

An insight into cyanobacterial genomics - a perspective

Palaniswamy Thanga Velan Lakshmi*

Phytomatics Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore 641 046;
E.mail: lakshmiptv@yahoo.co.in; * Corresponding author

received March 30, 2007; revised April 06, 2007; accepted April 10, 2007; published online May 20, 2007

Abstract:

At the turn of the millennium, cyanobacteria deserve attention to be reviewed to understand the past, present and future. The advent of post genomic research, which encompasses functional genomics, structural genomics, transcriptomics, pharmacogenomics, proteomics and metabolomics that allows a systematic wide approach for biological system studies. Thus by exploiting genomic and associated protein information through computational analyses, the fledging information that are generated by biotechnological analyses, could be well extrapolated to fill in the lacuna of scarce information on cyanobacteria and as an effort this paper attempts to highlights the perspectives available and awakens researcher to concentrate in the field of cyanobacterial informatics.

Keywords: cyanobacteria; genomics; bioinformatics; databases; phylogeny

Background:

Characteristics and importance of Cyanobacteria

The algae are the simplest members of the plant kingdom, and the cyanobacteria (blue-green algae) are the simplest of the algae, having considerable and increasing economic importance. They are relatives of the bacteria, not eukaryotes and are only the chloroplast in eukaryotic algae to which the cyanobacteria are related and like all eubacteria, their cell walls contain peptidoglycan. Studies of metabolic similarities and ribosomal RNA sequence suggest cyanobacteria to form a good monophyletic taxon. Although they are truly prokaryotic, cyanobacteria have an elaborate and highly organized system of internal membranes which function in photosynthesis. Chlorophyll *a* and several accessory pigments (phycoerythrin and phycocyanin) embedded in these photosynthetic lamellae are analogs of the eukaryotic thylakoid membranes. These are widely distributed over land and water, often in environments where no other vegetation can exist. Found in almost every conceivable habitat, from oceans to fresh water to bare rock to soil, cyanobacteria produce the compounds responsible for "earthy" odors we detect in soil. The greenish slime on the side of damp flower pot, the wall of house or the trunk of the big tree are more likely due to cyanobacteria and have also been found on the fur of polar bears, to which they impart a greenish tinge!. They have also been tremendously important in shaping the course of evolution and ecological change throughout earth's history. Heterocyst-forming species of cyanobacteria that "fix" atmospheric nitrogen have formed a niche in soil which has promoter effect on rice paddies of Asia [1] that feed about

75% of the world's human population. The coating of blue-greens on prairie soil binds the particles of the soil to their mucilage coating, maintains a high water content and reduces erosion. Humans also consume Spirulina, which contains all amino acids essential for humans forming a staple food in parts of Africa and Mexico. In China, Taiwan and Japan, several blue-greens are served as a side dish and a delicacy. World wide, it is cultured and commercially processed for various food and medicinal products such as vitamins, drug compounds, and growth factors and also for the production of hydrogen gas and fertilizers.

Status of cyanobacterial genomics

Genomics allows us to make new and unexpected links between the function of unrelated and hitherto uncharacterized genes and to suggest hypotheses and biochemistry. [2] The first complete genomic sequence obtained was that of *Haemophilus influenza*. [3] Since then, many genomes have been sequenced and analyzed. In Cyanobacteria, complete genome sequences have been reported in eight species and strains (*Synechocystis* sp. PCC 6803, *Anabaena* sp. PCC 7120, *Thermosynechococcus elongatus* BP-1, *Gloeobacter violaceus* PCC 7421, *Prochlorococcus marinus* SS120, *Prochlorococcus marinus* MED4, *Prochlorococcus marinus* MIT9313 and *Synechococcus* sp. WH8102).

A database [4] on *Nostoc punctiforme* computationally annotated by Frank Larimer at the Oak Ridge National Laboratory, yielded 7,281 open reading frames (ORFs) in a

9.25Mb genome, of which 73% was associated with a previously recognized ORF and 27% was unique to *N. punctiforme*. [5] This genome is among the largest in the microbial genomes that have been sequenced. However, the unicellular non-nitrogen fixing *Synechocystis sp.* Strain PCC 6803, with a genome of 3.57Mb was among the first microbial genomes sequenced. [6] It putatively encodes 3,215 proteins, 47% and 53% of which can be associated with known and hypothetical proteins respectively. The *Anabaena* PCC 7120 genome at 7.21Mb is about 78% of the size of *N. punctiforme* and encodes 6,132 ORFs. Approximately 79% of its ORFs find similarity with *N. punctiforme* genome. Reciprocal BLAST analyses indicate that 486 ORFs, which encode hypothetical proteins, are present in both *N. punctiforme* and *Anabaena* 7120 genomes and absent in the genomes of all other organisms (Table 1 in supplementary material). The number of ORFs encoding hypothetical proteins in *N. punctiforme* and *Anabaena* PCC 7120 are similar to the total coding capacity of a small cyanobacterial genome-*Prochlorococcus marinus* strain MED 4 [7] (Table 1 in supplementary material).

Comment on heterocyst coding genes

In *N. punctiforme* a heterocyst regulatory-gene product, *HetR* may [8] or may not [9] have been involved in the induction of akinite differentiation. However, a gene, *avaK*, whose gene products was found to be enriched in akinites of *Anabaena variabilis* was found in homologue to both *N. punctiforme* and *Anabaena* 7120 genomes. [10] Moreover, their associations with plants regulated hormogonium inducing factor (HIF), that induced the transcriptional genes *SigH* and *cpH*. [11] Perhaps, in a heterotrophic metabolic mode the rate of nitrogen fixation was elevated with the excess fixed nitrogen being released to the plant partner. Although the informatics comparison on different components of cyanobacteria has been initialized, work is a standstill due to unavailability of information on some of the other organisms (Table 1 in supplementary material).

Informatics on cyanobacterial phylogeny

Although, systematic classification of microbes in general are present, the phylogeny of the cyanobacteria is poorly understood till date because most classification schemes were organized by cell or colony shape, but recent efforts with the use of genomic tools have been widely used for molecular sequences that support to interpret the phylogeny for nitrogenase providing a true evolutionary schemes especially for *nifH*. Yes, hundreds of published and available nitrogenase gene sequences from many different species can now be interpreted to match with ribosomal genes to determine their consistency with multiple losses or with multiple transfers or with some combination of both processes to support genetic diversity analysis. Thus, in accordance to the above, the nitrogenase enzymes of all the nitrogen-fixing organisms (with few exceptions) were

found to be similar and were clearly derived from a common ancestor mainly through the process of horizontal gene transfer or independent loss in many lineages. However, the clades formed also indicated a large number of more distant relatives; some of these were well known proteins involved in the synthesis of photosynthetic pigment, namely protochlorophyllide reductase and chlorine reductase. As gene sequences became available, it was realized that *nifE* is a homolog of *nifD* and *nifN* is a homolog of *nifK*. Hence, it seems certain that *nifDK* and *nifEN* are the products of ancient gene duplication. Similarly, phylogenetic studies on the true branching cyanobacteria [12] revealed them to exhibit a high degree of morphological complexity. The phylogenetic trees constructed on the basis of the 16S rRNA, *rpoB* and *rbcLX* gene sequences of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* showed that the classification proposed [13] for these anabaenoids needed revision. With this being the current status, although the diversity of cyanobacteria is rich world wide requires more careful attention in the present scenario (Table 1 in supplementary material).

Conclusion:

It is too far for conclusion, because, although we are witnessing a remarkable change in the scale of molecular microbiological research and are entering an integrative "genomic science", the access to genome information is still very limited. Perhaps, the information on the sequences, gene organization and phylogeny reviewed in the present perspective would provide us the possibilities of enumerating the other potentials that clue about the origin of the nitrogenase genes in them. The location of the genes may indicate whether, they are long-term residents or recent intruders. Besides this, the genome information is most valuable in predicting the enzymatic machinery required for expression of proteins. Now with the availability of more complete sets of nitrogenase sequences in almost hundreds of microbial genomes, it is possible to reinforce the complexity and distinctness of the A, B and C types. Although, analyses of genomic context of the genes are at its infant level, still would promise to provide new insights into the distribution of nitrogenase genes and their integration into the metabolism of their host organisms.

References:

- [01] <http://fig.cox.miami.edu/Faculty/Dana/cyano.html>
- [02] D. Hughes, *Genome Biol.*, 1:6 (2000) [PMID: 11380986]
- [03] R. D. Fleischmann *et al.*, *Science*, 269:496 (1995) [PMID: 7542800]
- [04] <http://www.jgi.doe.gov/>
- [05] J. C. Meeks, *et al.*, *Photosynth Res.*, 70:85 (2001) [PMID: 16228364]
- [06] T. Kaneko, *et al.*, *DNA Res.*, 3:109 (1996) [PMID: 8905231]

-
- [07] W. R. Hess, *et al.*, *Photosynth. Res.* 70:53 (2001) [PMID: 16228362]
- [08] F. Leganes, *et al.*, *Mol. Microbiol.*, 12:679 (1994) [PMID: 7934891]
- [09] F. C. Wong and J. C. Meeks, *Microbiol.*, 148:315 (2002) [PMID: 11782524]
- [10] R. Zhou and C. P. Wolk, *J. Bacteriol.*, 184:2529 (2002) [PMID: 11948167]
- [11] E. L. Campbell, *et al.*, *J. Bacteriol.*, 180:4938 (1998) [PMID: 9733698]
- [12] A. Tomitani, *et al.*, *Proc Natl Acad Sci.*, 103:5442 (2006) [PMID: 16569695]
- [13] P. Rajaniemi, *et al.*, *Int J Syst Evol Microbiol.*, 55:11 (2005) [PMID: 15653847]

Edited by P. Kanguene

Citation: Lakshmi, *Bioinformatics* 2(1): 8-11 (2007)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material

Organism	Genome size	Contributor	Organization	Status
<i>Gloeobacter violaceus</i> PCC 7421	4.6 Mb	Satoshi Tabata	Kazusa DNA Research Institute, Japan http://www.kazusa.or.jp/eng/index.html	On going
<i>Microcystis aeruginosa</i>	4.8 Mb	Nicole Tandeau de Marsac	Institut Pasteur, Paris http://www.pasteur.fr/externe.html	On going
<i>Prochlorococcus marinus</i> MED4	1.66 Mb	Penny Chisholm	Joint Genome Institute	completed 2001
<i>Prochlorococcus marinus</i> MIT9313	2.4 Mb	Penny Chisholm	Joint Genome Institute	completed 2001
<i>Prochlorococcus marinus</i> SS120	1.8 Mb	Frédéric Partensky and Daniel Vault	Genoscope	On going
<i>Synechococcus</i> PCC 6301 (<i>Anacystis nidulans</i>)	unknown	Mamoru Sugita	Gene Research Center, Nagoya University http://www.gene.nagoya-u.ac.jp/index-e.html	No entry since December 2001
<i>Synechococcus</i> PCC 7002	unknown	Jindong Zhao	Beijing University	No entry October 2001
<i>Synechococcus</i> PCC 7942	unknown	Susan Golden	Originated by Texas A&M University but currently moved to Joint Genome Institute	On going
<i>Synechococcus</i> WH8102	2.4 Mb	Bianca Brahamsha, Brian Palenik, and John Waterbury	Joint Genome Institute	completed 2001
<i>Synechocystis</i> PCC 6803	3.6 Mb (a)	Tatsuo Omata	Kazusa DNA Research Institute http://cyano.genome.ad.jp/	completed 2002
<i>Thermosynechococcus elongatus</i> BP1	2.6 Mb	Tatsuo Omata	Kazusa DNA Research Institute, Community annotation project at TEORF	completed 2002
<i>Arthrospira (Spirulina) platensis</i>	unknown	Cheng-Cai Zhang	Human Genome Center, Beijing	No entry January 2001
<i>Trichodesmium erythraeum</i>	6.5 Mb	John Waterbury and Eric Webb	Joint Genome Institute	On going
<i>Anabaena</i> PCC 7120	7.2 Mb (b)	Tatsuo Omata	Kazusa DNA Research Institute, Community annotation project at ANORF http://cyano.genome.ad.jp/	completed 2001
<i>Nostoc punctiforme</i> ATCC 29133	9.25 Mb (c)	Jack Meeks	Joint Genome Institute	On going

Table 1: Numerical summary and workstations of cyanobacterial genomes is given. (a) 1,521 ORFs (47% of the total) encodes proteins that can be associated with probable known function; 1,694 ORFs (53%) encode conserved hypothetical proteins; 2,547 ORFs (80%) are present in *N. punctiforme*; 668 ORFs (20%) are unique to itself. (b) 4,157 ORFs (68% of the total) encode proteins with probable known functions; ii) 1,975 ORFs (32%) cannot be associated with a previously recognized ORF; 4,814 ORFs (79%) are present in *N. punctiforme*. 1,489 ORFs (24%) are unique to itself; 486 of 1,975 ORFs are present in *N. punctiforme*. (c) 5,314 ORFs (73% of the total) can be associated with a previously recognized ORF. 3,328 of the recognized ORFs (46% of the total) encode proteins with known or probable functions. 1,986 ORFs (27% of the total) encode conserved hypothetical proteins with no known functions. 3,965 ORFs (55%) are present in *Synechocystis* 6803. 1,967 ORFs (27%) cannot be associated with the previously recognized ORF, excluding ORFs in *Anabaena* 7120.