Ann Microbiol (2015) 65:799–807
DOI 10.1007/s13213-014-0920-1

ORIGINAL ARTICLE

Molecular phylogeny and evogenomics of heterocystous cyanobacteria using *rbcl* gene sequence data

Prashant Singh · Arshia Fatma · Arun Kumar Mishra

Received: 25 July 2013 / Accepted: 15 May 2014 / Published online: 30 May 2014

Abstract Taxonomic affiliations and molecular diversity of 41 heterocystous cyanobacteria representing 12 genera have been assessed on an evolutionary landscape using *rbcl* gene sequence data-based phylogenomics and evogenomics approaches. Phylogenetic affiliations have clearly demonstrated the polyphyly of the true branching cyanobacteria, along with a frequent intermixing amongst the heterocystous cyanobacteria. The monophyletic origin of the heterocystous cyanobacteria was also quite evident from maximum parsimony and neighbor joining analyses. Incongruency with the traditional scheme of cyanobacterial taxonomy was frequently observed, thus advocating towards some re-amendments in the cyanobacterial classificatory schemes. Evogenomics analyses of gene sequence data gave a clear indication about the greater evolutionary pace of the unbranched cyanobacteria as compared to the branched forms. It was evident that the order Nostocales would be controlling the future pace of evolution of heterocystous cyanobacteria. The cyanobacteria *Nostoc* was found to have the greatest genetic heterogeneity amongst the studied genera, along with some evidence towards events of lateral gene transfer amongst the heterocystous cyanobacteria in case of the *rbcl* gene. Thus, heterocystous cyanobacteria were found to be a fast evolving group, with estimates of gene conversion tracts pointing towards the unbranched heterocystous cyanobacteria being at the base of evolutionary diversifications of the complete heterocystous lineage.

P. Singh · A. Fatma · A. K. Mishra (✉)
Laboratory of Microbial Genetics, Department of Botany, Banaras Hindu University, Varanasi 221005, India
e-mail: akmishraau@rediffmail.com

A. K. Mishra
e-mail: akmishraau@hotmail.com

Keywords Heterocystous cyanobacteria · *rbcl* gene · Phylogeny · Nucleotide diversity · Gene conversion tracts · Evolutionary pace

Introduction

Cyanobacteria are one of the most morphologically and genetically diverse groups of prokaryotes and show significant efficiency for cellular and colonial differentiation. They can be grossly considered as one of the most prevalent groups of photosynthetic prokaryotes having an all-inclusive distribution and occupying a vast array of terrestrial and aquatic environments (Meeks and Elhai 2002). They are also responsible for converting the ancient Earth's anaerobic atmosphere to an aerobic one, which led to the creation of an aerobic environment that has allowed for the existence of life on earth and the diversity of life that exists today. They constitute one of the oldest and most ancient groups of organisms, with fossil records that date back to about 3.5 billion years (Henson et al. 2004; Singh et al. 2013). Molecular phylogenetic approaches have made it evident that all photoautotrophic eukaryotes (plants and algae) share a single origin, as well as a common endosymbiotic ancestry for cyanobacteria-derived chloroplasts (Yoon et al. 2004). Cyanobacteria are one of the premier photosynthetic prokaryotes, along with being very important nitrogen fixers through the nitrogenase gene complex (Lammers and Haselkorn 1984).

Cyanobacterial classification based on morphological characters has long been a matter of immense debate, because of the incongruency of molecular findings with the morphological schemes (Mollenhauer 1988; Mishra et al. 2013). A major drawback of this scheme was that it did not represent the true evolutionary relationships and phylogeny within the cyanobacterial lineage (Komárek and Anagnostidis 1989). Characteristics of vegetative cells, heterocysts and akinetes

are controversial, and are considerably altered due to phenotypic changes under different environmental conditions and selective culturing conditions (Desikachary 1959). Cyanobacteriologists all across the globe have done a lot of work to test and prove the inadequacy of the morphological scheme of classification, and it is estimated that around 50 % of the cyanobacterial strains found in culture collections all over the world have been wrongly marked (Komárek and Anagnostidis 1989).

Molecular parameters based on PCR based methods, such as the 16S rRNA sequences, *nif* gene(s), phycocyanin encoding locus and RNA polymerase gene (Giovannoni et al. 1988; Ludwig and Schleifer 1994; Nelissen et al. 1996; Lee et al. 1996; Zehr et al. 1997; Honda et al. 1999; Neilan et al. 2002; Hartmann and Barnum 2010; Mishra et al. 2013; Singh et al. 2013), have been used as genetic markers for assessing cyanobacterial diversity and phylogenetic designs. A very important functional gene present in cyanobacteria is the plastid-encoded *rbcl* gene. Apart from being a single copy gene, approximately 1,430 base pairs in length, it is also free from length mutations except at the far 3′ end, and is known to have a fairly conservative rate of evolution. The form I enzyme, found predominantly in plants, eukaryotic algae, and cyanobacteria, contains large (*rbcl*) and small (*rbcs*) subunits that have been shown to assemble into a complex hexadecameric structure, $(L_2)_4(S_4)_2$. Its catalytic site contains active amino acids from two neighboring large subunits, and studies of site-specific enzymes indicate that the small subunit is required for maximal catalysis and contributes to CO_2/O_2 specificity. Reports about the phylogenetic assessment of the *rbcl* gene are much fewer (Morden and Golden 1991; Gugger et al. 2002) and do not encompass many of the cyanobacterial genera. Hence, in this manuscript, the *rbcl* gene has been used as a molecular marker to establish the phylogenetic tendencies, taxonomic affiliations and evolutionary pace of heterocystous cyanobacteria representing 12 genera of subsections IV and V. This work is in continuation of our reports on the *nif*H gene (Singh et al. 2013) and 16S rRNA and *nif*D and *psb*A genes (communicated), using the same set of 41 heterocystous cyanobacteria collected and isolated from different parts of India.

Materials and methods

Growth and maintenance of cultures

Forty-one freshwater-dwelling heterocystous cyanobacterial strains were isolated, identified, purified and selected for the *rbcl*-gene-based genetic diversity and phylogenetic analysis in the present communication (Table 1). The strains were collected from many parts of India, with the constant feature being the fresh water habitat of the samples (Singh et al. 2013). The heterocystous cyanobacterial strains were grown axenically in 150 ml basal medium (BG-11_0 medium) (Rippka et al. 1979) in cotton-stoppered Erlenmeyer flasks (capacity 500 ml). pH of the medium was adjusted to 7.4 and the cultures were maintained in a culture room under illumination of approximately 50–55 μE m^{-2} s^{-1} with a 14/10 h light/dark cycle at 28±2 °C.

DNA isolation and PCR amplification of the *rbcl* gene

The DNA was isolated from 12-day-old cultures using Himedia Ultrasensitive Spin Purification Kit (MB505) with some modifications (Singh et al. 2013). The DNA was stored at −20 °C. The *rbcl* gene was amplified using the primer pairs *rbclf* (5′-GACTTCACCAAAGAYGACGAAAACAT-3′) and *rbclr* (5′-GAACTCGAACTTRATYTCTTTCCA-3′) (Halinen et al. 2007). The amplification reaction was carried out at an initial denaturation of DNA at 94 °C for 5 min, followed by 30 cycles of denaturation at 92 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 2 min. The final extension was done at 72 °C for 6 min, followed by incubation at 4 °C for 20 min. The appropriate sized bands of 700 bp were cut using sharp aseptic blades and sent for sequencing.

Phylogenomics analyses

The phylogenomic analysis was done using MEGA 5.1 Beta 2 version (Tamura et al. 2011). The phylogenetic tree was constructed using maximum parsimony and neighbor joining algorithms. The consensus tree was inferred from the 15 most parsimonious trees, and branches corresponding to partitions reproduced in less than 50 % of trees were collapsed. The neighbor joining tree was constructed with an aim to find an optimum tree with balanced minimum evolution. Also, since the algorithm does not assume that all the lineages evolved at the same rate, it was more appropriate to use the neighbor joining algorithm for estimating the phylogeny of investigated cyanobacterial strains (Gascuel and Steel 2006). The neighbor joining tree was computed using the number of differences method, and are in the units of the number of base differences per sequence.

Evogenomics analyses

Evolutionary tendencies of the heterocystous cyanobacteria were assessed by analyzing the gene sequences of the sampled taxa in perspective of calculating nucleotide diversity (level I), recombination frequencies (level II), and the DNA divergence (level III) that finally helped in studying the future course of evolution of the heterocystous cyanobacteria, based on *rbcl* gene sequences using the software DnaSP 5.10 (Librado and Rozas 2009).

Table 1 List of cyanobacterial strains used in present study, with habitats and accession numbers

S. No.	Cyanobacterial strain	Geographical location	Habitat	Accession number of *rbcl gene*
1.	*Anabaena doliolum* Ind1	Pond, Indian Institute of Technology Workshop, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JQ918770
2.	*Anabaena doliolum* Ind2	Pond, Varanasi, Uttar Pradesh, India	Freshwater	JX993226
3.	*Anabaena oryzae* Ind3	Dried water body, Vishwanath Temple, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JQ918771
4.	*Anabaena oryzae* Ind4	Dried water body, Vishwanath Temple, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JX993227
5.	*Anabaena* sp. Ind5	Pond, Indian Institute of Technology Workshop, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JX993228
6.	*Anabaena* sp. Ind6	Paddy field, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JQ918773
7.	*Anabaenopsis* sp. Ind8	Water tank, Mandapum Sea Beach, Chennai, Tamil Nadu, India	Freshwater	JQ918783
8.	*Calothrix brevissima* Ind9	Humid and moist rocky crevices as epiphytes on *Hydrodictyon*, Windham Falls, Barkachha, Mirzapur, Uttar Pradesh, India	Freshwater	JX993230
9.	*Calothrix brevissima* Ind10	Bee Falls, Panchamarhi, Madhya Pradesh, India	Freshwater	JQ918777
10.	*Calothrix* sp. Ind11	Panchamarhi, Madhya Pradesh, India	Freshwater	JX993231
11.	*Cylindrospermum muscicola* Ind12	Paddy field, Agricultural Farm, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JQ918784
12.	*Cylindrospermum muscicola* Ind13	Paddy Field, Raksaul, Bihar, India	Freshwater	JX993232
13.	*Cylindrospermum* sp. Ind14	Paddy field, Agricultural Farm, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JX993233
14.	*Cylindrospermum stagnale* Ind15	Stagnant waters of paddy field, Agricultural Farm, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JQ918785
15.	*Tolypothrix tenuis* Ind16	Pallikaranai Marsh Reserve Forest, Chennai, Tamil Nadu, India	Freshwater	JQ918778
16.	*Tolypothrix nodosa* Ind17	Nanmangalam Reserved Forest, Medavakkam, Tamil Nadu, India	Freshwater	JX993234
17.	*Westiellopsis* sp. Ind19	Humid and moist rocky crevices, Windham Falls, Barkachha, Mirzapur, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JX993235
18.	*Westiellopsis* sp. Ind20	Inside rocky and humid crevices, Arunachal Pradesh, India	Freshwater	JQ918780
19.	*Hapalosiphon welwitschii* Ind21	Centre for collection and utilization of Blue Green Algae (CCUBGA), IARI, PUSA, New Delhi, India	Culture Collection	JX993236
20.	*Hapalosiphon welwitschii* Ind22	Doimukh, Itanagar, Arunachal Pradesh, India	Freshwater	JQ918786
21.	*Hapalosiphon* sp. Ind23	Paddy field, Barkachha, Mirzapur, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JX993237
22.	*Scytonema bohnerii* Ind24	Near Pond, Indian Institute of Technology Workshop, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JQ918782
23.	*Scytonema bohnerii* Ind25	dripping rocks, Kodaikanal, Tamil Nadu, India	Freshwater	JX993238
24.	*Fischerella* sp. Ind26	Paddy field, Babatpur, Varanasi, Uttar Pradesh, India	Freshwater	JQ918779
25.	*Nostochopsis* sp. Ind28	Paddy field, Narayanpur, Varanasi, Uttar Pradesh, India	Freshwater	JX993239
26.	*Mastigocladus laminosus* Ind29	Muddy paddy field, Jalandhar, Punjab, India	Freshwater	JQ918781
27.	*Nostoc calcicola* Ind30	Dried Pond, Department of Botany, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JX993240
28.	*Nostoc calcicola* Ind31	Paddy field, Nashik, Maharashtra, India	Freshwater	JX993241
29.	*Nostoc calcicola* Ind32	Dried water body, Varanasi, Uttar Pradesh, India	Freshwater	JX993242
30.	*Nostoc muscorum* Ind33	Paddy field, Agricultural Farm, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JQ918774
31.	*Nostoc muscorum* Ind34	Paddy field, Agricultural Farm, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JX993243
32.	*Nostoc* sp. Ind36	Paddy field almost devoid of water, Arunachal Pradesh, India	Freshwater	JX993244
33.	*Nostoc* sp. Ind37	Fresh Water Pond, Arunachal Pradesh, India	Freshwater	JX993245
34.	*Nostoc* sp. Ind39	Paddy field, Nashik, Maharashtra, India	Freshwater	JX993247
35.	*Nostoc* sp. Ind40.1	Paddy field, Nashik, Maharashtra, India	Freshwater	JX993248
36.	*Nostoc* sp. Ind40.2	Paddy field, Nashik, Maharashtra, India	Freshwater	JX993249

Table 1 (continued)

S. No.	Cyanobacterial strain	Geographical location	Habitat	Accession number of *rbcl gene*
37.	*Nostoc* sp. Ind40.3	Paddy field, Nashik, Maharashtra, India	Freshwater	JX993250
38.	*Nostoc spongiforme* Ind41	Water body, Varanasi, Uttar Pradesh, India	Freshwater	JX993251
39.	*Nostoc spongiforme* Ind42	Botanical Garden, Department of Botany, Banaras Hindu University, Uttar Pradesh, India	Freshwater	JQ918776
40.	*Anabaena* sp. PCC 7120 Ind43	Laboratory of Prof. Peter Wolk, University of Michigan, USA	Culture Collection	JQ918772
41.	*Fischerella* sp. Ind81	Banaras Hindu University, Varanasi, Uttar Pradesh, India	Freshwater	JX993253

Results

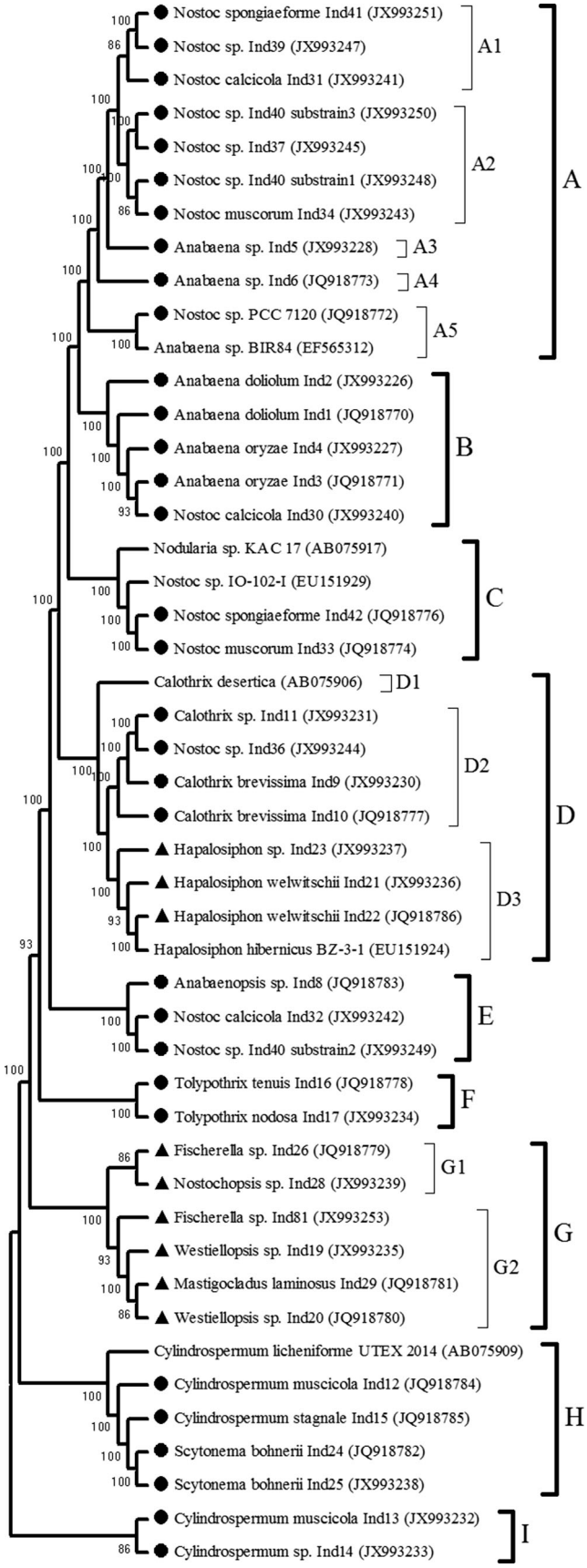
Insights into the molecular phylogeny

The maximum parsimony tree, constructed using the *rbcl* gene sequences, revealed nine interesting groups, viz., A, B, C, D, E, F, G, H and I (Fig. 1). Group A comprised of the closely related cyanobacterial genera *Nostoc* and *Anabaena*. All the branches had strong bootstrap values, with the minimum being 86. Group A was further subdivided into five subgroups, viz., A1, A2, A3, A4 and A5, with *Nostoc* and *Anabaena* species being evident at all positions. Group B comprised of four *Anabaena* strains along with the strain *Nostoc calcicola* Ind 30. Group C comprised of four members of the order Nostocales, in which the gene bank strain *Nodularia* sp. KAC 17 separated first from the gene bank strain *Nostoc* sp. IO-102-I, which further diverged to strains *Nostoc spongiforme* Ind 42 and *Nostoc muscorum* Ind 33. Group D was a larger, much complex and interesting group comprising of three subgroups, D1, D2 and D3, with representatives from both the unbranched heterocystous and the branched heterocystous orders. Group E was represented by the strain *Nostoc calcicola* Ind 32 pairing with *Nostoc* sp. Ind 40 substrain 2. This pairing was in a tight conjunction with the strain *Anabaenopsis* sp. Ind 8, with a bootstrap value of 100. Group F was solely represented by the false branching species *Tolypothrix tenuis* Ind 16 and *Tolypothrix nodosa* Ind 17. Group G comprised of the branching forms exclusively, with two subgroups, G1 and G2. Group H was represented by the strains *Cylindrospermum muscicola* Ind 12 and *Cylindrospermum stagnale* Ind 15, in conjunction with the false branching strains *Scytonema bohnerii* Ind 24 and Ind 25. Finally, Group I saw the clustering of strains *Cylindrospermum muscicola* Ind 13 and *Cylindrospermum* sp. Ind 14.

The neighbor joining tree showed two major groups, A and B (Fig. 2). Group A further subdivided into six subgroups: A1, A2, A3, A4, A5 and A6. Subgroup A1 exclusively comprised of the two nostocalean genera *Nostoc* and *Anabaena*, with the bootstrap values being satisfactorily robust enough at all the nodes. Noteworthy was the aligning of *Nostoc* sp. PCC 7120 with the gene bank strain *Anabaena* sp. BIR 84, thus supporting and providing validity to the findings of the maximum parsimony tree. Subgroup A2 comprised of four *Anabaena* strains sharing affinity with *Nostoc calcicola* Ind 30. Subgroup A3 comprised of two gene bank strains, *Nodularia* sp. KAC 17 and *Nostoc* sp. IO-102-I with *Nostoc spongiforme* Ind 42 and *Nostoc muscorum* Ind 33. Subgroup A4 had representatives of the true branching cyanobacteria *Hapalosiphon*, along with an initial divergence of the gene bank strain *Calothrix desertica*. Group A5 was composed of strains belonging to the genus *Calothrix*, along with the strain *Nostoc* sp. Ind 36. Finally, subgroup A6 was represented by *Anabaenopsis* sp. Ind 8, along with *Nostoc calcicola* Ind 32 and *Nostoc* sp. Ind 40.2 being in conjunction. Group B had two major subgroups, B1 and B2. Group B1 consisted entirely of the true branching forms, with the strains being *Mastigocladus laminosus* Ind 29, *Westiellopsis* sp. Ind 19 and Ind 20, *Fischerella* sp. Ind 26 and Ind 81, and finally, *Nostochopsis* sp. Ind 28. Sub-subgroup B2-1 was an assemblage of false branching cyanobacteria *Scytonema bohnerii* Ind 24 and Ind 25, while sub-subgroup B2-2 was represented by the false branching strains *Tolypothrix tenuis* Ind 16 and *Tolypothrix nodosa* Ind 17. Sub-subgroup B2-3 was an exclusive assemblage of *Cylindrospermum* strains, along with the gene bank reference strain *Cylindrospermum licheniforme* UTEX 2014.

Evogenomics analyses of *rbcl* gene sequences

We tested our phylogenomics results using intricate evogenomic tools so as to find whether any correlation existed between the two experiments (Table 2). The analyses of the heterocystous cyanobacteria on the basis of the nucleotide sequences resulted in very interesting findings. In the level I analyses, the nucleotide diversity per site (*Pi*) was found to be more in the unbranched Nostocales order than in the branched Stigonematales. In the level II analyses, we conducted recombination analyses in between the two intermixed orders and



◀ **Fig. 1** Maximum parsimony analysis of heterocystous cyanobacteria using *rbcl* gene sequence data. The evolutionary history was inferred using the maximum parsimony (MP) method. The consensus tree inferred from 15 most parsimonious trees is shown. Branches corresponding to partitions reproduced in less than 50 % of trees are collapsed. The consistency index is (0.672414), the retention index is (0.940470), and the composite index is 0.634145 (0.632385) for all sites and parsimony-informative sites. The percentages of parsimonious trees in which the associated taxa clustered together are shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm (Nei and Kumar 2000) with search level 1, in which the initial trees were obtained with the random addition of sequences (ten replicates). The analysis involved 47 nucleotide sequences. Codon positions included were 1st+2nd+3rd. All positions containing gaps and missing data were eliminated. There were a total of 166 positions in the final data set. Evolutionary analyses were conducted in MEGA5 (Tamura et al. 2011)

calculated the value of *R* per gene, which was found to be greater in the Nostocales in comparison to the Stigonematales. In the level III analyses, on studying DNA divergence between the orders, the hypothesis of the simpler Nostocales being more evolutionary progressive gained more evogenomic proof. Further, gene conversion tracts were also calculated individually in the 12 genera of heterocystous cyanobacteria in order to assess intergenic evolutionary pace, and that also supported the greater heterogeneity in the genus *Nostoc* (Table 3). Thus, our *rbcl* gene analysis helped in recognizing the importance of the unbranched cyanobacteria in deciding the course of evolution and origin of new species of cyanobacteria.

Discussion

Phylogenetic analyses

The phylogenetic analyses using the maximum parsimony and the neighbor joining methods revealed some crucial insights into the genetic diversity and the phylogenetic relatedness of heterocystous cyanobacteria. The genera *Nostoc* paired with the genera *Anabaena* in most of the cases in both the trees constructed. But, the immense genetic diversity of the *Nostoc* genera was reflected in some of the strains, even appearing in different clusters. Groups A1 and A2 in the maximum parsimony tree and group A1 in the neighbor joining tree had similar strains grouping with each other. The strains *Nostoc spongiforme* Ind 41 and *Nostoc* sp. Ind 39 clustered with each other in both the trees, irrespective of the fact that their sites of collection were far apart. Thus, from all of the small subgroups, it seemed pretty evident that the genotypes of the *Nostoc* genera were not being influenced by geographical constraints, as all the isolates were from freshwater, and hence had almost similar environmental checks.

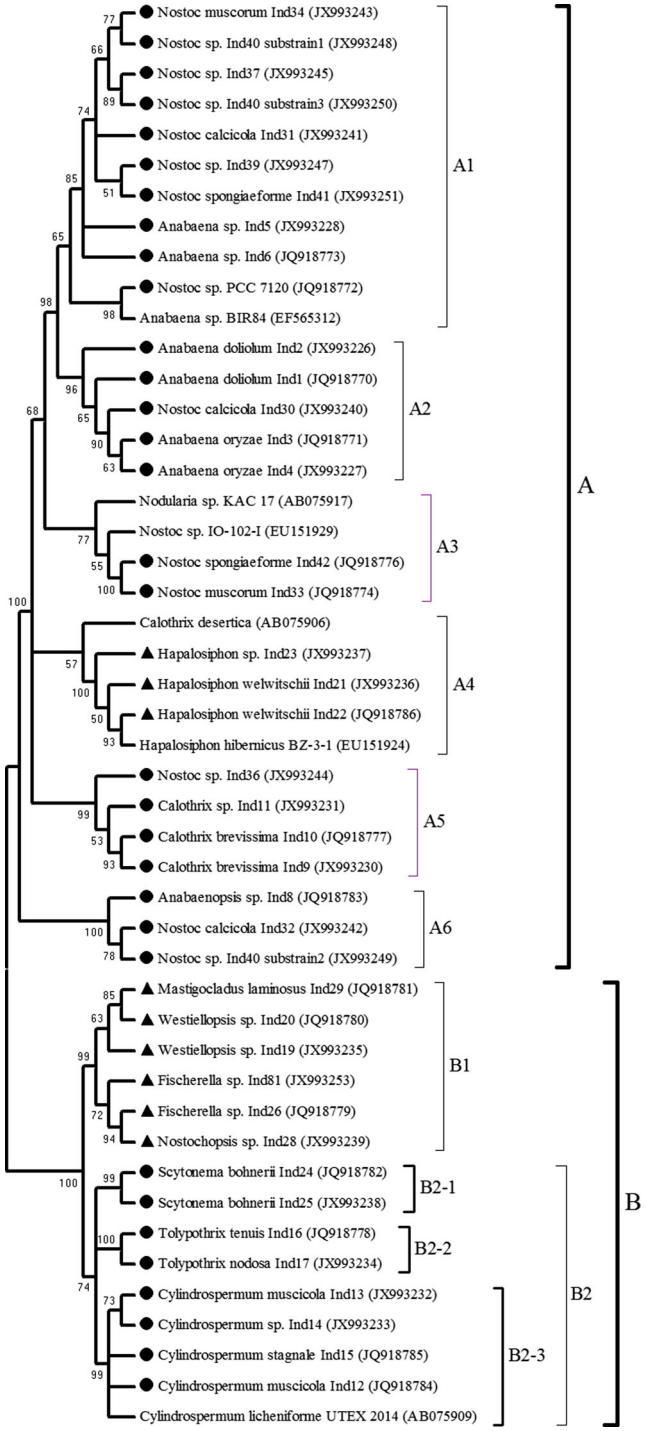


Fig. 2 The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The optimal tree with the sum of branch length=299.60745239 is shown. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches (Felsenstein 1985). The evolutionary distances were computed using the number of differences method (Nei and Kumar 2000) and are in the units of the number of base differences per sequence. The analysis involved 47 nucleotide sequences. Codon positions included were 1st+2nd+3rd. All positions containing gaps and missing data were eliminated. There were a total of 166 positions in the final data set. Evolutionary analyses were conducted in MEGA5 (Tamura et al. 2011)

The genera *Anabaena* also had overall similar clustering tendencies in both the trees. The alignment of *Nostoc* sp. PCC 7120 with the gene bank strain *Anabaena* sp. BIR84 in both the trees once again pointed towards the genetic relatedness of this important strain PCC 7120 with the genera *Anabaena*. Thus, the *rbcl* gene that we characterized has once again reignited the continuous debate over the placement of strain PCC 7120 in *Nostoc* or *Anabaena*. In some of our own reports using 16S rRNA, and *nif*D and *psb*A genes (communicated), we have already advocated the assigning of strain PCC 7120 into a new genus. When considering on an overall scale, it was still evident that *Nostoc* strains had greater genetic heterogeneity as compared to the *Anabaena* strains, which is in accordance with our own reports using the *nif*H gene as a molecular marker (Singh et al. 2013). It has already been reported that *Nostoc* and *Anabaena* show a high level of genetic heterogeneity (Rajaniemi et al. 2005), and our *rbcl* gene locus studies also adhere to the above-mentioned findings.

The genera *Cylindrospermum* represented an interesting case in our study, as in both the trees, it showed deviating affinity from the rest of the nostocalean members, thus pointing towards an evolutionary uncertainty (Rajaniemi et al. 2005) along with a high level of genetic heterogeneity in nostocalean genera. In the maximum parsimony tree, the *Cylindrospermum* strains aligned with the false branching *Scytonema bohnerii* strains in the group H with a good bootstrap value, thus pointing towards an evolutionary relatedness of unbranched and false branched cyanobacterial forms. It was also notable to observe the *Tolypothrix* strains in our study taking an intermediate position between the true branched and unbranched groups (group F of the maximum parsimony tree and sub-subgroup B2-2 in the neighbor joining tree) showing evidence of the false branched cyanobacteria to be intermediate in between the two groups. Further, it has been observed that both the false branching genera paired somewhat distantly in the maximum parsimony tree, while they were relatively closer in the neighbor joining tree. This shifting behavior also suggested having a re-look at the morphological classificatory schemes, as our genetic analyses failed to conclave with the traditional scheme of taxonomy, as has already been reflected by some other reports (Berrendero et al. 2008; Sihvonen et al. 2007; Mishra et al. 2013; Singh et al. 2013 Similar shifting tendencies were also observed in the genera *Calothrix*, where a relatively better proximity was observed with the true branching cyanobacteria *Hapalosiphon* in group D in the maximum parsimony tree along with *Nostoc* sp. Ind 36, suggesting once again the intermixing behavior of heterocystous cyanobacteria (Gugger and Hoffmann 2004). The placement of the *Calothrix* strains was not adhering with our findings (Singh et al. 2013), where a very tight clustering of this genus was observed using the *nif*H gene as a molecular marker. It was hence found impossible to assign this genus to its original taxonomic affiliations in the order Nostocales, as it

Table 2 Population genetics analysis of heterocystous cyanobacteria using the *rbcl* gene sequences

Parameter	Unbranched (members of Nostocales)	Branched (members of Stigonematales)	All strains
Monomorphic sites	0	0	0
Polymorphic sites (segregating sites; *S*)	270	345	270
Singleton variable sites	0	27	0
Parsimony informative sites	270	318	270
Nucleotide diversity per site (*Pi*)	0.71559	0.67552	0.72054
Theta per site from total number of mutations (*Eta*)	0.74125	0.88625	0.70031
Average number of nucleotide differences (*k*)	193.210	233.056	194.545
Recombinational analyses			
R per gene	277	77.7	428
Estimate of *R* between adjacent sites	0.5395	0.1519	0.8342
Average nucleotide distance between the most distant sites	513.44	511.67	513.05
Minimum number of recombination events detected	0	16	0
DNA divergence between the groups			
Average number of nucleotide differences between groups:		198.212	
Number of net nucleotide substitutions per site between the groups, *Da*:		0.03630	
Number of gene conversion tracts identified between groups:		45 (40 in Nostocales and 5 in Stigonematales)	
Mutations polymorphic in group 1, but monomorphic in group 2:		153	
Mutations polymorphic in group 2, but monomorphic in group 1:		3	

exhibited affinity towards the true branching and the unbranched cyanobacteria in the maximum parsimony tree. The results of the maximum parsimony tree were further confirmed in the neighbor joining tree, where the gene bank strain *Calothrix desertica* paired with *Hapalosiphon* genus, while the strains in subgroup A5 in our study formed a cluster with *Nostoc* sp. Ind 36, with good bootstrap values. The strain *Anabaenopsis* sp. Ind 8 exhibited relatedness with the genus *Nostoc* in both the trees constructed, and hence this finding also deviated from our own earlier reports using the *nif*H gene (Singh et al. 2013), along with reports using the 16S rRNA and RFLP markers (Iteman et al. 2002), where this strain had shown a distinct separating trend from the rest of the heterocystous cyanobacteria. On analysis of the true branching cyanobacteria, it was observed that the genus *Hapalosiphon* paired separately from the rest of the true branching cyanobacteria, thus pointing towards the suggested polyphyletic origin of the order Stigonematales (Zehr et al. 1997; Gugger and Hoffmann 2004; Henson et al. 2004; Singh et al. 2013). The alignment of this genus separately from the rest of the true branching forms was also in coherence with our own *nif*H analyses (Singh et al. 2013), thus suggesting a possible distinctiveness in the genetic architecture of the genus *Hapalosiphon*. The rest of the branched forms all merged into separate and distinct groups with intermixing amongst them, and thus the postulated intermixing of heterocystous cyanobacteria along with the polyphyly of true branching stigonematalean forms was found true.

Table 3 Number of gene conversion tracts in different genera of heterocystous cyanobacteria

S.No.	Cyanobacteria	Number of gene conversion tracts
1.	*Nostoc*	19
2.	*Anabaena*	2
3.	*Cylindrospermum*	7
4.	*Calothrix*	2
5.	*Anabaenopsis*	1
6.	*Scytonema*	6
7.	*Tolypothrix*	3
8.	*Hapalosiphon*	3
9.	*Westiellopsis*	1
10.	*Fischerella*	0
11.	*Mastigocladus*	0
12.	*Nostochopsis*	1

Thus, our phylogenetic analyses using the *rbcl* gene as a molecular marker failed to delineate the orders Nostocales and Stigonematales from each other.

Evolutionary tendencies of heterocystous cyanobacteria

Our phylogenetic analyses using the *rbcl* gene sequences gave a glimpse of the immense genetic diversity and the heterogeneity present in the heterocystous cyanobacteria we studied. Evogenomics analyses were done in order to assess the evolutionary pace of the orders Nostocales and Stigonematales in a comparative way so as to gain some insights into what the

DNA sequences could indicate in an evolutionary perspective. In the initial level analyses, it was evident that the order Nostocales had a greater evolutionary pace as compared to the order Stigonematales. The nucleotide diversity calculated in the unbranded heterocystous cyanobacteria was greater (0.71559) as compared to the branched forms (0.67552), and hence was a direct clue of a greater evolutionary pace of heterocystous cyanobacteria. But, the average number of nucleotide differences was found to be greater in the branched forms, thus creating an ambiguity in the results calculated. On close inspection, the reason for this was found to be the presence of 27 singleton variable sites in the order Stigonematales, while no such sites were found in the order Nostocales. The singleton variable sites had perhaps contributed to the average number of nucleotide differences being greater in the true branched forms, and as a result, the polymorphic sites and the parsimony informative sites were also greater in the branched forms. Still, in order to validate our results we performed recombination analyses and found the value of *R* per gene to be much greater in the order Nostocales in comparison to the order Stigonematales. A greater value of *R* per gene is directly proportional to the evolutionary pace of an organism or a group of organisms, and hence the results in the recombination analyses fully corroborated with our level I analyses. The estimate of *R* between adjacent sites was also found to be much greater in the unbranched forms, thus indicating a very high frequency of recombinational tendencies in comparison to the branched forms. Very surprisingly, the minimum number of recombination events detected was found to be 16 in the order Stigonematales, which had a lower value of *R* per gene. This finding holds importance, as it was being contributed by two factors: greater number of singleton variable sites in Stigonematales and the higher evolutionary rate of the Nostocales. To validate these recombination results, we conducted an in-depth recombination analysis of individual genera in our study, and found that the intergeneric recombination events were very rare (only four each in the genera *Anabaena* and *Cylindrospermum*), although there were 45 gene conversion tracts identified, with 40 being found in the order Nostocales and only five being evident in Stigonematales. Thus, this points towards the rate of evolutionary diversification being much greater in the unbranched cyanobacteria, along with a clear-cut bias towards more intrageneric recombination events than intergeneric recombination. Thus, the high number of recombination events detected in the order Stigonematales was in fact an indication of a more favorable horizontal gene transfer between the two orders Nostocales and Stigonematales and a significantly less favorable tendency for vertical gene transfer.

We also tested results on the basis of the Darwinian principles of natural selection for individual genera and found that the genus *Nostoc*, which is one of the most abundant and genetically heterogeneous groups, also has the maximum number of gene conversion tracts (19). The false branching genera *Tolypothrix* and *Scytonema* accounted for nine gene conversion tracts, thus also identifying their importance in being crucial links in between the unbranched and branched forms. In spite of very clear-cut level I and level II analyses, we again conducted level III analyses to calculate the DNA divergence between the orders, and found that there were 153 mutations that were polymorphic in the order Nostocales but monomorphic in Stigonematales, while there were only three mutations that were polymorphic in the order Stigonematales but monomorphic in the order Nostocales. Hence, our evogenomics analyses gave precise indications towards the order Nostocales being the flag bearers of evolutionary diversifications in heterocystous cyanobacteria. We have found crucial evidence that demarcates the origin of the Stigonematales from the Nostocales, with ample chances of this evolution being separated by some transitional false branching forms. Yet, clarifications are needed for this assumption to take shape, taking into account a greater number of genera and using a polyphasic and multi-locus approach. At the present moment, our *rbcl* gene analysis does suggest the role of the order Nostocales in constituting the skeleton of the course of evolution of heterocystous cyanobacteria. Our recombination analyses clearly show that the order Nostocales could be an essential milestone controlling the future course of evolution of heterocystous cyanobacteria, hence constituting the main route from which the other advanced cyanobacteria either would have evolved in the past or may evolve in the future.

The conditions in the non-heterocystous cyanobacteria remain to be tested, but certainly we have proven mathematically that the lower forms have greater and faster evolutionary tendencies than the higher and advanced forms. The calculations of the gene conversion tracts also adhere to the findings that lateral gene transfer events are a reality in case of the *rbcl* gene (Gugger et al. 2002).

Thus, phylogenomic and evogenomic analyses of the heterocystous cyanobacteria using the *rbcl* gene as a molecular marker have indicated the need for classificatory reamendments in the cyanobacterial taxonomy, the intermixing amongst both the orders and the polyphyletic origin of the Stigonematales. The unbranched Nostocales have proved to be more capable of evolutionary diversifications, and thus have a greater evolutionary pace as compared to the order of branched Stigonematales.

Acknowledgments We are thankful to the Council of Scientific and Industrial Research, New Delhi for providing financial support. The Head of the Department of Botany, Banaras Hindu University, Varanasi, India is gratefully acknowledged for providing laboratory facilities.

References

Berrendero E, Perona E, Mateo P (2008) Genetic and morphological characterization of *Rivularia* and *Calothrix* (Nostocales, Cyanobacteria) from running water. Int J Syst Evol Microbiol 58: 447–460

Desikachary TV (1959) Cyanophyta, part I and II. Indian Council of Agricultural Research, New Delhi

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Gascuel O, Steel M (2006) Neighbor-joining revealed. Mol Biol Evol 23: 1997–2000

Giovannoni SJ, Turner S, Olsen GJ, Barns S, Lane DJ, Pace NR (1988) Evolutionary relationships among cyanobacteria and green chloroplasts. J Bacteriol 170:3584–3592

Gugger M, Hoffmann L (2004) Polyphyly of the true branching cyanobacteria (Stigonematales). Int J Syst Evol Microbiol 54:349–357

Gugger M, Lyra C, Henrikson P, Coute A, Humbert JF, Sivonen K (2002) Phylogenetic comparison of the cyanobacterial genera *Anabaena* and *Aphanizomenon*. Int J Syst Evol Microbiol 52:1867–1880

Halinen K, Jokela J, Fewer DP, Wahlsten M, Sivonen K (2007) Direct evidence for production of microcystins by *Anabaena* strains from the Baltic Sea. Appl Environ Microbiol 73:6543–6550

Hartmann LS, Barnum SR (2010) Inferring the evolutionary history of mo-dependent nitrogen fixation from phylogenetic studies of *nif*K and *nif*DK. J Mol Evol 71:70–85

Henson BJ, Hesselbrock SM, Watson LE, Barnum SR (2004) Molecular phylogeny of the heterocystous cyanobacteria (subsections IV and V) based on *nif*D. Int J Syst Evol Microbiol 54:493–497

Honda D, Yokota A, Sugiyama J (1999) Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. J Mol Evol 48:723–739

Iteman I, Rippka R, Tandeau de Marsac N, Herdman M (2002) rdna analyses of planktonic heterocystous cyanobacteria, including members of the genera *Anabaenopsis* and *Cyanospira*. Microbiology 148:481–496

Komárek J, Anagnostidis K (1989) Modern approach to the classification system of cyanophytes 4 Nostocales. Arch Hydrobiol Suppl 82: 247–345

Lammers P, Haselkorn R (1984) Sequence of the *nif*D gene coding for the α subunit from the cyanobacterium *Anabaena*. Cell 44:905–911

Lee DH, Zo YG, Kim SJ (1996) Non-radioactive method to study genetic profiles of bacterial communities by PCR-single strand conformation polymorphism. Appl Environ Microbiol 62:3112–3120

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA Polymorphism data. Bioinformatics 25:1451–1452

Ludwig W, Schleifer KH (1994) Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS Microbiol Rev 15:155–173

Meeks JC, Elhai J (2002) Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. Microbiol Mol Biol Rev 66:94–121

Mishra AK, Shukla E, Singh SS (2013) Phylogenetic comparison among the heterocystous cyanobacteria based on a polyphasic approach. Protoplasma 250:77–94

Mollenhauer D (1988) *Nostoc* species in the field. Arch Hydrobiol Suppl 80:315–326

Morden CW, Golden SS (1991) Sequence analysis and phylogenetic reconstruction of the genes encoding the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase from the chlorophyll b-containing prokaryote *Prochlorothrix hollandica*. J Mol Evol 32:379–395

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Neilan BA, Burns BP, Relman DA, Lowe DR (2002) Molecular identification of cyanobacteria associated with stromatolites from distinct geographical locations. Astrobiology 2:271–280

Nelissen B, De Baere R, Wilmotte A, De Wachter R (1996) Phylogenetic relationships of nonaxenic filamentous cyanobacterial strains based on 16S rRNA sequence analysis. J Mol Evol 42:194–200

Rajaniemi P, Hrouzek P, Kastovska K, Willame R, Rantala A, Hoffmann L, Komárek J, Sivonen K (2005) Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (Nostocales, Cyanobacteria). Int J Syst Evol Microbiol 55:11–26

Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol 111:1–61

Saitou N, Nei M (1987) The neighbour joining method—a new method for reconstructing phylogenetic tree. Mol Biol Evol 4:406–425

Sihvonen LM, Lyra C, Fewer DP, Rajaniemi-Wacklin P, Lehtimáki JM, Wahlsten M, Sivonen K (2007) Strains of the cyanobacterial genera *Calothrix* and *Rivularia* isolated from the Baltic Sea display cryptic diversity and are distantly related to *Gloeotrichia* and *Tolypothrix*. FEMS Microbiol Ecol 61:74–84

Singh P, Singh SS, Mishra AK, Elster J (2013) Molecular phylogeny, population genetics and evolution of heterocystous cyanobacteria using *nif*H gene sequences. Protoplasma 250: 751–764

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809–818

Zehr JP, Mellon MT, Hiorns WD (1997) Phylogeny of cyanobacterial *nif*H genes: evolutionary implications and potential applications to natural assemblages. Microbiology 143:1443–1450