MICROBIAL GENOMICS

RESEARCH ARTICLE
Pratte and Thiel, *Microbial Genomics* 2021;7:000595
DOI 10.1099/mgen.0.000595

MICROBIOLOGY SOCIETY
OPEN DATA OPEN ACCESS

Comparative genomic insights into culturable symbiotic cyanobacteria from the water fern *Azolla*

Brenda S. Pratte and Teresa Thiel*

Abstract

Species of the floating, freshwater fern *Azolla* form a well-characterized symbiotic association with the non-culturable cyanobacterium *Nostoc azollae*, which fixes nitrogen for the plant. However, several cyanobacterial strains have over the years been isolated and cultured from *Azolla* from all over the world. The genomes of 10 of these strains were sequenced and compared with each other, with other symbiotic cyanobacterial strains, and with similar strains that were not isolated from a symbiotic association. The 10 strains fell into three distinct groups: six strains were nearly identical to the non-symbiotic strain, *Nostoc* (*Anabaena*) *variabilis* ATCC 29413; three were similar to the symbiotic strain, *Nostoc punctiforme*, and one, *Nostoc* sp. 2RC, was most similar to non-symbiotic strains of *Nostoc linckia*. However, *Nostoc* sp. 2RC was unusual because it has three sets of nitrogenase genes; it has complete gene clusters for two distinct Mo-nitrogenases and an alternative V-nitrogenase. Genes for Mo-nitrogenase, sugar transport, chemotaxis and pili characterized all the symbiotic strains. Several of the strains infected the liverwort *Blasia*, including *N. variabilis* ATCC 29413, which did not originate from *Azolla* but rather from a sewage pond. However, only *Nostoc* sp. 2RC, which produced highly motile hormogonia, was capable of high-frequency infection of *Blasia*. Thus, some of these strains, which grow readily in the laboratory, may be useful in establishing novel symbiotic associations with other plants.

DATA SUMMARY

Genome assemblies of the newly sequenced strains are available at NCBI under the following GenBank assembly accession numbers: GCA_014222145.1 (*Trichormus variabilis* 9RC), GCA_014222135.1 (*Trichormus variabilis* ARAD), GCA_014222245.1 (*Trichormus variabilis* FSR), GCA_014222155.1 (*Trichormus variabilis* N2B), GCA_014222225.1 (*Trichormus variabilis* PNB), GCA_014222125.1 (*Trichormus variabilis* V5), GCA_014222165.1 (*Nostoc* sp. 2RC), GCA_014222255.1 (*Nostoc* sp. UCD120), GCA_014222285.1 (*Nostoc* sp. UCD121) and GCA_014222275.1 (*Nostoc* sp. UCD122). The authors confirm that all comparative data, supporting data, code and protocols have been provided within the article or through supplementary data files.

INTRODUCTION

Species of the floating, freshwater fern *Azolla* are found throughout the world in Asia, Africa, and North and South America in temperate, tropical and subtropical habitats. Through a well-characterized symbiotic association with the cyanobacterium *Nostoc azollae*, which fixes nitrogen in the leaves, the *Azolla* plant can serve as a highly effective green manure for growing rice in Vietnam and China [1]. The cyanobiont, which is found exclusively in the periphery of an extracellular cavity on the dorsal side of the fern leaves, is transmitted during sexual reproduction vertically to the next generation via cyanobacterial filaments that are associated with the megasporocarp of the fern, without *de novo* infection [2–4]. Within the cavity, the cyanobiont differentiates a high percentage of heterocysts, which fix nitrogen using photosystem I and glycolysis for energy while relying on plant photosynthesis for fixed carbon.

A free-living, heterocyst-forming cyanobacterium, originally called *Anabaena azollae*, was isolated from *Azolla* and grown in culture as early as 1979 [5], but by 1986 and 1987 there was evidence, based on RFLPs, that the strain freshly harvested from *Azolla* was distinct from the free-living strains [6, 7]. Genetic differences in *nif* gene restriction sites were confirmed and it was also demonstrated that the major cyanobiont

Received 14 October 2020; Accepted 27 April 2021; Published 28 June 2021
Author affiliations: [1]Department of Biology, University of Missouri–St. Louis, One University Blvd, St. Louis, MO 63121, USA.
*Correspondence: Teresa Thiel, thiel@umsl.edu
Keywords: *Azolla*; cyanobacteria; heterocyst; nitrogen fixationn; symbiosis.
Abbreviations: ANI, average nucleotide identity; NCBI, National Center for Biotechnology Information; PCR, polymerase chain reaction.
Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Six supplementary figures and four supplementary tables are available with the online version of this article.
000595 © 2021 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License.

lacked the 11 kb *nifD* excision element that was present in the cultured strain [8]. Characterization of the freshly isolated cyanobiont cells, including sequencing of 16S rRNA genes, demonstrated that there were typically only minor differences in strains from different plants; however, other cyanobacteria have also been detected [9–13]. More recently, nearly 60 bacterial isolates, comprising nine genera outside the cyanobacteria, have also been cultured from *Azolla* [14].

The sequencing of the genome of a true, non-culturable cyanobiont, *Nostoc azollae* 0708, confirmed that it is different from the culturable forms, with reductive evolutionary degradation of the genome, such that the strain cannot grow outside of the plant [15]. Comparison of the phylogenies of six species of *Azolla* with those of their cyanobionts shows clear evidence of a pattern of co-speciation between the plant and its cyanobacterial partner [16, 17]. While *Nostoc azollae* 0708 is non-culturable [15, 18], the role of the culturable strains found in the symbiotic association remains unknown.

Culturable cyanobacteria form various types of symbiotic associations with plants, including intracellular associations (e.g. with *Gunnera*) and endophytic associations, within specialized cavities (e.g. *Azolla*, *Blasia* and *Anthoceros*) [19–23]. A culturable strain from *Azolla*, first named *Anabaena azolla,* was isolated in 1979 from *Azolla caroliniana*. The strain was difficult to isolate and propagate from *Azolla* and required disruption of the plant tissue to free the cyanobacteria from the fern leaf cavities, suggesting that it was not a surface contaminant [5]. The cultured strain was very similar to *Anabaena flos-aquae* ATCC 22664, a strain that showed cross-reactivity to antibodies made against Newton's culturable strain of *Anabaena azollae*. [5]. Zimmerman analysed 10 strains cultured from *Azolla* from multiple laboratories using morphology, enzymes and lectins to determine their similarity. Five of the strains appeared to be nearly indistinguishable from the free-living strain, *Nostoc* (*Anabaena*) *variabilis* ATCC 29413, while the other five were all quite divergent from the *N. variabilis* ATCC 29413-like strains [24]. Meeks isolated strains N1 (UCD120), A1 (UCD121) and A2 (UCD122) from an extract of cyanobacterial cells from *Azolla caroliniana* that was used to infect the hornwort, *Anthoceros*. The cyanobacteria that subsequently grew symbiotically in *Anthoceros* were excised from the plant tissue, and cultured on cyanobacterial medium [8].

Worldwide, several cyanobacterial strains have been isolated from *Azolla*, cultured in the laboratory and stored. We obtained and grew 10 strains from the Zimmerman and Meeks collections for comparisons of their genomes, to determine the similarities and differences among these isolates [8, 24]. We were particularly interested in the reported similarity of several of these strains, isolated from *Azolla*, to *N. variabilis* ATCC 29143 [24], a strain that was first isolated as *Anabaena flos-aquae* A-37 [25] from a sewage oxidation pond in Mississippi, with no known relationship to *Azolla*. Its name was later changed to *Anabaena variabilis* [26]; however, based on data presented here, we have called it *Nostoc variabilis* ATCC 29413. The strain was characterized by several laboratories but the early work by Wolk's laboratory on this strain led to its becoming a model strain for cyanobacterial physiology, nitrogen fixation and heterocyst formation [27–31].

Impact Statement

A variety of plants form beneficial symbiotic associations with cyanobacteria, allowing the plants to thrive in soils that lack key nutrients, such as fixed nitrogen. Genome sequences were determined for a group of symbiotic cyanobacteria that were isolated from the water fern *Azolla*, a plant that can provide nitrogenous fertilizer to rice plants, particularly in developing countries. Genes and pathways that confer key characteristics important in symbiosis include those for nitrogen fixation, motility and the ability to use sugars for growth. Some of these strains, which grow readily in the laboratory, can infect plants other than *Azolla* and, thus, may be useful for extending these ecologically important symbiotic associations to agriculturally important plants.

In this study we were primarily interested in answering three questions: (1) How similar are the 10 strains, isolated from *Azolla* in different locations worldwide, to each other and to well-characterized model strains? (2) Could they infect a plant? (3) What genes do they share that might shed light on the characteristics that define symbiotically competent cyanobacteria? We compared the genomes of the 10 strains to model strains *N. variabilis* ATCC 29413 and *Nostoc punctiforme* ATCC 29133, and to other cyanobacteria isolated from symbiotic associations with moss and lichens [21, 32]. *N. punctiforme*, isolated from an association with the cycad Macrozamia [33], forms associations with several plants and is a member of a clade that includes other cyanobacteria found in plant associations [34–36]. The phylogenetic relationships among the strains and among some of the genes thought to be associated with symbiosis are presented.

Another question was whether these strains, after years of storage and growth as axenic laboratory cultures, could infect a plant. Since there is no report of successful infection of *Azolla* with cyanobacteria, we determined whether the cyanobacteria could infect a more tractable model plant, the liverwort *Blasia pusilla* [20, 23, 37–39]. The endophytic infection of *Blasia* by motile hormogonia occurs via pores in the extracellular dome-shaped structures known as auricles on the surface of the *Blasia* plant thallus. After infection, cyanobacterial filaments with a very high frequency of heterocysts grow within the cavities, the pores close and the auricles produce mucilage as well as infiltrating plant structures thought to facilitate nutrient transfer between the cyanobacteria and the plant [20, 40]. Through nitrogen fixation in heterocysts, cyanobacteria provide fixed nitrogen to the plant [41–43] in exchange for sugar from the plant [39, 43, 44]. For the newly sequenced strains and the model strains *N. variabilis* ATCC 29413 and *N. punctiforme* ATCC 29133, we compared

genes associated with symbiosis, including those for nitrogen fixation, sugar transport, motile filaments called hormogonia used for infection of plants [45], and chemotaxis. The genes of interest are described in more detail in the Results.

METHODS

Genome sequencing and assembly

Genomic DNA was extracted from *N. variabilis* variants ARAD, 9RC, FSR, PNB, N2B and V5, *Nostoc* UCD strains (UCD120, UCD121 and UCD122), and *Nostoc* 2RC, grown in an eight-fold dilution of Allen and Arnon (AA/8) medium [46] containing 5 mM NH_4Cl and 10 mM *N*-[Tris(hydroxymethyl) methyl]-2-aminoethanesulfonic acid (TES), by vortexing cells with glass beads in the presence of phenol [47, 48]. Genomic DNA was treated with RNaseA, further purified with two phenol/chloroform/isoamyl alcohol extractions followed by a chloroform/isoamyl alcohol extraction before ethanol precipitation. DNA was additionally purified and concentrated using a Bio101 Gene Clean II kit. The concentration and purity of the genomic DNA was determined using a NanoDrop (Thermo Scientific) and 0.5 ng of genomic DNA from each cyanobacterial strain was fragmented and tagged with adapters using the protocol provided by the Nextera XT DNA Library Prep Kit (Illumina). Tagged DNA was amplified and index sequences were added using low-cycle Nextera Seq PCR (index Primers i5 and i7, annealing temperature at 55 °C, and 12 cycles). PCR products were purified using AMPure XP (Beckman Coulter) beads and normalized using a bead-based method provided in the Nextera XT DNA Library Prep Kit (Illumina). The library normalization process dilutes the genomic libraries to the same concentration before pooling, thus allowing all libraries to have consistent read depth. Normalized genomic libraries were sequenced on an Illumina Miseq using an Illumina MiSeq Reagent Kit v2 with 2×150 cycles. The Illumina Miseq paired-end reads were assembled within Pathosystems Resource Integration Center (PATRIC, www.patricbrc.org) [49] using Unicycler [50] and annotated using the RAST tool kit, RASTtk [51]. Information on the genomic sequences is provided in Table S1.

Phylogenetic trees and ANI calculations

The genomic phylogenetic tree was created in PATRIC using the Codon Trees pipeline, using a concatenated alignment of 500 randomly chosen amino acid and nucleotide sequences from PATRIC's global Protein Families (PGFams) [52]. A default setting of 0 was used for both Max Allowed Deletions and Duplications. Alignments were created using MUSCLE [53] for proteins and the Codon align function of BioPython [54] for nucleotides. Genomic maximum-likelihood trees were produced from the concatenated alignments by the program RAxML (Randomized Axelerated Maximum Likelihood) [55] with 100 rounds of rapid bootstrapping. BioNJ distance trees for *nifH*, *vnfR* and *frtA* phylogeny trees were constructed using the program SeaView [56] with 100-replicate bootstrap values. The genes used for these trees are provided in Table S2. The average nucleotide identity (ANI) values for two cyanobacterial genomes were determined using ChunLab's online ANI calculator (www.ezbiocloud.net/tools/ani) [57]. This calculator uses the OrthoANIu algorithm, which uses USEARCH. OrthoANIu breaks up two genomes into 1020 bp random fragments and identifies pairs of fragments with reciprocal best hits. It then calculates the ANI values for all reciprocal best hits using USEARCH, which serves a similar role as BLASTN but uses an algorithm that trades sensitivity for speed.

Genome and gene comparisons

The proteome comparisons, based on the deduced proteins for each genome, between a reference strain and closely related cyanobacterial strains (as determined by phylogeny, Fig. 1) were done using the Proteome Comparison tool in PATRIC [49], which uses BLASTP to calculate protein identity. The Sequence-Based Comparison tool marks each gene by colour as either unique, a unidirectional best hit or a bidirectional best hit compared to a reference genome, and the colour indicates the per cent identity between deduced proteins. The BLASTP results for all the genomes are displayed as a colour-based circular map, making it easier to identify different or deleted regions. The closest relative of *Nostoc* 2RC was determined using Similar Genome Finder in PATRIC [49], using Mash/MinHash [58]. Mash reduces a genome to smaller subsets of sequences to more rapidly determine mutation distances, while MinHash uses the mutational distances and *P* values to rapidly compare a genome to a massive collection of sequences.

Gene comparisons were made using the BLASTN or BLASTP algorithms [59] to determine nucleotide or amino acid identity, respectively. We used BLASTP to determine whether the 74 genes identified as 'symbiosis' genes in Supplement S2 of Warshan *et al.* [34] were present in the genomes analysed here. Because the percentage identity for some of the genes that had been compared among the strains shown in supplement S2 of Warshan *et al.* [34] was only around 50% (e.g. *Nostoc* Moss5 and Moss6 vs. *N. punctiforme*), we used a minimum cutoff score of at least 49% amino acid identity and an E value of at least −100 to identify homologues matching the 74 genes identified in supplement S2 of Warshan *et al.* [34]. These criteria for a match gave results identical to those presented in supplement S2 for the strains listed there. Accession numbers for the 'symbiosis' genes of representative strains studied here are provided in Table S3.

Symbiosis

In the *Blasia* symbiosis, cyanobacteria infect and multiply in auricles, which are dome-shaped structures on the ventral surface of the thallus. *Blasia* plants were maintained on agar-solidified BCD medium (1 mM $MgSO_4$, 1.84 mM KH_2PO_4, 10 mM KNO_3, 1 mM $CaCl_2$, 45 µM $FeSO_4 \cdot 7H_2O$, trace minerals, 0.7% agar) [60]. All cyanobacteria strains used to inoculate *Blasia* were grown in $Bgll_0$ liquid media [61]. *Blasia* was transferred to BCD medium lacking nitrate, a few drops of liquid cyanobacterial culture were placed atop the *Blasia*, and plates were incubated at 25 °C under 60 µE m^{-2} s^{-1} light for several weeks. *Blasia* was checked weekly for infection

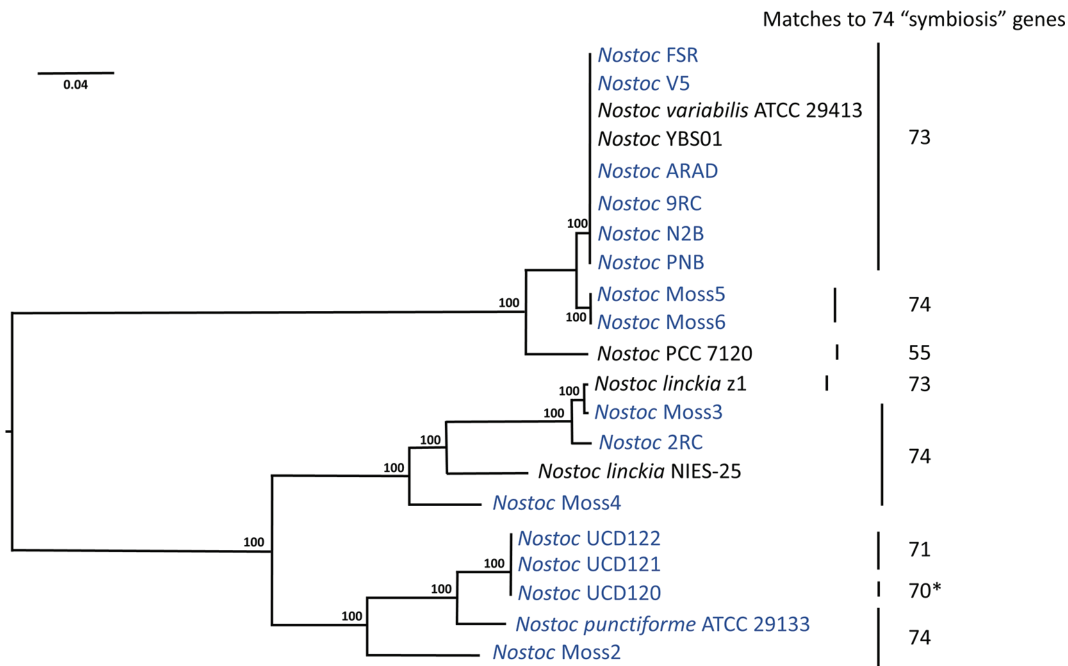


Fig. 1. Phylogenetic distance tree of genomes for the symbiotic cyanobacterial strains isolated from plants (in blue) and similar comparator strains that were not isolated from a plant (in black). Numbers at the branch points indicate bootstrap values. The value shown to the right of each strain indicates the number of genes that match a set of 74 genes identified as 'symbiosis genes' [34]. The criteria for a gene match are described in the Methods. Accession numbers for the 'symbiosis' genes of representative strains are provided in Table S3. *UCD120 has a partial copy of the gene that it is missing compared to UCD121 and UCD122. Accession numbers for these strains are provided in Table S1. Scale represents number of substitutions per site.

using a light microscope to screen several thalli (containing 20–30 auricles) to identify infected auricles. For most of the strains we found no infected auricles after screening several thalli (20–30 auricles). For poorly infectious strains, we typically found one or two infected auricles (out of 20–30 auricles). For highly infectious strains, typically about 75% of the auricles were infected. As can be seen in the light micrographs (Fig. 9), when infection was successful, *Blasia* auricles were tightly packed with cyanobacterial filaments. In contrast, when infection was not successful, cyanobacteria were loosely associated with the plant cells, but the auricles were empty.

RESULTS

Overall genome comparisons

The newly sequenced strains that were isolated from *Azolla* (the first 10 strains shown in Table 1) fall into three distinct clusters. A phylogenetic analysis of strains 9RC, ARAD, FSR, N2B, PNB and V5, isolated from *Azolla* indicated that they are very closely related to the model strain *N. variabilis* ATCC 29413 (Fig. 1). The ANI values for these strains compared to *N. variabilis* were greater than 99.9% (Table S4). The ANI value for the non-symbiotic strain isolated in India, *Nostoc* YBS01, is nearly identical to the others in this group. Phylogenetically, this group is also very closely related to the previously described symbiotic strains *Nostoc* Moss5 and Moss6 [21] (Fig. 1) with ANI values of 98% relative to members of the *N. variabilis* group (Table S4). A comparison of the circular proteome maps (based on the deduced proteins for each genome), shown as concentric rings with colour-coded amino acid identity values (Fig. 2a), confirms that *N. variabilis* ATCC 29413 is more similar to the strains ARAD, PNB, FSR, N2B, V5 and 9RC than to *Nostoc* Moss5 and Moss6 (which are virtually identical to each other). In contrast, *Nostoc* Moss5 and Moss6 show lower identity values throughout the genome, with several regions of much less than 90% identity. All of these strains, except *N. variabilis* and *Nostoc* YBS01, were isolated from a symbiotic association either with *Azolla* (ARAD, PNB, FSR, N2B, V5 and 9RC) or with feathermoss (Moss5 and Moss6) [21], and have geographically diverse origins (Table 1). The well-characterized non-symbiotic strain *Nostoc* sp. PCC 7120 has an ANI value of 92% compared to the *N. variabilis*-like strains.

Nostoc UCD120, UCD121 and UCD122, isolated from *Azolla*, are very closely related to the model strain *N. punctiforme* and *Nostoc* Moss2, their closest known relatives (Fig. 1). The three *Nostoc* UCD strains share ANI values greater than 99% compared to each other and 93% compared to *N. punctiforme* (Table 1). A comparison of the proteomes (Fig. 2b) shows that the three *Nostoc* UCD strains are more similar to each other than they are to *N. punctiforme* and somewhat less similar to the moss symbiont, *Nostoc* Moss2. The phylogenetic tree based on the genomes confirms these relationships (Fig. 1).

Table 1. Origins of cultured *Anabaena/Nostoc* strains

Strain designation	Plant source	Source of strain	Location	Citation
Strains grown and characterized in this study				
2RC	*Azolla pinnata*	R. Caudales	Rutgers University, USA	[24]
9RC	*Azolla caroliniana*	R. Caudales	Rutgers University, USA	[24]
A1 (UCD121)	*A. caroliniana*	J. Meeks	UC Davis, CA, USA	[8, 24]
A2 (UCD122)	*A. caroliniana*	J. Meeks	UC Davis, CA, USA	[8]
ARAD	*Azolla filiculoides*	E. Tel-Or	Hebrew Univ., Israel	[24, 118]
FSR	*A. pinnata*	R. Fisher	VA Commonwealth Univ., USA	[24]
N1 (UCD120)	*A. caroliniana*	J. Meeks	UC Davis, CA, USA	[8]
N2B	*A. caroliniana*	J. Newton	USDA, Peoria, IL, USA	[5, 24]
PNB	*A. pinnata*	S.Nierzwicki-Bauer	Rensselaer Poly. Inst., USA	[24]
V5	*A. pinnata*	I. N. Gogotov	IMPB RAS, Pushchino, Russia	[63]
Nostoc (*Anabaena*) *variabilis* ATCC 29413	None	C. P. Wolk	Michigan State Univ., USA	[62]
Nostoc (*Anabaena*) ATCC 29413 strain FD	None	C. P. Wolk	Michigan State Univ., USA	[62, 119]
Comparator strains for this study				
Nostoc punctiforme ATCC 29133	Gymnosperm cycad *Macrozamia sp.*		Australia	[120]
*Nostoc Moss*2	Moss		Sweden	[21]
*Nostoc Moss*3	Moss		Sweden	[21]
*Nostoc Moss*4	Moss		Sweden	[21]
*Nostoc Moss*5	Moss		Sweden	[21]
*Nostoc Moss*6	Moss		Sweden	[21]
Nostoc linckia z1	None		Israel	[121, 122]
Nostoc linckia NIES-25 (IAM M-251)	None		Japan	[123]
Nostoc (*Anabaena*) YBS01	None		Meghalaya, India	Accession number CP034058

In contrast, strains related to *Nostoc* 2RC form a third cluster, distinct from *N. punctiforme*. *Nostoc* 2RC is most closely related phylogenetically to the symbiotic strain, *Nostoc* Moss3, and *N. linckia* z1, with ANI values of 96.6 and 96.8%, respectively. It is less closely related to *N. punctiforme*, *Nostoc* UCD120, UCD121 and UCD122 with ANI values of about 82% (Table S4). The proteome map (Fig. 2c) indicates that *N. linckia* z1, which did not come from a plant (Table 1), is more similar to symbiotic strains *Nostoc* 2RC and *Nostoc* Moss3 than to the non-symbiotic strain *N. linckia* NIES-25 or the symbiotic strain *Nostoc* Moss4. *N. linckia* z1 has ANI values of 98% for Moss3 and 89% for *N. linckia* NIES-25 and *Nostoc* Moss4 strains (Table S4). Among the three clusters of strains, only the branch containing *N. punctiforme*, *Nostoc* Moss2 and the three *Nostoc* UCD strains comprises known symbiotic strains, while the other two branches include both known symbiotic and non-symbiotic strains. Notable are the nearly identical strains that comprise the *N. variabilis*-like group, of which only two strains were isolated as free-living cyanobacteria (Table 1). Similarly, *Nostoc* Moss3, which came from a plant, shares 98% ANI (Table S4) with *N. linckia* z1, which did not.

N. variabilis ATCC 29413 has four circular plasmids, A, B, C and D, as well as E, a 37 kb linear element. When the genome of *N. variabilis* was originally sequenced by JGI, a laboratory variant strain of *N. variabilis* called FD [62] was inadvertently the source of the DNA. When *N. variabilis* was sequenced in another laboratory [63] they found that *N. variabilis* had plasmid D, which was missing in the FD strain that was

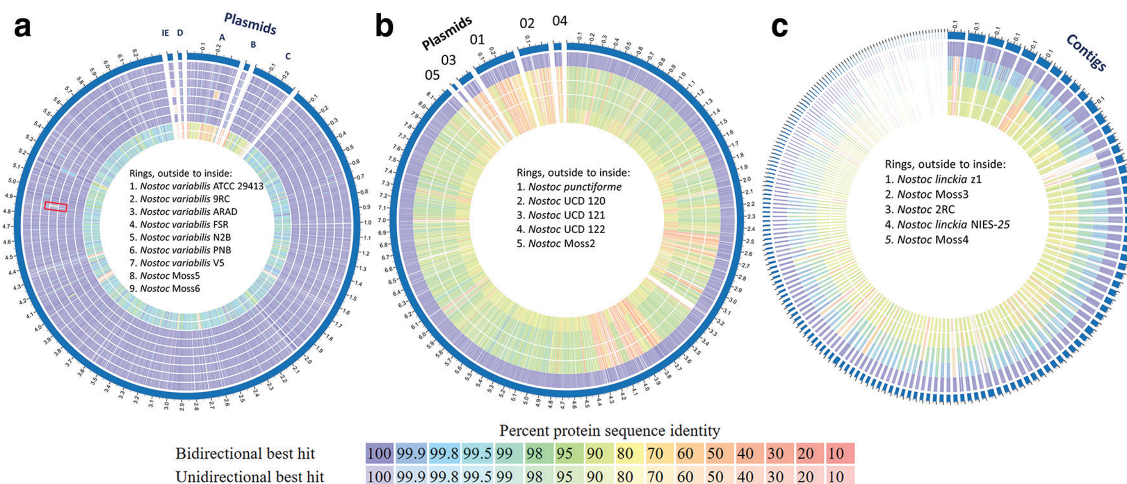


Fig. 2. Proteome amino acid identity maps based on the deduced proteins for each genome for the symbiotic cyanobacterial strains and similar comparator strains that were not isolated from a plant. (a) Strains similar to *N. variabilis* ATCC 29413. (b) Strains similar to *N. punctiforme*. (c) Strains similar to *N. linckia* z1. The red box in (a) indicates the absence of an 11 kb excision element in strain FSR.

sequenced by JGI. We have confirmed the lack of plasmid D in strain FD. Most of the strains that are nearly identical to *N. variabilis* have genes homologous to those in plasmids A, B, C and D and the linear element E; however, FSR lacks some genes homologous to those in part of plasmid A and both FSR and PNB lack genes homologous to those in the linear element (Fig. 2a). The free-living strain *Nostoc* YBS01, like FSR and PNB, also lacks genes homologous to those in the linear element. In contrast, *Nostoc* Moss5 and Moss6 only have genes homologous to those in plasmids A and C, and even those show regions of low similarity compared to *N. variabilis* ATCC 29413 (Fig. 2a). *Nostoc* UCD120, UCD121, UCD122 and *Nostoc* Moss2 have genes with some similarity to genes in the five plasmids present in *N. punctiforme*; however, the gene similarities are weak, indicating that these four strains probably lack the plasmids found in *N. punctiforme* (Fig. 2b).

Analyses of the genome sequences for several *Nostoc* strains isolated from lichens and feathermoss may provide new information on genes that are specifically associated with symbiosis [21, 34, 35]. These authors took a bioinformatics approach to identify genes and gene families that are associated specifically with strains isolated from lichens and feathermoss. In the case of feathermoss, 74 deduced proteins were identified in all the cyanobacterial plant isolates but not in a related non-symbiotic strain [34]. We determined how many of these 'symbiotic' genes were present in the genomes of the 10 strains that we sequenced here. Only *Nostoc* sp. 2RC had matches for all 74; however, the *N. variabilis*-like strains isolated from *Azolla* had 73 (Fig. 1). The *Nostoc* UCD strains had 71 genes (the missing gene in UCD120 is probably a sequencing artefact; the strain has a partial, identical 34 aa sequence on a small contig). In the data provided for the feathermoss symbionts, only one non-symbiotic strain, *Nostoc* sp. CALU996, was provided as a negative control [34]. Consistent with the lack of matches for *Nostoc* sp. CALU996, we found no matches for one of its closest relatives, *Nodularia spumigena* CCY9414; however, another strain of *Nodularia spumigena* CENA596 had 18/74 genes. Surprisingly, another close relative of *Nostoc* sp. CALU996, *Nostoc* sp. 7107, had matches for 51/74 genes. We found that several other strains that have no known symbiotic association also had many of these genes; the *Nostoc linckia* strains (73–74), *Anabaena cylindrica* PCC 7122 (46/74), *Nostoc* sp. PCC 7524 (45/74) and *Nostoc* sp. PCC 7120 (55/74). The relatively large number of these 'symbiosis' genes in *Nostoc* sp. PCC 7120 may reflect its similarity to *N. variabilis* ATCC 29413, a strain that we show here is symbiotic. As more symbiotic strains are sequenced, gene profiling techniques will probably be refined and improved, providing new information on the genes and metabolic pathways that are needed for infection and symbiosis.

nif genes and excision elements

Most cyanobacterial nitrogenase gene clusters are highly conserved both in gene sequence and in the organization of the *nif* genes in a single large cluster under the control of the primary *nifB* promoter [64–66]. One difference is *glbN*, cyanoglobin, a gene of unknown function located just upstream of *nifH* in some strains of *Nostoc*, including *N. punctiforme* [67, 68]. The *glbN* gene was present in *Nostoc* UCD120, UCD121 and UCD122, and in *Nostoc* 2RC, but was absent in all of the strains that are virtually identical to *N. variabilis* ATCC 29413.

There is considerable diversity among nitrogen-fixing strains in the excision element that interrupts the *nifD* gene in many cyanobacterial *nif* gene clusters (Table 2) [69]. This *nifD* element, which is excised during heterocyst development, leaving an intact *nifD* gene, was first identified in *Nostoc* PCC 7120 [47, 70–72]. The *nifD* excision element of *N. variabilis* ATCC 29413 is very similar to that in *Nostoc* PCC 7120 in the

Table 2. Excision elements in cyanobacterial strains

	nifD element	*fdxN* element	*hupL* element
Nostoc variabilis ATCC 29413	11074 bp	No	No
Nostoc PCC 7120	11289 bp	59428 bp	9435 bp
Nostoc punctiforme	23723 bp	No	No
Nostoc 2RC	58153 bp	No	5872 bp
Nostoc YBS01	11074 bp	No	No
Nostoc Moss5	11879 bp	38259 bp	No
Nostoc Moss6	11879 bp	38259 bp	No
Nostoc 9RC	11074 bp	No	No
Nostoc N2B	11074 bp	No	No
Nostoc ARAD	11074 bp	No	No
Nostoc FSR	No	No	No
Nostoc V5	11074 bp	No	No
Nostoc PNB	11074 bp	No	No
Nostoc UCD120	Yes*	No	No
Nostoc UCD12	24028 bp	No	No
Nostoc UCD122	Yes*	No	No
Nostoc linckia NIES-25	31130 bp	No	13051 bp
Nostoc linckia z1	32500 bp	No	No
Nostoc Moss2	25314 bp	No	No
Nostoc Moss3	32989 bp	No	No
Nostoc Moss4	20012 bp	No	No

*Located on multiple contigs.

middle region of the excision element but, except for *xisA*, shows no similarity at the end regions of the element (Fig. 3). Among the cyanobacterial strains isolated from *Azolla*, strains V5, ARAD, N2B, 9RC and PNB have *nifD* excision elements identical to the *nifD* element in *N. variabilis*. In contrast, the genome of strain FSR, which is otherwise virtually identical in sequence to *N. variabilis*, lacks the *nifD* excision element completely, which is also shown in the proteome comparison (red box in Fig. 2a). The free-living strain *Anabaena* YBS01 also has an excision element virtually identical to those in the *N. variabilis* ATCC 29413 group (data not shown). The *nifD* excision elements in the symbiotic strains, *Nostoc* Moss5 and Moss6, are similar to each other with respect to gene composition but differ from *N. variabilis* at the end closest to *nifH*. The *nifD* excision element in other *Nostoc* strains was more variable in size and composition than in the *N. variabilis*-like group (Fig. 3). The *nifD* excision element in *Nostoc* UCD121 was most like the element in *N. punctiforme*, with about half of the genes in the *nifD* element shared between them (Fig. 3). All the genes present in the *nifD* excision element of *Nostoc* UCD121 were also present in *Nostoc* UCD120 and UCD122 (data not shown); however, they were not all on a single contig. The similarity of the *nifD* elements in the *Nostoc* strains characterized here correlates with the phylogenetic relationship of the strains. Consistent with the phylogenetic tree (Fig. 1), the *nifD* element of *N. punctiforme* is most similar to that of *Nostoc* Moss2, while the *nifD* elements of the two *Nostoc linckia* strains and *Nostoc* Moss3 are most similar to each other (data not shown). *Nostoc* 2RC is unusual because it has a very large 58 kb *nifD* excision element; thus, it has many genes that are not similar to those in any of the other *nifD* elements. Its apparent difference compared to the other *nifD* elements is due primarily to the presence of an 18.4 kb insertion that spans the middle third of the *nifD* element in *Nostoc* 2RC (Fig. 3). Outside of this insertion, which includes an integrase gene, *Nostoc* 2RC and *N. linckia* z1 share about 75% of the genes in the *nifD* element. However, *N. linckia* z1, and *Nostoc* Moss3 and Moss4 have genomic regions outside the *nif* gene clusters with multiple genes that are similar in gene composition and organization to those in the 18.4 kb insertion in *Nostoc* 2RC, including the integrase gene. This implies that this 18.4 kb insertion element is mobile. Among all the *nifD* excision elements analysed here, the only genes that are shared by all the strains are the excisase gene, *xisA*,

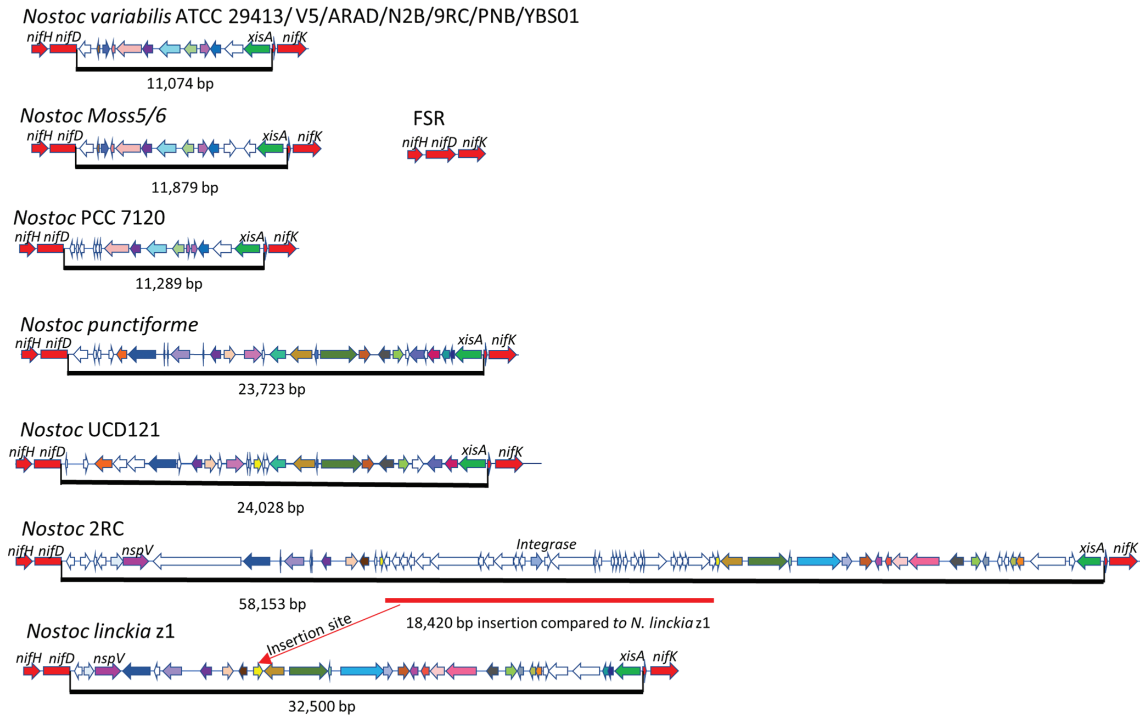


Fig. 3. Maps of excision elements interrupting the *nifD* gene (in red) in the symbiotic cyanobacterial strains isolated from plants and similar comparator strains that were not isolated from a plant. The excision element in *Nostoc* 2RC has an 18.4 kb insertion in a gene that is a homologue of a gene in *N. linckia*, indicated by the red arrow.

and a gene encoding a small hypothetical protein (Ava_3922, Npun_F0406) shown in purple in Fig. 3. The lack of genes of known function (except for the excisase) in the various excision elements and the absence of the element in FSR indicates that it does not have an essential function in nitrogen fixation because all the strains are able to grow in a medium lacking fixed nitrogen.

Among the cyanobacterial strains characterized here, none have the large excision element found in *Nostoc* PCC 7120 in the *nif* cluster gene, *fdxN* [73]; however, *Nostoc* Moss5 and Moss6 have a smaller element (Table 2) with a similar excisase gene in the same location. *Nostoc* PCC 7120 also has an excision element that interrupts the heterocyst-specific uptake hydrogenase, large-subunit gene, *hupL* [74]. Like the other excision elements, the *hupL* element is excised during heterocyst development by an excisase, XisC. This *hupL* element is absent in all the strains shown in Fig. 1 except *Nostoc* 2RC and *N. linckia* NIES-25 (Table 2), where it is found at the same location as in *Nostoc* PCC 7120 (Fig. 4). The *hupL* excision elements all have a similar excisase gene, *xisC*, and also

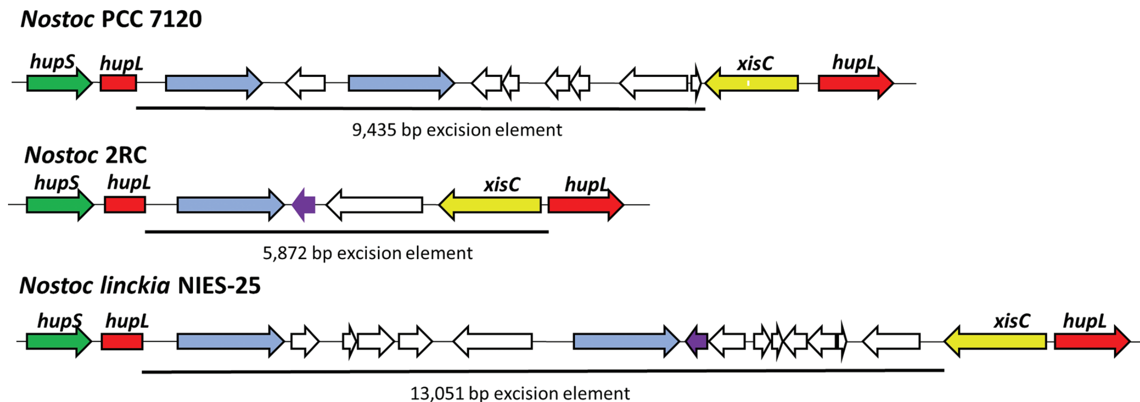


Fig. 4. Maps of excision elements interrupting the *hupL* gene (in red) in the symbiotic cyanobacterial strain *Nostoc* 2RC, isolated from *Azolla*, and comparator strains that were not isolated from a plant.

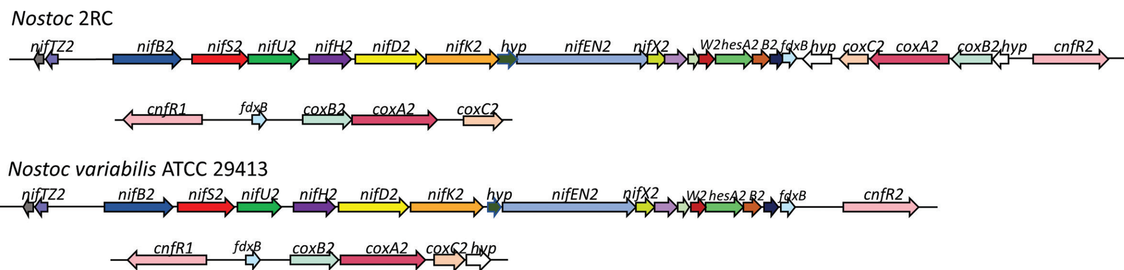


Fig. 5. Maps of *nif2-cnfR2* and the *cnfR1-cox2* gene regions in the symbiotic cyanobacterial strain *Nostoc* 2RC and the comparator strain not isolated from a plant, *N. variabilis*.

share one or two genes for hypothetical proteins. It appears that while excision elements have no vital function, they are potentially useful taxonomic markers.

In addition to the *nif1* gene cluster that makes the heterocyst-specific nitrogenase, *N. variabilis* has a large cluster of genes that encode a second Mo-nitrogenase that functions in vegetative cells under anoxic conditions [66, 75, 76]. All the *N. variabilis*-like strains have *nif2* genes (Fig. 5) with over 99% amino acid identity to the *nif2* gene cluster in *N. variabilis* ATCC 29413. *Nostoc* 2RC is the only strain not in the *N. variabilis* group that has the *nif2* cluster, which is identical in gene structure and organization to the *nif2* cluster in *N. variabilis* ATCC 29413 (Fig. 5); however, its *nif2* genes show only about 90% amino acid identity to the *N. variabilis* homologues. Downstream from the *nif2* cluster in *N. variabilis* is the gene for the *nifB2* transcriptional activator *cnfR2* [66]; however, between the *nif2* genes and *cnfR2* in *Nostoc* 2RC, there are *coxBAC2*-type genes, encoding a cytochrome oxidase that is important for nitrogen fixation [77]. These *coxBAC2* genes are absent in the *nif2* region in strains in the *N. variabilis* group. In *Nostoc* 2RC, the *coxBAC2* genes near *the nif2* cluster (Fig. 5) share 60–85% amino acid identity with the other *coxBAC2* genes near *cnfR1*. The *coxBAC2* genes close to *cnfR1* in *Nostoc* 2RC share 77–87% amino acid identity with the single set of *coxBAC2* genes located near *cnfR1* in *N. variabilis* while the *coxBAC2* genes close to *nif2* in *Nostoc* 2RC share 63–85% amino acid identity with the single set of *coxBAC2* genes in *N. variabilis* ATCC 29413. While it seems likely that in *Nostoc* 2RC the *coxBAC2* genes near the *nif2* genes function to support Nif2, another cytochrome oxidase must function for Nif2 in *N. variabilis*, which has five sets of *cox* genes, but only one *cox2*-like set.

In addition to the second Mo-nitrogenase, *N. variabilis* has an alternative V-nitrogenase that is made only in the absence of Mo and functions in heterocysts [78, 79]. While the V-nitrogenase is not common in cyanobacteria, the genes for the V-nitrogenase have been found in cyanobacteria in symbiotic association with lichens in boreal and arctic ecosystems where they are expressed in lichens growing in soils deficient in Mo [80, 81]. All the *N. variabilis*-like strains, as well as *Nostoc* 2RC, have *vnf* genes. Except for *Nostoc* 2RC, the *vnf* genes described here are at least 97% identical (Table 3). In contrast, the *Nostoc* 2RC *vnf* genes show only about 75–80% nucleotide identity to the *vnf* genes of *N. variabilis*, consistent with the fact that *Nostoc* 2RC is not closely related to *N. variabilis*. The two *N. linckia* strains, which are most similar to *Nostoc* 2RC as well as to *Nostoc* Moss2, Moss3 and Moss4, lack both the *vnf* and *nif2* genes.

N. punctiforme has homologues of only two *vnf* genes, *vnfH* and *vnfR2* (the regulator of *vnfH* [79]), and they are nearly identical to *vnfH* and *vnfR2* in *Nostoc* 2RC; however, unlike *Nostoc* 2RC, *N. punctiforme* lacks the genes for the V-nitrogenase alpha and beta subunits (*vnfDG* and *vnfK*), as well as the vanadate transport genes, *vupABC* [82], so it cannot make a V-nitrogenase (Table 3). The presence of the entire *vnf* gene cluster in *Nostoc* 2RC and the similarity of *vnfR2* and *vnfH* to homologues in *N. punctiforme* suggests that an ancestor of *N. punctiforme* might have had the entire *vnf* cluster. *Nostoc* UCD120, UCD121 and UCD122 are most similar to *N. punctiforme* in terms of having only *vnfH*-like and *vnfR2*-like genes (Table 3); however, the *vnfR* gene in the three UCD strains is truncated to only about 15% of the size of the typical *vnfR* gene.

The true *vnfH* genes in the *N. variabilis*-like strains that have complete *vnf* gene clusters are distantly related to the *vnfH* copies present in the *Nostoc* strains that lack major *vnf* genes (Fig. 6). *Nostoc* 2RC has all the *vnf* structural genes, although their organization is different from the cluster in the *N. variabilis*-like strains. The *vnfH* gene of *Nostoc* 2RC clusters with the *vnfH* genes of the *Nostoc* strains that lack the rest of the *vnf* genes. In contrast to the *N. variabilis*-like strains, *Nostoc* 2RC has a second copy of *vnfH* located near the V-nitrogenase structural gene *vnfDG* but divergently transcribed (Fig. 6). Although we have named it *vnfH2* based on its location, its function as part of the V-nitrogenase in this strain will need to be confirmed experimentally. This second *vnfH* gene in *Nostoc* 2RC clusters with a group of *nifH5* genes of unknown function that, in these other strains, is found downstream of the *cydAB* genes (Fig. 6). This group of *nifH* genes is phylogenetically distinct from all the other *nifH* copies. The gene most closely related to *vnfH2* in *Nostoc* 2RC is *nifH4* in an uncharacterized strain, *Microchaete diplosiphon* NIES-3275; however, that *nifH4* is located near the *cydAB* genes like the *nifH5* copies of *N. variabilis* (Fig. 6).

Table 3. V-nitrogenase and V-transport genes (*vupABC*) in symbiotic and non-symbiotic strains

	Percentage nucleotide identity based on shared regions of similarity with ATCC 29413									
	vnfR1	***vnfR2***	***vnfR3***	***vnfH***	***vnfH2***	***vnfDG***	***vnfK***	***vupA***	***vupB***	***vupC***
Nostoc variabilis ATCC 29413	100	100	–	100	–	100	100	100	100	100
Nostoc sp. YBS01	100	100	–	100	–	100	100	100	100	100
Nostoc 9RC	100	100	–	100	–	100	100	100	100	100
Nostoc N2B	100	100	–	100	–	100	100	100	100	100
Nostoc ARAD	100	100	–	100	–	100	100	100	100	100
Nostoc FSR	100	100	–	100	–	100	100	100	100	100
Nostoc V5	100	100	–	100	–	100	100	100	100	100
Nostoc PNB	100	100	–	100	–	100	100	100	100	100
Nostoc Moss5	97	100	–	99	–	99	99	94	96	95
Nostoc Moss6	97	100	–	99	–	99	99	94	96	95
Nostoc punctiforme	–	75	–	82	–	–	–	–	–	–
Nostoc UCD120	–	82†	–	82	–	–	–	–	–	–
Nostoc UCD121	–	82†	–	82	–	–	–	–	–	–
Nostoc UCD122	–	82†	–	82	–	–	–	–	–	–
Nostoc 2RC	70	73	+‡	85	+	80*	76	82	84	77
Nostoc linckia z1	–	80†	+	84	–	–	–	–	–	–
Nostoc linckia NIES-25	–	75	–	85	–	–	–	–	–	–

*Although *vnfDG* is a fused gene, *vnfD* in 2RC has 80% identity with *vnfD* in *N. variabilis* ATCC 29413 but *vnfG* has only 68% identity with *vnfG* in *N. variabilis* ATCC 29413; a 330 bp region between *vnfD* and *vnfG* in 2RC has no matching region in *N. variabilis* ATCC 29413.
†Truncated *vnfR*, about 15% of the length compared to homologues.
‡+ = Present in the strain. Accession numbers for these genes are provided in Table S2.

The *vnf* genes in *N. variabilis* ATCC 29413 are repressed by VnfR1 or VnfR2 when Mo is present. In the absence of Mo, the repressor cannot bind and the *vnf* genes are expressed [79]. *N. punctiforme*, the two *N. linckia* strains and *Nostoc* Moss3, all lacking the structural genes for the V-nitrogenase, *vnfDGK*, have a *vnfH* homologue with a single copy of *vnfR* upstream of *vnfH* (Fig. 6). This *vnfR* gene is homologous to the *vnfR2* gene in *N. variabilis* ATCC 29413, which is located just upstream of the functional *vnfH* gene; however, there is no information on the expression or function of the *vnfH* homologue in *N. punctiforme*. In addition to the *vnfR1* and *vnfR2* genes present in the *N. variabilis*-like group, *Nostoc* 2RC and *Microchaete diplosiphon* NIES-3275 have an additional copy, *vnfR3*, located just upstream of the vanadate transport genes, *vupABC* [82] (Fig. 7). This *vnfR3* gene clusters with the *vnfR1* genes that are present in all the strains that have a functional V-nitrogenase. It is interesting that while there are redundant copies of *vnfR*, none is located close to the V-nitrogenase genes, *vnfDG* and *vnfK*, which are regulated by VnfR. Similarly, *Nostoc* 2RC has a copy of *vnfH* upstream of *vnfDG* and *vnfK* and has three copies of *vnfR*, but none is near the *vnfDGK* structural genes, suggesting that the regulatory genes have always been distant from the major *vnf* structural genes.

Many cyanobacteria, even those with only one *nif* system, have additional *nifH* copies (Fig. 6) that have not been studied. *N. variabilis* has five copies of *nifH* genes (*nifH5* is mentioned above) of which only three, *nifH1*, *nifH2* and *vnfH*, function as part of a complete nitrogenase [83]. A phylogeny of all the *nifH* genes in the strains described in this study shows that the *nifH* copies that are most closely related to each other also share similar genes surrounding them, even (in the case of *nifH5*) when these nearby genes have no similarity to known nitrogenase genes. The *nifH1* copies of all the *N. variabilis*-like strains, as well *Nostoc* 2RC and the two *N. linckia* strains, are most closely related to each other. The *nifH1* genes of *N. punctiforme* and its close relatives *Nostoc* UCD120, UCD121 and UCD122 are closely related to each other and to *Nostoc* Moss2 and Moss4, but less closely related to the *N. variabilis* group (Fig. 6).

The *nifH* copy most closely related to *nifH1* is called *nifH2* in strains that lack the *nif2* system and *nifH4* in strains that have the *nif2* system (Fig. 6). Although the function of *nifH2* (*nifH4*)

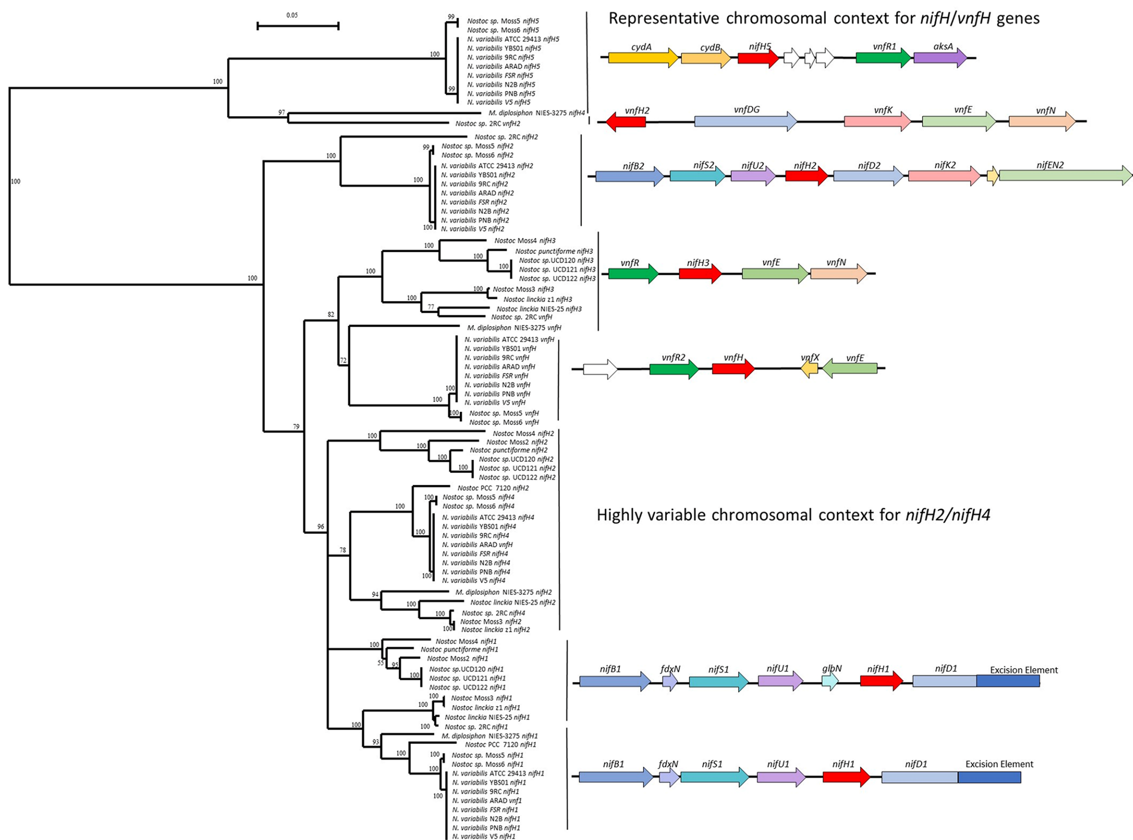


Fig. 6. Phylogenetic tree of *nifH/vnfH* genes for the symbiotic cyanobacterial strains isolated from plants and similar comparator strains that were not isolated from a plant. The *N. variabilis* group represents *N. variabilis* ATCC 29413 and the other nearly identical strains, V5, ARAD, FSR, N2B, PNB, 9RC and YBS01, while *Nostoc* UCD strains represent *Nostoc* UCD120, UCD121 and UCD122. Accession numbers for *nifH/vnfH* genes are provided in Table S2. Scale represents the number of substitutions per site.

is unknown, the gene, like *nifH1*, is expressed exclusively in heterocysts [84]. Like the *nifH1* genes, the *nifH4* genes of the *N. variabilis*-like strains (*nifH2* in strains lacking a *nif2* system) form a cluster that is related to their homologues in *Nostoc* 2RC, the two *N. linckia* strains and *Nostoc* Moss3. The *nifH1* and *nifH4* groups are related to each other but are distinct from the *vnfH* group (Fig. 6). Only *vnfH* from *N. variabilis* ATCC 29413 has been shown to function as part of the V-nitrogenase; however, its similarity in sequence and gene context to *vnfH* genes in other cyanobacteria suggests that either the other strains lost the rest of the *vnf* genes or possibly that this *nifH* copy gained its VnfH function later in the evolution of the *vnf* genes, since the *nifH1* gene of *N. variabilis* ATCC 29413 functions well in place of *vnfH* [85].

Sugar transport genes

Cyanobacteria in symbiotic associations obtain their carbon primarily from the plant in the form of sugars; hence, they must be able to take up sugar from the plant [39, 44]. Both *N. variabilis* and *N. punctiforme*, but not *Nostoc* PCC 7120, are capable of using fructose to grow heterotrophically in the dark [27, 86]. If the fructose transport genes of *N. variabilis* are expressed in *Nostoc* PCC 7120, the latter strain gains the ability to grow on fructose heterotrophically in the dark, indicating that it is the lack of a fructose transport system that limits the use of fructose in *Nostoc* PCC 7120 [87]. *N. punctiforme* has both fructose and glucose transporters [86]; however, most other filamentous cyanobacteria have only fructose transport genes. In *N. variabilis* ATCC 29413 and the nearly identical strains V5, ARAD, FSR, N2B, PNB, 9RC and YBS01 the *frtA-frtB-frtC* genes comprise a typical ABC-transporter [87]; however, *N. punctiforme*, *Nostoc* UCD120, UCD121 and UCD122, *Nostoc* Moss2, Moss3 and Moss4, and *Nostoc* 2RC all have two copies of *frtA* (Fig. 8). The *frtA1* genes are most closely related to each other as are the *frtA2* genes, but the single *frtA* gene in *N. variabilis* clusters with the *frtA2* genes (Fig. S1). In *N. punctiforme*, *Nostoc* UCD120, UCD121 and UCD122, and *Nostoc* Moss2, the glucose transporter, *glcP*, is just downstream from the fructose transport genes (Fig. 8) but is elsewhere in *Nostoc* strains NIES-25, Moss3 and 2RC. The glucose transporter, *glcP*, is absent in the symbiont *Nostoc* Moss4 and the non-symbiont, *N. linckia* z1. Although *N. punctiforme* grows poorly using glucose as a carbon source, loss of the glucose transporter in *N. punctiforme* prevented infection of the hornwort *Anthoceros*, but the absence of the fructose transporter (in a strain overexpressing the glucose transporter) did not affect infection [86].

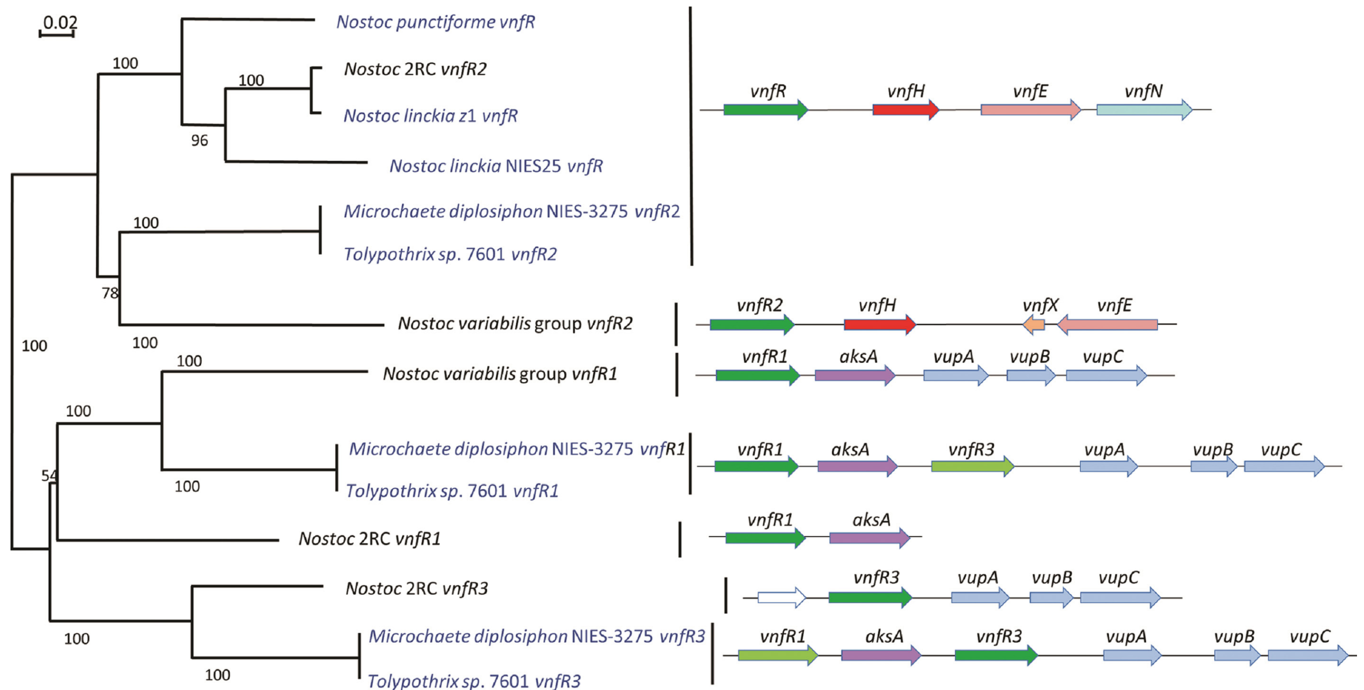


Fig. 7. Phylogenetic tree of *vnfR* genes for the cyanobacterial strains that have the structural V-nitrogenase genes (black) versus comparator strains that lack these genes (blue). The *N. variabilis* group represents *N. variabilis* ATCC 29413 and the other nearly identical strains, V5, ARAD, FSR, N2B, PNB, 9RC and YBS01. Accession numbers for *vnfR* genes are provided in Table S2. Scale represents the number of substitutions per site.

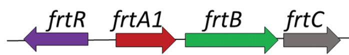
All of the strains studied here have a gene, *frtR* (*hrmR*), that makes a LacI-like repressor that regulates the expression of the *frt* operon [87]. While all the strains isolated from plants have sugar transport genes, other non-symbiotic strains, notably *N. variabilis* ATCC 29413, *Nostoc* PCC 7107 and the two *N. linckia* strains, have sugar transport genes that are very similar to those in the symbionts.

The *hrm* genes are involved in the regulation of hormogonia, which are plant-responsive, motile cyanobacterial filaments that initiate plant infection [20, 33]. HrmR (FrtR) regulates its own transcription and that of *hrmE* while *hrmA* and *hrmU* are important for the regulation of the level of hormogonia production [45, 88]. The *hrm* genes were absent in *N. variabilis* and the nearly identical strains V5, ARAD, FSR, N2B, PNB, 9RC and YBS01 and were also absent in *Nostoc* Moss5 and Moss6, but were present in *N. punctiforme*, the two *N. linckia* strains, *Nostoc* UCD120, UCD121 and UCD122, *Nostoc* 2RC, and *Nostoc* Moss 2, Moss3 and Moss4. Although the organization of the sugar transport and *hrm* genes is very similar among the strains that have these genes [35], in *Nostoc* UCD120, UCD121 and UCD122 the *hrmA* and *hrmU* genes are some distance from the rest of the *hrm* cluster (on a different contig) (Fig. 8).

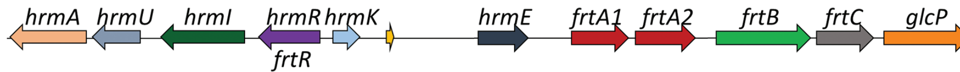
Chemotaxis

N. punctiforme has multiple clusters of genes with similarity to chemotaxis (*che*) genes from other bacteria [89] but few have been well characterized. A deletion of the entire locus of *cheR*-like genes (NpR0244–NpR0250) has no effect on hormogonia formation or phototaxis and these genes do not respond to the addition of a hormogonium-inducing factor [90]; however, it has also been reported that a strain with a mutation of one of the genes in this cluster, NpR0248, is impaired in hormogonia formation and motility and is unable to infect the symbiotic liverwort host, *Blasia pusilla* [37]. A *che* region with a high degree of sequence similarity to this large cluster is not present in any of the *N. variabilis*-like strains, the *Nostoc* UCD strains, the two strains of *N. linckia* nor any of the *Nostoc* Moss strains. However, several genes in the cluster, with about 50% amino acid identity based on shared regions of similarity to the *N. punctiforme* homologues, are present in many of the *Nostoc* strains. The best-characterized chemotaxis genes in cyanobacteria are the *hmpBCDE* genes in *N. punctiforme* [91–93]. Mutants in *hmpB*, *hmpC*, *hmpD* or *hmpE* prevent the formation of hormogonia and, therefore, are not motile and fail to establish a symbiotic relationship with the hornwort *Anthoceros punctatus* [92]. This gene cluster is well conserved in filamentous cyanobacteria, including the *N. variabilis*-like strains, *Nostoc* UCD120, UCD121 and UCD122, *Nostoc* 2RC, the two strains of *N. linckia,* and all the *Nostoc* Moss strains (Fig. S2). The *hmpE* gene is highly variable in size in these strains due to repeat regions of variable lengths, and *hmpD* in *Nostoc* Moss4 is split into two genes. The nucleotide identity for

Nostoc variabilis ATCC 29413/V5/ARAD/N2B/9RC/PNB/YBS01

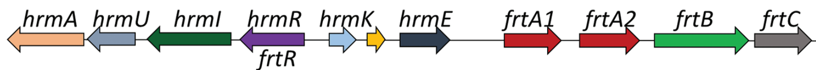


Nostoc punctiforme (similar in *Nostoc* Moss2)



Nostoc 2RC (similar in *Nostoc* Moss3 and *N. linckia* NIES-25)

Nostoc Moss4 and *N. linckia* z1



Nostoc UCD120-122

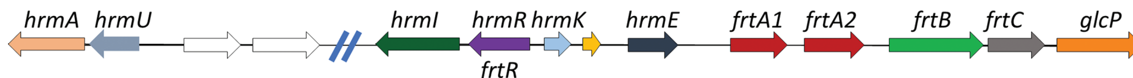


Fig. 8. Maps of fructose transport (*frtABC*) and glucose transport (*glcP*) gene regions, including genes, *hrm*, implicated in hormogonia regulation.

these genes between pairs of strains was similar to the ANI between pairs of strains. Although chemotaxis is surely important in symbiosis, we did not identify any *che*-like genes that distinguished symbiotic strains from many other non-symbiotic cyanobacteria.

Pili have been studied in unicellular cyanobacteria where they mediate twitching motility [94] and in *N. punctiforme* where a type IV pilus-like system powers the gliding motility and polysaccharide secretion of hormogonia [93]. Mutants in NpR0117 (*pilT*-like) and NpR2800 (*pilD*-like) of *N. punctiforme* have very low rates of infection of *Blasia* while a mutant in NpF0069 (*pulG, or pilA-like*) has somewhat reduced levels of infection of *Blasia* [95], suggesting that pili are important for the motile hormogonia that lead to infection of plants. Another putative *pilA* gene (NpF0676), which did not yield segregated mutants [95], is localized to rings at hormogonia cell junctions [92]. All of these putative *pilA*, *pilT* and *pilD* genes have homologues in all the *N. variabilis*-like strains, *Nostoc* YBS01, *Nostoc* strains UCD120, UCD121 and UCD122, *Nostoc* 2RC, *Nostoc* Moss2–6, and the two strains of *N. linckia* (Figs S3–S6). The nucleotide identity between pairs of strains for these genes was similar to the ANI between pairs of strains.

Symbiosis

Individual plant strains associate with a variety of *Nostoc* strains, and, similarly, individual *Nostoc* strains associate with a variety of plant strains [10, 96]. However, there is no evidence of successful reconstitution of the symbiosis of any cyanobacterial strain with the water fern *Azolla* [97]; therefore, we attempted to infect the more tractable liverwort, *Blasia pusilla,* with the newly sequenced cyanobacterial strains. Cyanobacteria infect structures known as auricles on the surface of the *Blasia* plant thallus. Within the auricle, the endophytic cyanobacteria grow and differentiate a high percentage of heterocysts, filling the cavity. Because the auricles on the thallus are colourless, it is easy to distinguish the dense green infected auricles from uninfected auricles and from cyanobacterial filaments that are loosely associated with the plant surface (Fig. 9a). We infected axenic *Blasia* plants on agar plates with axenic liquid cultures of *N. punctiforme*, *N. variabilis* ATCC 29413 and the newly sequenced strains. Despite the near identities of the *N. variabilis*-like strains V5, ARAD, FSR, N2B, PNB and 9RC, only *N. variabilis* ATCC 29413, PNB and V5 infected *Blasia*, and the infection rate for all was poor with few infected auricles (determined semi-quantitatively to be about 5%). Among the *Nostoc* strains, *N. punctiforme* and *Nostoc* 2RC infected *Blasia* easily with about 70–75% infected auricles (Fig. 9a). *Nostoc* 2RC was unusual among all these strains in that it produced abundant, highly motile hormogonia, causing the strain to swarm over the entire surface of an agar plate. It seems likely that its proficiency in producing motile hormogonia aids in its ability to infect *Blasia*. Among the three nearly identical *Nostoc* strains, UCD120, UCD121 and UCD122, only *Nostoc* UCD122 infected *Blasia* and the frequency was low, similar to *N. variabilis* ATCC 29413 (Fig. 9a). Although these three strains are

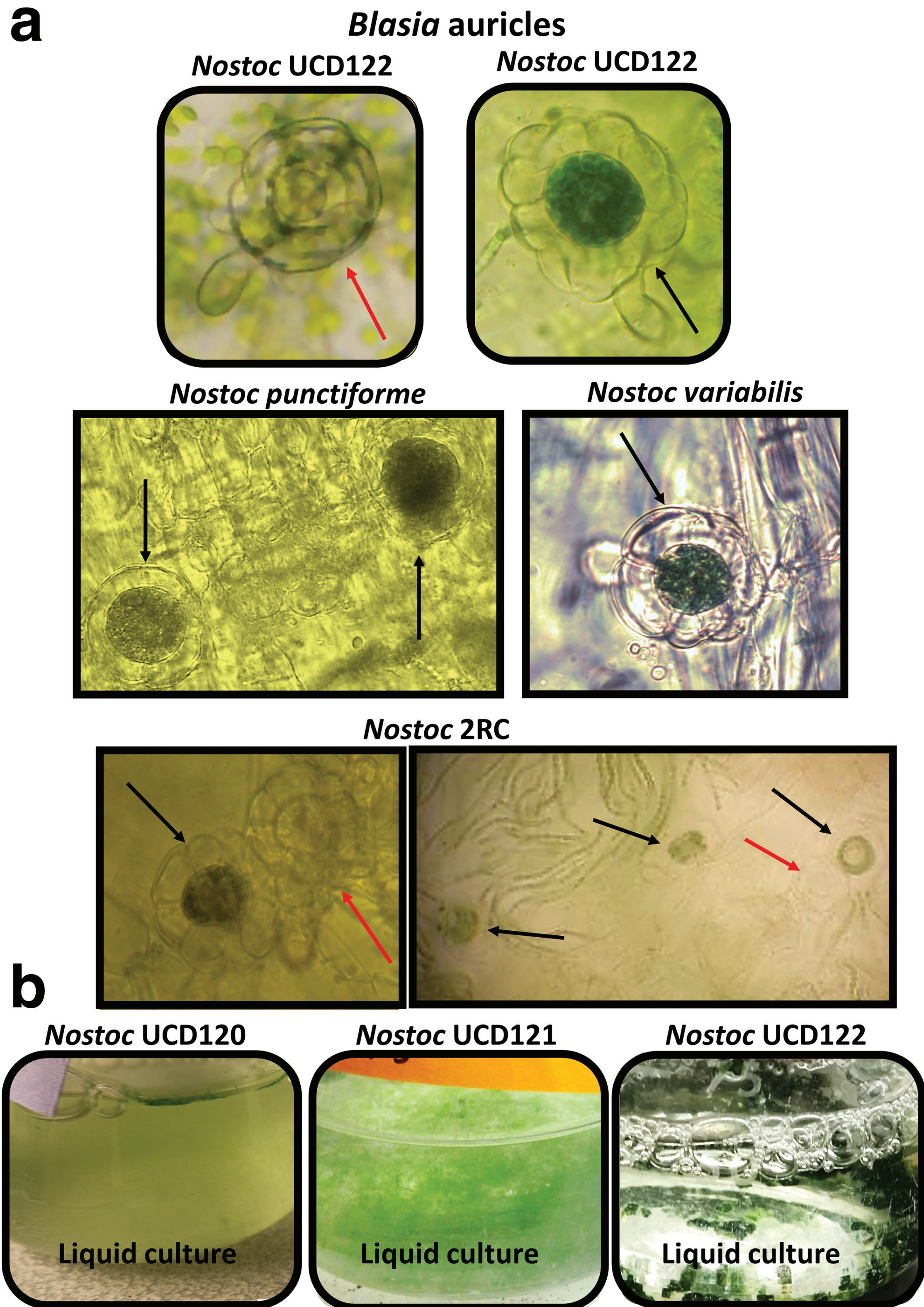


Fig. 9. *Blasia* infection and *Nostoc* UCD strain growth. (a) Light micrographs of uninfected *Blasia* auricles and *Blasia* auricles infected by *Nostoc* UCD122, *N. punctiforme*, *N. variabilis* ATCC 29413 and *Nostoc* 2RC. Black arrows indicate green infected auricles (about 120 µm in diameter), packed with cyanobacteria, while red arrows indicate nearly colourless uninfected auricles. Infected auricles were determined as described in the methods. (b) Liquid growth characteristics of *Nostoc* UCD strains.

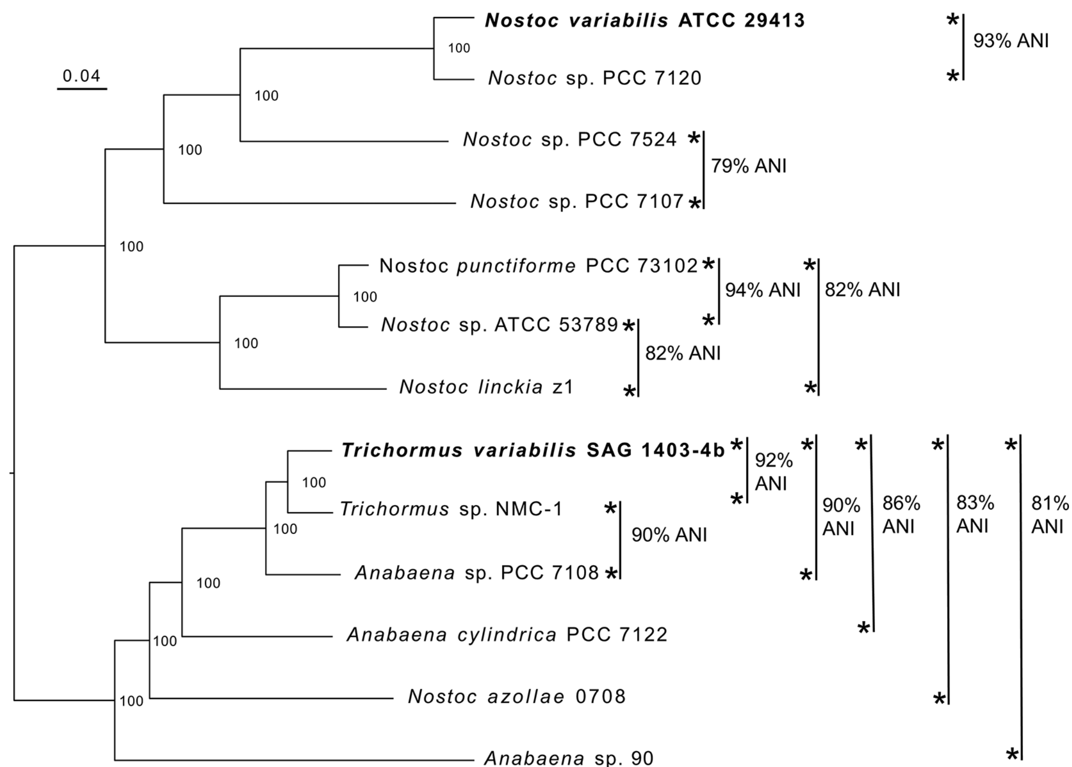


Fig. 10. Phylogenetic distance tree based on genomes for representative cyanobacterial strains. ANI values are included for comparison between strains indicated by asterisks. Accession numbers for these strains are provided in Table S1. Scale represents the number of substitution per site.

nearly identical genetically, their morphology in liquid culture was different. *Nostoc* UCD122 filaments aggregated into clumps, while liquid cultures of the other two strains had a smooth and homogeneous appearance (Fig. 9b). Thus, for *Nostoc* UCD122, aggregation and hormogonia formation were correlated with its ability to infect *Blasia*. *N. punctiforme* normally grows in clumps [37, 90, 98]; however, there is an uncharacterized smooth variant that is easier to work with but lacks hormogonia [92]. An *hmp* mutant that could not differentiate hormogonia grew as dispersed suspensions [92], while a *pks2* mutant that produced many highly motile hormogonia showed increased aggregation [99]. The smooth variant of *N. punctiforme* infects *Blasia* very poorly [37], similar to *N. variabilis*, which also grows as a homogeneous suspension.

Taxonomy and nomenclature

'*Anabaena variabilis* ATCC 29413' is now called '*Nostoc variabilis* ATCC 29413', a change that is supported by both the characteristics of the strain and its phylogeny. Like other symbiotic *Nostoc* strains, *N. variabilis* ATCC 29413 and its very close relatives that were isolated from *Azolla* produce hormogonia [100], which are required for the infection of plants. Also, the symbiotic *Nostoc* strains can transport and use sugars, notably fructose, consistent with their symbiotic lifestyle. While there are strains that are now considered to be *Trichormus variabilis* [101], *N. variabilis* ATCC 29413 is not closely related to them (see Discussion). A phylogenetic tree based on 500 shared genes as well as the whole genome ANI values (Fig. 10) indicates that *N. variabilis* ATCC 29413 is closely related to other *Nostoc* strains and is more distantly related to the only two *Trichormus* strains that have a sequenced genome [102, 103]. In fact, the two sequenced *Trichormus* strains cluster together with *A. cylindrica* PCC 7122 and *Anabaena* sp. PCC 7108 (Fig. 10), not with *N. variabilis* ATCC 29413.

DISCUSSION

Genome characterization of newly sequenced strains isolated from *Azolla*

Although the non-culturable strain *N. azollae* 0708 has been identified as the primary symbiont in *Azolla* [15], other cyanobacteria have also been found associated with the plant [9–13]. We sequenced the genomes of 10 cyanobacteria that were isolated from *Azolla* strains by different laboratories in different parts of the world (Table 1). Most were nearly identical to *N. variabilis*, which was not itself isolated from *Azolla*, and were very similar to *Nostoc* Moss5 and Moss6 strains that were isolated from feathermoss [21]. Strains ARAD from *Azolla filiculoides* in Israel and V5 from *Azolla pinnata* in Russia were nearly indistinguishable from strains 9RC and N2B isolated from *Azolla caroliniana* in the USA. However, there were some differences, including the absence

of the 37 kb linear plasmid in strains PNB and FSR and the lack of the *nifD* 11 kb excision element in FSR. In contrast, *Nostoc* UCD120, UCD121 and UCD122 were all isolated from *A. caroliniana* originally collected in Ohio [8, 104]. Unlike the other strains that we sequenced, the UCD strains were first collected from *Azolla* and then isolated after selecting for strains that were able to form an association with *Anthoceros*, prior to growing them as free-living cultures [8]. These strains are nearly identical genetically but are morphologically different, especially *Nostoc* UCD122, which aggregates into clumps in liquid culture.

One strain isolated from *Azolla*, *Nostoc* 2RC, was different from the other characterized strains. *Nostoc* 2RC is brownish in colour, aggregates into clumps in liquid culture and spreads rapidly over the surface of agar plates via its very motile hormogonia. Its genome is most similar to two *N. linckia* strains, which have not been reported to be symbiotic, and to *Nostoc* Moss3, a symbiont that was isolated from feathermoss [21]. However, *Nostoc* 2RC differs significantly from these other strains because it has two additional sets of nitrogenase genes, also present in all of the *N. variabilis*-like strains and in *Nostoc* Moss5 and Moss6. These are the *nif2* genes, encoding an Mo-nitrogenase that functions in vegetative cells under anoxic conditions [66, 83] and the *vnf* genes that make the alternative V-nitrogenase [78, 79], which are absent in all the *Nostoc* strains closely related to *Nostoc* 2RC, except for the *vnfH*-like and *vnfR*-like genes. The presence of these two conserved *vnf*-like genes suggests that these *Nostoc* strains may have once had the full cluster of *vnf* genes.

Characteristics of many symbiotic strains shared by the newly sequenced strains

Culturable symbiotic strains isolated from *Azolla* and other plants share several physiological characteristics. All form heterocysts and fix nitrogen, which is supplied to the plant [41, 43, 105, 106], a characteristic shared with *N. azollae* 0708, the non-culturable *Azolla* symbiont [15, 107, 108]. All culturable symbionts can transport sugars, typically fructose, although many of the symbiotic *Nostoc* strains described here have a gene for glucose transport, a characteristic shared with several non-symbiotic strains including unicellular cyanobacteria [109, 110]. In contrast, orthologues of the glucose and fructose transport genes in *N. punctiforme* and *N. variabilis* [86, 87] are absent in *N. azollae* 0708 [15]; however, other transporters may mediate sugar transport in *N. azollae* 0708 [111].

Nitrogenase genes are common in cyanobacteria, especially the *nif* genes that encode the heterocyst-specific Mo-nitrogenase, and all the strains studied here had these genes. In contrast, complete sets of *vnf* genes are comparatively rare but are present in the genomes of about a dozen strains of the genera *Nostoc*, *Fischerella*, *Chlorogleopsis*, *Calothrix* and *Tolypothrix*/*Aulosira* (accessed via the JGI genomes database). The *vnf* genes have also been found in a *Nostoc* strain associated with the lichen *Peltigera* [32, 80] and in three *Nostoc* strains isolated from hornworts *Phaeoceros carolinanus* and *Leiosporoceros dussii*, and *Blasia pusilla,* some on plasmids [112]. The V-nitrogenase may confer a selective advantage to the plants in environments where Mo is limiting and they may be laterally transferred [112], perhaps even between strains in the same plant.

Several strains that have the *vnf* genes, including *Aulosira laxa* NIES-50, *Nostoc carneum* NIES-2107, *Calothrix brevissima* NIES-22 and *Tolypothrix tenuis* PCC 7101, also have the *nif2* gene cluster. There appear to be no strains that have the *nif2* genes but not the *vnf* genes, suggesting that the *nif2* genes, which form a single tight cluster, may have been lost from strains that once had both. The significance of the association of the *vnf* and *nif2* gene clusters in these strains as well as in symbiotic strains closely related to *N. variabilis* is unknown.

Infection of *Blasia* by the strains isolated from *Azolla*

The culturable, free-living strains that infect plants probably do so through the differentiation of motile, non-growing hormogonia [106, 113]. In the context of symbiosis, hormogonia have been studied only in *N. punctiforme* [36, 44], although they are also made in *N. variabilis* ATCC 29413 in cells subjected to starvation for fixed nitrogen [100]. Among the newly sequenced strains described here, hormogonia were abundant and highly motile only in *Nostoc* 2RC. Although several of the strains were capable of infecting the liverwort *Blasia*, only *N. punctiforme*, *Nostoc* UCD122 and *Nostoc* 2RC did so readily. These strains shared the characteristic of clumpy growth, probably a result of hormogonia formation [92, 99].

All the strains had similar chemotaxis genes and pili-related genes that are likely to be involved in infection [37, 92, 93, 114]; however, many cyanobacteria that have no known association with symbiosis also have homologes of these genes. Analysis of the genes and metabolic pathways that correlate with symbiotic cyanobacteria suggests that many of these pathways function in a coordinated manner to allow plant detection, infections and maintenance of the symbiotic state [21, 34]. However, we found that even genes identified from several cyanobacteria as associated with symbiosis in feathermoss [34] have homologues in strains not known to be symbiotic (Fig. 1, Table S3). Among the well-characterized genes analysed here, there appear to be no genes that clearly distinguish symbiotic strains from many other related, non-symbiotic strains of the genus *Nostoc*.

N. variabilis ATCC 29413 has been reported not to infect *Anthoceros* [106] but we found that it infected *Blasia*, as do its close relatives *Nostoc* Moss5 and Moss6 in a clade called Extra II [34]. However, *Nostoc* Moss5 and Moss6 do not infect *Gunnera* [34], and there is one report that *N. variabilis* N2B (also known as *Anabaena azollae* N1) also does not infect *Gunnera* [96]. Although *Nostoc* UCD122 infected *Blasia* (Fig. 9), as does its close relative *N. punctiforme* [34], the nearly genetically identical strain *Nostoc* UCD120 (ANI between UCD120 and UCD122 is 99.9%) did not. While we were not successful in infecting *Blasia* with *Nostoc* UCD120, it has been shown to infect *Anthoceros* [115]. Therefore, it seems likely that in different laboratories, variations in the physiological conditions of the plants and the cyanobacteria lead to differences in success in reconstituting

a symbiosis. Further, natural ecosystems with their plant-associated biota are likely to favour symbiosis in ways that cannot be reproduced in the laboratory.

Taxonomy and naming of *Nostoc* (*Anabaena*) *variabilis* ATCC 29413

N. variabilis and its close relatives described here share an ANI of 98% with feathermoss isolates *Nostoc* Moss5 and Moss6, which form symbiotic associations with *Pleurozium schreberi* and *Blasia pusilla* [34]. The latter strains are virtually identical to each other and somewhat distant from the other feathermoss *Nostoc* isolates, Moss2, Moss3 and Moss 4. *N. variabilis*, *Nostoc* Moss5 and *Nostoc* Moss6, like *N. variabilis* ATCC 29413, belong to a clade called Extra II, indicating their symbiosis is extracellular [34]. However, while infection of *Blasia* is extracellular but endophytic, infection of feathermoss is extracellular and epiphytic. Among characterized strains, only those in the clade called Extra/Intra are capable of intracellular symbiosis with *Gunnera* and extracellular symbiosis with bryophytes The Extra II clade, including *Nostoc* Moss5 and Moss6, which diverged about 1500 Ma appears to have gained its symbiotic capacity by horizontal gene transfer with genes coming almost equally from the strains in the Extra/Intra clade (which includes *N. punctiforme* and *Nostoc* Moss 2), from *N. azollae*, and from the Extra I clade (which includes *Nostoc* Moss3 and Moss4) [34].

Komárek and Anagnostidis created the genus name *Trichormus* and placed a strain, *Anabaena variabilis* SAG 1403 4b (ATCC 29211; PCC 6309), in that genus [101]. In just the last few years NCBI has renamed any *Anabaena variabilis* strain as *Trichormus variabilis*, including ATCC 29413. However, based on multiple lines of evidence, we have called *Anabaena variabilis* ATCC 29413 *Nostoc variabilis* ATCC 29413 instead. As early as 1979, *Anabaena* strains were defined as obligate photoautotrophs [61]; however, by then it was known that *A. variabilis* ATCC 29413 is capable of growing in the dark using fructose [27]. Bergey's Manual states that the 'Wolk strain' (C.P. Wolk) of *A. variabilis* (i.e. ATCC 29413) was misidentified, commenting that strains of *A. variabilis* that are closely related to *Anabaena* sp. PCC 7108 (a photoautotroph) do not include the strain *A. variabilis* ATCC 29413. Also, Bergey's Manual states that *A. variabilis* ATCC 29413 differs from the true *A. variabilis* strains in the morphology of the trichomes, the akinetes and the hormogonia [116].

This distinction among *Trichormus (Anabaena) variabilis* strains is supported by our data that demonstrate the *Nostoc*-like characteristics of *N. variabilis* ATCC 29413 and by the phylogenetic trees provided here and in the online cyanobacterial taxonomy database (http://cyanophylogeny.scienceontheweb.net/). Earlier phylogenetic trees, based on either 16S rRNA or *rpoB* gene sequences, showed that *T. variabilis* HINDAK 2001/4 and *T. variabilis* GREIFSWALD/92 are most closely related to *Anabaena* sp. PCC 7108, but are distant from *Nostoc* sp. PCC 7120, the closest relative to *N. variabilis* ATCC 29413 shown in those trees [117]. Similarly, phylogenetic trees published for *T. variabilis* SAG 1403 4b (ATCC 29211; PCC 6309) [102] and for *Trichormus* sp. NMC-1 [103] show that these *Trichormus* strains are not closely related to *N. variabilis* ATCC 29413. *N. azollae* 0708, the non-culturable *Azolla* symbiont, has also recently been renamed *Trichormus azollae* by NCBI. Phylogenetic trees based on whole-genome data indicate that *T. azollae* clusters with *T. variabilis* SAG 1403 4b and *Trichormus* sp. NMC-1, along with *Anabaena* sp. PCC 7108 and *A. cylindrica* PCC 7122 [102, 103]. With more sequenced genomes the taxonomic relationships among the *Nostoc*/*Anabaena*/*Trichormus* genera may become clearer.

CONCLUSIONS

Free-living *Nostoc* strains that were isolated from *Azolla* from many parts of the world share many genes with other characterized symbiotic cyanobacteria cultured from a variety of plants. However, these genes are also found in closely related cyanobacteria that have no known associations with plants, such as two strains of *Nostoc linckia*. The accepted method for demonstrating symbiosis is the ability of a purified axenic cyanobacterial strain to infect an axenic plant. In this study, we found that some axenic strains isolated from *Azolla* could readily infect axenic *Blasia*, but others did so poorly or not at all, despite having 'symbiosis' genes and, in some cases, nearly identical genomes with strains that could infect. This may not be surprising for two reasons. First, mutations in genes or regulatory regions that are critical for producing motile hormogonia and for communicating with the plant would be expected to drastically reduce infection. These would be difficult to identify because most such genes are unknown or poorly characterized. Strains maintained in a sterile laboratory setting are very likely to lose characteristics that are important for infection and symbiosis. Second, normal infection and establishment of symbiosis occur in an ecological environment rich in bacteria, fungi, insects, and plants that probably play important roles in symbiotic infections. Thus, we speculate that in the natural environment other organisms aid cyanobacteria in establishing associations with plants that will never be reproduced in a laboratory setting. Future research on symbiosis should consider and incorporate the natural plant biome as a possible contributor to successful symbiosis.

Funding information
This work was supported by the National Science Foundation grant MCB-1818298.

Acknowledgements
We are grateful to John Meeks for suggestions on the taxonomy of *Nostoc* and to undergraduate research students Sara Abramowski, Kaylyn Bauer, Rajesh Dhimal, Abigail Eaker, Fei Fu, Rahmah Ghazal, Christopher Harper, Kent Morgan, Erica Nelson, Lindsey Probst and Thomas Smith for their contributions to the research.

Author contributions
T.T. was responsible for the project conceptualization, funding acquisition, project supervision and oversight, data analysis, and writing and editing the manuscript. B.P. was responsible for methodology, experimentation, genome assembly, sequence and data analysis, production of figures, and reviewing and editing the manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Ventura W, Watanabe I. Green manure production of *Azolla microphylla* and *Sesbania rostrata* and their long-term effects on rice yields and soil fertility. *Biol Fertil Soils* 1993;15:241–248.
2. Perkins SK, Peters GA. The *Azolla-Anabaena* symbiosis: Endophyte continuity in the *Azolla* life-cycle is facilitated by epidermal trichomes. I. Partitioning of the endophytic *Anabaena* into developing sporocarps. *New Phytol* 1993;123:53–64.
3. Nagalingum NS, Schneider H, Pryer KM. Comparative morphology of reproductive structures in heterosporous water ferns and a reevaluation of the sporocarp. *Int J Plant Sci* 2006;167:805–815.
4. Lumpkin TA, Plucknett DL. *Azolla*: botany, physiology, and use as a green manure. *Econ Bot* 1980;34:111–153.
5. Newton JW, Herman AI. Isolation of cyanobacteria from the aquatic fern, *Azolla*. *Arch Microbiol* 1979;120:161–165.
6. Franche C, Cohen-Bazire G. Evolutionary divergence in the *nifHDK* gene region among nine symbiotic *Anabaena azollae* and between *Anabaena azollae* and some free-living heterocystous cyanobacteria. *Symbiosis* 1987;3:159–178.
7. Nierzwicki-Bauer SA, Haselkorn R. Differences in mRNA levels in Anabaena living freely or in symbiotic association with *Azolla*. *Embo J* 1986;5:29–35.
8. Meeks JC, Joseph CM, Haselkorn R. Organization of the nif genes in cyanobacteria in symbiotic association with *Azolla* and *Anthoceros*. *Arch Microbiol* 1988;150:61–71.
9. Sood A, Prasanna R, Prasanna BM, Singh PK. Genetic diversity among and within cultured cyanobionts of diverse species of *Azolla*. *Folia Microbiol* 2008;53:35–43.
10. Papaefthimiou D, Hrouzek P, Mugnai MA, Lukesova A, Turicchia S *et al*. Differential patterns of evolution and distribution of the symbiotic behaviour in Nostocacean cyanobacteria. Int J Syst Evol Microbiol 2008;58:553–564.
11. Papaefthimiou D, Van Hove C, Lejeune A, Rasmussen U, Wilmotte A. Diversity and host specificity of *Azolla* cyanobionts. *J Phycol* 2008;44:60–70.
12. Dijkhuizen LW, Brouwer P, Bolhuis H, Reichart G-J, Koppers N, *et al*. Is there foul play in the leaf pocket? The metagenome of floating fern *azolla* reveals endophytes that do not fix N2 but may denitrify. *New Phytol* 2018;217:453–466.
13. Kumar U, Nayak AK, Panneerselvam P, Kumar A, Mohanty S, *et al*. Cyanobiont diversity in six *azolla* spp. And relation to *azolla*-nutrient profiling. *Planta* 2019;249:1435–1447.
14. Banach A, Kuźniar A, Mencfel R, Wolińska A. The study on the cultivable microbiome of the aquatic fern *Azolla filiculoides* L. As new source of beneficial microorganisms. *Appl Sci* 2019;9:2143.
15. Ran L, Larsson J, Vigil-Stenman T, Nylander JAA, Ininbergs K. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. *PLoS One* 2010;5:e11486.
16. Pereira AL, Vasconcelos V. Classification and phylogeny of the cyanobiont *Anabaena azollae* Strasburger: an answered question? *Int J Syst Evol Microbiol* 2014;64:1830–1840.
17. Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J *et al*. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants* 2018;4:460–472.
18. Tang LF, Watanabe I, Liu CC. Limited multiplication of symbiotic cyanobacteria of *Azolla* spp. *Appl Environ Microbiol* 1990;56:3623–3626.
19. Bergman B, Rai A, Rasmussen U. Cyanobacterial Associations. In: Elmerich C, Newton W (editors). *Associative and Endophytic Nitrogen-fixing Bacteria and Cyanobacterial Associations Nitrogen Fixation: Origins, Applications, and Research Progress*. Dordrecht: Springer; 2007. pp 257–301.
20. Adams DG, Duggan PS. Cyanobacteria-bryophyte symbioses. *J Exp Bot* 2008;59:1047–1058.
21. Warshan D, Espinoza JL, Stuart RK, Richter RA, Kim S-Y *et al*. Feathermoss and epiphytic Nostoc cooperate differently: expanding the spectrum of plant–cyanobacteria symbiosis. *ISME J* 2017;11:2821–2833.
22. Rikkinen J, Virtanen V. Genetic diversity in cyanobacterial symbionts of thalloid bryophytes. *J Experiment Bot* 2008;59:1013–1021.
23. Stewart WDP, Rodgers GA. Studies on the symbiotic blue-green algae of anthoceros, *Blasia and Peltigera. Ecological Bulletins* 1978;26:247–259.
24. Zimmerman WJ, Rosen B, Lumpkin T. Enzymatic, lectin, and morphological characterization and classification of presumptive cyanobionts from *Azolla* lam. *New Phytol* 1989;113:497–503.
25. Tischer RG. Pure Culture of Anabaena flos-aquae A-37. *Nature* 1965;205:419–420.
26. Healey FP. Characteristics of phosphorus deficiency in *Anabaena*. *J Phycology* 1973;9:383–394.
27. Wolk CP, Shaffer PW. Heterotrophic micro- and macrocultures of a nitrogen-fixing cyanobacterium. *Arch Microbiol* 1976;110:145–147.
28. Currier TC, Haury JF, Wolk CP. Isolation and preliminary characterization of auxotrophs of a filamentous cyanobacterium. *J Bacteriol* 1977;129:1556–1562.
29. Haury JF, Wolk CP. Classes of *Anabaena variabilis* mutants with oxygen-sensitive nitrogenase activity. *J Bacteriol* 1978;136:688–692.
30. Peterson RB, Wolk CP. High recovery of nitrogenase activity and of ^{55}Fe-labeled nitrogenase in heterocysts isolated from *Anabaena variabilis*. *Proc Natl Acad Sci USA* 1978;75:6271–6275.
31. Meeks JC, Wolk CP, Lockau W, Schilling N, Shaffer PW, *et al*. Pathways of assimilation of [13N]N2 and 13NH4+ by cyanobacteria with and without heterocysts. *J Bacteriol* 1978;134:125–130.
32. Hodkinson BP, Allen JL, Forrest LL, Goffinet B, Sérusiaux E *et al*. Lichen-symbiotic cyanobacteria associated with *Peltigera* have an alternative vanadium-dependent nitrogen fixation system. *Eur J Phycol* 2014;49:11–19.
33. Meeks JC. Symbiosis between nitrogen-fixing cyanobacteria and plants: the establishment of symbiosis causes dramatic morphological and physiological changes in the cyanobacterium. *BioScience* 1998;48:266–276.
34. Warshan D, Liaimer A, Pederson E, Kim S-. Y, Shapiro N, *et al*. Genomic changes associated with the evolutionary transitions of Nostoc to a plant symbiont. *Mol Biol Evol* 2018;35:1160–1175.
35. Gagunashvili AN, Andrésson ÓS. Distinctive characters of *Nostoc* genomes in cyanolichens. *BMC Genomics* 2018;19:434:434.
36. Wong FCY, Meeks JC. Establishment of a functional symbiosis between the cyanobacterium *Nostoc punctiforme* and the bryophyte *Anthoceros punctatus* requires genes involved in nitrogen control and initiation of heterocyst differentiation. *Microbiol* 2002;148:315–323.
37. Duggan PS, Thiel T, Adams DG. Symbiosis between the cyanobacterium *Nostoc* and the liverwort *Blasia* requires a CheR-type MCP methyltransferase. *Symbiosis* 2013;59:111–120.
38. Rodgers GA, Stewart WDP. The cyanophyte-hepatic symbiosis I. morphology and physiology. *New Phytol* 1977;78:441–458.
39. Adams DG. Cyanobacteria in symbiosis with hornworts and Lliverworts. In: Rai AN BB, Rasmussen U (editors). *Cyanobacteria in Symbiosis*. Dordrecht: Springer; 2002.
40. Kimura J, Nakano T. Reconstitution of a *Blasia-Nostoc* symbiotic association under axenic conditions. *Nova Hedwigia* 1990;50:191–200.
41. Bond G, Scott GD. An examination of some symbiotic systems for fixation of nitrogen. *Ann Bot* 1955;19:67–77.
42. Meeks JC. Cyanobacterial-Bryophyte associations. In: Rai A (editor). *CRC Handbook of Symbiotic Cyanobacteria*. Boca Raton, FL: CRC Press; 1990. pp 43–64.
43. Adams D, Bergman B, Nierzwicki-Bauer S, Duggan PS, Rai AN, *et al*. Cyanobacterial-plant symbioses. In: *The prokaryotes: Prokaryotic biology and symbiotic aassociations*. 2013. pp. 359–400.

44. Meeks JC, Elhai J. Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiol Mol Biol Rev* 2002;66:94–121.
45. Cohen MF, Meeks JC. A hormogonium regulating locus, hrmUA, of the cyanobacterium Nostoc punctiforme strain ATCC 29133 and its response to an extract of a symbiotic plant partner Anthoceros punctatus. *Mol Plant-Microbe Interact* 1997;10:280–289.
46. Allen MB, Arnon DI. Studies on nitrogen-fixing blue-green algae. I. Growth and nitrogen fixation by Anabaena cylindrica Lemm. *Plant Physiol* 1955;30:366–372.
47. Golden JW, Robinson SJ, Haselkorn R. Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. *Nature* 1985;314:419–423.
48. Golden SS, Brusslan J, R Haselkorn. Genetic engineering of the cyanobacterial chromosome. *Methods Enzymol* 1987;153:215–231.
49. Antonopoulos DA, Assaf R, Aziz RK, Brettin T, Bun C, *et al*. PATRIC as a unique resource for studying antimicrobial resistance. *Brief Bioinform* 2019;20:1094–1102.
50. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol* 2017;13:e1005595.
51. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, *et al*. RASTTK: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 2015;5:8365.
52. Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, *et al*. PATTYFAMS: Protein families for the microbial genomes in the PATRIC database. *Front Microbiol* 2016;7:118.
53. Edgar RC. Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:113.
54. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, *et al*. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;25:1422–1423.
55. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
56. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010;27:221–224.
57. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 2017;110:1281–1286.
58. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, *et al*. MASH: Fast genome and metagenome distance estimation using minhash. *Genome Biol* 2016;17:132.
59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
60. Cove DJ, Perroud P-F, Charron AJ, McDaniel SF, Khandelwal A *et al*. Culturing the moss Physcomitrella patens. *Cold Spring Harbor Protocols* 2009;2009:pdb.prot5136.
61. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* 1979;111:1–61.
62. Currier TC, Wolk CP. Characteristics of Anabaena variabilis influencing plaque formation by cyanophage N-1. *J Bacteriol* 1979;139:88–92.
63. Mardanov AV, Beletskii AV, Gumerov VM, Karbysheva EA, Mikheeva LE. New low-copy plasmid in cyanobacterium Anabaena variabilis. *Russ J Genet* 2013;49:798–805.
64. Ungerer JL, Pratte BS, Thiel T. RNA processing of nitrogenase transcripts in the cyanobacterium Anabaena variabilis. *J Bacteriol* 2010;192:3311–3320.
65. Pratte BS, Ungerer J, Thiel T. Role of RNA secondary structure and processing in stability of the nifH1 transcript in the cyanobacterium Anabaena Variabilis. *J Bacteriol* 2015;197:1408–1422.
66. Thiel T. Organization and regulation of cyanobacterial nif gene clusters: implications for nitrogenase expression in plant cells. *FEMS Microbiol Lett* 2019;366.
67. Potts M, Angeloni SV, Ebel RE, Bassam D. Myoglobin in a cyanobacterium. *Science* 1992;256:1690–1692.
68. Hill DR, Belbin TJ, Thorsteinsson MV, Bassam D, Brass S, *et al*. GlbN (cyanoglobin) is a peripheral membrane protein that is restricted to certain Nostoc spp. *J Bacteriol* 1996;178:6587–6598.
69. Hilton JA, Meeks JC, Zehr JP. Surveying DNA elements within functional genes of heterocyst-forming cyanobacteria. *PLoS One* 2016;11:e0156034.
70. Lammers PJ, Golden JW, Haselkorn R. Identification and sequence of a gene required for a developmentally regulated DNA excision in Anabaena. *Cell* 1986;44:905–911.
71. Golden JW, Mulligan ME, Haselkorn R. Different recombination site specificity of two developmentally regulated genome rearrangements. *Nature* 1987;327:526–529.
72. Golden JW, Carrasco CD, Mulligan ME, Schneider GJ, Haselkorn R. Deletion of a 55-kilobase-pair DNA element from the chromosome during heterocyst differentiation of Anabaena sp. strain PCC 7120. *J Bacteriol* 1988;170:5034–5041.
73. Golden JW, Mulligan ME, Haselkorn R. Different recombination site specificity of two developmentally regulated genome rearrangements. *Nature* 1987;327:526–529.
74. Carrasco CD, Buettner JA, Golden JW. Programmed DNA rearrangement of a cyanobacterial hupL gene in heterocysts. *Proc Natl Acad Sci USA* 1995;92:791–795.
75. Thiel T, Lyons EM, Erker JC, Ernst A. A second nitrogenase in vegetative cells of a heterocyst-forming cyanobacterium. *Proc Natl Acad Sci USA* 1995;92:9358–9362.
76. Vernon SA, Pratte BS, Thiel T. Role of the nifB1 and nifB2 Promoters in cell-type-specific expression of two Mo nitrogenases in the cyanobacterium Anabaena variabilis ATCC 29413. *J Bacteriol* 2017;199.
77. Valladares A, Herrero A, Pils D, Schmetterer G, Flores E. Cytochrome c oxidase genes required for nitrogenase activity and diazotrophic growth in Anabaena sp. PCC 7120. *Mol Microbiol* 2003;47:1239–1249.
78. Thiel T. Characterization of genes for an alternative nitrogenase in the cyanobacterium Anabaena variabilis. *J Bacteriol* 1993;175:6276–6286.
79. Pratte BS, Sheridan R, James JA, Thiel T. Regulation of V-nitrogenase genes in Anabaena variabilis by RNA processing and by dual repressors. *Mol Microbiol* 2013;88:413–424.
80. Darnajoux R, Zhang X, McRose DL, Miadlikowska J, Lutzoni F, *et al*. Biological nitrogen fixation by alternative nitrogenases in boreal cyanolichens: Importance of molybdenum availability and implications for current biological nitrogen fixation estimates. *New Phytologist* 2016.
81. Darnajoux R, Magain N, Renaudin M, Lutzoni F, Bellenger J-P *et al*. Molybdenum threshold for ecosystem scale alternative vanadium nitrogenase activity in boreal forests. *Proc Natl Acad Sci USA* 2019;116:24682–24688.
82. Pratte BS, Thiel T. High-affinity vanadate transport system in the cyanobacterium Anabaena variabilis ATCC 29413. *J Bacteriol* 2006;188:464–468.
83. Thiel T, Pratte B. Regulation of three nitrogenase gene clusters in the cyanobacterium Anabaena variabilis ATCC 29413. *Life* 2014;4:944–967.
84. Kumar K, Ota M, Taton A, Golden JW. Excision of the 59-kb fdxN DNA element is required for transcription of the nifD gene in Anabaena PCC 7120 heterocysts. *N Z J Bot* 2018;57:76–92.
85. Pratte BS, Eplin K, Thiel T. Cross-functionality of nitrogenase components NifH1 and VnfH in Anabaena variabilis. *J Bacteriol* 2006;188:5806–5811.

86. Ekman M, Picossi S, Campbell EL, Meeks JC, Flores E. A Nostoc punctiforme sugar transporter necessary to establish a cyanobacterium-plant symbiosis. *Plant Physiol* 2013;161:1984–1992.
87. Ungerer JL, Pratte BS, Thiel T. Regulation of fructose transport and its effect on fructose toxicity in *Anabaena* spp. *J Bacteriol* 2008;190:8115–8125.
88. Campbell EL, Wong FC, Meeks JC. DNA binding properties of the HrmR protein of Nostoc punctiforme responsible for transcriptional regulation of genes involved in the differentiation of hormogonia. *Mol Microbiol* 2003;47:573–582.
89. Wuichet K, Zhulin IB. Molecular evolution of sensory domains in cyanobacterial chemoreceptors. *Trends Microbiol* 2003;11:200–203.
90. Campbell EL, Hagen KD, Chen R, Risser DD, Ferreira DP, *et al*. Genetic analysis reveals the identity of the photoreceptor for phototaxis in hormogonium filaments of Nostoc punctiforme. *J Bacteriol* 2015;197:782–791.
91. Risser DD, Meeks JC. Comparative transcriptomics with a motility-deficient mutant leads to identification of a novel polysaccharide secretion system in *Nostoc punctiforme. Mol Microbiol, Comparative Study Research Support, US Gov't, Non-PHS* 2013;87:884–893.
92. Risser DD, Chew WG, Meeks JC. Genetic characterization of the hmp locus, a chemotaxis-like gene cluster that regulates hormogonium development and motility in Nostoc punctiforme. *Mol Microbiol* 2014;92:222–233.
93. Khayatan B, Meeks JC, Risser DD. Evidence that a modified type IV pilus-like system powers gliding motility and polysaccharide secretion in filamentous cyanobacteria. *Mol Microbiol* 2015;98:1021–1036.
94. Bhaya D, Bianco NR, Bryant D, Grossman A. Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp. *Mol Microbiol* 2000;37:941–951.
95. Duggan PS, Gottardello P, Adams DG. Molecular analysis of genes in *Nostoc punctiforme* involved in pilus biogenesis and plant infection. *J Bacteriol* 2007;189:4547–4551.
96. Johansson C, Bergman B. Reconstitution of the symbiosis of *Gunnera manicata* Linden: cyanobacterial specificity. *New Phytol* 1994;126:643–652.
97. Becking JH. (n.d.) Ecology and physiological adaptations of *Anabaena* in the *Azolla-Anabaena azollae* symbiosis. *Ecological Bulletins*;1978:266–281.
98. Splitt SD, Risser DD. The non-metabolizable sucrose analog sucralose is a potent inhibitor of hormogonium differentiation in the filamentous cyanobacterium Nostoc punctiforme. *Arch Microbiol* 2016;198:137–147.
99. Liaimer A, Jenke-Kodama H, Ishida K, Hinrichs K, Stangeland J, *et al*. A polyketide interferes with cellular differentiation in the symbiotic cyanobacterium Nostoc punctiforme. *Environ Microbiol Rep* 2011;3:550–558.
100. Thiel T. Nitrogen fixation in heterocyst-forming cyanobacteria. In: Klipp W, Masepohl B, Gallon J, Newton W (editors). *Genetics and Regulation of Nitrogen Fixing Bacteria*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2004. pp 73–110.
101. Komárek J, Anagnostidis K. Modern approach to the classification system of Cyanophytes 4 - Nostocales. *Algological Studies/Archiv fu r Hydrobiologie, Supplement Volumes* 1989;56:247–345.
102. Will SE, Henke P, Boedeker C, Huang S, Brinkmann H, *et al*. Day and night: Metabolic profiles and evolutionary relationships of six axenic non-marine cyanobacteria. *Genome Biol Evol* 2019;11:270–294.
103. Qiao Q, Huang Y, Qi J, Qu M, Jiang C, *et al*. The genome and transcriptome of Trichormus sp. NMC-1: Insights into adaptation to extreme environments on the Qinghai-Tibet plateau. *Sci Rep* 2016;6:29404.
104. Peters GA, Mayne BC. The *Azolla*, *Anabaena azollae* relationship: I. Initial characterization of the association. *Plant Physiol* 1974;53:813–819.
105. Meeks JC, Enderlin CS, Joseph CM, Chapman JS, Lollar MW. Fixation of [(13)N]N2 and transfer of fixed nitrogen in the Anthoceros-Nostoc symbiotic association. *Planta* 1985;164:406–414.
106. Enderlin CS, Meeks JC. Pure culture and reconstitution of the Anthoceros-Nostoc symbiotic association. *Planta* 1983;158:157–165.
107. Braun-Howland EB, Lindblad P, Nierzwicki-Bauer SA, Bergman B. Dinitrogenase reductase of nitrogenase in the cyanobacterial symbionts of three *Azolla* species: localization and sequence of appearance during heterocyst differentiation. *Planta* 1988;176:319–332.
108. Ekman M, Tollback P, Bergman B. Proteomic analysis of the cyanobacterium of the *Azolla* symbiosis: identity, adaptation, and NifH modification. *J Exp Bot* 2008;59:1023–1034.
109. Osanai T, Kanesaki Y, Nakano T, Takahashi H, Asayama M, *et al*. Positive regulation of sugar catabolic pathways in the cyanobacterium Synechocystis sp. PCC 6803 by the group 2 sigma factor Sige. *J Biol Chem* 2005;280:30653–30659.
110. Osanai T, Imamura S, Asayama M, Shirai M, Suzuki I, *et al*. Nitrogen induction of sugar catabolic gene expression in Synechocystis sp. PCC 6803. *DNA Res* 2006;13:185–195.
111. Nieves-Morión M, Flores E. Multiple ABC glucoside transporters mediate sugar-stimulated growth in the heterocyst-forming cyanobacterium Anabaena sp. strain PCC 7120. *Environ Microbiol Rep* 2018;10:40–48.
112. Nelson JM, Hauser DA, Gudiño JA, Guadalupe YA, Meeks JC, *et al*. Complete genomes of symbiotic cyanobacteria clarify the evolution of vanadium-nitrogenase. *Genome Biol Evol* 2019;11:1959–1964.
113. Campbell EL, Meeks JC. Characteristics of hormogonia formation by symbiotic Nostoc spp. in response to the presence of Anthoceros punctatus or Its extracellular products. *Appl Environ Microbiol* 1989;55:125–131.
114. Risser DD, Meeks JC. Comparative transcriptomics with a motility-deficient mutant leads to identification of a novel polysaccharide secretion system in Nostoc punctiforme. *Mol Microbiol* 2013;87:884–893.
115. Campbell EL, Hagen KD, Cohen MF, Summers ML, Meeks JC. The *devR* gene product is characteristic of receivers of two-component regulatory systems and is essential for heterocyst development in the filamentous cyanobacterium Nostoc sp. strain ATCC 29133. *J Bacteriol* 1996;178:2037–2043.
116. Castenholz RW, Wilmotte A, Herdman M, Rippka R, Waterbury JB, *et al*. Phylum BX. Cyanobacteria. Boone D, Castenholz R and Garrity G (eds). In: *Bergey's Manual® of Systematic Bacteriology: Volume One: The Archaea and the Deeply Branching and Phototrophic Bacteria*. New York: Springer; 2001. pp. 473–599.
117. Rajaniemi P, Hrouzek P, Kastovska K, Willame R, Rantala A, *et al*. Phylogenetic and morphological evaluation of the genera Anabaena, Aphanizomenon, Trichormus and Nostoc (Nostocales, Cyanobacteria). *Int J Syst Evol Microbiol* 2005;55:11–26.
118. Arad H, Keysari A, Tel-or E, Kobiler D. A comparison between cell antigens in different isolates of *Anabaena azollae*. *Symbiosis* 1985;1:195–203.
119. Thiel T, Pratte BS, Zhong J, Goodwin L, Copeland A, *et al*. Complete genome sequence of Anabaena variabilis ATCC 29413. *Stand Genomic Sci* 2014;9:562–573.
120. Meeks JC, Elhai J, Thiel T, Potts M, Larimer F *et al*. An overview of the genome of Nostoc punctiforme, a multicellular, symbiotic cyanobacterium. *Photosynth Res* 2001;70:85–106.
121. Dvornyk V, Nevo E. Evidence for multiple lateral transfers of the circadian clock cluster in filamentous heterocystic cyanobacteria Nostocaceae. *J Mol Evol* 2004;58:341–347.
122. Dvornyk V, Nevo E. Genetic polymorphism of cyanobacteria under permanent natural stress: A lesson from the "Evolution Canyons". *Res Microbiol* 2003;154:79–84.
123. Seo P-. S, Yokota A. The phylogenetic relationships of cyanobacteria inferred from 16S rRNA, *gyrB*, *rpoC1* and *rpoD1* gene sequences. *J Gen Appl Microbiol* 2003;49:191–203.