Functional profiling of cyanobacterial genomes and its role in ecological adaptations

Ratna Prabha [a], Dhananjaya P. Singh [a], Pallavi Somvanshi [b], Anil Rai [c]

[a] ICAR-National Bureau of Agriculturally Important Microorganisms, Indian Council of Agricultural Research, Kushmaur, Maunath Bhanjan 275103, India
[b] Department of Biotechnology, TERI University, New Delhi, India
[c] Center for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Indian Council of Agricultural Research, Library Avenue, New Delhi 110012, India

A R T I C L E   I N F O

A B S T R A C T

With the availability of complete genome sequences of many cyanobacterial species, it is becoming feasible to study the broad prospective of the environmental adaptation and the overall changes at transcriptional and translational level in these organisms. In the evolutionary phase, niche-specific competitive forces have resulted in specific features of the cyanobacterial genomes. In this study, functional composition of the 84 different cyanobacterial genomes and their adaptations to different environments was examined by identifying the genomic composition for specific cellular processes, which reflect their genomic functional profile and ecological adaptation. It was identified that among cyanobacterial genomes, metabolic genes have major share over other categories and differentiation of genomic functional profile was observed for the species inhabiting different habitats. The cyanobacteria of freshwater and other habitats accumulate large number of poorly characterized genes. Strain specific functions were also reported in many cyanobacterial members, of which an important feature was the occurrence of phage-related sequences. From this study, it can be speculated that habitat is one of the major factors in giving the shape of functional composition of cyanobacterial genomes towards their ecological adaptations.

1. Introduction

Cyanobacteria are oxygenic photosynthetic bacteria widely used as a model organism for the study of numerous biological processes including photosynthesis, nitrogen fixation, environmental stress tolerance and molecular evolution [1]. Cyanobacteria show broad diversity in terms of their morphology, habitat and functionalities [2,3], which are reflected in their genome and/or other pertinent genomic features (e.g., genome size, coding region, GC content). Cyanobacterial species are subjected to a variety of niche-specific competitive forces in the environment that result in unique characteristics of each genome, in the course of evolution [2]. They are reported to follow two types of trends in terms of their genome size; certain species involve a mix of expansion and reduction, e.g. genus *Acaryochloris* [4], whereas closely related marine picocyanobacteria in general tends to follow genome reduction or genome streamlining and maintain only a minimal gene repertoire [5–7]. Earlier report suggests that, in billion years of evolution, significant changes occur in the genomes of cyanobacteria possibly due to adaptation towards their habitats and this might have resulted in redundancy of duplicated genes and increased mutation rates [2].

With respect to genome size and gene content, prokaryotic genomes are constantly shifting, owing to the factors like gene duplication, gene loss, horizontal gene transfer, and *de novo* origin of the genes (gene genesis) [8]. Overall, the genome size of the bacteria is maintained in equilibrium between the duplication or horizontal transfer, and mutations leading to elimination of function(s) followed by deletions (gene loss) [9]. It has been already reported that larger genomes preferentially accumulate genes involved in the regulation (e.g., genes involved in secondary metabolism along with those related to energy conversion are ecologically more adaptable to the environments, where the resources are varied and poor) and complete understanding of this will provide the detailed insight into the interaction between ecology and genome evolution [10]. Furthermore, microbial genomes are subjected to horizontal gene transfer related events and acquire foreign DNA from the surroundings more frequently in comparison to the higher organisms [11]. The events of horizontal gene transfer in combination with intragenomic rearrangements and duplication are accountable for bacterial adaptations in different environmental niches and variance in closely related species. Phylogenomics studies have shown a complex evolutionary pattern for the microbes that undergo not only vertical descent or lateral gene transfer but also include a mix of recombination, gene duplication, gene invention, gene loss, gene degradation, convergence and selection processes [12,13].

Keeping aforesaid facts in view and the availability of a large number of cyanobacterial sequenced genomes, here in this study we identified the functional composition of 84 cyanobacterial genomes and based

upon the findings a correlation has been deduced with the adaptation of these organisms to various ecological niches. We attempted to compare the cyanobacterial species for the identification of the part/genes of their genome contributing to a particular cellular process as it may give reflection of both the cellular and ecological strategies associated with genome expansion and adaptation to the environmental conditions.

2. Material and methods

2.1. Genome sequences

The complete genome sequence of 84 cyanobacterial strains available at Integrated Microbial Genomes (IMG) database (https://img.jgi.doe.gov/cgi-bin/w/main.cgi) were retrieved for this study. Among these 84 genomes, 69 species occupy aquatic habitats (freshwater: 33 sps.; marine: 30 sps.; hot spring: 06 sps.) (genome size in a range of 1.44–8.73 Mb), 06 terrestrial species (genome size: 6.69–8.27 Mb), 05 members occupy multiple habitats (genome size: 4.66–9.06 Mb) while 04 species acquire other habitats (genome size: 4.68–5.13 Mb) (Table S1). The major characteristics of all the cyanobacteria (e.g., habitat, genome size, no. of genes etc.) included in this study have been given in Table S1.

2.2. Functional characterization and Clusters of Orthologous Groups (COG) assignment

Functional characterization of cyanobacterial genomes was performed by using the Clusters of Orthologous Groups (COG) database [14]. For each cyanobacterial genome, all the genes were subjected to COG assignment using the Function Profile tool (IMG database) [15]. The Function Profile tool assists in the identification of the genes associated with a particular function in 'query genome' and thus, genes are expected to share at least the same functions associated with their COG matches. Once the genes were assigned to the COGs, they were clustered into 23 functional categories, which were further grouped into four major classes (Table 1).

2.3. COG categorization

Practical Extraction and Report Language (PERL) scripts were used for arranging all the COGs respective of their functional category for each cyanobacterial species. PERL scripts were also used for the analysis of the distribution pattern of each COG category across the members of the dataset.

2.4. Statistical analysis

All the statistical analyses involved in this study were performed by using the SPSS Version 16.0 software.

3. Results and discussion

3.1. General characteristics

The size of cyanobacterial genomes lies in a range of 1.44 Mb (CAt_ALOHA) to 9.06 Mb (Np_PCC_73102) (Table S1). All the genomes contain single circular chromosome as their major genetic material, whereas, an additional chromosome was present in three cyanobacterial species i.e., Cs_ATCC51142, Av_ATCC29413 and As_90. Members of the dataset occupied diverse habitats, i.e., aquatic (marine, freshwater or hot springs), terrestrial, and some represent multiple habitats (Table S1). It was also found that cyanobacteria exhibiting marine habitats tend to have lower genome size in comparison with those residing in other habitats (Table S1).

Table 1
Functional categories identified in the cyanobacterial genomes.

S. no.	Function class	Identified COG categories
1	Metabolism	Amino acid transport and metabolism (E)
		Carbohydrate transport and metabolism (G)
		Nucleotide transport and metabolism (F)
		Energy production and conversion (C)
		Coenzyme transport and metabolism (H)
		Lipid transport and metabolism (I)
		Inorganic ion transport and metabolism (P)
		Secondary metabolite biosynthesis, transport and catabolism (Q)
2	Cellular processes and signalling	Cell wall/membrane/ envelope biogenesis (M)
		Cell motility (N)
		Cell cycle control, cell division, chromosome partitioning (D)
		Posttranslational modification, protein turnover, chaperones (O)
		Signal transduction mechanisms (T)
		Intracellular trafficking, secretion, and vesicular transport (U)
		Defense mechanisms (V)
		Cytoskeleton (Z)
3	Information storage & processing	RNA processing and modification (A)
		Chromatin structure and dynamics (B)
		Translation, ribosomal structure and biogenesis (J)
		Transcription (K)
		Replication, recombination and repair (L)
4	Poorly categorized	Function unknown (S)
		General function prediction only (R)

3.2. COG assignment and functional genomics profiling of cyanobacteria

We were able to assign on an average of ~53% of the genes in any genome to a particular COG functional category by using the Function Profile tool. Among the cyanobacterial species, 2616 individual COGs from 23 different functional categories of four major classes, i.e., Metabolism, Cellular processes and signalling, Information storage & processing, and Poorly categorized were identified (Table 1). Only 400 COGs (15%) were common in all the 84 cyanobacteria, while remaining represent the functions of pan-genome for these organisms, in which ~06.68% functions are strain-specific, i.e., owned by only one member. Among all these 23 categories, Cell motility (N), Chromatin structure and dynamics (B), Cytoskeleton (Z), and RNA processing and modification (A) categories were not present in all the cyanobacteria (Table 1). While analyzing the distribution of each functional category, it was observed that across all cyanobacterial species, genes associated with metabolic functions gained maximum share (Fig. 1). The next most abundant functional category in most of the cyanobacteria (specifically those inhabiting freshwater and multiple habitat) was that of poorly categorized genes, i.e., Function unknown (S) and General function prediction only (R). Marine cyanobacteria preferred genes for Information Storage and Processing over Cellular Processes and Signalling, whereas, the later one is mostly preferred by the cyanobacteria from other habitats (Fig. 2). Further, hierarchical clustering on the basis of functional categories reflected two different groups among the cyanobacteria, the first group (group I) included the members from diverse habitats (freshwater, multiple, other), while the second group (group II) specifically included marine and hot-spring cyanobacteria with only three freshwater species (Fig. 2). Both the above mentioned groups showed different functional profiles. In general, habitat seems to influence the functional profile as members from similar habitats possess similar kind of functional profile (Fig. 2). Earlier, it has been reported that bacterial genomes contain specific functional gene inventories, which are in concurrence with their survival in the particular ecological niche [16–18]. In the bi-variate correlation analysis, the larger genomes showed a significant strong positive correlation with the functional categories