between the three species. This project has prepared the way for a more in-depth study on photosynthetic microbes on a genetic level.

I. INTRODUCTION

Cyanobacteria are photosynthetic organisms capable of fast growth and represent a promising bio-platform for harnessing the sun's energy to make biofuel. They have relatively simple genomes and can be easily cultured in the laboratory, making them ideal for genetic and biochemical analyses. Their photosynthetic capability allows them to absorb CO_2 from their environment and fix it into a biologically useful chemical, such as the sugar form sucrose. These characteristics make cyanobacteria prime candidates for biotechnical applications. Improving our understanding of cyanobacteria gene expression, especially with regards to its metabolic pathways, is the next step in using them as a solution to the rise of CO_2 concentration in our atmosphere as well as to synthesize biofuels at a commercial scale. More research on the transcriptional regulation of these cells is needed to learn how to harness the untapped potential of cyanobacteria for these applications.

RNA sequencing, or RNA-Seq, is a widely used method for measuring gene expression levels and identifying differentially expressed genes in various biological samples. This technique involves the conversion of RNA molecules into complementary DNA (cDNA) fragments. These fragments are then sequenced using high-throughput sequencing technologies such as Illumina. Transcriptional analysis via RNA-seq provides an understanding of how gene expression changes at the mRNA level under diverse growing conditions. It is commonly used to examine how cyanobacterial genes respond to perturbations in light/dark cycles, added stress conditions, or nutrient-limited media. This paper presents our analysis of RNA-Seq data currently available for 3 cyanobacteria species: *Synechocystis sp.* PCC 6803 (hereafter PCC 6803), *Synechococcus sp.* PCC 7002 (hereafter PCC 7002), and *Synechococcus elongatus sp.* PCC 7942 (hereafter PCC 7942). PCC 6803, PCC 7002, and PCC 7942 are each considered to be model organisms due to the high amount of completed and continuing research being performed with them. Combining data collected world-wide into a single dataset for each of these species forms a clearer picture of how they express their genes under a variety of growing conditions.

II. METHODS

A. Data collection

Samples were found and obtained using the NCBI: Sequence Read Archive (SRA) online, open-source database for the 3 cyanobacteria species of interest- PCC 6803,[1-39] PCC 7002,[40-58] and PCC 7942.[59-77] All bio-projects that contained RNASeq data for *S.* elongatus- PCC7942 were collected. Most were downloaded directly as fastq files, however the few that were unavailable in that format were extracted from the database using the "fasterq-dump" command from the SRA Toolkit from NCBI. Each sample was then compared to the reference genome (.fastq/.gff) file using the Rsubread library[78] for the programming language R. The reference sequences used: for PCC 6083 it was (GCF_000009725.1_ASM972v1) generated in 2004, for PCC 7002 it was (GCF_000019485.1_ASM1948v1) generated in 2008, and for PCC 7942 it was (GCF_022984195.1_ASM2298419v1) generated in 2021. The output from Rsubread is a list of counts for each gene in the respective genome.

The datasets were then checked for quality. It was determined that samples with too few overall counts provided a biased, low-resolution gene expression profile. To increase our confidence in our findings from down-stream analysis, samples that mapped less than 10^5 counts were removed. I also did a correlation check to mark any samples that had an expression profile extremely different from the rest of the samples. I examined the correlation between samples that were labeled as biological or technical replicates on the NCBI: SRA run selector. Any samples that had a correlation of less than 0.9 with a replicate sample, including samples that were not sequenced in replicate, were removed from the dataset. While performing the replicate correlation check, adjacent samples that were used to study a time series during a perturbation condition were considered as replicates. An exception was made for samples from the PCC 6803 project tag "First_Transcriptome" and condition "Kai_dark_11.5h" from project tag "kaiABC", and from the PCC 7002 project tag "B12", "ccmRdel", and "Acclimation" (see Table 1) since they were a variety of conditions that I wanted to retain in their respective datasets. The only identified reason to throw them out was the lack of replicates. The PCC 7942 dataset used was taken directly from Johnson, et al. (2023)- paper not yet published.

B. Generating networks

To create the gene co-expression network, I used the Genie3 package[79,80] for R on the normalized counts. Genie3 uses an “ensemble of trees” approach to compare relative levels of gene expression and connect genes that are commonly expressed together. I used the FastGreedy algorithm[81] to determine cluster grouping (min # = 12). Each network was filtered such that all edges below a specified cutoff weight were removed. This simplified the network by removing connections between less-associated genes. I could then set the cutoff value based on how many edges would be remaining after applying the filter. The networks were analyzed using Cytoscape[82]. Edges and nodes not connected to the main network were trimmed. To facilitate further analysis of the networks, the networks selected for further analysis were made up of a similar number of edges- about 4,300 edges per network. The visualizations of these networks are included here in appendix A as figures 4, 5, and 6.

III. RESULTS AND DISCUSSION

A total of 728 samples covering 250 growth conditions for PCC 6803 were collected from the online databases (Table1). Samples with <105 counts tended to be poorly correlated to their replicates and other like conditions which could be due to poor quality of the reads preventing their mapping to the reference genome. I removed 89 samples for generating less than 10^5 counts and an additional 91 samples that were not available in replicate or were poorly correlated to their replicate samples. The final PCC 6803 dataset was made up of 515 samples. Figure 1 shows a PCA plot of the samples retained in the dataset.

A total of 269 samples covering 108 growth conditions for PCC 7002 were collected from the online databases (Table2). I removed 12 samples for generating less than 10^5 counts and an additional 36 samples that were not available in replicate or were poorly correlated to their

replicate samples. The final PCC 7002 dataset was made up of 215 samples. Figure 2 shows a PCA plot of the samples retained in the dataset.

A total of 416 samples covering 238 growth conditions for PCC 7942 were collected from the online databases (Table 3). I removed 12 samples for generating less than 10^5 counts and an additional 66 samples that were not available in replicate or were poorly correlated to their replicate samples. The final PCC 7942 dataset was made up of 333 samples. Figure 3 shows a PCA plot of the samples retained in the dataset.

Further analysis of these networks is presently underway.

IV. CONCLUSION

In conclusion, PCC 6803, PCC 7002, and PCC 7942 are a model cyanobacteria with a bunch of gene expression data publicly available. I gathered all RNA-seq data available from the NCBI: SRA database and performed a quality check to remove samples that might skew downstream analysis. I then generated three co-expression networks, one from each cyanobacterial species' dataset. An analysis of these networks will help us to understand which genes are most important in regulating gene expression.

A. Future work

The datasets and corresponding co-expression networks generated for this SULI project are under further analysis. The genes with the highest betweenness-centrality are being listed and compared. These genes are called "hub genes" and are thought to play important roles in the regulation of gene expression. The analysis can include comparing the overall network topology: characteristics such as the degree distribution and clustering coefficient among others. Understanding the network topology can provide insights into the underlying biological processes and functional modules. An iModulon analysis for functional enrichment can be performed using the datasets reported in this work. This type analysis will give more context to the biological importance of the gene interactions represented in the co-expression networks.

Additionally, the next aim of this research is to annotate all homologous genes between these three cyanobacteria species and merge all three datasets to build a single, core gene co-expression network that will be more generalizable for cyanobacteria. This network can be used to identify conserved biological pathways shared across the different species provide insights into the core genetic machinery of cyanobacteria and help identify potential targets for biotechnological applications. Conversely, a core gene co-expression network can also help identify species-specific pathways and processes that may be unique to individual species. This can provide insights into the adaptation and diversification of cyanobacteria in different environments.

V. ACKNOWLEDGEMENTS

The author would like to thank Dr. Pavlo Bohutskyi for the project design and guidance in this research. Also, thanks go to Ryan McClure, Zach Johnson, Sarah Bogart, and Johanna Owen for their help and support in troubleshooting the bioinformatics pipeline and completing the research for this project.

VI. APPENDECIES

A. Tables and Figures

Project Tag	NCBI: Bioproject ID	NCBI: SRA ID	GEO ID	# of Samples	Raw Read Format	# Unique Conditions	Submitting Entity	Citation #
Ethylene	PRJNA361291	SRP096747	GSE93614	6	S	2	University of Tennessee	24
IsiC&D	PRJNA692056	SRP301698	NA	18	P	6	Huazhong Agricultural University	4
MV_toxicity	PRJNA725979	SRP316820	NA	18	P	6	Huazhong Agricultural University	16
Glycogen	PRJNA827942	SRP371250	NA	9	S	3	Chulalongkorn University	28
Salt	PRJNA587823	SRP228526	NA	3	P	3	Shenzhen University	15
Alcohol	PRJNA485521	SRP157096	GSE118423	6	S	2	Russian Academy of Science	22
Hydrazones	PRJNA800046	SRP356469	GSE194275	9	P	3	Central China Normal University	39
CRISPRi	PRJEB35238	ERP118266	NA	15	S	5	Swedish Royal Institute of Technology	37
eCarrier	PRJNA649552	SRP274142	GSE155385	18	P	6	Universidad Politecnica de Madrid	10
rNucleaseE	PRJNA747814	SRP328835	GSE180316	36	S	4	University of Freiburg	12
NtcA	PRJNA381210	SRP102822	GSE97289	4	S	2	Centro Nacional de Biotecnologia	11
RNApoly	PRJNA473849	SRP149383	GSE115134	3	S	1	Natural History Museum	29
First_Transcriptome	PRJNA224696	SRP032228	NA	12	S	12	University of Freiburg	23
RNAhelicase	PRJNA626347	SRP257225	NA	18	P	7	University of Freiburg	27
RNAhelicase	PRJNA627099	SRP257665	NA	8	P	5	University of Freiburg	27
Topo_batch	PRJEB47621_batch	ERP131906	NA	76	P	19	University of Bielefeld	1
Topo_turbidostat	PRJEB47621	ERP131906	NA	48	P	4	University of Bielefeld	1
MV_glycogen	PRJNA803019	SRP358297	NA	6	S	2	Chulalongkorn University	32
Grad-seq	PRJNA608723	SRP250715	NA	48'	P	1	Albert-Ludwigs University	30
RNaseE_5	PRJNA766607	SRP338861	GSE184824	8	S	3	University of Freiburg	13
KpsM	PRJNA693188	SRP302332	GSE165073	6	P	2	Instituto de Investigação e Inovação em Saúde	31
PS-	PRJNA666973	SRP286154	NA	14	S	7	Korea Advanced Institute of Science and Technology	6
IsiA	PRJNA624961	SRP256110	NA	12	P	4	Huazhong Agricultural University	5
RibosomeProfiling	PRJEB28203	ERP110380	NA	7	S	7	Swedish Royal Institute of Technology	20
Butanol	PRJNA318146	SRP073279	NA	18	P	18	Tianjin university	14
PHA	PRJNA218538	SRP029697	GSE50688	6	P	3	Riken Yokohama Institute	25
DeepSeq	PRJNA213661	SRP028387	NA	1	S	1	Chinese Academy of Sciences	36
PndbA600	PRJEB40560	ERP124213	NA	39	S/P	13	Glasgow Polyomics	26
CAHS	PRJNA888863	SRP401767	NA	4	P	4	Shenzhen University	38
RNaseE&J	PRJNA611475	SRP252009	NA	8	S	3	Institut de Biologie Physico-Chimique	3
Type1-RM	PRJNA628017	SRP258660	NA	12	P	4	Chinese Academy of Sciences	35
Oscillations	PRJEB42778	ERP126684	NA	21	S	10	Swedish Royal Institute of Technology	19
vs7338	PRJNA629670	SRP281791	NA	15	S	1	Korea Advanced Institute of Science and Technology	18
SigB&D	PRJNA791245	SRP351868	GSE192357	111	S	15	University of Turku	34
Fe-	PRJNA315016	SRP072019	NA	6	S	2	Arizona State University	21
asRNA	PRJNA257500	SRP045272	GSE60109	8	S	4	Korea Advanced Institute of Science and Technology	33
vs7942	PRJNA79775	SRP003530	NA	15	P	15	Joint Genome Institute	2
CrhR	PRJNA554812	SRP214789	NA	6	S	3	University of Freiburg	9
HeatAcclimation	PRJNA772179	SRP341882	GSE186038	72	P	24	University of Essex	7
kaiABC	PRJNA547483	SRP200618	GSE132275	24	P	12	Friedrich-Schiller University	17
RNaseE	PRJNA564715	SRP221190	NA	2	S	2	University of Freiburg	8
			Total # of Samples:	728	Total # of Conditions:	250		

Table 1. *Synechocystis sp.* PCC 6803 datasets collected from the NCBI database by bio-project. The 'Project Tag' column was used as a higher-level identifier during the quality check of the data. 'Bioproject ID', 'SRA ID', and 'GEO ID' columns are identifiers used in the NCBI, NCBI: SRA, and GEO databases respectively. I collected 728 RNA-Seq samples from 41 bio-projects representing 250 unique growth conditions.

Project Tag	NCBI: Bioproject ID	NCBI: SRA ID	GEO ID	# of Samples	Raw Read Format	# Unique Conditions	Submitting Entity	Citation #
NrrA	PRJDB6648	DRP004943	NA	8	P	4	Tokyo University of Agriculture	55
Acclimation	PRJNA169550	SRP013965	NA	10	S	8	Pennsylvania State University	48
ChlR	PRJNA198203	SRP021200	NA	8	S	8	Pennsylvania State University	51
OrganicH+	PRJNA212552	SRP027559	GSE48981	4	S	2	University of Wisconsin-Madison	40
CoCulture	PRJNA231839	SRP034523	GSE53360	42	S	7	Environmental Molecular Sciences Laboratory	41
Vipp1	PRJNA235952	SRP035555	NA	2	S	2	Pennsylvania State University	58
ChemProduction	PRJNA283622	SRP058241	NA	2	S	2	Pennsylvania State University	57
ccmRdel	PRJNA288806	SRP060290	NA	2	S	2	Pennsylvania State University	46
Zn	PRJNA292903	SRP062732	NA	10	S	10	Pennsylvania State University	50
Network_1	PRJNA294693	SRP063309	GSE72691	36	S	6	Boston University	52
NutrientResponse	SRP007372	SRP007372	NA	8	S	8	Pennsylvania State University	49
Profiling	PRJNA294837	SRP066851/SRP004049	GSE72691	11	S	9	Pennsylvania State University	47
Network_2	PRJNA295259	SRP063549	GSE72880	48	S	8	Boston University	52
Fe	PRJNA310120	SRP069025	GSE77354	12	S	4	University of Geneva	42
B12	PRJNA325540	SRP076516	NA	2	S	2	Pennsylvania State University	54
Doc	PRJNA342321	SRP087652	NA	2	S	2	Pennsylvania State University	53
RNase3	PRJNA387916	SRP107964	GSE99279	24	S	8	University of Wisconsin-Madison	44
RNAdecay	PRJNA429921	SRP130967	GSE109174	21	S	7	University of Wisconsin-Madison	43
Photosynthesis	PRJNA777890	SRP344575	NA	12	P	4	University of Southampton	56
JGI	PRJNA80125	SRP007768	NA	5	S/P	5	Joint Genome Institute	45
			Total # of Samples:	269	Total # of Conditions:	108		

Table 2. *Synechococcus sp.* PCC 7002 datasets collected from the NCBI database by bio-project. The 'Project Tag' column was used as a higher-level identifier during the quality check of the data. 'Bioproject ID', 'SRA ID', and 'GEO ID' columns are identifiers used in the NCBI, NCBI: SRA, and GEO databases respectively. I collected 269 RNA-Seq samples from 20 bio-projects representing 108 unique growth conditions.

Project Tag	NCBI: Bioproject ID	NCBI: SRA ID	GEO ID	# of Samples	Raw Read Format	# Unique Conditions	Submitting Entity	Citation #
StressR	PRJNA10645	SRP003368	Gp0000422	14	P	14	Joint Genome Institute	59
Hi_Res	PRJNA140271	SRP006795	GSE29264	3	S	1	Harvard University	75
FFA	PRJNA196229	SRP020509	GSE45762	17	S	6	Sandia National Laboratories	70
RpaA	PRJNA221220_aws	SRP030395	GSE51112	18	S	18	Harvard University	65
PRJNA259562	PRJNA259562	SRP045863	NA	1	S	1	Joint Genome Institute	59
Nstarv	PRJNA315938	SRP072154	GSE79726	6	P	3	Korea Institute of Science and Technology	60
ClockRes	PRJNA354335	SRP093663	GSE89999	24	S	24	Harvard University	67
ppGpp	PRJNA401742	SRP117070	GSE103462	8	S	4	Harvard University	68
ppGpp	PRJNA401777	SRP117071	GSE103463	8	S	4	Harvard University	68
ppGpp	PRJNA403840	SRP117169	GSE103644	6	S	6	Harvard University	68
ppGpp	PRJNA404081	SRP117265	GSE103704	36	S	18	Harvard University	68
Light	PRJNA412032	SRP118803	GSE104203	60	S	30	Harvard University	66
ppGpp	PRJNA415380	SRP120955	GSE105774	36	S	18	Harvard University	68
Sigma	PRJNA472248	SRP148555	GSE114693	72	S	34	Harvard University	62
H2O2	PRJNA506580	SRP170141	GSE122841	4	P	4	Kyungpook National University	63
DHAR	PRJNA588336	SRP229334	GSE140121	4	P	4	Kyungpook National University	64
CoCulture	PRJNA642094	SRP269594	NA	6	P	2	National Renewable Energy Laboratory	77
Mixotroph	PRJNA729175	SRP319593	NA	9	S	3	Shandong University	73
Glucose	PRJNA740138	SRP325478	NA	4	P	4	Chinese Academy of Sciences	69
Biofilm	PRJNA845529	SRP378421	GSE205444	21	S	7	University of California San Diego	71
HS199	PRJNA847037	SRP379051	NA	12	P	4	Chinese Academy of Sciences	72
UVTol	PRJNA854269	SRP384271	NA	10	P	4	University of California San Diego	76
HS	PRJNA888938	SRP401786	NA	6	P	2	Shandong University	74
SaltTol	PRJNA917560	SRP415570	GSE222067	12	P	4	Tianjin University	61
			Total # of Samples:	397	Total # of Conditions:	219		

Table 3. *Synechococcus elongatus sp.* PCC 7942 datasets collected from the NCBI database by bio-project. The 'Project Tag' column was used as a higher-level identifier during the quality check of the data. 'Bioproject ID', 'SRA ID', and 'GEO ID' columns are identifiers used in the NCBI, NCBI: SRA, and GEO databases respectively. I collected 397 RNA-Seq samples from 24 bio-projects representing 219 unique growth conditions.

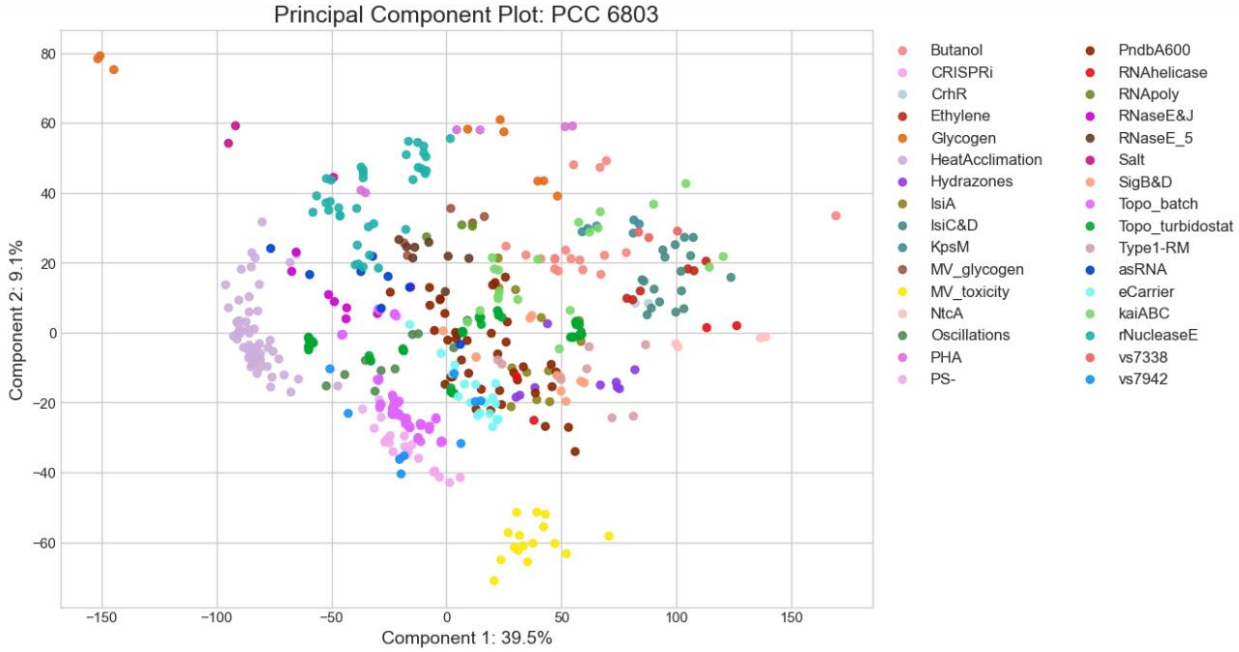

Figure 1. Principle component analysis of the 515 samples of *Synechocystis* PCC 6803 from 32 bio-projects used in this study. Each bio-project was given a tag unique tag. Samples are grouped and colored by bio-project.

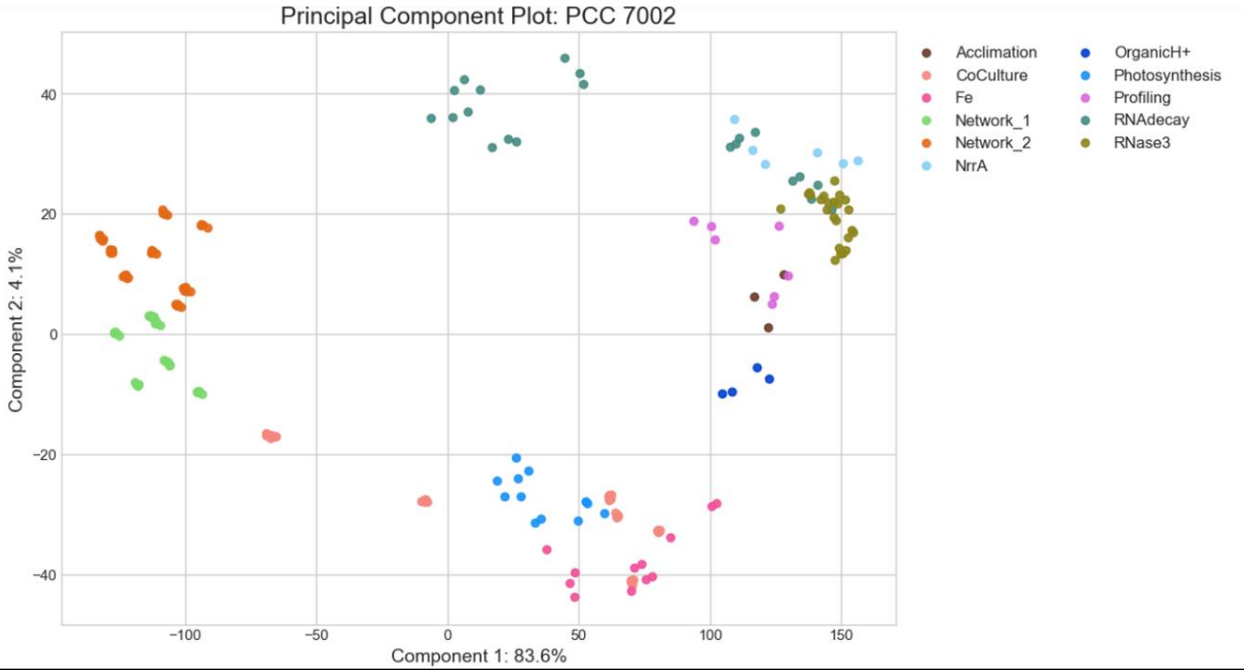

Figure 2. Principle component analysis of the 215 samples of *Synechococcus* PCC 7002 from 11 bio-projects used in this study. Each bio-project was given a tag unique tag. Samples are grouped and colored by bio-project.

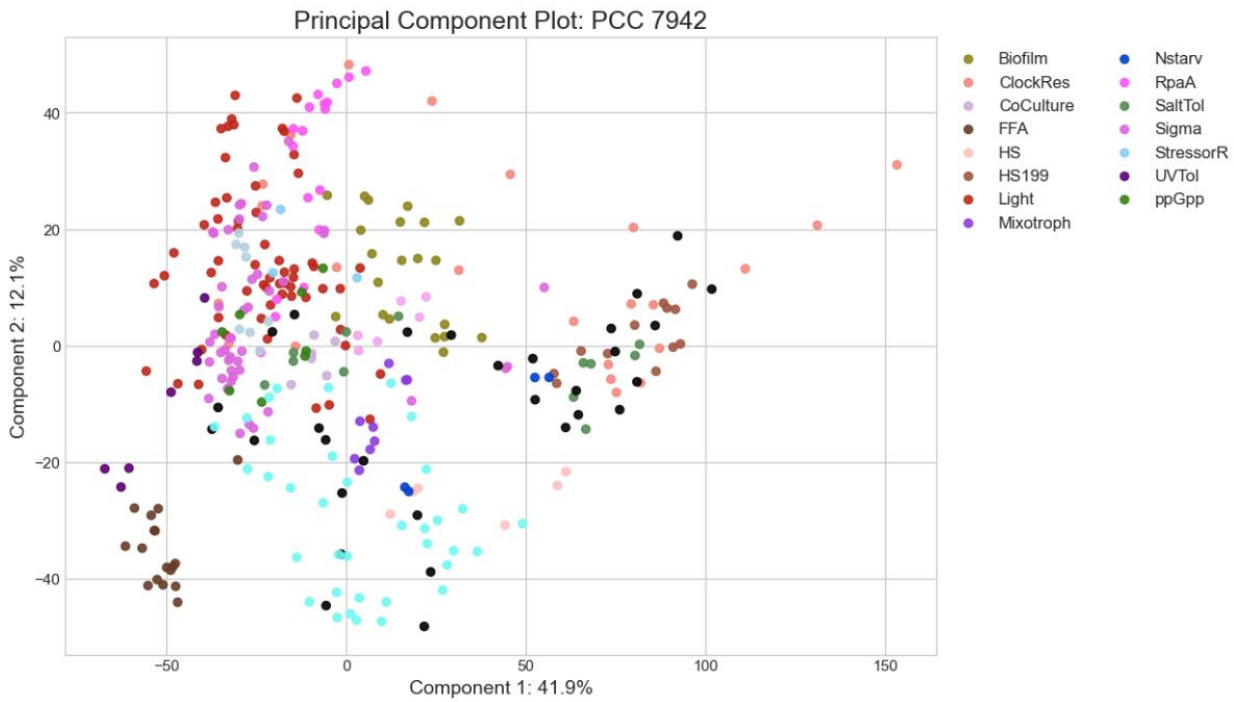

Figure 3. Principle component analysis of the 333 samples of *Synechococcus elongatus* PCC 7942 from 15 bio-projects used in this study. Each bio-project was given a tag unique tag. Samples are grouped and colored by bio-project.

Figure 4. Gene co-expression network for *Synechocystis* PCC 6803. Colors show clusters detected by the FastGreedy algorithm (n_{min} = 12 nodes). Sizing of the nodes and edges represents the betweenness centrality of that node or edge. The network was filtered to have the strongest ~6000 edges and then edges not connected to the main network were discarded. This network contains 4252 edges.

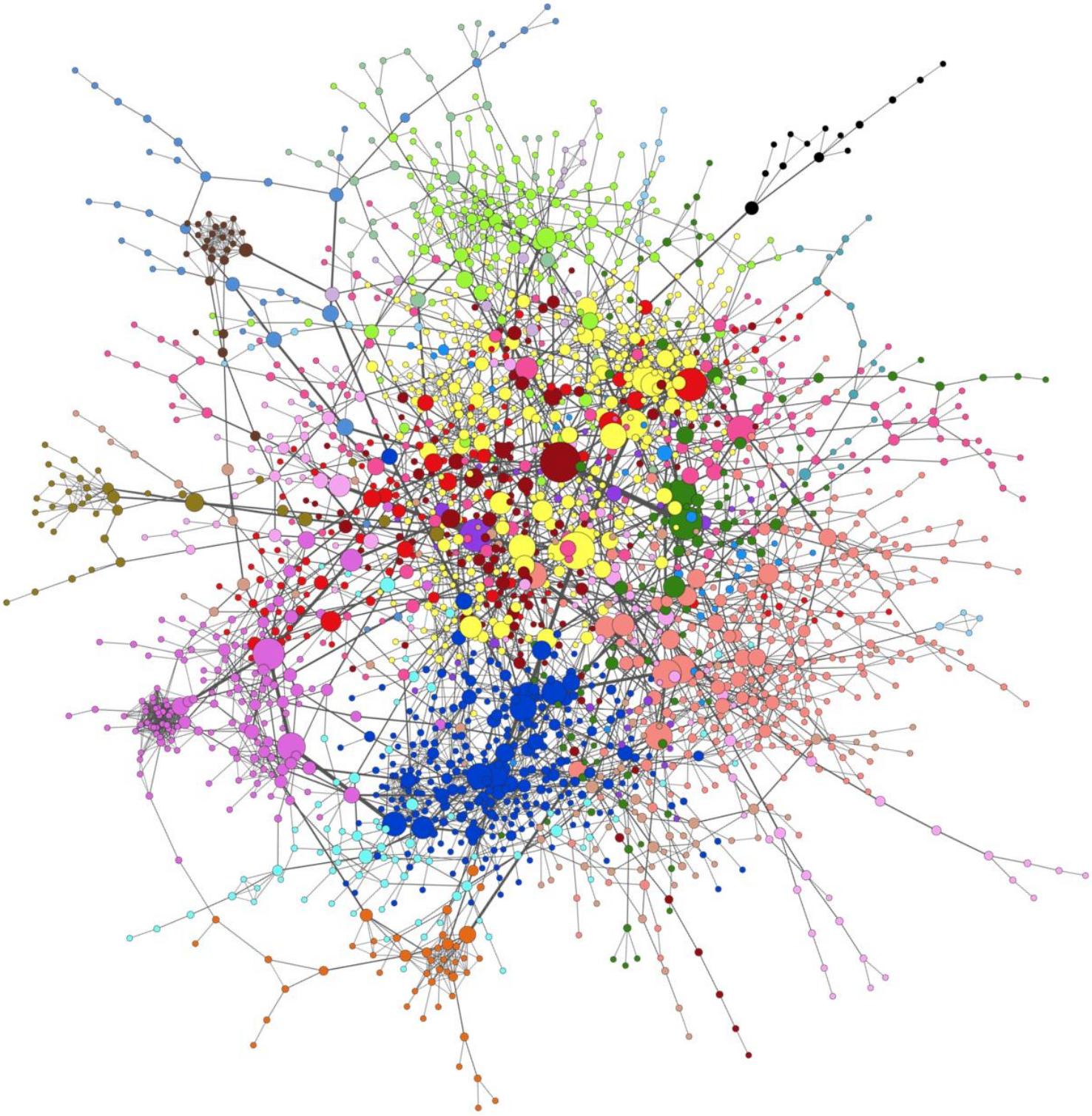

Figure 5. Gene co-expression network for *Synechococcus* PCC 7002. Colors show clusters detected by the FastGreedy algorithm (n_{min} = 12 nodes). Sizing of the nodes and edges represents the betweenness centrality of that node or edge. The network was filtered to have the strongest ~5000 edges and then edges not connected to the main network were discarded. This network contains 4380 edges.

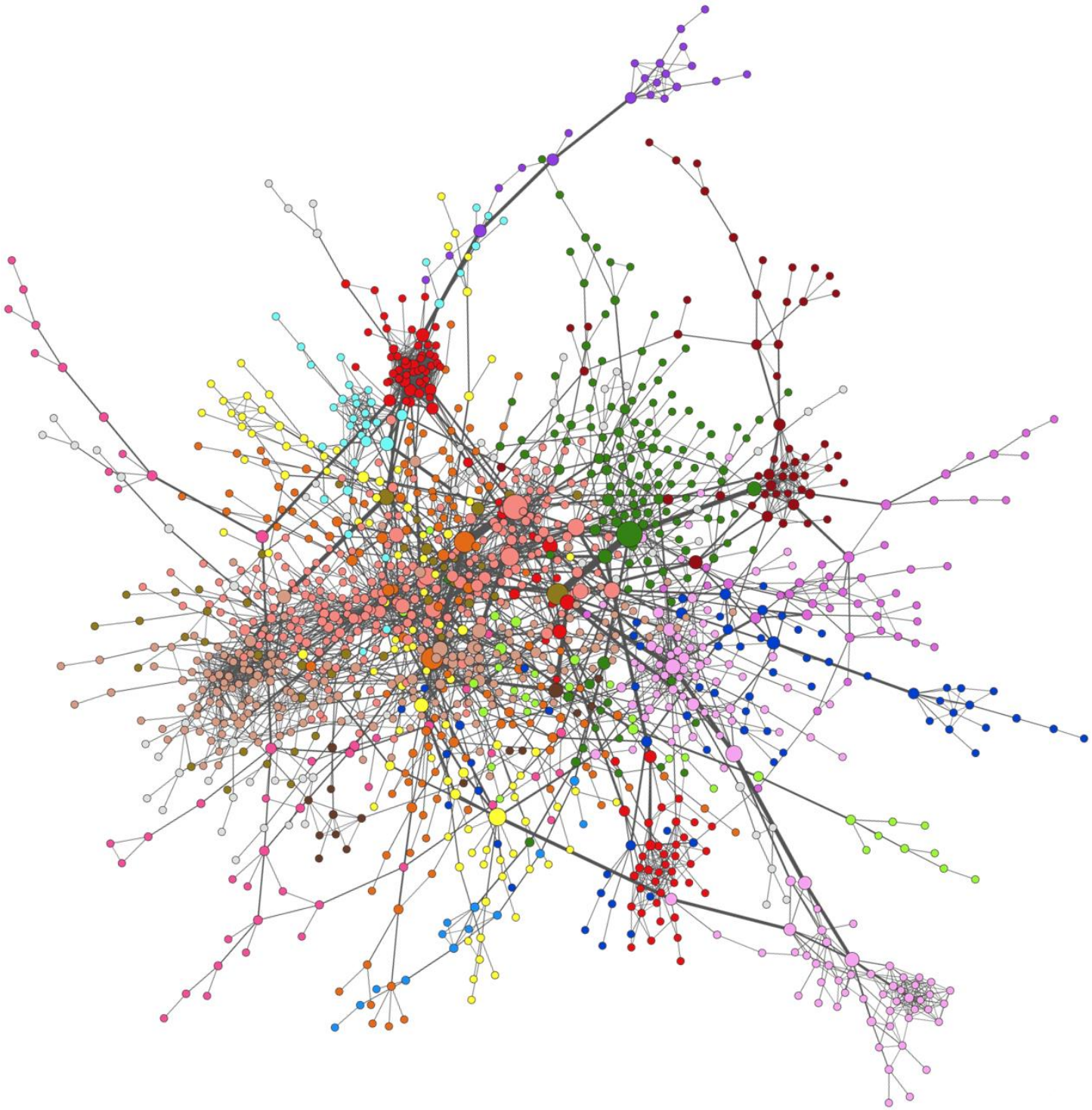

Figure 6. Gene co-expression network for *Synechococcus elongatus* PCC 7942. Colors show clusters detected by the FastGreedy algorithm (n_{min} = 12 nodes). Sizing of the nodes and edges represents the betweenness centrality of that node or edge. The network was filtered to have the strongest ~6000 edges and then edges not connected to the main network were discarded. This network contains 4255 edges.

References

1. A. Behle, M. Dietsch, L. Goldschmidt, W. Murugathas, L. C. Berwanger, J. Burmester, L. Yao, D. Brandt, T. Busche, J. Kalinowski, E. P. Hudson, O. Ebenhöh, I. M. Axmann and R. Machné, Nucleic Acids Research 50 (22), 12790–12808 (2022).
2. K. Billis, M. Billini, H. J. Tripp, N. C. Kyrpides and K. Mavromatis, PLoS ONE 9 (10), e109738 (2014).
3. M. Cavaiuolo, C. Chagneau, S. Laalami and H. Putzer, Frontiers in Microbiology 11 (2020).
4. Y. Cheng, T. Zhang, Y. Cao, L. Wang and W. Chen, Applied Microbiology and Biotechnology 105, 4693-4707 (2021).
5. Y. Cheng, T. Zhang, L. Wang and W. Chen, Applied and Environmental Microbiology 86 (13), e00517-00520 (2020).
6. S.-H. Cho, Y. Jeong, S.-J. Hong, H. Lee, H.-K. Choi, D.-M. Kim, C.-G. Lee, S. Cho and B.-K. Cho, mSystems 6 (6), e00943-00921 (2021).
7. U. o. Essex, (NCBI, 2021).
8. U. o. Freibrug, (NCBI, 2019).
9. U. o. Freiburg, (NCBI, 2019).
10. R. García-Cañas, J. Giner-Lamia, F. J. Florencio and L. López-Maury, PNAS 118 (5), e2017898118 (2021).
11. J. Giner-Lamia, R. Robles-Rengel, M. A. Hernández-Prieto, M. I. Muro-Pastor, F. J. Florencio and M. E. Futschik, Nucleic Acids Research 45 (20), 11800-11820 (2017).
12. U. A. Hoffmann, F. Heyl, S. N. Rogh, T. Wallner, R. Backofen, W. R. Hess, C. Steglich and A. Wilde, Nucleic Acids Research 49 (22), 13075–13091 (2021).
13. U. A. Hoffmann, E. Lichtenberg, S. N. Rogh, R. Bilger, V. Reimann, F. Heyl, R. Backofen, C. Steglich, W. R. Hess and A. Wilde, BIORXIV- PREPRINT (2023).
14. Hongji Zhu, Xiaoyue Ren, Jiangxin Wang, Zhongdi Song, Mengliang Shi, Jianjun Qiao, Xiaoxu Tian, Jie Liu, L. Chen and W. Zhang, Biotechnology for Biofuels 6, 106 (2013).
15. L. Hu, J. He, M. Dong, X. Tang, P. Jiang, A. Lei and J. Wang, Algal Research 47, 101856 (2020).
16. X. Hu, T. Zhang, K. Ji, K. Luo, L. Wang and W. Chen, Applied Microbiology and Biotechnology 105, 8377–8392 (2021).
17. F.-S.-U. Jena, (NCBI, 2019).
18. Y. Jeong, S.-J. Hong, S.-H. Cho, S. Yoon, H. Lee, H.-K. Choi, D.-M. Kim, C.-G. Lee, S. Cho and B.-K. Cho, Frontiers in Microbiology (2021).
19. J. Karlsen, J. Asplund-Samuelsson, M. Jahn, D. Vitay and E. P. Hudson, Frontiers in Microbiology 12 (2021).
20. J. Karlsen, J. Asplund-Samuelsson, Q. Thomas, M. Jahn and E. P. Hudson, mSystems 3 (5), e00126-00118 (2018).
21. M. Kellom, Arizona State University (2017).

22. E. V. K. Kirill S. Mironov, Maria Shumskaya, Dmitry A. Los,, Gene 764, 145055 (2021).
23. M. Kopf, S. Klähn, I. Scholz, J. K. F. Matthiessen, W. R. Hess and B. Voß, DNA Research 21 (5), 527-539 (2014).
24. R. F. Lacey, C. J. Allen, A. Bakshi and B. M. Binder, Plant Direct 2 (3), e00048 (2018).
25. N. S. Lau, C. P. Foong, Y. Kurihara, K. Sudesh and M. Matsui, PLoS One 9 (1), e86368 (2014).
26. M. A. Madsen, G. Hamilton, P. Herzyk and A. Amtmann, Frontiers in Bioengineering and Biotechnology 8 (2021).
27. A. Migur, F. Heyl, J. Fuss, A. Srikumar, B. Huettel, C. Steglich, J. S. S. Prakash, R. Reinhardt, R. Backofen, G. W. Owttrim and W. R. Hess, Journal of Experimental Botany 72 (21), 7564-7579 (2021).
28. P. Pichaiyotinkul, N. Ruankaew, A. Incharoensakdi and T. Monshupanee, World Journal of Microbiology and Biotechnology 39, 27 (2022).
29. A. Riaz-Bradley, K. James and Y. Yuzenkova, Nucleic Acids Research 48 (3), 1241-1352 (2020).
30. M. Riediger, P. Spät, R. Bilger, K. Voigt, B. Maček and W. R. Hess, The Plant Cell 33 (2), 248-269 (2021).
31. M. Santos, S. B. Pereira, C. Flores, C. Príncipe, N. Couto, E. Karunakaran, S. M. Cravo, P. Oliveira and P. Tamagnini, mSphere 6 (1), e00003-00021 (2021).
32. N. Sukkasam, A. Incharoensakdi and T. Monshupanee, Plant and Cell Physiology 63 (9), 1253-1272 (2022).
33. K. Systems and Synthetic Biology Lab, (2014).
34. O. Turunen, S. Koskinen, J. Kurkela, O. Karhuvaara, K. Hakkila and T. Tyystjärvi, Life 12 (2), 162 (2022).
35. D. Wu, Y. Wang and X. Xu, Frontiers in Microbiology 11 (11) (2020).
36. W. Xu, H. Chen, C.-L. He and Q. Wang, PLoS ONE 9 (3), e92711 (2014).
37. L. Yao, K. Shabestary, S. M. Björk, J. Asplund-Samuelsson, H. N. Joensson, M. Jahn and E. P. Hudson, Nature Communications 11, 1666 (2020).
38. H. Zhang, Q. Liu, Q. Liang, B. Wang, Z. Chen and J. Wang, Frontiers in Microbiology 13 (2023).
39. Y. Zhou, Y. Qin, H. Zhou, T. Zhang, J. Feng, D. Xie, L. Feng, H. Peng, H. He and M. Cai, Pesticide Biochemistry and Physiology 184, 105098 (2022).
40. M. B. Begemann, E. K. Zess, E. M. Walters, E. F. Schmitt, A. L. Markley and B. F. Pfleger, PLoS One 8 (10), e76594 (2013).
41. 2. A. S. Beliaev, M. F. Romine, M. Serres, H. C. Bernstein, B. E. Linggi, L. M. Markillie, N. G. Isern, W. B. Chrisler, L. A. Kucek, E. A. Hill, G. E. Pinchuk, D. A. Bryant, H. S. Wiley, J. K. Fredrickson and A. Konopka, ISME J 8 (11), 2243-2255 (2014).
42. S. Blanco-Ameijeiras, C. Cosio and C. S. Hassler, Frontiers in Marine Science 4 (2017).
43. G. C. Gordon, J. C. Cameron, S. T. P. Gupta, M. D. Engstrom, J. L. Reed and B. F. Pfleger, mSystems 5 (4), e00224-00220 (2020).
44. G. C. Gordon, J. C. Cameron and B. F. Pfleger, Nucleic Acids Res 46 (4), 1984-1997 (2018).
45. J. G. Institute, (NCBI, 2011).

46. A. Krishnan, S. Zhang, Y. Liu, K. A. Tadmori, D. A. Bryant and C. G. Dismukes, Biotechnol Bioeng 113 (7), 1448-1459 (2016).
47. M. Ludwig and D. A. Bryant, Front Microbiol 2, 41 (2011).
48. M. Ludwig and D. A. Bryant, Front Microbiol 3, 354 (2012).
49. M. Ludwig and D. A. Bryant, Front Microbiol 3, 145 (2012).
50. M. Ludwig, T. T. Chua, C. Y. Chew and D. A. Bryant, Front Microbiol 6, 1217 (2015).
51. M. Ludwig, M.-E. Pandelia, C. Y. Chew, B. Zhang, J. H. Golbeck, C. Krebs and D. A. Bryant, J Biol Chem 289 (24), 16624-16639 (2014).
52. R. S. McClure, C. C. Overall, J. E. McDermott, E. A. Hill, L. M. Markillie, L. A. McCue, R. C. Taylor, M. Ludwig, D. A. Bryant and A. S. Beliaev, Nucleic Acids Res 44 (18), 8810-8825 (2016).
53. A. A. Pérez, Pennsylvania State University (2016).
54. A. A. Pérez, D. A. Rodionov and D. A. Bryant, Journal of Bacteriology 198 (19), 2753-2761 (2016).
55. Y. Shimmori, Y. Kanesaki, M. Nozawa, H. Yoshikawa and S. Ehira, Plant Cell Physiol 59 (6), 1225-1233 (2018).
56. A. Torrado, H. M. Connabeer, A. Rottig, N. Pratt, A. J. Baylay, M. J. Terry, C. M. Moore and T. S. Bibby, Plant Physiol 189 (4), 2554-2566 (2022).
57. S. Zhang, Y. Liu and D. A. Bryant, Metab Eng 32, 174-183 (2015).
58. S. Zhang, G. Shen, Z. Li, J. H. Golbeck and D. A. Bryant, J Biol Chem 289 (23), 15904-15914 (2014).
59. K. Billis, M. Billini, H. J. Tripp, N. C. Kyrpides and K. Mavromatis, PLoS ONE 9 (10), e109738 (2014).
60. S. Y. Choi, B. Park, I.-G. Choi, S. J. Sim, S.-M. Lee, Y. Um and H. M. Woo, Scientific Reports 6 (1), 30584 (2016).
61. E. L. Weiss, M. Fang, A. Taton, R. Szubin, B. Ø. Palsson, B. G. Mitchell and S. S. Golden, Proceedings of the National Academy of Sciences 119 (45), e2211789119 (2022).
62. Z. Dong, T. Sun, W. Zhang and L. Chen, Frontiers in Microbiology 14 (1123081) (2023).
63. K. E. Fleming and E. K. O'Shea, Cell Reports 25 (11), 2937-2945.e2933 (2018).
64. Y.-S. Kim, J.-J. Kim, S.-I. Park, S. Diamond, J. S. Boyd, A. Taton, I.-S. Kim, J. W. Golden and H.-S. Yoon, Frontiers in Plant Science 9 (2018).
65. Y.-S. Kim, S.-I. Park, J.-J. Kim, J. S. Boyd, J. Beld, A. Taton, K.-I. Lee, I.-S. Kim, J. W. Golden and H.-S. Yoon, Frontiers in Plant Science 11 (2020).
66. S. J. Markson, R. J. Piechura, M. A. Puszynska and K. E. O'Shea, Cell 155 (6), 1396-1408 (2013).
67. J. R. Piechura, K. Amarnath and E. K. O'Shea, eLife 6 (2017).
68. A. M. Puszynska and E. K. O'Shea, eLife 6 (2017).
69. A. M. Puszynska and E. K. O'Shea, Cell Reports 21 (11), 3155-3165 (2017).
70. C. A. o. S. Qingdao Institute of Bioenergy and Bioprocess Technology, (NCBI, 2022).
71. A. M. Ruffing, Biotechnology for Biofuels 6 (1), 113 (2013).

72. R. Simkovsky, R. Parnasa, J. Wang, E. Nagar, E. Zecharia, S. Suban, Y. Yegorov, B. Veltman, E. Sendersky, R. Schwarz and S. S. Golden, Frontiers in Microbiology 13 (2022).
73. H. Sun, G. Luan, Y. Ma, W. Lou, R. Chen, D. Feng, S. Zhang, J. Sun and X. Lu, Nature Communications 14 (2023).
74. L.-R. Tan, Y.-Q. Cao, J.-W. Li, P.-F. Xia and S.-G. Wang, Microbial Cell Factories 21 (1), 31 (2022).
75. S. University, (2023).
76. V. Vijayan, I. H. Jain and E. K. O'Shea, Genome Biol 12 (5), R47 (2011).
77. C. Zuñiga, T. Li, M. T. Guarnieri, J. P. Jenkins, C.-T. Li, K. Bingol, Y.-M. Kim, M. J. Betenbaugh and K. Zengler, Nature Communications 11 (1), 3803 (2020).
78. Y. Liao, G. K. Smyth and W. Shi, Nucleic Acids Research 47 (8), e47-e47 (2019).
79. V. Huynh-Thu, A. Irrthum, L. Wehenkel and P. Geurts, PLoS ONE 5(9): e12776 (2010).
80. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J. C. Marine, P. Geurts, J. A., J. van den Oord, Z. K. Atak, J. Wouters and S. Aerts Nature Methods 14: 1083-1086 (2017).
81. A. Clauset, M. E. J. Newman and C. Moore, Phys. Rev. E 70, 066111 (2004)
82. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Genome Research 13(11): 2498-504 (2003)

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov