

2014-07-02

Toward a systems-level understanding of gene regulatory, protein interaction, and metabolic networks in cyanobacteria

Hernandez-Prieto, MA

<http://hdl.handle.net/10026.1/8415>

10.3389/fgene.2014.00191

Frontiers in Genetics

Frontiers Media SA

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.



Toward a systems-level understanding of gene regulatory, protein interaction, and metabolic networks in cyanobacteria

Miguel A. Hernández-Prieto^{1*}, Trudi A. Semeniuk¹ and Matthias E. Futschik^{1,2}

¹ Systems Biology and Bioinformatics Laboratory, IBB-CBME, University of Algarve, Faro, Portugal

² Centre of Marine Sciences, University of Algarve, Faro, Portugal

Edited by:

Thierry Tonon, CNRS-UPMC, France

Reviewed by:

Brian Palenik, Scripps Institution of Oceanography, USA

Laurence Garczarek, CNRS-INSU-UPMC, France

*Correspondence:

Miguel A. Hernández-Prieto, University of Algarve, FCT, Ed. 8, Campus de Gambelas, 8005-139 Faro, Portugal
e-mail: mprieto@ualg.pt

Cyanobacteria are essential primary producers in marine ecosystems, playing an important role in both carbon and nitrogen cycles. In the last decade, various genome sequencing and metagenomic projects have generated large amounts of genetic data for cyanobacteria. This wealth of data provides researchers with a new basis for the study of molecular adaptation, ecology and evolution of cyanobacteria, as well as for developing biotechnological applications. It also facilitates the use of multiplex techniques, i.e., expression profiling by high-throughput technologies such as microarrays, RNA-seq, and proteomics. However, exploration and analysis of these data is challenging, and often requires advanced computational methods. Also, they need to be integrated into our existing framework of knowledge to use them to draw reliable biological conclusions. Here, systems biology provides important tools. Especially, the construction and analysis of molecular networks has emerged as a powerful systems-level framework, with which to integrate such data, and to better understand biological relevant processes in these organisms. In this review, we provide an overview of the advances and experimental approaches undertaken using multiplex data from genomic, transcriptomic, proteomic, and metabolomic studies in cyanobacteria. Furthermore, we summarize currently available web-based tools dedicated to cyanobacteria, i.e., CyanoBase, CyanoEXpress, ProPortal, Cyanorak, CyanoBIKE, and CINPER. Finally, we present a case study for the freshwater model cyanobacteria, *Synechocystis* sp. PCC6803, to show the power of meta-analysis, and the potential to extrapolate acquired knowledge to the ecologically important marine cyanobacteria genus, *Prochlorococcus*.

Keywords: meta-analysis, cyanobacteria, systems biology, networks, metabolic pathways

INTRODUCTION

Cyanobacteria have a unique position in the living world, as they are the only prokaryotes capable of performing oxygenic photosynthesis. This confers on them unique roles in the carbon and nitrogen cycles on this planet, as well as an important primary ecological function. They thrive in diverse habitats, ranging from desert crusts to open sea (Partensky et al., 1999; Garcia-Pichel and Pringault, 2001; Garcia-Pichel et al., 2003).

Ecologically, marine cyanobacteria are at the bottom of the food pyramid and, thus, provide organic matter directly or indirectly to virtually every other marine organism (Falkowski, 2012). Specifically, marine cyanobacteria play a role of global importance, accounting for approximately 25% of the total carbon fixation in oceans through their photosynthetic activity (Flombaum et al., 2013); this is a remarkable quota, considering that approximately half of the atmospheric carbon is fixed in oceans (Field et al., 1998). The main part of this budget is fixed by planktonic picocyanobacteria, which are the most abundant photosynthetic organisms on earth (Li et al., 1983). Their most prominent representatives are from the genera *Prochlorococcus*

and *Synechococcus*, which occur mainly in the euphotic zone of tropical and subtropical oceans (Flombaum et al., 2013). Notably, *Prochlorococcus* is the smallest known photosynthetic organism, ranging from 0.6 to 1 μm in diameter. Its abundance in oligotrophic marine areas reaches up to a million cells per milliliter, making them a key component of the global carbon cycle (Chisholm et al., 1988). Other planktonic marine genera, such as *Trichodesmium* and *Crocospaera*, play a key role in nitrogen fixation in oceans (Duce et al., 2008; Langlois et al., 2008). Finally, marine benthic cyanobacteria are less abundant and have a smaller impact on the global carbon and nitrogen cycle, but are more phylogenetically diverse than the planktonic cyanobacteria. Representative examples of nitrogen-fixing marine benthic genera are the heterocyst-forming *Calothrix* and *Scytonema*, the non-heterocyst forming *Lyngbya*, *Microcoleus*, *Phormidium*, *Schizothrix* (Hoffmann, 1999), and the unicellular *Cyanothece* (Reddy et al., 1993).

Besides their global ecological importance, cyanobacteria have attracted the interest of researchers for the study of photosynthesis (Sun et al., 1999; Nelson, 2011). Several characteristics,

such as their faster life cycle, simple nutrient requirements, smaller genomes, and in some cases, the ease with which genetic manipulation can be carried out (Frigaard et al., 2004; Heidorn et al., 2011; Ruffing, 2011), make them the preferred photosynthetic organism for laboratory studies. In addition to this basic research, cyanobacteria are the focus of biotechnological applications (Abed et al., 2009), especially related to energy production (Wang et al., 2012).

Their capacity for photosynthesis, along with their potential for valuable product production and bioremediation—in both ecological and technological contexts—has motivated intense research on their genomes, as a first step toward understanding these organisms. Their small genome size together with both the development and decreasing cost of new sequencing technologies have facilitated the annotation of a wide range of cyanobacterial genomes (Nakao et al., 2010; Shih et al., 2013). This plethora of genomes confers scientists with a great resource for comparative analysis (Kettler et al., 2007; Dufresne et al., 2008; Stanley et al., 2013), which is further fueled by the ongoing discovery of new genes, functions, and applications (reviewed in Scanlan et al., 2009; Partensky and Garczarek, 2010). It should be noted, however, that data generated, using these new genomic techniques, do not equally cover different cyanobacteria species. Historic biases toward model organisms, as well as the difficulty in culturing many free-living species, have both resulted in great quantities of data for specific organisms, and under-representation of others. Thus, in spite of the abundance of genomes, the vast majority of downstream studies have focused on relatively few genera, including *Synechocystis*, *Prochlorococcus*, *Synechococcus*, *Anabaena*, and *Nostoc* (Table S1).

However, the existence of a core set of genes, present in all sequenced cyanobacteria (Shi and Falkowski, 2008), might permit extrapolation from well-studied species to newly sequenced organisms, using systems-level approaches (Albert, 2007). In this context, tools from systems biology can integrate existing knowledge acquired from phylogenetically-related model organisms, like freshwater *Synechocystis* sp. PCC6803 (hereafter *Synechocystis* 6803), with new data from recently sequenced marine cyanobacteria. For example, genome comparison studies, together with *in silico* models of molecular networks provide researchers with a powerful framework to build new knowledge into, without starting from scratch for each new species. Similarly, network representations of relationships between genes (leading to regulatory networks), proteins (interaction networks), and metabolites (metabolic pathways) are intuitive systems-level tools to integrate and consolidate data from different studies, or even different organisms; such networks have been successfully applied to many diverse biological groups to date. While networks constructed directly from medium to large-scale experimental data are viewed with greater confidence than those obtained through computational extrapolation of reference networks, the later approach can still provide important supplementary information for less well-studied species, as is the case for many of the marine cyanobacteria. Thus, system-tools can provide an attractive framework for research into marine cyanobacteria.

“OMIC” DATA: AN OVERVIEW OF THE IMPACT OF NEW TECHNOLOGIES

We can date the beginning of the “omic” era in cyanobacteria to the sequencing of the first cyanobacterium genome, i.e., freshwater cyanobacterium *Synechocystis* 6803 (Kaneko et al., 1996). Its early sequencing was in part due to the relative small genome size of *Synechocystis* 6803, with a chromosome length of 3573 kb (47.72% GC content), plus 383 kb distributed between seven megaplasms (http://genome.microbedb.jp/cyanobase/*Synechocystis*); and in part to its established presence as a model organism in many laboratories (e.g., there are 431 entries for *Synechocystis* 6803 in PubMed Central between 1950 and 1995). Notably, the availability of its genome further enhanced a bias in the literature toward *Synechocystis* 6803, making it a preferred organism for biotechnological studies, illustrated by the fact that there are almost twice as many PubMed Central records from 1996 to 2013 for *Synechocystis* 6803 than for the next most cited cyanobacterium genus *Anabaena* (2623 compared with 1580 for *Anabaena*). This bias in the literature is remarkable given that the completely sequenced genome of *Anabaena* sp. PCC 7120 was published in 2001 (Kaneko et al., 2001).

Nevertheless, in the last decade, advances in high-throughput techniques (HTT) for genetic analysis have compounded the data available for marine and other cyanobacteria on genomic, transcriptomic, proteomic and metabolomic levels (Ow and Wright, 2009). Importantly, this technological revolution in “omic” data has paved the way to the development of methods to collate and integrate such data into a systems-level framework (Figure 1). We highlight some of the major advancements in these fields for cyanobacteria—with a focus on marine cyanobacteria—below.

FROM GENOMICS TO METAGENOMICS: FROM THE OCEAN TO THE LABORATORY AND BACK TO THE OCEAN

Assemblage of the genome sequences of members of the two most abundant marine cyanobacteria genera took place almost a decade later than for *Synechocystis* 6803. In Dufresne et al. (2003) presented the genome of the low-light adapted strain *Prochlorococcus marinus* SS120, while Rocap et al. (2003) published the genome of the high-light adapted strains *Prochlorococcus marinus* MED4 and MIT9313; in the same issue as Rocap et al. (2003), the genome sequence of *Synechococcus* sp. WH8102 was described by Palenik et al. (2003). More recently, genomic data acquisition has been driven by the development of so-called next generation sequencing techniques (to differentiate them from the sequencing technique, originally developed by Sanger) and a drastic decrease in cost (Liu et al., 2012). Currently, data are generated in the pursuit of scientific and biotechnological objectives in multiple species-specific genome projects, as well as in global metagenomic projects, in which cyanobacteria are also identified (Figure 2). The plethora of data resulting from these projects is commonly available through public repositories for the benefit of the scientific community. Genomic data for cyanobacteria are accessible through several general as well as organism-specific repositories (Table 1). Specific repositories include CyanoBase (Nakamura et al., 1998) with 39 cyanobacterial genomes available, as well as repositories focussed exclusively on marine picocyanobacteria, i.e.,

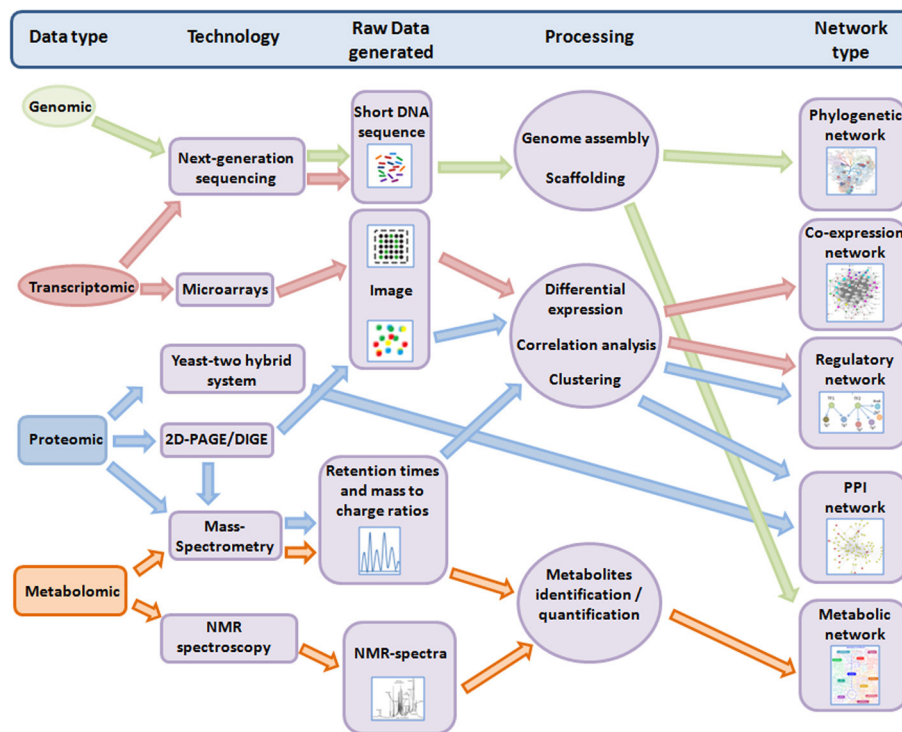


FIGURE 1 | Overview of the workflow from different “omic” methods to different systems-level networks. Only technologies discussed in this review are shown. For more details, the manually-curated meta-database OMICtools (<http://omictools.com/>)

provides tools and platforms for multi-omic data analysis. Note: 2-D PAGE, two-dimensional polyacrilamide gels; DIGE, difference gel electrophoresis; NMR, nuclear magnetic resonance; PPI, protein-protein interaction.

Cyanorak (Dufresne et al., 2008; Scanlan et al., 2009) and ProPortal (Kelly et al., 2012). While more general repositories are the integrated microbial genomes (IMG) database from the Department of Energy (DOE, USA) (Markowitz et al., 2012) with 89 complete cyanobacterial genomes, and the National Center for Biotechnological Information (NCBI) database with more than 100 cyanobacterial genome sequences available. A great part (54) of the genomes included in NCBI database resulted from the CyanoGEBA project (Shih et al., 2013), which aims to enhance the phylogenetic diversity available in public repositories by providing information on cyanobacterial taxa that were previously under-represented (Table S1). In addition to laboratory-based efforts, global-scale metagenomic projects, such as the Global Ocean Survey (Rusch et al., 2007), have sequenced populations of marine microorganisms collected around the globe, and vastly extended the number of sequences available to the research community and industry (reviewed in Lorenz and Eck, 2005). For instance, metagenomic data for marine habitats are available through CAMERA (<http://camera.calit2.net/>). Thus, our research efforts over the last decade have significantly expanded our knowledge of cyanobacteria compared with the status for earlier reviews in this field (e.g., Burja et al., 2003).

Clearly, the public availability of genome sequences for different cyanobacteria has had considerable impact on research directions. It helped laboratory-based researchers to direct molecular

and biochemical work toward specific genes, e.g., toward those identified as novel, or toward those conserved in other organisms. For instance, new genome sequences can be exploited to identify “orphan pathways,” in which metabolites were previously detected, but not the gene clusters responsible for their biosynthesis (Gross, 2007). Indeed, this was the case for the pathway responsible for the biosynthesis of patellamides (didemnid peptides with potential medical applications) (Ireland et al., 1982; Williams and Jacobs, 1993). This metabolite was initially detected in the symbiotic cyanobacteria *Prochloron didemni* (Degnan et al., 1989), and later in the bloom-forming cyanobacterium *Trichodesmium erythraeum* IMS101. Sequencing of the gene set responsible for its synthesis in *P. didemni* and function confirmation through cloning in *Escherichia coli* facilitated the identification of the counterpart gene cluster, based on similarity, in *T. erythraeum* IMS101 (Schmidt et al., 2005).

Unfortunately, newly identified functions of genes or proteins do not necessarily result in the update of annotations in genome repositories, which typically lag behind our current knowledge. To address this problem, a community-database named CYORF (<http://cyano.genome.ad.jp/>) was set up to annotate newly described functions with their corresponding genes, and to allow scientists to actively curate these annotations. Just the initial effort within the Japanese scientific community resulted in about 1000 gene function re-annotations (Furumichi et al., 2002). This database has been superseded by a social genome annotation



FIGURE 2 | World map showing genomic records for cyanobacteria generated using an interface powered by google maps, available on the Genomes Online Database (GOLD) website (<http://genomesonline.org/cgi-bin/GOLD/index.cgi>). Red labels indicate the original location of a specifically sequenced strain. Labels direct the user to information on the

organism, genome characteristics (i.e., GC content, size), sequencing method used, specific coordinates of the origin of the strain, as well as links to external databases (as shown for *Prochlorococcus marinus* NATL1A). The GOLD also describes the status of each record in tabular format.

Table 1 | Publicly available full genome sequences for cyanobacteria in various repositories as at April 2014.

Name	Type	No. full genomes	# <i>P</i> records	* <i>S</i> records	Weblink
Cyanorak	MC	14	3	11	http://www.sb-roscoff.fr/cyanorak
ProPortal	MC	68 (44 phages)	13	11	http://proportal.mit.edu/
CyanoBase	C	39	12	10	http://genome.microbedb.jp/cyanobase/
CyanoBIKE	C	27	11	12	http://biobike-8003.csbc.vcu.edu/biologin
NCBI	G	103	12	13	http://www.ncbi.nlm.nih.gov/
IMG (JGI)	G	208	17	15	http://img.jgi.doe.gov
KEGG	G	75	12	13	http://www.genome.jp/kegg/
Microbes online	G	50	17	12	http://www.microbesonline.org/
SEED	G	161	28	35	http://pubseed.theseed.org/

#*Prochlorococcus*; **Synechococcus*; MC, marine cyanobacteria; C, cyanobacteria; G, general.

tool called TogoAnnotation (<http://togo.annotation.jp/>) open to the entire scientific community (Fujisawa et al., 2014).

Despite public accessibility to genomic data, only few comparative studies have been carried out to date (Scanlan et al., 2009). For example, a comparison of *Prochlorococcus* genomes distinguished between “core” genes present in all, and “flexible” genes that were not conserved in all the branches of the phylogenetic tree (Kettler et al., 2007). This approach was expanded to include other cyanobacteria species, including fresh-water types, reducing the number of core genes from 1273 to 323. These core genes are significantly enriched in key photosynthetic (12%) and ribosomal proteins (7%) over other functional categories (Shi and Falkowski, 2008) (Figure 3). The presence of core and flexible genes also served to estimate the importance and relative contribution of vertical inheritance vs. horizontal gene transfer for each of these gene fractions in 11 *Synechococcus* strains. The

estimated number of gene families present in the core genome of these *Synechococcus* strains was 1572, adding three *Prochlorococcus* strains to this comparative analysis reduced the number to 1228 gene families. The number of unique genes varied greatly between strains, from 91 to 845 genes, and was correlated with genome size. Two *Synechococcus* subclusters WH5701 and RCC307 were an exception, since they presented a higher number of unique genes than expected for their genome size. The presence of these unique genes in genomic islands and their horizontal transfer likely confers advantages for adaption to narrow ecological niches (Dufresne et al., 2008). Avrani et al. (2011) also reinforced the importance of genomic variability in adaptation of natural populations. Their work revealed the importance of genomic variability within a bacterial community for viral resistance. Starting with four high-light adapted *Prochlorococcus* strains, 77 sub-strains resistant to infection by 10 different podoviruses were

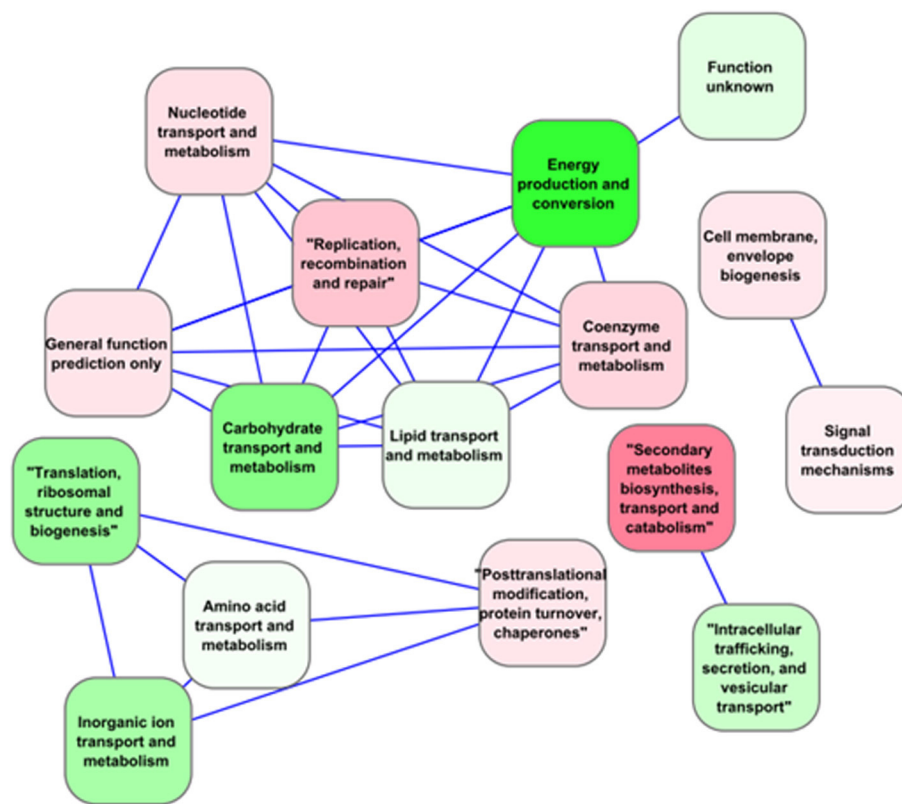


FIGURE 3 | Correlation network for functional categories (as defined by Falkowski, 2012) based on the expression of “core” genes in *Synechocystis* sp. PCC 6803 under multiple environmental conditions extracted from CyanoExpress. Nodes represent Gene Ontology (GO)

pathways, colored based on their average differential expression in the dark such that gradients of red indicate induction, while green indicate repression. Only categories with an absolute Spearman correlation value greater than 0.95 were connected by an edge.

selected to further characterize the genomic region responsible for this resistance. Using next-generation sequencing technologies, 27 resistant and eight control sub-strains were sequenced to identify resistance specific mutations. Other sub-strains were screened by sequencing of PCR amplicons. These results demonstrated that phage infection promotes enhanced diversity. It also revealed the importance of genomic islands (regions in the genome acquired by horizontally transfer) in phage resistance, since most mutations accumulated in these regions. The preferential location of viral-attachment genes in genomic islands imposes a selective pressure on genomic islands, and island gene-content diversity. These findings have direct implications on our understanding of the ecology of *Prochlorococcus* communities and of the mechanisms supporting the long-term coexistence of this cyanobacterium with its phages (Avrani et al., 2011).

Data from metagenomic studies have been very useful for clarification of the genetic content of cyanobacteria. Sampling and carrying out genomic DNA isolation directly from a particular environmental niche circumvents problems related to the difficulty in culturing many of these organisms. Such was the case for the unicellular N₂-fixing UCYN-A cyanobacteria, where attempts to cultivate have so far failed. UCYN-A nitrogenase genes are maximally expressed during the light period (Church et al., 2005; Zehr et al., 2007), contrary to the expression

pattern observed for *Cyanothece* nitrogenase genes, which separate temporally oxygen evolution from nitrogen fixation to avoid inhibitory effects (Stockel et al., 2008; Toepel et al., 2008; Welsh et al., 2008). The use of enriched fractions obtained by fluorescence-activated cell sorting helped to overcome the difficulty to separate UCYN-A cells from other small phototrophic and heterotrophic populations using traditional flow cytometry. Metagenomic analysis of this enriched fraction resulted in at least a 10-fold genomic coverage of the UCYN-A genome. Comparison of these results with the cyanobacterial core genome indicated that at least 79% of the core genes, as well as the nitrogenase genes were present in the UCYN-A genome. Strikingly, no UCYN-A sequences encoding proteins involved in CO₂ concentration, CO₂ fixation or Photosystem II were detected. The absence of Photosystem II and thus, of light-driven oxygen evolution, can explain the expression pattern of the nitrogenase genes in this strain (Zehr et al., 2008). Flow cytometry-based sorting was also used to enrich marine samples with cyanobacteria of the genus *Synechococcus*. This enrichment facilitated the assembly of contigs and the identification of at least three distinct plasmids lacking in genomes from model strains. The data obtained from a natural population showed a great genomic diversity compared with model *Synechococcus* strains isolated from the same environment, stressing the importance

of horizontal gene transfer in natural populations (Palenik et al., 2009).

In general, genomic data from free-living organisms give researchers a much wider view of their molecular and ecological diversity, as well as their population dynamics. For example, the number of *Prochlorococcus* species estimated from ocean metagenomic data is in the order of 35 (Thompson et al., 2013). A good example of how metagenomic data from the global ocean survey collection have been exploited, is in the search for specific gene families. This was the case for phosphate acquisition (Martiny et al., 2009a) and nitrate/nitrite assimilation genes (Martiny et al., 2009b). Regarding phosphate acquisition, the authors established a correlation between ortho-phosphate availability in different oceanic regions and the presence or absence of phosphate uptake genes (Martiny et al., 2009a). A similar approach was taken in the later study to locate genes involved in nitrate/nitrite assimilation. Sequences related to nitrogen assimilation were sorted phylogenetically, based on their characteristic GC content and on detected homology (using both blastx and blastn) of their corresponding paired-end sequences. This distinction allowed the authors to assign sequences to specific clades or species. Importantly, this study served to establish the existence of *Prochlorococcus* strains, harboring genes that encode transporters and reductases for both nitrite and nitrate, and thus, capable of capturing and using these forms of nitrogen. This is a remarkable finding; since it was previously thought that genes for nitrate assimilation were absent in *Prochlorococcus* strains, based on the available genome sequences from cultured strains. It also indicates that generalizations based only on sequenced genomes from cultured strains should be treated with caution.

Despite the great potential of metagenomic data, identifying novel molecular mechanisms remains difficult due to inherent ambiguities of sequence assembly. For instance, the work undertaken by two research groups, using metagenomic data obtained from high-nutrient low-chlorophyll oceanic regions, revealed new clades of high-light adapted *Prochlorococcus*, lacking culture representatives. The identification of these clades was based on phylogenetic analysis of “core” functional genomes (Rusch et al., 2010), or through identification of 16S rRNA sequences (West et al., 2011). Due to assembly problems, these studies did not describe any previously unidentified gene. To circumvent this problem, Malmstrom and co-workers combined analysis of metagenomic data with single cell genomics of selected strains. Their analysis of just 10 single cells identified 394 genes, not previously described in *Prochlorococcus* strains (Malmstrom et al., 2013). In particular, they characterized genes encoding new siderophore-mediated iron scavenging mechanisms employed by *Prochlorococcus*. These genes represent an adaptation to the low iron concentrations in the ocean region, where these sequences were obtained. Curiously, the genes involved in siderophore transport are located within a genomic island, indicating that their origin was through horizontal transfer, probably mediated by phages (Kettler et al., 2007; Lindell et al., 2007).

Metagenomic data have also been used to predict functional relationships between conserved protein domains with unknown function, and those participating in a particular cellular function. This approach assumes that functional domains are only retained

in the genome, when their presence gives an organism some competitive advantage in a particular environment. Under this premise, if the presence of an unknown protein domain correlates with that of domains of known function, it is inferred that it is involved in a same function; this principle of inference is commonly known as “guilt-by-association.” Formally, it can be treated using graphs, where functional domains or genes are represented by nodes. The presence of two nodes in similar niches (high correlation) is represented by an edge. The graphical association of unknown domains with those related to a particular function is taken as indicative of their functional role (Buttigieg et al., 2013). This “guilt-by-association” principle is also used in gene functional prediction based on clustering of RNA expression data, as discussed in the section on web-based tools.

TRANSCRIPTOMICS: GENE EXPRESSION PROFILING FROM MICROARRAYS TO RNAseq

A great advantage of having the full genome sequence is that it facilitates the development of comprehensive microarrays for transcriptome profiling. Thus, it is not surprising that the early publication of the *Synechocystis* 6803 genome resulted in the advent of the first whole genome microarray platform for a cyanobacterium, denoted as cyanoCHIP, and commercialized by Takara Bio. The first articles using this platform were published in 2001, dealing with changes in gene expression, as a result of changes in light intensity (Hihara et al., 2001) and temperature (Suzuki et al., 2001). Other works soon followed these initial experiments in the field of whole genome transcriptomics, and a number of alternative microarray platforms for *Synechocystis* 6803 were developed using different approaches (Postier et al., 2003; Singh et al., 2008; Zhang et al., 2008; Georg et al., 2009; Dickson et al., 2012). To date, more than 700 microarray experiments have been contributed to three main public repositories: ArrayExpress (Rustici et al., 2013), Kyoto Encyclopedia of Genes and Genomes (KEGG) Expression (Goto et al., 2000) and the Gene Expression Omnibus (GEO) database (Edgar et al., 2002). Such data contain valuable information on gene regulation under multiple environmental conditions, and genetic backgrounds. They are essential to the compilation of robust regulatory networks, using systems-level approaches (Figure 1).

Despite the bias toward *Synechocystis* 6803, abundant transcriptomics data also exist for other cyanobacteria (Figure 4). In the study of marine cyanobacteria, genome-level microarray analyses have frequently focused on limiting growth factors in oceans, such as iron (Thompson et al., 2011), nickel (Dupont et al., 2012), copper (Stuart et al., 2009), phosphate (Tetu et al., 2009; Ostrowski et al., 2010), and nitrogen (Su et al., 2006; Tolonen et al., 2006). Furthermore, studies were carried out to assess the effect of UV on the cell cycle (Kolowrat et al., 2010), the protective capacity to reactive oxygen species of cells acclimated to high- or low-intensity light (Blot et al., 2011), and to compare the transcriptome structures of the *Prochlorococcus* strains MED4 and MIT9313 (Voigt et al., 2014).

The earliest genome-level transcriptomic study carried out on marine cyanobacteria was on nitrogen metabolism. Su and co-workers predicted genes controlled by the global regulator NtcA of nitrogen metabolism in *Synechococcus* sp. WH8102 based on

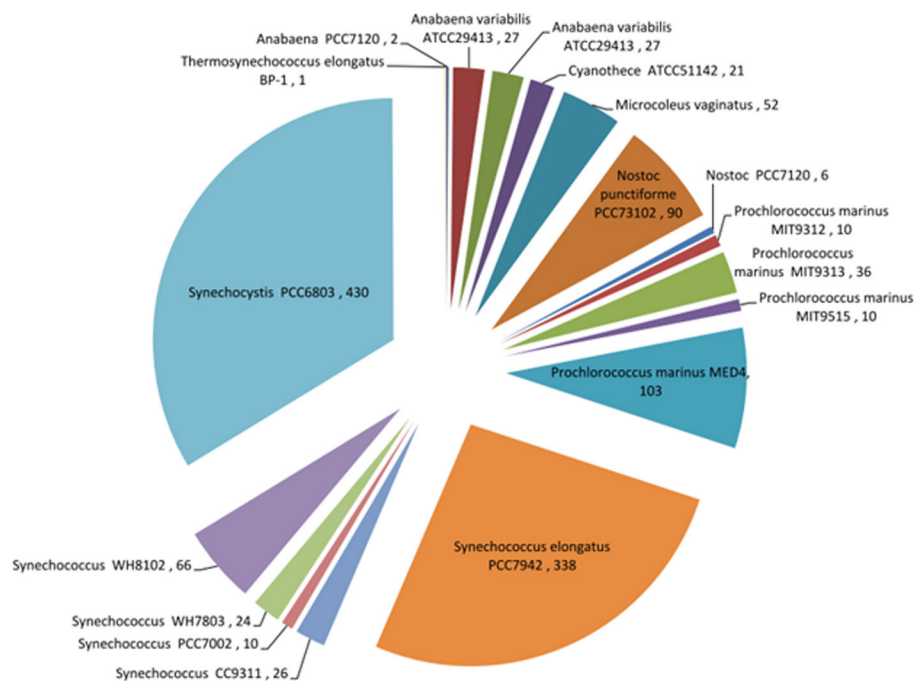


FIGURE 4 | Pie chart showing records of expression data for different cyanobacteria that are currently available in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>).

The bias toward *Synechocystis* 6803 can be clearly seen in this species breakdown of transcriptomic data. Data from April 2014.

comparative genomics analysis, and mining of experimental data from related organisms (Su et al., 2005, 2006). To validate their predictions, and to incorporate specific data for *Synechococcus* sp. WH8102, they profiled RNA extracted from cultures grown on nitrate and ammonium, as the sole nitrogen source. As a result, they compiled an extended regulatory network, constituting 429 genes, of which 338 were differentially expressed in the microarray data (Su et al., 2006). Elements of this network were also found in *Prochlorococcus*. Here, microarray data served to compare two distinct ecotypes, MED4 and MIT9313, and to identify the genes differentially regulated during nitrogen stress (Tolonen et al., 2006). Temporal expression profiling revealed a faster and shorter transcriptional response for MED4 than for MIT9313. The authors argued that such differences reflected their niche-specific requirements (Moore et al., 1998; West et al., 2001; Johnson et al., 2006). The MED4 ecotype occupies the top layers of the water column, where nitrogen sources (mainly ammonium and urea) fluctuate due to high-mixing levels, and the amount of incident light is higher than for deeper layers occupied by the low-light adapted MIT9313. A faster transcriptional response in MED4 would therefore avoid damage to their photosystems under nitrogen-limited conditions (Tolonen et al., 2006). In conclusion, their respective depth distributions have resulted in different adaptations not only to light, but also to nitrogen levels.

A large amount of work has been carried out to characterize the regulatory network underlying the circadian rhythm in both planktonic and benthic marine cyanobacteria (Zinser et al., 2009; Yang et al., 2010; Axmann et al., 2014). A tight

circadian expression regulation can help photosynthetic organisms to optimize cellular processes through anticipating and synchronizing transcription of photosynthetic-related genes with daylight hours. In the case of the unicellular *Cyanospora* sp. ATCC 51142, this regulation ensures that incompatible processes, such as oxygen-sensitive nitrogen fixation and evolution of oxygen through photosynthesis, do not occur at the same time (Reddy et al., 1993). Expression profiling of this *Cyanospora* during dark-light cycles showed cyclic regulation, with different clusters of genes peaking in dark or light conditions (Stockel et al., 2008; Toepel et al., 2008). The profiling served as a basis to model the regulatory network underlying this cyclic expression pattern, inferring regulatory influences for gene clusters. To do this, the authors employed a previously developed algorithm (the Inferelator) for derivation of gene regulatory networks using observed expression data (Bonneau et al., 2006). The algorithm predicts models for the expression of a single gene or cluster of genes, as a function of the expression levels of transcription factors (TFs), environmental conditions, and interactions between these factors. The model was used to infer potential key TFs in gene cluster regulation and to predict the behavior of co-expressed clusters under conditions not been used for calibration (McDermott et al., 2011). A comparable cyclic regulation was also identified in *Prochlorococcus* MED4 through comparison of mRNAs and protein expression levels of 312 genes (Waldbauer et al., 2012).

Although microarrays are powerful tools to capture transcriptional activity on a comprehensive scale, small sample

size, insufficient control of false positives, and poor method description can result in findings with limited reproducibility (Ntzani and Ioannidis, 2003; Jafari and Azuaje, 2006). To overcome such limitations, the statistical combination of multiple studies (or meta-analysis) is used to enhance the significance and validity of new findings. Such meta-analyses have great potential, as studies in various other biological fields have demonstrated (Steele and Tucker, 2008; Sun et al., 2009; Genini et al., 2011). Intriguingly, only two studies have systematically combined available gene expression data for cyanobacteria to obtain a systems level overview. The first study integrated 163 microarray data sets generated using *Synechocystis* 6803 as the model organism, with genetic, as well as environmental perturbations (Singh et al., 2010). They concluded that up to 12% of the genes in *Synechocystis* 6803 responded similarly to different perturbations, indicative of a similar regulatory program. A Bayesian network, connecting different KEGG pathways based on gene expression correlation between set of genes and genes with regulatory genes, was generated. Using this network, the authors proposed a functional connection between regulatory elements that control nitrogen and carbon metabolism.

Independently, our group collated all available raw microarray data for *Synechocystis* 6803 and pre-processed them in a standardized manner. We chose to re-analyze all experiments to minimize variations in expression levels due to the use of different normalization methods by independent researchers. Our dataset currently contains more than 700 microarray measurements, and is regularly updated with newly available data. These expression profiles were clustered and made available to researchers through CyanoExpress (<http://cyanoexpress.sysbiolab.eu/>), a web-based user-friendly visualization tool, for exploration of expression profiles (Hernandez-Prieto and Futschik, 2012). Gene-directed search enables rapid detection of co-expression patterns, which can be highly indicative for the annotation of less-studied genes, and in the prediction of new gene functions.

As an alternative to microarray analyses, which are based on hybridization of cDNA or mRNA in probes attached to a solid surface, the development of cost-effective and efficient next generation sequencing technologies has promoted the direct sequencing of RNA for expression profiling (Flaherty et al., 2011; Mitschke et al., 2011; Vijayan et al., 2011; Ludwig and Bryant, 2012; Ruffing, 2013). Together with the use of tiling arrays, this approach (commonly referred to as RNA-Seq) has revealed the existence of numerous non-protein coding transcripts (ncRNAs) in cyanobacteria. Previously, these ncRNAs remained undetected in microarrays, which only comprised probes for protein-coding genes. Thus, it was surprising when more comprehensive profiling approaches revealed that ncRNAs can make up a substantial part of the transcriptome in cyanobacteria. For instance, RNA-Seq data for *Prochlorococcus* indicate that up to three quarters of the genes have an antisense RNA (ncRNA transcribed from the opposed strand to a protein encoding gene) (Waldbauer et al., 2012; Voigt et al., 2014). While the detection of new ncRNA has been facilitated through RNA-Seq, their functional characterization poses a major task for the future (Haas et al., 2012). Interestingly, ncRNAs seem also to contribute to a substantial part of the sequences detected by meta-transcriptomics analyses of

RNA from natural environments (Voigt et al., 2014). In fact, comparison of four meta-transcriptomic studies carried out in oceans showed that approximately 25% of the detected RNA sequences could not be assigned to protein-coding genes or rRNAs; approximately 16% of these sequences displayed structural similarity to sRNA families represented in the Rfam database (<http://rfam.xfam.org/>) (Shi et al., 2009; Gardner et al., 2011).

Hence, RNA-Seq has highlighted the importance of ncRNAs in cyanobacteria and provided an important method for characterizing transcription start sites. Two major drawbacks of RNA-Seq analyses compared with microarray analyses are the limited number of tools available for downstream analyses, and the larger data files, which require more powerful computers for their analysis. A further typical feature of RNA-Seq is that a small number of highly expressed genes can make up the vast majority of sequence data, hindering the detection of many lowly expressed genes. For instance, it is often necessary to remove rRNA making up the vast majority of total RNA prior to sequencing. Also, a sufficiently large number of reads need to be generated to obtain a faithful profile of the transcriptome (Haas et al., 2012). Nevertheless, RNA-Seq is rapidly becoming the standard method for cyanobacterial transcriptomics.

PROTEOMICS: THE CHALLENGE TO PROVIDE GENOME-LEVEL PROTEIN PROFILING

The lack of knowledge about post-transcriptional and post-translational regulatory processes makes it difficult to extrapolate from a gene's transcript level to the activity level of its corresponding protein. This lack of understanding of the post-transcriptional regulatory process in cyanobacteria was patent when RNA and protein levels were monitored in parallel during the diel-cycle in *Cyanothece* (Aryal et al., 2011; Stockel et al., 2011) and *Prochlorococcus* MED4 (Waldbauer et al., 2012). Whole-cell proteomics seeks to overcome this limitation by directly studying protein structures and functions, their expression levels at different cellular stages, as well as their protein-protein interactions; these aspects are essential to understand biological processes in cells (Figure 1).

Traditionally, two-dimensional polyacrilamide gels (2D-PAGE) combined with the use of different fluorescence dyes (difference gel electrophoresis; 2D-DIGE) were conceived to estimate concentration differences for each protein between two different physiological states in a similar manner to two color microarrays (Arruda et al., 2011). More recent strategies take advantage of the sensitivity of liquid chromatography (LC), coupled with tandem mass spectrometry (MS), known as LC-MS/MS, for quantitative proteomic analysis, using different tags (as reviewed in Rotilio et al., 2012).

In marine cyanobacteria, global protein studies were carried out to evaluate organism response to stress, or their variability with respect to ecotype. The most recent studies on marine cyanobacteria looked at the effects of phosphate, cadmium and zinc stress using LC-MS/MS technology on *Synechococcus* WH8102 (Cox and Saito, 2013), or nitrogen regimes using 2-D/MALDI-TOF-MS on *T. erythraeum* IMS101 (Sandh et al., 2011). The genome shrinkage process, which was identified for *Prochlorococcus* strains (Dufresne et al., 2005), makes the study

of the proteome of different ecotypes in this genera of special interest to detect niche specific adaptations. Fuszard et al. (2012) investigated changes in protein levels induced upon phosphate depletion in three *Prochlorococcus* strains representing the high-light (MIT9312) and low-light ecotypes (NATL2A and SS120). Protein levels and growth rate changes were larger in the high-light ecotype than in the two strains belonging to the low-light ecotype. These results confirmed previous differences observed through genome comparisons, reflecting niche-driven molecular diversity, not only at specific gene divergences, but also at the codon usage level between different ecotypes (Paul et al., 2010).

For proteins with undefined functions, protein-protein interaction studies can assist researchers to determine their roles through their interaction with proteins (or protein complexes) with known functions (Prommeenate et al., 2004; Yao et al., 2007; Nixon et al., 2010). With this purpose in mind, *Synechocystis* 6803 protein extracts under native conditions (BN-PAGE) were separated by splitting the gel into small slices, and identifying the proteins in each slice by LC-MS/MS (Tanaka et al., 2010). These data were used to compile the protein co-migration database (PCoM-DB; <http://pcomdb.lowtem.hokudai.ac.jp/proteins/top>). In this way, the authors identified proteins of unknown function co-migrating with known proteins or protein-complexes.

In addition to experiments directed toward specific proteins, large-scale protein-protein interactions are crucial to post-genomic systems biology. HTT applied to proteomics, such as the yeast two-hybrid system, assisted scientists in different fields in the construction of large protein-protein interaction networks (Wallach et al., 2013; Ngounou Wetie et al., 2014). Networks generated from these studies serve to identify potential targets for future biochemical and bioinformatics studies (Kaçar and Gaucher, 2013; Yu et al., 2013). To fill this gap in the field of cyanobacteria, Sato et al. (2007) undertook the first systematic identification of protein interactions in *Synechocystis* 6803. Using yeast two-hybrid assays, they screened 1825 genes, discovering 3236 independent two-hybrid interactions (Sato et al., 2007). Such interaction data are important for functional analyses of genes in *Synechocystis* 6803, as well as for those conserved in marine cyanobacteria. They are accessible through the CyanoBase website (Nakao et al., 2010) (Table 1). Nowadays, several resources that model protein-protein interaction networks for cyanobacteria are available for researchers to examine, i.e., generic databases such as STRING, as well as specialized databases such as SynechoNET or InteroPORC. STRING covers more than 1000 organisms, containing experimentally-validated interactions, predicted and transferred interactions, together with interactions obtained through text mining (Franceschini et al., 2013). In contrast, the SynechoNET database is dedicated to *Synechocystis* 6803, and covers 2930 proteins (i.e., 79% of all predicted proteins in *Synechocystis* 6803). It includes 109,532 predicted protein-protein interactions extracted from the databases STRING (2658 proteins, 26,805 interactions), PSIMAP-based (1028 proteins, 12,748 interactions), InterDom (1760 proteins, 80,319 interactions) and iPFAM (1541 proteins, 13,448 interactions) (Kim et al., 2008). The SynechoNET visualization interface also permits the exploration of a “high confidence” sub-network

composed of 509 proteins, common to all databases, connected by 1591 interactions, where each interaction has an attached score value, obtained using the arithmetic value of the scores provided by STRING and InterDom. Some of the predicted high-confidence interactions were supported by the yeast two-hybrid data in Sato et al. (2007). Another set of predicted protein-protein interactions for *Synechocystis* 6803 was obtained using a method called InteroPORC (Michaut et al., 2008). Here, interactions were computationally inferred based on orthology of proteins which are known to interact in other organisms. The protein-protein interaction network for *Synechocystis* 6803 available on the InteroPORC interface contains 2259 interactions for 807 proteins. A core of 222 interactions were supported by experimental data obtained from three curated databases IntAct (Orchard et al., 2014), MINT (Licata et al., 2012), or DIP (Salwinski et al., 2004), as well as the yeast two-hybrid data generated by Sato et al. (2007).

METABOLOMICS: FROM PATHWAYS TO WHOLE GENOME FLUXOMICS

Similar to other “omic” approaches, recent advances in HTT also have had their impact on the study of metabolites in cyanobacteria, enabling simultaneous measurement of hundreds of metabolites. The quality of these results and the breadth of metabolite coverage greatly depend on the analytical flow rate of the experimental setup. In this regard, the development of enhanced compound separation techniques, coupled to MS or nuclear magnetic resonance (NMR) spectroscopy has played a key role (Zhang et al., 2012) (Figure 1). In general, we can differentiate between three types of metabolic studies: (i) target analysis, where the goal is the identification of specific or bioactive metabolites; (ii) metabolomic profiling, where the goal is to identify as many metabolites as possible; and (iii) flux analysis, where the goal is to define fluxes through specific biochemical pathways, mainly by isotope labeling. All these approaches have been successfully applied to study metabolites in cyanobacteria and were recently reviewed in Schwarz et al. (2013).

Here, we will concentrate on approaches that deal with genome-level modeling of cyanobacteria metabolism. Such genome-scale metabolic models consider cellular metabolism in its entirety, instead of focusing on individual pathways in isolation, providing a more realistic view of the interconnection and interdependence of cellular processes. A general pre-requisite for such modeling is the existence of the whole genome sequence. In comparison to the number of available genomes, however, the number of species-specific genome-level metabolic models is still scarce (less than a 3%). This is partially due to the lack of automated modeling tools. Despite some progress in this direction (Arakawa et al., 2006; Devoid et al., 2013; Krishnakumar et al., 2013), modeling still requires intensive supervision to avoid pitfalls due to different database annotations, and there are important limitations inherent to each method currently available (Ginsburg, 2009). These limitations make modeling a cumbersome task. Nevertheless, metabolic networks are an important key to assist researchers in: the design of strains to overproduce a desired product, the identification of potential enzymes responsible for orphan reactions, the determination of optimal growth

conditions to favor a reaction, the identification of coupled reaction sets, as well as evolutionary studies (Ma et al., 2013).

Once again, abundant experimental information on *Synechocystis* 6803, coupled with its use in many biotechnological applications, has supported the publication of several metabolic models for this strain (Montagud et al., 2011; Nogales et al., 2012; Knoop et al., 2013). The flow of metabolites through these models has also been determined by Flux Balance Analysis (FBA), a constraint-based method that has given rise to the new “omic” term: Fluxomics (Winter and Kromer, 2013). One advantage of FBA is that no previous knowledge of the kinetic parameters for individual metabolic reactions is required. FBA can predict the optimal steady-state fluxes required to maximize the synthesis of biomass or a product of interest, e.g., the synthesis of fatty acids in *Synechocystis* 6803.

The earliest FBA published for *Synechocystis* 6803 was carried out early in 2005 (Shastri and Morgan, 2005) and only included the evaluation of central metabolic pathways, under different light and media conditions (e.g., photoautotrophic, mixotrophic, and heterotrophic). In this study, cyclic and non-cyclic electron transport chains were considered as non-interacting events, although they share multiple components (Vermaas, 2001). This disconnection disappeared only in more recent metabolic models (Nogales et al., 2012), indicating that alternative electron flow pathways maximize growth under diverse environmental conditions, particularly when light or carbon are limiting factors. Progress has also been made in the reconstruction of central metabolic pathways. Knoop et al. (2013) were the first ones to incorporate a complete tricarboxylic acid cycle (TCA) cycle (Zhang and Bryant, 2011), while in the previous works either the glyoxylate shunt (Shastri and Morgan, 2005) or the GABA shunt were used to close the cycle (Knoop et al., 2010). Their core reconstruction encompassed 677 genes that encoded for 495 enzymes or enzyme-complexes. The annotated enzymes gave rise to 759 metabolic reactions, involving 601 metabolites. In an attempt to consolidate these networks, and at the same time to provide a tool for biologists to model their results, a user friendly-interface with a *Synechocystis*-specific section was added to web-based FAME resource (<http://f-a-m-e.org/synechocystis/>) (Boele et al., 2012; Maarleveld et al., 2014). The model, iTM686, is based on that of Nogales et al. (2012), but amended to include later findings, such as the complete TCA cycle (Zhang and Bryant, 2011), arginine metabolism (Schriek et al., 2009), and the proline metabolism (Knoop et al., 2010). A remarkable feature of FAME is that it permits the user to visualize flux analysis results together with gene expression data; to assist users with this objective, it has preloaded part of the CyanoEXpress microarray data.

It is important to note that these models only offer a pseudo-steady state condition, while the metabolism of cyanobacteria growing in the environment is generally under circadian control. This temporal separation is of special importance in the non-heterocyst forming nitrogen-fixing *Cyanothece* sp. ATCC 51142 (Reddy et al., 1993). A step forward in this direction was the modeling of the circadian cyclic behavior by developing separate (light/dark) biomass equations (Vu et al., 2012), adding constraints on the metabolic fluxes based on gene and protein expression (Stockel et al., 2008, 2011). A later study also incorporated

constraints on the flux through alternative pathways based on ^{13}C assisted metabolic flux analysis. Using the software OpenFLUX (Quek et al., 2009) the authors predicted network behavior from data obtained by the alternate use of labeled glycerol vs. unlabeled CO_2 . Rates of reactions in the carbon metabolic pathway of this *Cyanothece* strain showed that incorporation of labeled glycerol vs. unlabeled CO_2 into amino acids under nitrogen-deficient and nitrogen-sufficient conditions was distinctly different. These experiments suggested that two distinct metabolic programs were active (Alagesan et al., 2013).

In some cases, the partial sequence of a genome is sufficient to build an organism metabolic network, as shown for the industrialized cyanobacterium *Spirulina* (*Arthrospira*) *platensis* (Klanchui et al., 2012). A model was derived, using a semi-automated process based on the algorithms available through the Pathway Tools software (Karp et al., 2010), and cross-comparison with data stored in the MetaCyc database (Caspi et al., 2012). This work—besides providing an exceptional model to improve the industrial use of this strain—can serve as guideline for researchers with limited programming knowledge to develop an initial metabolic draft for their organism of interest.

WEB-BASED TOOLS FOR SYSTEM BIOLOGY ANALYSIS

While empowering basic and applied research in cyanobacteria, the rapidly increasing volume of data generated by HTT also provides formidable challenges to individual researchers. Data treatment is often computational intensive and needs to cover a range in raw data types, such as images of microarrays or protein gels, set sequences, and mass spectra (Figure 1). Furthermore, raw data frequently requires quality control, pre-processing, and normalization procedures prior to analysis to ensure reliability of results. To facilitate the use of available data for scientists, several web-based or community resources have been established specifically for cyanobacteria. They allow users to custom analyze data repositories as part of their experimental design protocol, and to compare their data with other research results (Table 2).

The central and most widely-used resource in the field is CyanoBase (<http://genome.microbedb.jp/cyanobase/>; Nakao et al., 2010). Started in 1995, this database includes currently sequenced and annotated genomes for 39 species of cyanobacteria. Although it contains only a very limited number of analysis tools (i.e., Blast2 for genes and genomes similarities searches, and KazusaMart to quickly convert between IDs in different formats, powered by the free software BioMart), it allows the user to explore genomes, as well as to obtain gene annotations. For instance, after querying for a gene, the user can obtain information on relevant publications, number of predicted transmembrane regions, putative protein-protein interactions, orthologous genes in other cyanobacteria, and more importantly download sequence data. The current version of CyanoBase was updated by the Kazusa DNA Research Institute (Nakao et al., 2010) to include a full text search, gene indexing, and mutants. It is currently maintained by Nakamura's Laboratory at the National Institute of Genetics.

The Virginia Commonwealth University hosts the BioBike Server for the Public Cyanobacterial Edition v5.2. CyanoBike (<http://biobike-8003.csb.cvu.edu/biologin>) is a web-based,

Table 2 | Web-based analytical tools with a specific focus on cyanobacteria as at April 2014.

Website	In-built data	Analysis tools	Visualization	Other advantages
CyanoBase	Genetic	Blast2, KazusaMart	YES (Genome view)	Downloadable data
CyanoBIKE	Genetic transcriptomic metabolic	In-built and custom tools	NO	Programmable
CyanoExpress	Transcriptomic	Co-regulation and cluster analysis	YES (GeneXplorer)	Special datasets
Cyanorak	Genetic	Blast	NO	Downloadable data
ProPortal	Genetic transcriptomic ecologic	Blast, cluster, enrichment analysis	NO	Species-specific data
CINPER	Genetic data from biosubsystems	Network mapping, enrichment analysis	YES	Exports networks, user input
FAME	Metabolic networks	FBA Gene expressionImage manipulation	YES	Export images and networks

programmable knowledge base, designed to make genomic, metabolic, and experimental data available to the public. It has built-in tools to manipulate and analyze these data (Massar et al., 2005; Elhai et al., 2009), although they require some basic programming skills for their use. Researchers can select databases devoted to marine cyanobacteria or to other specific organisms. Drop down menus are available to select functions, e.g., to compare proteins or gene-strings. It currently has 13 fully sequenced cyanobacterial strains in its dataset, and expects to include another 20 over the next year. Twelve microarray datasets also are integrated, including nine for *Synechocystis* PCC 6803 (corresponding to normalized data available through the KEGG EXPRESSION Database). It is updated regularly and is funded by the National Science Foundation (USA).

CyanoExpress is a web-served dedicated to transcriptomics data for cyanobacteria (<http://cyanoexpress.sysbiolab.eu/>; Hernandez-Prieto and Futschik, 2012). Currently, it includes expression data only for *Synechocystis* PCC 6803, with 718 microarray measurements compiled from 33 independent studies with both environmental and genetic perturbations. As such, there are 177 expression entries for 3073 genes. It aims to assist researchers in the characterization and functional annotation of genes using the guilty-by-association principle. Its visualization tool is a modified version of GeneXplorer (Rees et al., 2004). Different data subsets can be selected, with genes associated to different functions, based on whether all perturbations (genetic and environmental) or only environmental ones are included. CyanoExpress is regularly updated with new data retrieved from the public repositories: GEO, ArrayExpress, and KEGG. Future versions will also include expression data for other cyanobacteria, such as *Prochlorococcus* and *Synechococcus*.

Cyanorak (<http://www.sb-roscoff.fr/cyanorak>) is a dedicated resource for curation and annotation of clusters of orthologous sequences from marine picocyanobacteria. Cyanorak v.1 contains three *Prochlorococcus* and 11 *Synechococcus* genomes (Dufresne et al., 2008); Cyanorak v.2 containing 14 *Prochlorococcus*, 3 *Cyanobium*, and 40 *Synechococcus* genomes will soon be released. The current version allows the user to export individual Genbank files, as well as protein/gene sequences as FASTA files. More importantly, Cyanorak is manually curated, providing an updated version of the genomes available. The curation results in detailed descriptions for many clusters of gene orthologs, which can be retrieved using a search engine with different options (fast/advanced) set by the user.

ProPortal (<http://proportal.mit.edu/>) aims to provide easy access to the growing genetic database devoted to the cyanobacterium *Prochlorococcus* and its phages; to facilitate its use as a model system for systems biology (Kelly et al., 2012). The database includes genomes of cultured isolates of *Prochlorococcus* and their phages (and 11 strains of *Synechococcus*), as well as processed expression data from microarray experiments. Users can search for orthologous gene clusters, compare genomes from different populations, and identify up- or down-regulated genes under different environmental stressors (light, nitrogen, phosphate and iron) using various modules. The ProPortal database (Version 3) was recently updated with new gene cluster definitions (Kelly et al., 2013), so that it now contains 68 (24 host, 44 phage) genomes and 55,622 genes.

CINPER (<http://csbl.bmb.uga.edu/cinper/>) is one of the newer websites devoted to prokaryotes, developed by the Computations Systems Biology Laboratory at the University of Georgia. It focuses on networks, mapping well-known genes from multiple template genomes to a target genome and has been applied to study osmoregulation in *Synechococcus* sp. WH8102 (Mao et al., 2012). Networks can be validated with gene expression data (provided by the user), and can be visualized and explored using the Cytoscape Web interface (Lopes et al., 2010). One of its main advantages is that networks can be exported as images in various formats or as xml-based formats, i.e., svg and xgmml.

Finally, FAME (<http://f-a-m-e.org/synechocystis/>), previously discussed in the metabolomics section, provides the user with the option to download editor-friendly image files of the *Synechocystis* 6803 metabolic network. This model can be used as a framework to modify and build other organism-specific networks.

CASE STUDY: INFERRING CONSERVED REGULATORY INTERACTIONS FROM INTER-SPECIES GENE EXPRESSION DATA

The identification of conserved elements in regulatory networks is a suitable task to demonstrate some of the capabilities of the above web-resources. For this, we compare two time-series of expression data under iron limitation for *Prochlorococcus* MED4 (Thompson et al., 2011) (available through the ProPortal database) and *Synechocystis* 6803 (Hernandez-Prieto et al., 2012) (available through the GEO database). Here, we assume if a gene is present in two given organisms, then its function also is likely to be conserved. Our first step is therefore to identify putative orthologous genes between these cyanobacteria, using the tools

provided in the IMG database (<https://img.jgi.doe.gov/>) to compare genomes, under the option “Genome Gene Best Homologs” (Table S2). After selecting the genomes to compare, and deciding on a minimum threshold for identification, this option generates a tabulated file with the gene IDs from the reference genome, and their corresponding homologs for the given organism. In our case, to maximize the number of identified homologs, we use the lowest identity-percent allowed (20%). The output file had a total of 1050 unique *Prochlorococcus* MED4 genes (59.8% of the total genes) with identified homologs in *Synechocystis* 6803. This straight-forward step results in a list of *Synechocystis* 6803 and *Prochlorococcus* MED4 genes, annotated in a different format (i.e., SYNGTS_0535; PMM1032) to that used in the microarray annotation (i.e., slr2043; PMED4_11751), following the CyanoBase and ProPortal nomenclature, respectively. In order to merge both nomenclatures, we use equivalent identifiers provided for each genome through the databases: UniProt (<http://www.uniprot.org/>), CYORF (<http://cyano.genome.jp/>), and NCBI. A key conversion table is available in the supplementary material (Table S2).

Once differences in annotation have been resolved, we can compare differentially-expressed genes upon iron limitation in *Prochlorococcus* MED4, using the criteria provided in the corresponding publication (Thompson et al., 2011). In the case of *Synechocystis* 6803, we limit the list of genes to those in which expression was highly correlated ($r_s \geq 0.8$) with that of the iron-stress induced A and B (*isiA* and *isiB*) genes, known to be regulated by the ferric uptake regulator (Fur) (Kunert et al., 2003). Surprisingly, of these 97 genes in *Synechocystis* 6803, only 6 have homologs (based on at least a 20% homology) in the 74 differentially-expressed genes in *Prochlorococcus* MED4. To facilitate the visualization of these results, we built a small network with the identified genes connected to Fur, the main transcription factor involved in iron adaptation for *Synechocystis* and *Prochlorococcus* (Figures 5A,B). Fur homologs are present in both organisms (sll0567 in *Synechocystis* 6803 and PMM0637 in *Prochlorococcus* MED4). In the presence of sufficient iron, Fur binds upstream of genes involved in iron acquisition, blocking their transcription. In iron-limited conditions, Fur loses its affinity to the promoter regions, permitting the binding of specific sigma factors, and thus, transcription of downstream genes. In *E. coli* a second level of regulation is exerted by the ncRNA RyhB (Masse and Gottesman, 2002), which is under the control of Fur in iron-limited conditions. Upon transcription, RyhB promotes degradation of mRNAs, encoding non-essential iron-using proteins during iron limitation (Massé et al., 2003). Similar ncRNAs induced under iron-limiting conditions were identified in *Synechocystis* 6803 (Hernandez-Prieto et al., 2012), and candidates also exist in *Prochlorococcus* (Steglich et al., 2008). Therefore, to complete the network, we connect the genes repressed by iron, to this hypothetical regulatory element (ncRNA), even though this connection is not yet supported by experimental data. Thus, the generated regulatory networks are only hypothetical. Finally, we visualize the networks using the Cytoscape environment (Shannon et al., 2003); connections between genes and regulatory elements were established based on whether expression was higher or lower under iron-limiting conditions. Inspection of the two network shows that the direction of

expression changes is conserved in both cyanobacteria upon iron limitation; only one discrepancy associated with CmpA exists (a subunit of the bicarbonate transporter). This discrepancy appears to be related to the homology of the gene product of PMM0370 (annotated as putative cyanate ABC transporter) to two different transporter subunits, NrtC and CmpA. To resolve this discrepancy, we specifically looked for homologous genes to PMM0370, using the local blastp tool in the NCBI database. The top scoring protein in this case, was not CmpA, but NrtC (sll1452, a subunit of the nitrate transporter), which is also up-regulated in *Synechocystis* 6803 (Hernandez-Prieto et al., 2012). In fact, PMM0370 is part of the *cynABDS* operon and displays a common expression response to nitrogen limitation together with co-localized genes (Kamennaya and Post, 2011). To compare our regulatory network based on conservation of gene expression, with an automatically generated network, we created one with the keyword set to Fur, using CINPER with *Synechocystis* 6803 as the reference organism, and *Prochlorococcus* MED4 as the target (Figure 5C). The predicted protein-protein interactions related to the keyword were extracted from different sources, e.g., SEED, and the STRING database (Mao et al., 2012). In the CINPER network, Fur is identified with two *Synechocystis* 6803 proteins, FurA (sll0567) and PerR (slr1738), both related to Fur in *E. coli*. Of these two proteins, only FurA, is involved in iron regulation; while PerR is related to gene control during redox stress (Li et al., 2004; Scholnick et al., 2009).

This illustrative example, using available online tools and organism-specific data, shows how information from well-studied organisms can be used to interpret results obtained for phylogenetically-related organisms.

CONCLUSIONS

In this review, we have illustrated the breadth and depth of currently available “omic” data in public repositories. In addition, we have emphasized the value and advantages of these types of data for systems-level analyses, and the novel research that can result from integrating species-specific data into a wider context. Various examples were given for systems-level analysis of marine cyanobacteria, including the identification of class-specific genes located in genomic islands in *Prochlorococcus* using metagenomics, development of regulatory networks for nitrogen metabolism in *Synechococcus* through transcriptomics, the effects of phosphate limitation on *Synechococcus* in proteomics, and identification of two distinct metabolic pathways for different nitrogen conditions in *Cyanothece* in metabolomics. These results illustrate the power and potential of systems-level approaches in biological research.

In addition, we provided an overview of currently available user-friendly tools for researchers to use to manipulate “omic” data for marine cyanobacteria, illustrating their scope with a case study. The case study emphasizes both the types of networks that can be generated from online data, as well as their dependence on data quality. It also highlights some of the present limitations in automatically-generated networks, suggesting that there is still a vital role to supervised curation of data analysis.

At present, the number of researchers with the bioinformatic skills is growing, in response to the need to deal with data generated by HTT. This enhanced-skills base will continue

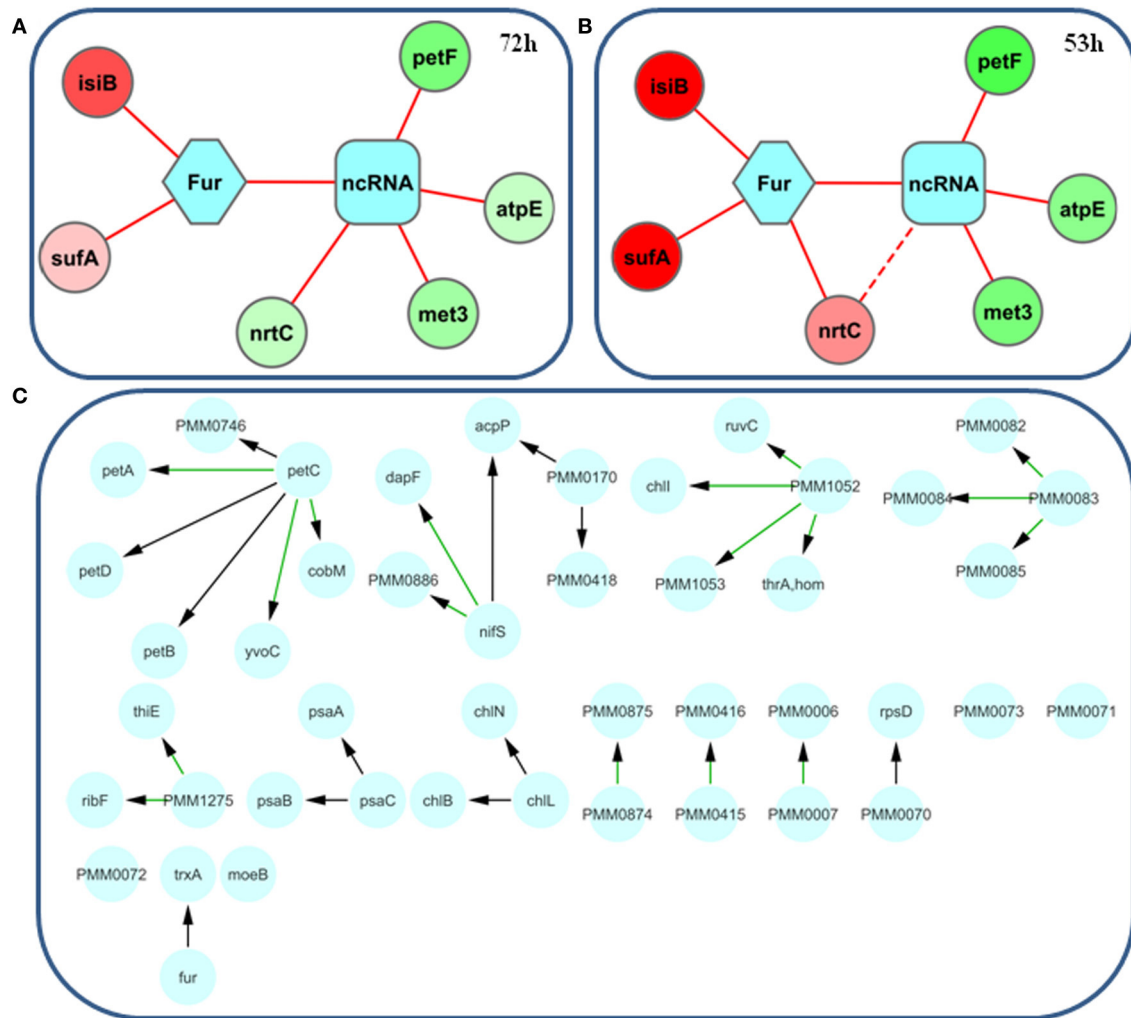


FIGURE 5 | Iron regulatory networks based on the differential expression upon iron limitation of genes conserved (identity larger than 20%) between *Synechocystis* sp. PCC6803 and *Prochlorococcus* MED4. The ferric uptake regulator (Fur) and a hypothetical ncRNA were set as central regulatory elements (in blue) following the well-described iron regulatory network of *E. coli*. Circular nodes were colored using a gradient from green to red, reflecting

differential expression upon iron depletion at 72h for *Synechocystis* 6803 (A) and 53h for *Prochlorococcus* MED4 (B). Red edges indicate a putative repression by the regulatory element. The apparent differential regulation of *cmpA/nrtC* is discussed in more detail in the case study. (C) CINPER-generated network created using the keyword set to Fur, with *Synechocystis* 6803 as the reference organism, and *Prochlorococcus* MED4 as the target.

to improve our capacity to connect physiological aspects with “omic” data, and to develop better bioinformatics tools to process large complex data sets. It is these system-level approaches that will support the future of meta-studies, and allow us to offset the quantity of HTT data with quality biological interpretations. New insights gained from system-level approaches will further research in both science and industry related to cyanobacteria.

ACKNOWLEDGMENTS

We would like to thank the reviewers for the constructive comments and suggestions. This work is supported by national Portuguese funding through FCT—Fundação para a Ciência e a Tecnologia, project ref. PEst-OE/EQB/LA0023/2013, IF/00881/2013 and PTDC/BIA-MIC/4418/2012 granted to Matthias E. Futschik.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2014.00191/abstract>

REFERENCES

- Abed, R. M. M., Dobretsov, S., and Sudesh, K. (2009). Applications of cyanobacteria in biotechnology. *J. Appl. Microbiol.* 106, 1–12. doi: 10.1111/j.1365-2672.2008.03918.x
- Alagesan, S., Gaudana, S. B., Sinha, A., and Wangikar, P. P. (2013). Metabolic flux analysis of *Cyanothece* sp. ATCC 51142 under mixotrophic conditions. *Photosynth. Res.* 118, 191–198. doi: 10.1007/s11120-013-9911-5
- Albert, R. (2007). Network inference, analysis, and modeling in systems biology. *Plant Cell Online* 19, 3327–3338. doi: 10.1105/tpc.107.054700
- Arakawa, K., Yamada, Y., Shinoda, K., Nakayama, Y., and Tomita, M. (2006). GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics* 7:168. doi: 10.1186/1471-2105-7-168

- Arruda, S. C. C., Barbosa, H. D. S., Azevedo, R. A., and Arruda, M. A. Z. (2011). Two-dimensional difference gel electrophoresis applied for analytical proteomics: fundamentals and applications to the study of plant proteomics. *Analyst* 136, 4119–4126. doi: 10.1039/c1an15513j
- Aryal, U., Stockel, J., Krovvidi, R., Gritsenko, M., Monroe, M., Moore, R., et al. (2011). Dynamic proteomic profiling of a unicellular cyanobacterium *Cyanothece* ATCC51142 across light-dark diurnal cycles. *BMC Syst. Biol.* 5:194. doi: 10.1186/1752-0509-5-194
- Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., and Lindell, D. (2011). Genomic island variability facilitates *Prochlorococcus-virus* coexistence. *Nature* 474, 604–608. doi: 10.1038/nature10172
- Axmann, I. M., Hertel, S., Wiegand, A., Dorrich, A. K., and Wilde, A. (2014). Diversity of KaiC-based timing systems in marine Cyanobacteria. *Mar. Genomics* 14, 3–16. doi: 10.1016/j.margen.2013.12.006
- Blot, N., Mella-Flores, D., Six, C., Le Corguille, G., Boutte, C., Peyrat, A., et al. (2011). Light History influences the response of the marine cyanobacterium *Synechococcus* sp. WH7803 to oxidative stress. *Plant Physiol.* 156, 1934–1954. doi: 10.1104/pp.111.174714
- Boele, J., Olivier, B. G., and Teusink, B. (2012). FAME, the flux analysis and modeling environment. *BMC Syst. Biol.* 6:8. doi: 10.1186/1752-0509-6-8
- Bonneau, R., Reiss, D., Shannon, P., Facciotti, M., Hood, L., Baliga, N., et al. (2006). The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.* 7:R36. doi: 10.1186/gb-2006-7-5-r36
- Burja, A. M., Dhamwichukorn, S., and Wright, P. C. (2003). Cyanobacterial postgenomic research and systems biology. *Trends Biotechnol.* 21, 504–511. doi: 10.1016/j.tibtech.2003.08.008
- Buttigieg, P. L., Hankeln, W., Kostadinov, I., Kottmann, R., Yilmaz, P., Duhaime, M. B., et al. (2013). Ecogenomic perspectives on domains of unknown function: correlation-based exploration of marine metagenomes. *PLoS ONE* 8:e50869. doi: 10.1371/journal.pone.0050869
- Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., et al. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40, D742–D753. doi: 10.1093/nar/gkr1014
- Chisholm, S. W., Olson, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B., and Welschmeyer, N. A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* 334, 340–343. doi: 10.1038/334340a0
- Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., and Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl. Environ. Microbiol.* 71, 5362–5370. doi: 10.1128/AEM.71.9.5362-5370.2005
- Cox, A. D., and Saito, M. A. (2013). Proteomic responses of oceanic *Synechococcus* WH8102 to phosphate and zinc scarcity and cadmium additions. *Front. Microbiol.* 4:387. doi: 10.3389/fmicb.2013.00387
- Degnan, B. M., Hawkins, C. J., Lavin, M. F., McCaffrey, E. J., Parry, D. L., Van Den Brenk, A. L., et al. (1989). New cyclic peptides with cytotoxic activity from the ascidian *Lissoclinum patella*. *J. Med. Chem.* 32, 1349–1354. doi: 10.1021/jm00126a034
- Devoid, S., Overbeek, R., Dejongh, M., Vonstein, V., Best, A., and Henry, C. (2013). “Automated genome annotation and metabolic model reconstruction in the SEED and model SEED,” in *Systems Metabolic Engineering*, ed H. S. Alper (New York, NY: Humana Press), 17–45. doi: 10.1007/978-1-62703-299-5_2
- Dickson, D. J., Luterra, M. D., and Ely, R. L. (2012). Transcriptomic responses of *Synechocystis* sp. PCC 6803 encapsulated in silica gel. *Appl. Microbiol. Biotechnol.* 96, 183–196. doi: 10.1007/s00253-012-4307-6
- Duce, R. A., Laroche, J., Altieri, K., Arrigo, K. R., Baker, A. R., Capone, D. G., et al. (2008). Impacts of atmospheric anthropogenic nitrogen on the open ocean. *Science* 320, 893–897. doi: 10.1126/science.1150369
- Dufresne, A., Garczarek, L., and Partensky, F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol.* 6:R14. doi: 10.1186/gb-2005-6-2-r14
- Dufresne, A., Ostrowski, M., Scanlan, D. J., Garczarek, L., Mazard, S., Palenik, B. P., et al. (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* 9:R90. doi: 10.1186/gb-2008-9-5-r90
- Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., et al. (2003). Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10020–10025. doi: 10.1073/pnas.1733211100
- Dupont, C. L., Johnson, D. A., Phillippy, K., Paulsen, I. T., Brahamsha, B., and Palenik, B. (2012). Genetic identification of a high-affinity Ni transporter and the transcriptional response to Ni deprivation in *Synechococcus* sp. strain WH8102. *Appl. Environ. Microbiol.* 78, 7822–7832. doi: 10.1128/AEM.01739-12
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207
- Elhai, J., Taton, A., Massar, J. P., Myers, J. K., Travers, M., Casey, J., et al. (2009). BioBIKE: a Web-based, programmable, integrated biological knowledge base. *Nucleic Acids Res.* 37, W28–W32. doi: 10.1093/nar/gkp354
- Falkowski, P. (2012). OCEAN SCIENCE The power of plankton. *Nature* 483, S17–S20. doi: 10.1038/483S17a
- Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237
- Flaherty, B. L., Van Nieuwerburgh, F., Head, S. R., and Golden, J. W. (2011). Directional RNA deep sequencing sheds new light on the transcriptional response of *Anabaena* sp. strain PCC 7120 to combined-nitrogen deprivation. *BMC Genomics* 12:332. doi: 10.1186/1471-2164-12-332
- Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9824–9829. doi: 10.1073/pnas.1307701110
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094
- Frigaard, N.-U., Sakuragi, Y., and Bryant, D. (2004). “Gene inactivation in the cyanobacterium *Synechococcus* sp. PCC 7002 and the green sulfur bacterium *Chlorobium tepidum* using *in vitro*-made DNA constructs and natural transformation,” in *Photosynthesis Research Protocols*, ed R. Carpentier (Totowa, NJ: Humana Press), 325–340. doi: 10.1385/1-59259-799-8:325
- Fujisawa, T., Okamoto, S., Katayama, T., Nakao, M., Yoshimura, H., Kajiyama-Kanegae, H., et al. (2014). CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. *Nucleic Acids Res.* 42, D666–D670. doi: 10.1093/nar/gkt1145
- Furumichi, M., Sato, Y., Omata, T., Ikeuchi, M., and Kanehisa, M. (2002). CYORF: community annotation of cyanobacteria genes. *Genome Informatics* 13, 402–403.
- Fuszard, M. A., Wright, P. C., and Biggs, C. A. (2012). Comparative quantitative proteomics of *prochlorococcus* ecotypes to a decrease in environmental phosphate concentrations. *Aquat. Biosyst.* 8:7. doi: 10.1186/2046-9063-8-7
- Garcia-Pichel, F., Belnap, J., Neuer, S., and Schanz, F. (2003). Estimates of global cyanobacterial biomass and its distribution. *Arch. Hydrobiol.* 109, 213–227. doi: 10.1127/1864-1318/2003/0109-0213
- Garcia-Pichel, F., and Pringault, O. (2001). Microbiology: cyanobacteria track water in desert soils. *Nature* 413, 380–381. doi: 10.1038/35096640
- Gardner, P. P., Daub, J., Tate, J., Moore, B. L., Osuch, I. H., Griffiths-Jones, S., et al. (2011). Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Res.* 39, D141–D145. doi: 10.1093/nar/gkq1129
- Genini, S., Badaoui, B., Sclep, G., Bishop, S., Waddington, D., Pinard Van Der Laan, M.-H., et al. (2011). Strengthening insights into host responses to mastitis infection in ruminants by combining heterogeneous microarray data sources. *BMC Genomics* 12:225. doi: 10.1186/1471-2164-12-225
- Georg, J., Voss, B., Scholz, I., Mitschke, J., Wilde, A., and Hess, W. R. (2009). Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol. Syst. Biol.* 5:305. doi: 10.1038/msb.2009.63
- Ginsburg, H. (2009). Caveat emptor: limitations of the automated reconstruction of metabolic pathways in Plasmodium. *Trends Parasitol.* 25, 37–43. doi: 10.1016/j.pt.2008.08.012
- Goto, S., Kawashima, S., Okuji, Y., Kamiya, T., Miyazaki, S., Numatani, Y., et al. (2000). KEGG/EXPRESSION: a database for browsing and analysing microarray expression data. *Genome Informatics* 11, 222–223.
- Gross, H. (2007). Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl. Microbiol. Biotechnol.* 75, 267–277. doi: 10.1007/s00253-007-0900-5

- Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W., and Livny, J. (2012). How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* 13:734. doi: 10.1186/1471-2164-13-734
- Heidorn, T., Camsund, D., Huang, H. H., Lindberg, P., Oliveira, P., Stensjo, K., et al. (2011). Synthetic biology in cyanobacteria: engineering and analyzing novel functions. *Methods Enzymol.* 497, 539–579. doi: 10.1016/B978-0-12-385075-1.00024-X
- Hernandez-Prieto, M. A., and Futschik, M. E. (2012). CyanoEXpress: a web database for exploration and visualization of the integrated transcriptome of cyanobacterium *Synechocystis* sp. PCC6803. *Bioinformatics* 8, 629–633. doi: 10.6026/97320630008634
- Hernandez-Prieto, M. A., Schon, V., Georg, J., Barreira, L., Varela, J., Hess, W. R., et al. (2012). Iron deprivation in *Synechocystis*: inference of pathways, non-coding RNAs, and regulatory elements from comprehensive expression profiling. *G3 (Bethesda)* 2, 1475–1495. doi: 10.1534/g3.112.003863
- Hihara, Y., Kamei, A., Kanehisa, M., Kaplan, A., and Ikeuchi, M. (2001). DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. *Plant Cell* 13, 793–806. doi: 10.1105/tpc.13.4.793
- Hoffmann, L. (1999). Marine cyanobacteria in tropical regions: diversity and ecology. *Eur. J. Phycol.* 34, 371–379. doi: 10.1080/09670269910001736432
- Ireland, C. M., Durso, A. R., Newman, R. A., and Hacker, M. P. (1982). Antineoplastic cyclic peptides from the marine tunicate *Lissoclinum patella*. *J. Org. Chem.* 47, 1807–1811. doi: 10.1021/jo00349a002
- Jafari, P., and Azuaje, F. (2006). An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors. *BMC Med. Inform. Decis. Mak.* 6:27. doi: 10.1186/1472-6947-6-27
- Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., and Chisholm, S. W. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311, 1737–1740. doi: 10.1126/science.1118052
- Kaçar, B., and Gaucher, E. A. (2013). Experimental evolution of protein–protein interaction networks. *Biochem. J.* 453, 311–319. doi: 10.1042/BJ20130205
- Kamennaya, N. A., and Post, A. F. (2011). Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. *Appl. Environ. Microbiol.* 77, 291–301. doi: 10.1128/AEM.01272-10
- Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., et al. (2001). Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* 8, 205–213. doi: 10.1093/dnares/8.5.205
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., et al. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 3, 109–136. doi: 10.1093/dnares/3.3.109
- Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., Dale, J. M., Lee, T. J., et al. (2010). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief. Bioinform.* 11, 40–79. doi: 10.1093/bib/bbp043
- Kelly, L., Ding, H. M., Huang, K. H., Osburne, M. S., and Chisholm, S. W. (2013). Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J.* 7, 1827–1841. doi: 10.1038/ismej.2013.58
- Kelly, L., Huang, K. H., Ding, H., and Chisholm, S. W. (2012). ProPortal: a resource for integrated systems biology of *Prochlorococcus* and its phage. *Nucleic Acids Res.* 40, D632–D640. doi: 10.1093/nar/gkr1022
- Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., et al. (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* 3:e231. doi: 10.1371/journal.pgen.0030231
- Kim, W.-Y., Kang, S., Kim, B.-C., Oh, J., Cho, S., Bhak, J., et al. (2008). SynechoNET: integrated protein-protein interaction database of a model cyanobacterium *Synechocystis* sp. PCC 6803. *BMC Bioinformatics* 9:S20. doi: 10.1186/1471-2105-9-S1-S20
- Klanchui, A., Khannapho, C., Phodee, A., Cheevadhanarak, S., and Meechai, A. (2012). iAK692: a genome-scale metabolic model of *Spirulina platensis* C1. *BMC Syst. Biol.* 6:71. doi: 10.1186/1752-0509-6-71
- Knoop, H., Grundel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., et al. (2013). Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput. Biol.* 9:e1003081. doi: 10.1371/journal.pcbi.1003081
- Knoop, H., Zilliges, Y., Lockau, W., and Steuer, R. (2010). The metabolic network of *Synechocystis* sp. PCC 6803: systemic properties of autotrophic growth. *Plant Physiol.* 154, 410–422. doi: 10.1104/pp.110.157198
- Kolowrat, C., Partensky, F., Mella-Flores, D., Le Corguille, G., Boutte, C., Blot, N., et al. (2010). Ultraviolet stress delays chromosome replication in light/dark synchronized cells of the marine cyanobacterium *Prochlorococcus marinus* PCC9511. *BMC Microbiol.* 10:204. doi: 10.1186/1471-2180-10-204
- Krishnakumar, S., Durai, D. A., Wangikar, P. P., and Viswanathan, G. A. (2013). SHARP: genome-scale identification of gene-protein-reaction associations in cyanobacteria. *Photosyn. Res.* 118, 181–190. doi: 10.1007/s11120-013-9910-6
- Kunert, A., Vinnemeier, J., Erdmann, N., and Hagemann, M. (2003). Repression by Fur is not the main mechanism controlling the iron-inducible isiAB operon in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiol. Lett.* 227, 255–262. doi: 10.1016/S0378-1097(03)00689-X
- Langlois, R. J., Hummer, D., and Laroche, J. (2008). Abundances and distributions of the dominant nifH phylotypes in the Northern Atlantic Ocean. *Appl. Environ. Microbiol.* 74, 1922–1931. doi: 10.1128/AEM.01720-07
- Li, H., Singh, A. K., McIntyre, L. M., and Sherman, L. A. (2004). Differential gene expression in response to hydrogen peroxide and the putative PerR regulon of *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 186, 3331–3345. doi: 10.1128/JB.186.11.3331-3345.2004
- Li, W. K., Rao, D. V., Harrison, W. G., Smith, J. C., Cullen, J. J., Irwin, B., et al. (1983). Autotrophic picoplankton in the tropical ocean. *Science* 219, 292–295. doi: 10.1126/science.219.4582.292
- Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., et al. (2012). MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 40, D857–D861. doi: 10.1093/nar/gkr930
- Lindell, D., Jaffe, J. D., Coleman, M. L., Futschik, M. E., Axmann, I. M., Rector, T., et al. (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449, 83–86. doi: 10.1038/nature06130
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of Next-Generation Sequencing Systems. *J. Biomed. Biotechnol.* 2012, 11. doi: 10.1155/2012/251364
- Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., and Bader, G. D. (2010). Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26, 2347–2348. doi: 10.1093/bioinformatics/btq430
- Lorenz, P., and Eck, J. (2005). Metagenomics and industrial applications. *Nat. Rev. Microbiol.* 3, 510–516. doi: 10.1038/nrmicro1161
- Ludwig, M., and Bryant, D. A. (2012). *Synechococcus* sp. strain PCC 7002 transcriptome: acclimation to temperature, salinity, oxidative stress and mixotrophic growth conditions. *Front. Microbiol.* 3:354. doi: 10.3389/fmicb.2012.00354
- Ma, C.-Y., Lin, S.-H., Lee, C.-C., Tang, C. Y., Berger, B., and Liao, C.-S. (2013). Reconstruction of phyletic trees by global alignment of multiple metabolic networks. *BMC Bioinformatics* 14:S12. doi: 10.1186/1471-2105-14-S2-S12
- Maarleveld, T. R., Boele, J., Bruggeman, F. J., and Teusink, B. (2014). A data integration and visualization resource for the metabolic network of *Synechocystis* sp. PCC 6803. *Plant Physiol.* 164, 1111–1121. doi: 10.1104/pp.113.224394
- Malmstrom, R. R., Rodrigue, S., Huang, K. H., Kelly, L., Kern, S. E., Thompson, A., et al. (2013). Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J.* 7, 184–198. doi: 10.1038/ismej.2012.89
- Mao, X., Chen, X., Zhang, Y., Pangle, S., and Xu, Y. (2012). CINPER: an interactive web system for pathway prediction for prokaryotes. *PLoS ONE* 7:e1252. doi: 10.1371/journal.pone.0051252
- Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40, D115–D122. doi: 10.1093/nar/gkr1044
- Martiny, A. C., Huang, Y., and Li, W. (2009a). Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ. Microbiol.* 11, 1340–1347. doi: 10.1111/j.1462-2920.2009.01860.x
- Martiny, A. C., Kathuria, S., and Berube, P. M. (2009b). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10787–10792. doi: 10.1073/pnas.0902532106

- Massar, J. P., Travers, M., Elhai, J., and Shrager, J. (2005). BioLingua: a programmable knowledge environment for biologists. *Bioinformatics* 21, 199–207. doi: 10.1093/bioinformatics/bth465
- Massé, E., Escorcia, F. E., and Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev.* 17, 2374–2383. doi: 10.1101/gad.1127103
- Masse, E., and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4620–4625. doi: 10.1073/pnas.032066599
- McDermott, J. E., Oehmen, C. S., McCue, L. A., Hill, E., Choi, D. M., Stockel, J., et al. (2011). A model of cyclic transcriptomic behavior in the cyanobacterium *Cyanothece* sp. ATCC 51142. *Mol. Biosyst.* 7, 2407–2418. doi: 10.1039/c1mb05006k
- Michaut, M., Kerrien, S., Montecchi-Palazzi, L., Chauvat, F., Cassier-Chauvat, C., Aude, J. C., et al. (2008). InterOPORC: automated inference of highly conserved protein interaction networks. *Bioinformatics* 24, 1625–1631. doi: 10.1093/bioinformatics/btn249
- Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., et al. (2011). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2124–2129. doi: 10.1073/pnas.1015154108
- Montagud, A., Zelezniak, A., Navarro, E., De Córdoba, P. F., Urchueguía, J. F., and Patil, K. R. (2011). Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803. *Biotechnol. J.* 6, 330–342. doi: 10.1002/biot.201000109
- Moore, L. R., Rocap, G., and Chisholm, S. W. (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* 393, 464–467. doi: 10.1038/30965
- Nakamura, Y., Kaneko, T., Hirose, M., Miyajima, N., and Tabata, S. (1998). CyanoBase, a www database containing the complete nucleotide sequence of the genome of *Synechocystis* sp. strain PCC6803. *Nucleic Acids Res.* 26, 63–67. doi: 10.1093/nar/26.1.63
- Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., et al. (2010). CyanoBase: the cyanobacteria genome database update 2010. *Nucleic Acids Res.* 38, D379–D381. doi: 10.1093/nar/gkp915
- Nelson, N. (2011). Photosystems and global effects of oxygenic photosynthesis. *Biochim. Biophys. Acta* 1807, 856–863. doi: 10.1016/j.bbabi.2010.10.011
- Ngounou Wetie, A., Sokolowska, I., Woods, A., Roy, U., Deinhardt, K., and Darie, C. (2014). Protein–protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. *Cell. Mol. Sci.* 71, 205–228. doi: 10.1007/s00018-013-1333-1
- Nixon, P. J., Michoux, F., Yu, J., Boehm, M., and Komenda, J. (2010). Recent advances in understanding the assembly and repair of photosystem II. *Ann. Bot.* 106, 1–16. doi: 10.1093/aob/mcq059
- Nogales, J., Gudmundsson, S., Knight, E. M., Pálsson, B. O., and Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2678–2683. doi: 10.1073/pnas.1117907109
- Ntzani, E. E., and Ioannidis, J. P. (2003). Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 362, 1439–1444. doi: 10.1016/S0140-6736(03)14686-7
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42, D358–D363. doi: 10.1093/nar/gkt1115
- Ostrowski, M., Mazard, S., Tetu, S. G., Phillippy, K., Johnson, A., Palenik, B., et al. (2010). PtrA is required for coordinate regulation of gene expression during phosphate stress in a marine *Synechococcus*. *ISME J.* 4, 908–921. doi: 10.1038/ismej.2010.24
- Ow, S. Y., and Wright, P. C. (2009). Current trends in high throughput proteomics in cyanobacteria. *FEBS Lett.* 583, 1744–1752. doi: 10.1016/j.febslet.2009.03.062
- Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., et al. (2003). The genome of a motile marine *Synechococcus*. *Nature* 424, 1037–1042. doi: 10.1038/nature01943
- Palenik, B., Ren, Q., Tai, V., and Paulsen, I. T. (2009). Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ. Microbiol.* 11, 349–359. doi: 10.1111/j.1462-2920.2008.01772.x
- Partensky, F., and Garczarek, L. (2010). *Prochlorococcus*: advantages and limits of minimalism. *Ann. Rev. Mar. Sci.* 2, 305–331. doi: 10.1146/annurev-marine-120308-081034
- Partensky, F., Hess, W. R., and Vaulot, D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* 63, 106–127.
- Paul, S., Dutta, A., Bag, S. K., Das, S., and Dutta, C. (2010). Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria *Prochlorococcus*. *BMC Genomics* 11:103. doi: 10.1186/1471-2164-11-103
- Postier, B. L., Wang, H. L., Singh, A., Impson, L., Andrews, H. L., Klahn, J., et al. (2003). The construction and use of bacterial DNA microarrays based on an optimized two-stage PCR strategy. *BMC Genomics* 4:23. doi: 10.1186/1471-2164-4-23
- Prommeeate, P., Lennon, A. M., Markert, C., Hippler, M., and Nixon, P. J. (2004). Subunit composition of NDH-1 complexes of *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* 279, 28165–28173. doi: 10.1074/jbc.M401107200
- Quek, L.-E., Wittmann, C., Nielsen, L., and Kromer, J. (2009). OpenFLUX: efficient modelling software for ¹³C-based metabolic flux analysis. *Microb. Cell Fact.* 8:25. doi: 10.1186/1475-2859-8-25
- Reddy, K. J., Haskell, J. B., Sherman, D. M., and Sherman, L. A. (1993). Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus *cyanothece*. *J. Bacteriol.* 175, 1284–1292.
- Rees, C., Demeter, J., Matese, J., Botstein, D., and Sherlock, G. (2004). GeneXplorer: an interactive web application for microarray data visualization and analysis. *BMC Bioinformatics* 5:141. doi: 10.1186/1471-2105-5-141
- Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., et al. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424, 1042–1047. doi: 10.1038/nature01947
- Rotilio, D., Della Corte, A., D'Imperio, M., Coletta, W., Marcone, S., Silvestri, C., et al. (2012). Proteomics: bases for protein complexity understanding. *Thromb. Res.* 129, 257–262. doi: 10.1016/j.thromres.2011.12.035
- Ruffing, A. (2013). RNA-Seq analysis and targeted mutagenesis for improved free fatty acid production in an engineered cyanobacterium. *Biotechnol. Biofuels* 6, 1–15. doi: 10.1186/1754-6834-6-113
- Ruffing, A. M. (2011). Engineered cyanobacteria: teaching an old bug new tricks. *Bioeng. Bugs* 2, 136–149. doi: 10.4161/bbug.2.3.15285
- Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooshep, S., et al. (2007). The *Sorcerer II* global ocean sampling expedition: northwest atlantic through eastern tropical pacific. *PLoS Biol.* 5:e77. doi: 10.1371/journal.pbio.0050077
- Rusch, D. B., Martiny, A. C., Dupont, C. L., Halpern, A. L., and Venter, J. C. (2010). Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16184–16189. doi: 10.1073/pnas.1009513107
- Rustici, G., Kolesnikov, N., Brandizi, M., Burdett, T., Dylag, M., Emam, I., et al. (2013). ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res.* 41, D987–D990. doi: 10.1093/nar/gks1174
- Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004). The database of interacting proteins: 2004 update. *Nucleic Acids Res.* 32, D449–D451. doi: 10.1093/nar/gkh086
- Sandh, G., Ran, L., Xu, L., Sundqvist, G., Bulone, V., and Bergman, B. (2011). Comparative proteomic profiles of the marine cyanobacterium *Trichodesmium erythraeum* IMS101 under different nitrogen regimes. *Proteomics* 11, 406–419. doi: 10.1002/pmic.201000382
- Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y., and Tabata, S. (2007). A large-scale protein–protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res.* 14, 207–216. doi: 10.1093/dnares/dsm021
- Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* 73, 249–299. doi: 10.1128/MMBR.00035-08
- Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., et al. (2005). Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7315–7320. doi: 10.1073/pnas.0501424102

- Schriek, S., Kahmann, U., Staiger, D., Pistorius, E. K., and Michel, K. P. (2009). Detection of an L-amino acid dehydrogenase activity in *Synechocystis* sp. PCC 6803. *J. Exp. Bot.* 60, 1035–1046. doi: 10.1093/jxb/ern352
- Schwarz, D., Orf, I., Kopka, J., and Hagemann, M. (2013). Recent applications of metabolites toward cyanobacteria. *Metabolites* 3, 72–100. doi: 10.3390/metabo3010072
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shastri, A. A., and Morgan, J. A. (2005). Flux balance analysis of photoautotrophic metabolism. *Biotechnol. Prog.* 21, 1617–1626. doi: 10.1021/bp050246d
- Shcolnick, S., Summerfield, T. C., Reyman, L., Sherman, L. A., and Keren, N. (2009). The Mechanism of iron homeostasis in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803 and its relationship to oxidative stress. *Plant Physiol.* 150, 2045–2056. doi: 10.1104/pp.109.141853
- Shi, T., and Falkowski, P. G. (2008). Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2510–2515. doi: 10.1073/pnas.0711165105
- Shi, Y., Tyson, G. W., and Delong, E. F. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* 459, 266–269. doi: 10.1038/nature08055
- Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., et al. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1053–1058. doi: 10.1073/pnas.1217107110
- Singh, A. K., Elvitigala, T., Bhattacharyya-Pakrasi, M., Aurora, R., Ghosh, B., and Pakrasi, H. B. (2008). Integration of carbon and nitrogen metabolism with energy production is crucial to light acclimation in the cyanobacterium *Synechocystis*. *Plant Physiol.* 148, 467–478. doi: 10.1104/pp.108.123489
- Singh, A. K., Elvitigala, T., Cameron, J. C., Ghosh, B. K., Bhattacharyya-Pakrasi, M., and Pakrasi, H. B. (2010). Integrative analysis of large scale expression profiles reveals core transcriptional response and coordination between multiple cellular processes in a cyanobacterium. *BMC Syst. Biol.* 4:105. doi: 10.1186/1752-0509-4-105
- Stanley, D. N., Raines, C. A., and Kerfeld, C. A. (2013). Comparative analysis of 126 cyanobacterial genomes reveals evidence of functional diversity among homologs of the redox-regulated CP12 protein. *Plant Physiol.* 161, 824–835. doi: 10.1104/pp.112.210542
- Steele, E., and Tucker, A. (2008). Consensus and Meta-analysis regulatory networks for combining multiple microarray gene expression datasets. *J. Biomed. Inform.* 41, 914–926. doi: 10.1016/j.jbi.2008.01.011
- Steglich, C., Futschik, M. E., Lindell, D., Voss, B., Chisholm, S. W., and Hess, W. R. (2008). The Challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genet.* 4:e1000173. doi: 10.1371/journal.pgen.1000173
- Stockel, J., Jacobs, J. M., Elvitigala, T. R., Liberton, M., Welsh, E. A., Polpitiya, A. D., et al. (2011). Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium *Cyanothece* 51142. *PLoS ONE* 6:e16680. doi: 10.1371/journal.pone.0016680
- Stockel, J., Welsh, E. A., Liberton, M., Kunnavakkam, R., Aurora, R., and Pakrasi, H. B. (2008). Global transcriptomic analysis of *Cyanothece* 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6156–6161. doi: 10.1073/pnas.0711068105
- Stuart, R. K., Dupont, C. L., Johnson, D. A., Paulsen, I. T., and Palenik, B. (2009). Coastal strains of marine *synechococcus* species exhibit increased tolerance to copper shock and a distinctive transcriptional response relative to those of open-ocean strains. *Appl. Environ. Microbiol.* 75, 5047–5057. doi: 10.1128/AEM.00271-09
- Su, Z., Mao, F., Dam, P., Wu, H., Olman, V., Paulsen, I. T., et al. (2006). Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium *Synechococcus* sp. WH 8102. *Nucleic Acids Res.* 34, 1050–1065. doi: 10.1093/nar/gkj496
- Su, Z. C., Olman, V., Mao, F. L., and Xu, Y. (2005). Comparative genomics analysis of NtcA regulons in cyanobacteria: regulation of nitrogen assimilation and its coupling to photosynthesis. *Nucleic Acids Res.* 33, 5156–5171. doi: 10.1093/nar/gki817
- Sun, J., Heck, D., Xu, W., Chitnis, V., and Chitnis, P. (1999). "Abundance of photosystem I proteins in cyanobacteria and chloroplasts," in *The Chloroplast: From Molecular Biology to Biotechnology*, eds J. Argyroudi-Akoyunoglou and H. Senger (Netherlands: Springer), 227–232. doi: 10.1007/978-94-011-4788-0_36
- Sun, R., Fu, X., Guo, F., Ma, Z., Goulbourne, C., Jiang, M., et al. (2009). A strategy for meta-analysis of short time series microarray datasets. *Front. Biosci. (Landmark Ed)* 14, 4058–4070. doi: 10.2741/3512
- Suzuki, I., Kanesaki, Y., Mikami, K., Kanehisa, M., and Murata, N. (2001). Cold-regulated genes under control of the cold sensor Hik33 in *Synechocystis*. *Mol. Microbiol.* 40, 235–244. doi: 10.1046/j.1365-2958.2001.02379.x
- Tanaka, R., Rothbart, M., Oka, S., Takabayashi, A., Takahashi, K., Shibata, M., et al. (2010). LIL3, a light-harvesting-like protein, plays an essential role in chlorophyll and tocopherol biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16721–16725. doi: 10.1073/pnas.1004699107
- Tetu, S. G., Brahamsha, B., Johnson, D. A., Tai, V., Phillippy, K., Palenik, B., et al. (2009). Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J.* 3, 835–849. doi: 10.1038/ismej.2009.31
- Thompson, A. W., Huang, K., Saito, M. A., and Chisholm, S. W. (2011). Transcriptome response of high- and low-light-adapted *Prochlorococcus* strains to changing iron availability. *ISME J.* 5, 1580–1594. doi: 10.1038/ismej.2011.49
- Thompson, C., Silva, G. Z., Vieira, N., Edwards, R., Vicente, A., and Thompson, F. (2013). Genomic taxonomy of the genus *prochlorococcus*. *Microb. Ecol.* 66, 752–762. doi: 10.1007/s00248-013-0270-8
- Toepel, J., Welsh, E., Summerfield, T. C., Pakrasi, H. B., and Sherman, L. A. (2008). Differential transcriptional analysis of the cyanobacterium *Cyanothece* sp. strain ATCC 51142 during light-dark and continuous-light growth. *J. Bacteriol.* 190, 3904–3913. doi: 10.1128/JB.00206-08
- Tolonen, A. C., Aach, J., Lindell, D., Johnson, Z. I., Rector, T., Steen, R., et al. (2006). Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol. Syst. Biol.* 2:53. doi: 10.1038/msb4100087
- Vermaas, W. F. (2001). "Photosynthesis and respiration in cyanobacteria," in *eLS* (Chichester: John Wiley and Sons Ltd). doi: 10.1038/npg.els.0001670
- Vijayan, V., Jain, I., and O'Shea, E. (2011). A high resolution map of a cyanobacterial transcriptome. *Genome Biol.* 12:R47. doi: 10.1186/gb-2011-12-5-r47
- Voigt, K., Sharma, C. M., Mitschke, J., Joke Lambrecht, S., Voss, B., Hess, W. R., et al. (2014). Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. *ISME J.* doi: 10.1038/ismej.2014.57. [Epub ahead of print].
- Vu, T. T., Stolyar, S. M., Pinchuk, G. E., Hill, E. A., Kucek, L. A., Brown, R. N., et al. (2012). Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium *Cyanothece* sp. ATCC 51142. *PLoS Comput. Biol.* 8:1002460. doi: 10.1371/journal.pcbi.1002460
- Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisholm, S. W. (2012). Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS ONE* 7:e43432. doi: 10.1371/journal.pone.0043432
- Wallach, T., Schellenberg, K., Maier, B., Kalathur, R. K. R., Porras, P., Wanker, E. E., et al. (2013). Dynamic circadian protein-protein interaction networks predict temporal organization of cellular functions. *PLoS Genet.* 9:e1003398. doi: 10.1371/journal.pgen.1003398
- Wang, B., Wang, J., Zhang, W., and Meldrum, D. R. (2012). Application of synthetic biology in cyanobacteria and algae. *Front. Microbiol.* 3:344. doi: 10.3389/fmicb.2012.00344
- Welsh, E. A., Liberton, M., Stockel, J., Loh, T., Elvitigala, T., Wang, C., et al. (2008). The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15094–15099. doi: 10.1073/pnas.0805418105
- West, N. J., Lebaron, P., Strutton, P. G., and Suzuki, M. T. (2011). A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J.* 5, 933–944. doi: 10.1038/ismej.2010.186
- West, N. J., Schonhuber, W. A., Fuller, N. J., Amann, R. I., Rippka, R., Post, A. F., et al. (2001). Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by *in situ* hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* 147, 1731–1744.

- Williams, A. B., and Jacobs, R. S. (1993). A marine natural product, patellamide D, reverses multidrug resistance in a human leukemic cell line. *Cancer Lett.* 71, 97–102. doi: 10.1016/0304-3835(93)90103-G
- Winter, G., and Kromer, J. O. (2013). Fluxomics—connecting 'omics analysis and phenotypes. *Environ. Microbiol.* 15, 1901–1916. doi: 10.1111/1462-2920.12064
- Yang, Q., Pando, B. F., Dong, G., Golden, S. S., and Van Oudenaarden, A. (2010). Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* 327, 1522–1526. doi: 10.1126/science.1181759
- Yao, D., Kieselbach, T., Komenda, J., Promnares, K., Prieto, M. A., Tichy, M., et al. (2007). Localization of the small CAB-like proteins in photosystem II. *J. Biol. Chem.* 282, 267–276. doi: 10.1074/jbc.M605463200
- Yu, D., Kim, M., Xiao, G., and Hwang, T. H. (2013). Review of biological network data and its applications. *Genomics Inform.* 11, 200–210. doi: 10.5808/GI.2013.11.4.200
- Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niazi, F., Shi, T., et al. (2008). Globally distributed uncultivated oceanic N₂-fixing cyanobacteria lack oxygenic photosystem II. *Science* 322, 1110–1112. doi: 10.1126/science.1165340
- Zehr, J. P., Bench, S. R., Mondragon, E. A., McCarren, J., and Delong, E. F. (2007). Low genomic diversity in tropical oceanic N₂-fixing cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17807–17812. doi: 10.1073/pnas.0701017104
- Zhang, A., Sun, H., Wang, P., Han, Y., and Wang, X. (2012). Modern analytical techniques in metabolomics analysis. *Analyst* 137, 293–300. doi: 10.1039/c1an15605e
- Zhang, S. Y., and Bryant, D. A. (2011). The tricarboxylic acid cycle in cyanobacteria. *Science* 334, 1551–1553. doi: 10.1126/science.1210858
- Zhang, Z., Pendse, N. D., Phillips, K. N., Cotner, J. B., and Khodursky, A. (2008). Gene expression patterns of sulfur starvation in *Synechocystis* sp. PCC 6803. *BMC Genomics* 9:344. doi: 10.1186/1471-2164-9-344
- Zinser, E. R., Lindell, D., Johnson, Z. I., Futschik, M. E., Steglich, C., Coleman, M. L., et al. (2009). Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *prochlorococcus*. *PLoS ONE* 4:e5135. doi: 10.1371/journal.pone.0005135

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 April 2014; accepted: 11 June 2014; published online: 02 July 2014.

Citation: Hernández-Prieto MA, Semeniuk TA and Futschik ME (2014) Toward a systems-level understanding of gene regulatory, protein interaction, and metabolic networks in cyanobacteria. *Front. Genet.* 5:191. doi: 10.3389/fgene.2014.00191

This article was submitted to *Systems Biology*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Hernández-Prieto, Semeniuk and Futschik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.