Genomic Changes Associated with the Evolutionary Transitions of *Nostoc* to a Plant Symbiont

Denis Warshan,[1] Anton Liaimer,[2] Eric Pederson,[1] Sea-Yong Kim,[1] Nicole Shapiro,[3] Tanja Woyke,[3] Bjørn Altermark,[4] Katharina Pawlowski,[1] Philip D. Weyman,[5] Christopher L. Dupont,[6] and Ulla Rasmussen*,[1]

[1]Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden

[2]Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics, UiT–The Arctic University of Norway, Tromsø, Norway

[3]US Department of Energy Joint Genome Institute, Walnut Creek, CA

[4]Department of Chemistry, Faculty of Science and Technology, UiT–The Arctic University of Norway, Tromsø, Norway

[5]Department of Synthetic Biology and Bioenergy, J. Craig Venter Institute, La Jolla, CA

[6]Department of Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA

*Corresponding author: E-mail: ulla.rasmussen@su.se.

Associate editor: Nicole Perna

The data have been deposited in https://gold.jgi.doe.gov/studies?id=Gs0110198, and GenBank with accession numbers NNBT00000000, NNBU00000000, NNBV00000000, and LSSA00000000.

Article

Abstract

Cyanobacteria belonging to the genus *Nostoc* comprise free-living strains and also facultative plant symbionts. Symbiotic strains can enter into symbiosis with taxonomically diverse range of host plants. Little is known about genomic changes associated with evolutionary transition of *Nostoc* from free-living to plant symbiont. Here, we compared the genomes derived from 11 symbiotic *Nostoc* strains isolated from different host plants and infer phylogenetic relationships between strains. Phylogenetic reconstructions of 89 Nostocales showed that symbiotic *Nostoc* strains with a broad host range, entering epiphytic and intracellular or extracellular endophytic interactions, form a monophyletic clade indicating a common evolutionary history. A polyphyletic origin was found for *Nostoc* strains which enter only extracellular symbioses, and inference of transfer events implied that this trait was likely acquired several times in the evolution of the Nostocales. Symbiotic *Nostoc* strains showed enriched functions in transport and metabolism of organic sulfur, chemotaxis and motility, as well as the uptake of phosphate, branched-chain amino acids, and ammonium. The genomes of the intracellular clade differ from that of other *Nostoc* strains, with a gain/enrichment of genes encoding proteins to generate L-methionine from sulfite and pathways for the degradation of the plant metabolites vanillin and vanillate, and of the macromolecule xylan present in plant cell walls. These compounds could function as C-sources for members of the intracellular clade. Molecular clock analysis indicated that the intracellular clade emerged ca. 600 Ma, suggesting that intracellular *Nostoc* symbioses predate the origin of land plants and the emergence of their extant hosts.

Key words: **cyanobacteria, symbiosis, evolution, plant–microbe interaction.**

Introduction

A primary endosymbiotic event between a cyanobacterium and a unicellular eukaryote giving rise to photosynthetic eukaryotes dated to >1.2 Ba (Parfrey et al. 2011; Keeling 2013; Zimorski et al. 2014; Archibald 2015a; Sánchez-Baracaldo et al. 2017) is considered to be one of the most life-changing events in history. From a single or multiple events originated the Archaeplastida: red algae, glaucophyte, green algae, and their descendent land plants (Archibald 2009; Keeling 2013; Burki 2017). Later transfers of photosynthesis to other lineages by multiple independent endosymbiotic events led to the emergence of many photosynthetic protists and macroalgae (Gould et al. 2008; Archibald 2015b; Burki 2017). However, this was not the last symbiosis between cyanobacteria and eukaryotes (single-celled and multicellular). Extant cyanobacteria can enter into symbiosis with a highly diverse taxonomic host range distributed in both aquatic and terrestrial environments. For instance, in aquatic environments these symbioses can involve hosts, such as diatoms, corals, sponges, and even another case of cyanobacterial primary endosymbiosis with the rhizarian *Paulinella* (Carpenter and Foster 2002; Foster et al. 2006; Usher et al. 2007; Foster et al. 2011; Zehr et al. 2016). In terrestrial environments, they can involve fungi, *Geosiphon pyriformis* and lichens, as well as representatives from major lineages of the plant kingdom (Rikkinen 2017). The cyanobacterial host plants include taxa from the division Bryophyta (mosses, liverworts, and hornworts), Pteridophyta (aquatic

ferns of the genus *Azolla*), the Gymnosperm family Cycadaceae, and the Angiosperm family Gunneraceae (Rikkinen 2017).

The cyanobacteria can inhabit different plant organs or tissues either intracellularly or externally (reviewed in Santi et al. 2013). Cyanobacterial-plant symbioses are mutualistic where the benefit to the plant is reduced nitrogen (N) from cyanobacterial N_2-fixation, whereas the symbiotic cyanobacteria have a shelter and are supplied with reduced carbon and other nutrients by the host (Rai et al. 2000; Liaimer et al. 2016). The only symbiosis where a clear contribution of the host is still not demonstrated is the cyanobacterial-feathermoss symbiosis. Nevertheless, recent discoveries indicate that the feathermoss symbiosis is also mutualistic since it was shown that in symbiosis with cyanobacteria, the moss tissue is getting enriched in N over time (Bay et al. 2013), and that the cyanobacterial sulfur demands might be provided by the moss (Warshan et al. 2017). With the exception of the water fern *Azolla* (dated to 90 Ma; Metzgar et al. 2007) and the angiosperm *Gunnera* (dated to 115 Ma; Vekemans et al. 2012), other cyanobacterial host plants, specifically cycads (265–290 Ma; Brenner et al. 2003; Condamine et al. 2015) and bryophytes (470 Ma; Edwards et al. 2014; Laenen et al. 2014; Brown et al. 2015), are ancient lineages, implying potentially a long coevolutionary history between plants and cyanobacteria.

Symbiotic cyanobacteria have different types of physical interactions with their host plants. They may be extracellular, living epiphytically or in specialized compartments (leaf cavities of *Azolla*, cavities in the gametophytic thalli of liverworts/hornworts and in coralloid roots of cycads) but they can also be intracellularly localized in stem glands of *Gunnera* sp. (Santi et al. 2013). When compartmentalized intracellularly in *Gunnera* sp., the cyanobacteria are still separated from the plant cytosol by the plant plasmalemma (Bergman 2002). The intracellular *Gunnera-Nostoc* symbiosis is unique within the plant phylum but there is one other *Nostoc* intracellular symbiosis, the interaction between the glomeromycotinan fungus *Geosiphon pyriformis* and *Nostoc punctiforme* (Schüßler et al. 1994).

Almost all symbioses between cyanobacteria and plants are facultative with the exception of the obligate *Azolla* symbiosis (Rikkinen 2017) and the *G. pyriformis*/*N. punctiforme* symbiosis which is obligate for the fungus, but not for the cyanobacterium (Schüßler 2012). All plant-symbiotic cyanobacteria belong to the filamentous heterocystous cyanobacteria (Group IV and V; Rippka et al. 1979). *Nostoc* sp. is the main genus found in symbiosis with plants from all lineages, and to date, genomic information for symbiotic cyanobacteria were only available for the facultative *Nostoc punctiforme* ATCC 29133, isolated from the cycad *Macrozamia* sp., and *Nostoc azollae*, the obligate partner of *Azolla* sp. (Meeks et al. 2001; Ran et al. 2010). Interestingly, symbiosis reconstruction experiments with *N. punctiforme* ATCC 29133 showed that this strain has a broad host-range and can establish all facultative symbiotic relationships known between plants and cyanobacteria, ranging from the "loose" epiphytic symbiosis with feathermosses to the intracellular symbiosis with *Gunnera* sp. (table 1; Campbell and Meeks 1989; Johansson and Bergman 1994; Warshan et al. 2017). Moreover, phylogenetic inference of cyanobacteria associated with bryophytes (feathermosses, liverworts, and hornworts) and *Gunnera* species has shown that most of the cyanobacteria are closely related to *N. punctiforme* (Svenning et al. 2005), suggesting that a broad host-range is more common than thought previously.

Table 1. Host Range of *Nostoc* sp. and Symbiotic Localization.

Host	Symbiotic Tissue	Type of Symbiosis
Angiosperm		
Gunnera	Stem glands	Intracellular
Gymnosperms		
cycads	Collaroid roots	Extracellular
Pteridophytes		
Azolla	Leaf cavities	Extracellular
Bryophytes		
Mosses	Epiphytic	Extracellular
Liverworts	Slime cavities	Extracellular
Hornworts	Slime cavities	Extracellular
Fungi		
Geosiphon pyriformis	Bladder	Intracellular

We have recently isolated and sequenced the genome of five representative *Nostoc* strains isolated from the feathermosses *Hylocomium splendens* and *Pleurozium schreberi* and four strains from the liverwort *Blasia pusilla* (Liaimer et al. 2016; Warshan et al. 2017). Phylogenetic analysis of *Nostoc* symbionts of feathermosses revealed two sister lineages (Warshan et al. 2017). *Nostoc* strains comprising the first lineage showed high similarity at the genomic scale to the nonsymbiotic strain *Anabaena variabilis* ATCC 29413, whereas the second lineage was more related to *N. punctiforme* and the nonsymbiotic strain *Nostoc* sp. CALU 996 (Warshan et al. 2017). This suggests that the two divergent subclades within Nostocales have different host ranges, one restricted to epiphytic interactions with feathermosses and the other having a broader host range that includes both epiphytic and endophytic interactions. A set of questions arises: whether there are genomic differences associated with different host ranges, how these differences substantiated over the evolution of cyanobacterial-plant symbioses, and the functional gains and losses associated with these differences.

In this study, we compared sequenced genomes of nine facultative symbiotic *Nostoc* strains isolated from bryophytes (feathermosses and *B. pusilla*), one obligate symbiotic *Nostoc* strain (*N. azollae*) and one nonsymbiotic strain *Nostoc* sp. CALU 996. The close phylogenetic relationship between some *Nostoc* isolates from feathermosses and the *N.* CALU 996 enabled us to investigate the evolutionary history of symbiotic competence present within the genus *Nostoc*. We also investigated the genomic retention and enrichment patterns of these lineages. Functional annotations were conducted to identify the cyanobacterial biological processes and/or pathways likely associated with the evolution of cyanobacterial-plant symbioses. Based on these results, we were able to infer the emergence and evolutionary history of these symbioses and provide novel insights into the origin of symbiotic *Nostoc* strains.

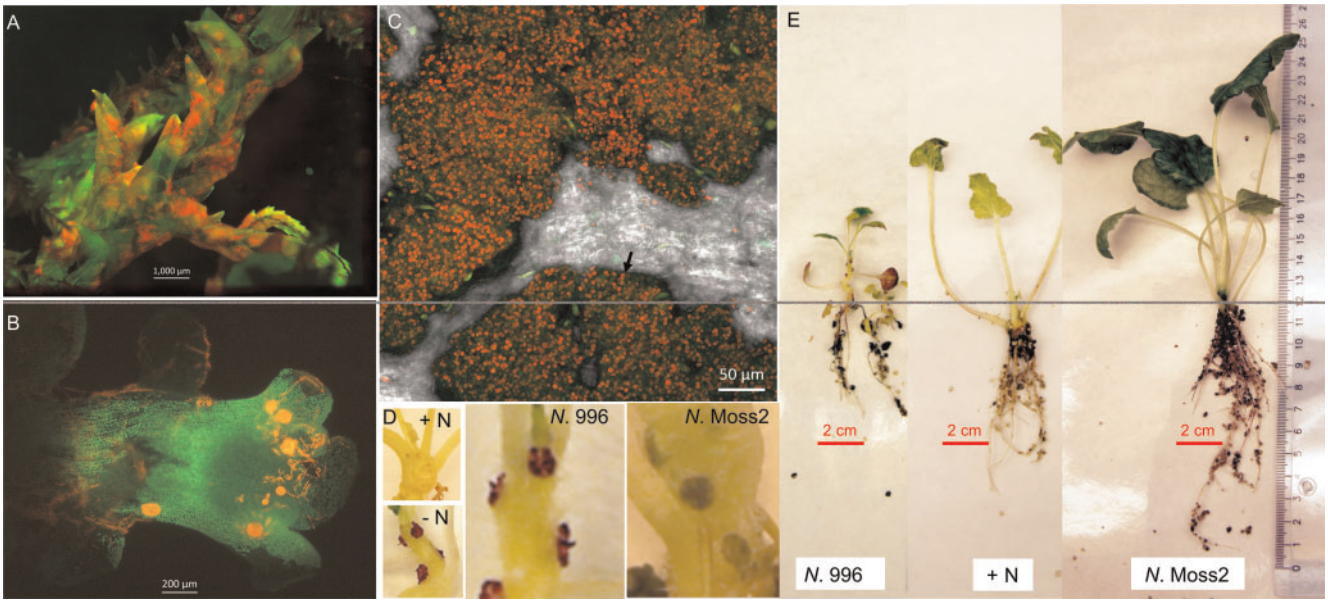

Fig. 1. Microscopy observation of multihost competent symbiotic *Nostoc* Moss2 colonization of (*A*) *Pleurozium schreberi* its original host where the cyanobacteria is epiphytic; (*B*) *Blasia pusilla* where the symbiont remains extracellular but enclosed in cavities and (*C*) *Gunnera manicata* where *N.* Moss2 (in red) is intracellularly located in the host's stem gland cells (black arrow indicate the plant cell wall). (*D*) *G. manicata* stem gland when treated with and without N addition, and in contact with *N.* CALU 996 and with *N.* Moss2 (green spot inside the gland). (*E*) Effect of *N.* CALU 996 and *N.* Moss2 inoculum on the growth of *G. manicata* compared with N addition (10 mM $NaNO_3$).

Results and Discussion

Analysis of *Nostoc* Isolates Originating from Different Host Plants Leads to the Identification of Two Host Specific Groups

To test if the symbiotic *Nostoc* isolates have a broad host-range, we performed reconstitution experiments with 11 *Nostoc* isolates and the hosts *Blasia pusilla* and *Pleurozium schreberi* (forming both extracellular symbioses) and *Gunnera manicata* (intracellular symbiosis). We found that all nine bryophyte derived isolates (i.e., those from feathermosses and *Blasia*) were capable of establishing an extracellular symbiosis with both *P. schreberi* and *B. pusilla* (fig. 1*A* and *B*; table 2). However, the capacity to form an intracellular symbiosis was only observed for *N. punctiforme*, the four isolates from *B. pusilla* and a single feathermoss isolate, *N.* Moss2 (fig. 1*C*–*E* and table 2). It is notable that all strains capable of establishing a symbiosis with *G. manicata* acted as beneficial symbionts, for instance *N.* Moss2 promoted the growth of the host at a greater extent than N addition (fig. 1*E*). The difference between the two groups of feathermoss isolates with regard to symbiotic competence with *G. manicata* suggests that the feathermoss symbionts encompass strains that can enter only extracellular symbioses, and strains that can be both intra- and extracellular symbionts.

Nostoc Strains Able to Form Both Extra- and Intracellular Symbioses Form a Monophyletic Clade

The phylogenetic analysis of the concatenated protein alignments using maximum likelihood grouped all *Nostoc* strains able to enter an intracellular symbiosis with *G. manicata* into a strongly supported monophyletic clade that diverged after the radiation of the *Anabaena*/*Nostoc* lineages (fig. 2). Since the difference in host specificity is consistent with the phylogenetic grouping, the clade containing strains able to establish extracellular symbioses with bryophytes as well as an intracellular symbiosis with *G. manicata* will be termed the "Intra/Extra clade," whereas the other clades of symbiotic *Nostoc* strains only able to enter extracellular symbioses will be termed "Extra clades" in this study. We found that *N. punctiforme* is basal within the Intra/Extra clade. To help to resolve this sister lineage, sequencing the genomes of additional *Nostoc* taxa isolated from cycads, which were shown to be closely related to *N. punctiforme* (Svenning et al. 2005; Gehringer et al. 2010), should be performed. Moreover, we found two strongly supported "Extra" clades, composed of the strains *N.* Moss3/Moss4 and *N.* Moss5/Moss6, respectively (fig. 2). The Extra I clade originated after the radiation between *Anabaena* and *Nostoc* and is a basal branch of the clade encompassing the *B. pusilla* isolates and the *N.* Moss2 isolate. This could indicate that strains *N.* Moss3 and *N.* Moss4 might be in a transitional stage, expanding their host range to include intracellular symbioses. The Extra II clade comprises the free-living strains *Anabaena variabilis* ATCC 29413 and *Nostoc* sp. PCC 7120, and is phylogenetically divergent from all other symbiotic *Nostoc* clades (fig. 2). This result implies phenotypic convergent evolution via gene gain of feathermoss symbionts, and that they are more genetically diverse than the symbionts of other host plants. *Nostoc azollae* belongs to another intermixed lineage, composed of nonsymbiotic *Anabaena*/*Aphanizomenon* strains, and does not fall into any of the symbiotic clades identified in this study (fig. 2). This is consistent with previous results on the phylogeny of *N. azollae* (Ran et al. 2010; Larsson et al. 2011). Likewise, the

Table 2. *Nostoc* Isolates, Their Origin of Isolation, and Their Symbiotic Competence.

Taxon	Plant Host	*Pleurozium schreberi*	*Blasia pusilla*	*Gunnera manicata*
***Nostoc punctiforme* ATCC 29133**	***Macrozamia* sp.**	+	+	+
***Nostoc* sp. Moss2**	***P. schreberi*[a]**	+	+	+
***Nostoc* sp. Moss3**	***P. schreberi*[a]**	+	+	−
***Nostoc* sp. Moss4**	***P. schreberi*[a]**	+	+	−
***Nostoc* sp. Moss5**	***Hylocomium splendens*[a]**	+	+	−
***Nostoc* sp. Moss6**	***H. splendens*[a]**	+	+	−
***Nostoc* sp. KVJ2**	***B. pusilla*[b]**	+	+	+
***Nostoc* sp. KVJ10**	***B. pusilla*[b]**	+	+	+
***Nostoc* sp. KVJ20**	***B. pusilla*[b]**	+	+	+
***Nostoc* sp. KVS11**	***B. pusilla*[b]**	+	+	+
***Nostoc* sp. CALU 996**	**Free-living Guinea, Central Africa**	−	−	−

[a]Feathermosses collected in Ruttjeheden, Reivo forest, Northern Sweden (65°80′N–19°10′E).
[b]*Blasia pusilla* collected at Kvaløya island (69.64°N 18.73°E), Northern Norway.

relationship between *N. azollae* and *Cylindrospermopsis raciborskii* CS-505 as well as *Raphidiopsis brookii* D9, two cyanobacterial strains which have undergone genome reduction similar to that of *N. azollae*, is consistent with previous results (Ran et al. 2010; Stucken et al. 2010). The polyphyly of four cyanobacterial clades forming symbioses with plants, suggests that this capability has evolved at least four times in the Nostocales order, once in an ancestor of *N. azollae*, in the two Extra clades, and in the Intra/Extra clade.

Symbiotic *Nostoc* Strains Share a Small Number of Genes Missing in the Nonsymbiotic Strain

The comparison of gene content of the nonsymbiotic strain *N.* CALU 996 and the three groups of symbiotic *Nostoc* strains (Extra I, Extra II, and Intra/Extra) as well as the obligate symbiont *N. azollae* was conducted to identify genes potentially required for a symbiosis (fig. 3). Using a gene flux analysis, we inferred which gene families were gained/expanded and lost/contracted along a species tree composed of symbiotic and nonsymbiotic *Nostoc* strains (fig. 3). In agreement with previous studies (Ran et al. 2010; Larsson et al. 2011), this analysis revealed that *N. azollae* underwent extensive gene loss after the split from other *Nostoc* strains (fig. 3*A* and *B*). Our analysis predicted that 610/35 gene families were lost/contracted after the divergence of *N. azollae* and 178/57 were gained/expanded after the split. Among the gene families lost by *N. azollae*, 28 were retained in the other symbiotic *Nostoc* strains (fig. 3*B*). COG functional categorization of the gene families lost in *N. azollae* but retained in other symbiotic strains revealed that the majority belongs to the COG categories "General function prediction only," "Inorganic ion transport and metabolism," "Cell motility," and "Coenzyme transport and metabolism" (fig. 3*B*). Specifically, gene families lost by the obligate symbiont *N. azollae* and retained in other symbiotic strains were dominated by genes encoding proteins involved in chemotaxis and organic sulfur (alkane and aliphatic sulfonate) transport and metabolism. A relatively small set of 178 gene families (6.3% of the total gene families found in the genome of *N. azollae*) were uniquely gained by *N. azollae* and are not present in the other symbiotic *Nostoc* strains. They encode proteins involved in "Defense mechanisms," "Function unknown," "Transcription," "Translation, ribosomal structure and biogenesis," and "Carbohydrate transport and metabolism" (fig. 3*B*). Whether any of the genes gained by *N. azollae* are involved in the interaction with *Azolla* remains to be investigated. The nonsymbiotic strain *N.* CALU 996 lost/contracted 268/35 gene families after diverging from the symbiotic *Nostoc* strains (fig. 3*A* and *C*). Among those gene families, 170 (55%) were retained in the genomes of symbiotic strains excluding *N. azollae* (fig. 3*C*). COG functional analysis of those gene families reveals that they are mainly categorized in "Inorganic ion transport and metabolism," "General function prediction only," and "Signal transduction mechanisms" (fig. 3*C*), indicating that a large part of the proteins putatively involved in cyanobacterial symbioses are only loosely classified functionally. The relatively small set of genes not retained in the nonsymbiotic strain *N.* CALU 996 suggest that only a few molecular pathways are involved in the establishment and maintenance of cyanobacterial-plant symbiosis.

The last common ancestor (LCA) of the Intra/Extra *Nostoc* strains gained/expanded 171/41 gene families which represent a small part of the total number of gene families in the LCA's genome (4,997 gene families) (fig. 3*D*). Those gene families gained by the Intra/Extra LCA were involved in "General function prediction only," "Cell wall/membrane/envelope biogenesis" (mainly in exopolysaccharide modification), "Secondary metabolites biosynthesis, transport and catabolism," and "Signal transduction mechanisms," suggesting an adaptation to the intracellular lifestyle by gaining sensing mechanisms and the ability to modify the plant-derived cell wall-like matrix in which intracellular symbionts are embedded.

There is a relationship between genomic signatures and the nature of a symbiosis where some obligate symbiotic, parasitic or commensal microorganisms underwent genome streamlining due to mechanisms such as low effective population size causing genetic drift which can be materialized by reduction to a core metabolic repertoire, pseudogenization, and low number of duplicated genes (Giovannoni et al. 2014). Interestingly, *N. azollae* is the only clear example of a plant symbiotic cyanobacteria genome that underwent genome streamlining (Ran et al. 2010). By contrast the genomes of the other plant symbiotic *Nostoc* strains are among the largest

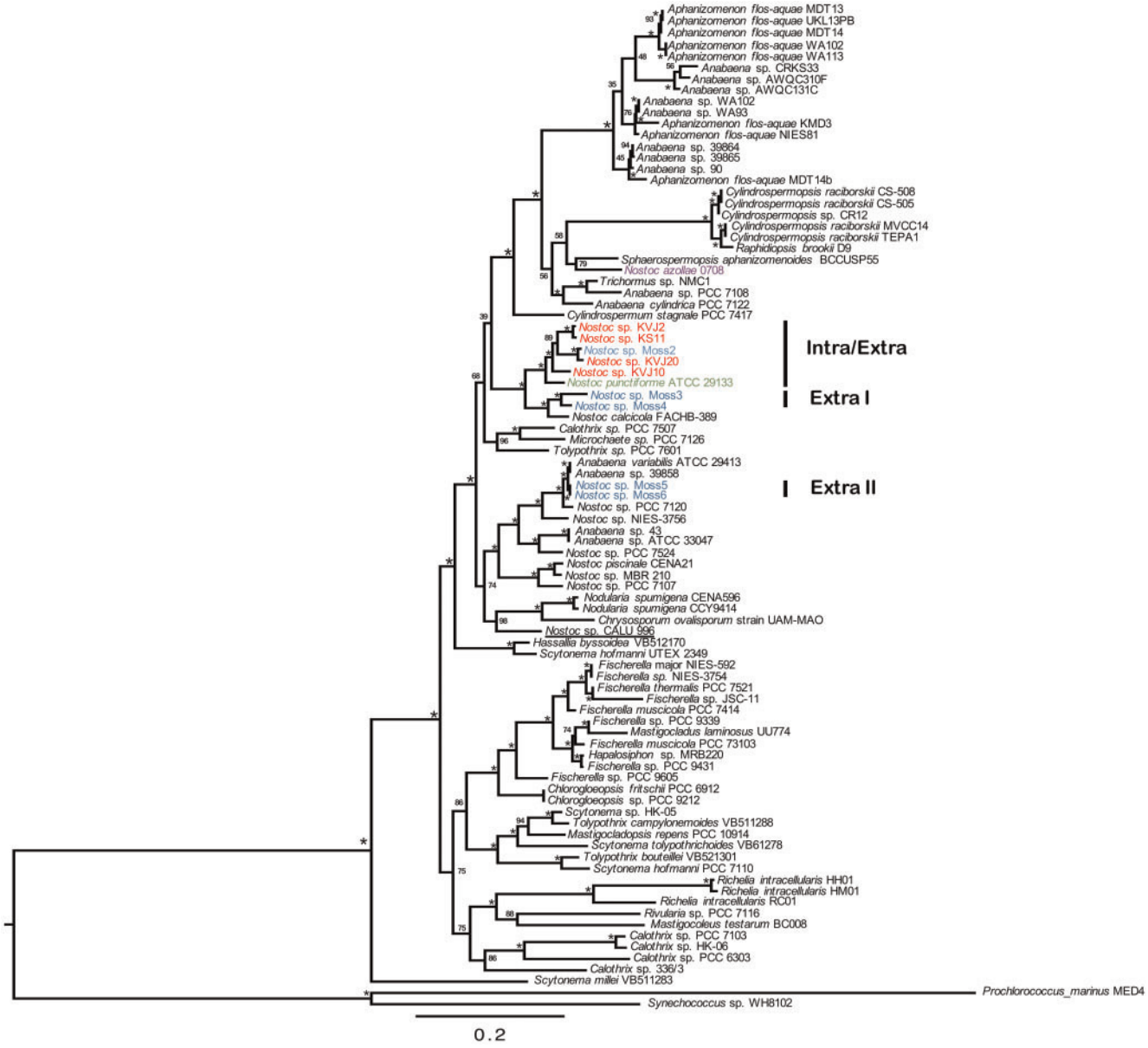

Fig. 2. Genome-wide phylogeny of the order Nostocales including symbiotic *Nostoc* strains. Tree topology and bootstrap support values (calculated from 100 resampling) were inferred with RaxML using 65 concatenated proteins for 89 species of cyanobacteria (18,440 positions). *Nostoc* strains isolated from feathermosses are highlighted in blue, in red from *Blasia pusilla* isolates, in green from cycad and in purple the obligate symbiont *N. azollae* from *Azolla filiculoides*. The symbiotic incompetent *N.* CALU 996 is underlined. Values next to the nodes show bootstrap support values, * indicates a bootstrap value of 100. The genomes of *Prochlorococcus marinus* MED4 and *Synechococcus* sp. WH8102 were included as outgroups. Clades of extracellular-exclusive *Nostoc* isolates (Extra I and II) and the clade formed by intracellular competent strains (Intra/Extra) are shown.

Nostocales genomes, ranging from 7.1–10.4 Mb (supplementary table S1, Supplementary Material online). Larsson et al. (2011) showed that the genome of *N. punctiforme* is expanding compared with the estimated genome size of the most recent common ancestor of cyanobacteria, and observed that *N. punctiforme* has accumulated more gene duplication events than most cyanobacteria. Consistently, the gene flux analysis revealed that facultative symbiotic cyanobacteria seems to have underwent genome expansion and have been accumulating more gene family gains than losses (fig. 3*A*). Altogether these results indicate that genome expansion in cyanobacteria might be associated with facultative symbiosis.

A Set of Transporters Is Shared in Symbiotic *Nostoc* Strains

To obtain insights into possible function conserved in facultative symbiotic *Nostoc* able to enter extra- or intracellular symbioses, we investigated the 170 gene families retained and/or expanded in the genome of facultative strains compared with the nonsymbiotic strain *N.* CALU 996 (237 genes)

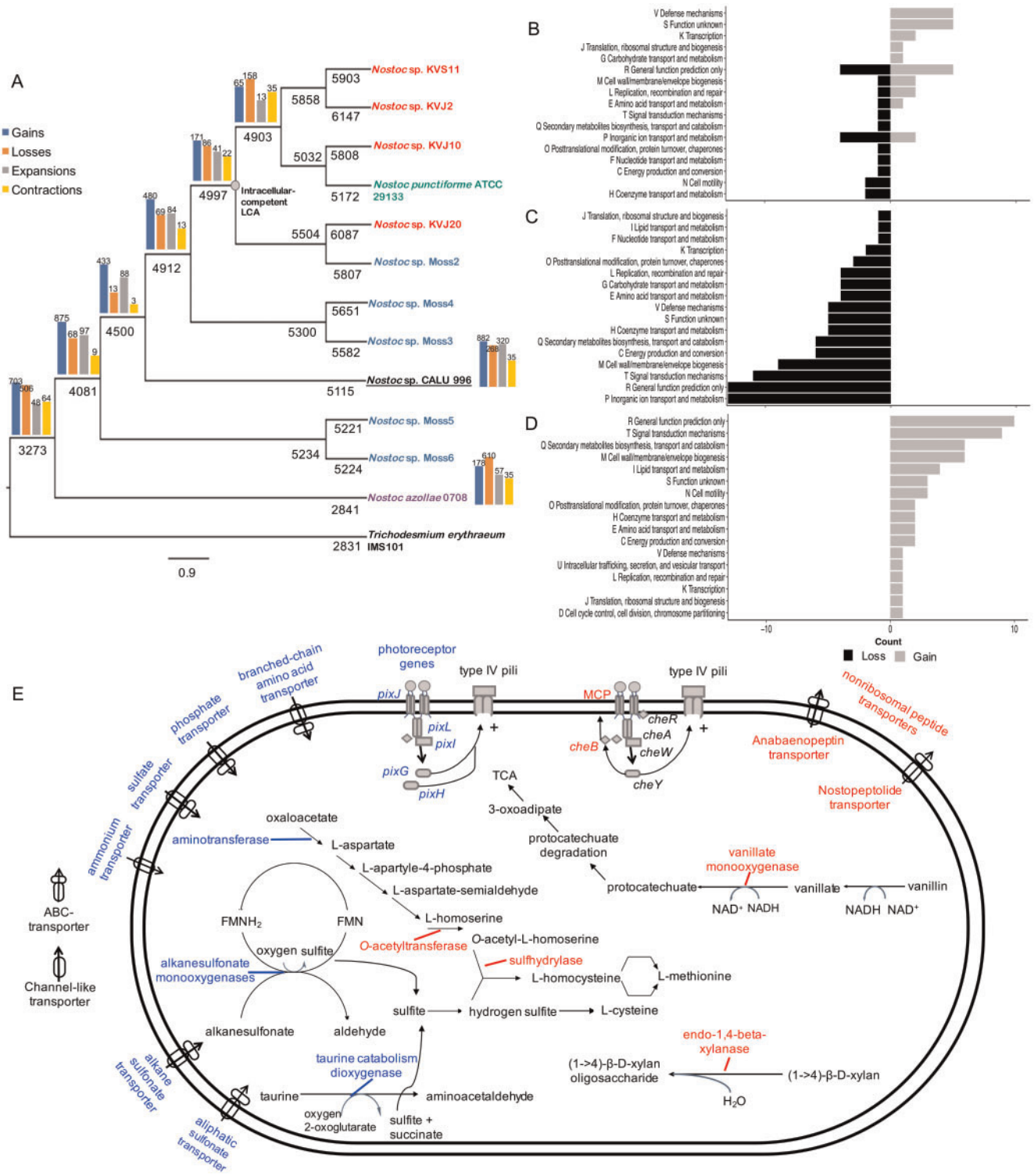

Fig. 3. Ancestral genome content reconstruction and functional annotation. (*A*) The reconstruction of ancestral genome content based on phylogenetic inference of 11 genomes of symbiotic *Nostoc* (feathermoss isolates in blue, *Blasia* isolates in red, *N. punctiforme* in green and *N. azollae* in purple), the genome of the symbiotic incompetent *N.* CALU 996 (underlined) and the genome of *Trichodesmium erythraeum* MS1 as outgroup. The numbers of reconstructed gain/loss and expansion/contraction events of selected lineage are represented by a bar plot (numbers are expressed in log10). Numbers next to the branches are predicted gene numbers for ancestral nodes and observed gene numbers for extant lineages. Functional classification of (*B*) gene gained in *N. azollae* and gene families lost in this strain but retained in symbiotic *Nostoc*, (*C*) gene families lost in *N.* CALU 996 and retained in symbiotic *Nostoc* (*D*) gene families gained in the lowest common ancestor of intracellular competent *Nostoc*. Numbers indicate count, positive value indicates gain of gene families and negative value indicates loss of gene families. (*E*) Overview of the enriched pathway in all facultative symbiotic *Nostoc* (blue) and specifically gained by the LCA of intracellular competent *Nostoc* (red). FMN, flavin mononucleotide; MCP, modular cyanobacteriochrome–methyl-accepting chemotaxis protein; TCA, tricarboxylic acid cycle.

and the 212 gene families (231 genes) specifically gained and/or expanded in the intracellular competent LCA (fig. 3*E*). The genomes of all ten symbiotic *Nostoc* strains were enriched in gene families encoding substrate-binding periplasmic proteins associated with transport of aliphatic and alkane sulfonates, sulfate, phosphate, branched-chain amino acids and a gene encoding an ammonium importer (fig. 3*E*). The acquisition of these additional transporters suggests an adaptation to the environment offered by the host. Deficiency in micro- and macronutrients is known to limit nitrogen fixation activity in terrestrial ecosystems (Barron et al. 2009; Vitousek et al. 2013) and in at least one study, phosphorus was shown to be a limiting factor for nitrogen fixation activities of cyanobacteria associated with feathermosses (Rousk et al. 2017). Several nutrients are needed in sufficient quantities in the host environment to sustain cyanobacterial N_2-fixation. Additional support for the role of genes encoding sulfate and phosphate transporters in the *Nostoc*-feathermoss symbiosis are the earlier findings that the transcript abundance of such genes was higher in the strains *N.* Moss2 and *N. punctiforme* when in symbiosis with the feathermoss *P. schreberi* (Warshan et al. 2017).

A gene family encoding substrate-binding periplasmic proteins associated with transport of branched-chain amino acids (BCAAs) was found to be conserved among all symbiotic *Nostoc* strains but missing in *N.* CALU 996 (fig. 3*E*). To date, no results have been reported on the role of BCAA transporters in cyanobacterial-plant symbioses, yet it is known that in legume root nodules induced by rhizobia, the plant supplies the intracellular symbiotic bacteria (bacteroids) with BCAAs while rhizobial BCAA biosynthesis is downregulated (Prell et al. 2010). Transcriptome analyses of *N. punctiforme* and strain *N.* Moss2 shows that BCAA biosynthesis is down-regulated during symbiosis (supplementary fig. S1, Supplementary Material online; Warshan et al. 2017). This feature seems to be shared between two independently evolved symbioses between plants and nitrogen-fixing prokaryotes, feathermoss-cyanobacterial and rhizobial symbioses. This might be due to the fact that BCAA biosynthesis can work as an electron sink (Shimizu et al. 2010) while nitrogenase activity requires a low redox potential as nitrogenase is supplied with electrons by ferredoxin.

The *amtB* gene, which encodes an ammonium importer, was also found to be conserved among all facultative symbiotic cyanobacteria (fig. 3*E*). In *Nostoc* PCC 7120, the ammonium importer genes are organized in a cluster of three *amt* genes (*amt1*, *amt4*, and *amtB*) that are transcribed independently (Paz-Yepes et al. 2008). Under N deprivation, *amtB* gene expression was lower than *amt1* expression (Paz-Yepes et al. 2007, 2008). These results suggest that the *amtB* might be regulated in response to conditions other than N starvation; it might also be specifically regulated during the symbiosis with plants.

Interestingly, we also found a gene family encoding transporters of nonribosomal peptides, such as nostopeptolides and anabaenopeptins, which was specifically expanded in the LCA of the Intra/Extra clade (fig. 3*E*). This result is consistent with a role of these secondary metabolites in cyanobacterial-plant symbioses, and more specifically in the establishment of an intracellular symbiosis with *Gunnera* sp. It was shown that nostopeptolides produced by *N. punctiforme* are tightly connected to its life cycle as regulators of the motile stage (Liaimer et al. 2011, 2015); closely related nostocyclopeptides produced by the *Blasia* isolates used in this study are also suggested to have this function (Liaimer et al. 2016). Cyanobacteria that form symbioses with plants require the differentiation of vegetative filaments into a transient motile stage, hormogonia, in order to colonize the host (Rai et al. 2000). Nostopeptolides were suggested as hormogonium-repressing factors or as chemoattractants, depending on their extracellular concentration (Liaimer et al. 2015). The expression of nostopeptolide gene cluster was strictly downregulated in symbiosis with *B. pusilla*, and the peptides content was reduced in symbiosis with *G. manicata* whereas the production of other metabolites was up-regulated (Liaimer et al. 2015).

Shared and Specific Pathways for Organic Sulfur Metabolism

All facultative symbiotic *Nostoc* strains were enriched in gene families involved in transport of alkane sulfonate and other aliphatic sulfonates but also retained pathways specific to the assimilatory catabolism of these compounds to generate sulfite (fig. 3*E*). Alkane sulfonate monooxygenases and a taurine catabolism dioxygenase were specifically retained in the genomes of all facultative symbiotic *Nostoc* strains, allowing the anabolic incorporation of sulfonate-derived sulfite into L-cysteine (fig. 3*E*). The putative importance of these functions for the symbiosis is underlined by the fact that gene families of aliphatic sulfonate transporters and monooxygenases were found to be highly transcriptionally upregulated in *N. punctiforme* and strain *N.* Moss2 when in symbiosis with feathermoss (Warshan et al. 2017). This may indicate a parallel with the symbiosis between a marine *Roseobacter* clade bacterium and the diatom *Thalassiosira pseudonana* where an aliphatic sulfonate is supplied by the diatom to the bacterium (Durham et al. 2015). So, the feathermosses might transfer aliphatic sulfonates to their cyanobacterial symbionts in exchange for fixed N, and this exchange might be shared by other eukaryote–prokaryote symbioses, although with a different purpose—the *Roseobacter* clade bacterium uses the aliphatic sulfonate as carbon source for respiration, not as source of sulfur. Interestingly, the LCA of the Intra/Extra clade specifically gained genes encoding proteins involved in the assimilation of sulfite, a homoserine-*O*-acetyltransferase and an *O*-acetyl-L-homoserine sulfhydrylase to synthetize L-methionine from hydrogen sulfide and *O*-acetyl-L-homoserine (fig. 3*E*). This is another parallel with legume/rhizobia symbioses, where methionine biosynthesis has been shown to be essential for *Sinorhizobium meliloti* to survive in root nodules (Campbell et al. 2006; Taga et al. 2007; Taga and Walker 2010).

Phototaxis, Chemotaxis, and Motility Pathways

All facultative symbiotic *Nostoc* strains retained the *pix* locus which contains phototaxis and motility genes (fig. 3*E*). There is evidence suggesting that cyanobacterial interactions with

their host plants are mediated by chemotaxis-related signal transduction systems (Duggan et al. 2013; Campbell et al. 2015; Warshan et al. 2017). For the feathermoss-cyanobacteria symbiosis the *pix* locus and six other loci containing chemotaxis and motility-related genes were shown to be upregulated during the establishment and/or maintenance of the symbiosis (Warshan et al. 2017). Moreover, the *pix* genes were induced during both the early phase when chemical signals are exchanged between the partners, and during the later phase when the cyanobacteria are physically associated with *P. schreberi* (Warshan et al. 2017). The LCA of the Intra/Extra clade gained one gene family encoding a modular cyanobacteriochrome–methyl-accepting chemotaxis protein (MCP) and expanded another MCP gene family and a *cheB*–like gene family which catalyzes the demethylation of specific methylglutamate residues introduced into chemoreceptors by the chemotaxis protein methyltransferase CheR (fig. 3*E*; Campbell et al. 2015). Deletion of those two genes as well as other linked *cheAWR* and MCP genes from the genome of *N. punctiforme* showed that this locus is not involved in either hormogonium differentiation, motility, or in phototaxy (Campbell et al. 2015). This result together with our findings suggests that this locus might be specifically involved in the establishment or maintenance of a cyanobacterial-plant symbiosis.

The LCA of the Intra/Extra Clade Gained/Enriched Pathways for the Degradation of Vanillin and Hemicellulose

Among the gene families specifically enriched by the LCA of the Intra/Extra clade were genes involved in the degradation of vanillin, which is a phenolic compound commonly found in plant exudates (Li et al. 2010). All members of the Intra/Extra clade contain genes encoding enzymes in the pathway for the degradation of vanillin and vanillate (fig. 3*E*). The vanillate monooxygenase which generates protocatechuate from vanillate was gained by the LCA. In addition, the complete pathway for the degradation of protocatechuate to 3-oxoadipate is present in all symbiotic strains, where 3-oxoadipate can be further broken down to acetyl-CoA which enters the TCA cycle (fig. 3*E*). Vanillate can be an important source of energy and carbon for microorganisms (Nardi et al. 2003; MacLean et al. 2006; Varman et al. 2016); the symbiotic root nodule bacterium *Bradyrhizobium japonicum* is able to aerobically catabolize vanillate and protocatechuate (Ito et al. 2006). One degradation pathway for hemicellulose was specifically gained in the LCA of the Intra/Extra clade in that an endo-1,4-beta-xylanase involved in xylan degradation was acquired (fig. 3*E*). Altogether the ability to break down parts of the plant cell wall-hemicelluloses such as xylan may be a key feature in the intracellular colonization of *Gunnera* sp. by *Nostoc*. In addition, xylans can represent a substantial source of nutrition for bacteria able to degrade this substrate (Kulkarni et al. 1999; Saha 2003; Dodd and Cann 2009). Therefore, being able to derive C-sources from xylan and vanillate might give an advantage to the Intra/Extra clade strains over the members

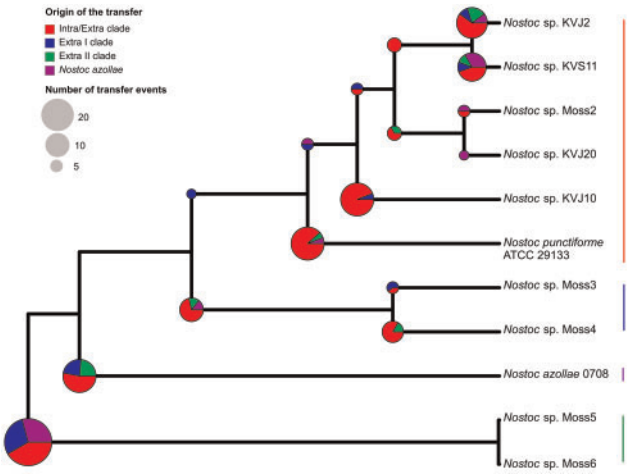

Fig. 4. HGTs detected by reconciliation of gene and species trees, and number of transfer events projected on a species tree for 25 gene families shared by symbiotic *Nostoc* strains. Pie charts at the nodes represent the number of transfers that arrived on the descending lineage. The size of each pie chart represents the number of HGTs and the proportion of transfers coming from the Intra/Extra clade are indicated in red, in blue from the Extra I clade, in green from the Extra II clade and in purple for *Nostoc azollae*. The species tree is a subtree of the genome-wide phylogeny presented in (fig. 6).

of the Extra clades. Their ability might not only expand their host range but also improve their competitiveness in epiphytic interactions. For instance, *nifH* gene quantification and phylotyping of the cyanobacterial community associated with feathermosses revealed that phylotypes closely related to the Intra/Extra strain *N. punctiforme* were dominating the community (Warshan et al. 2016). Consistent with these results, the vanillate monooxygenase and the endo-1,4-beta-xylanase for xylan degradation were identified as highly upregulated in *N.* Moss2 during physical contact with *P. schreberi* (Warshan et al. 2017), suggesting vanillate and xylan as C-sources for the Intra/Extra strains when in symbiosis with different host plants. Furthermore, the presence of β-1,4-linked xylan in the cell wall of the moss *Physcomitrella patens* has been confirmed (Moller et al. 2007).

Detection of Horizontal Gene Transfers between Symbiotic *Nostoc* Strains

Horizontal gene transfers (HGTs) were inferred by reconciliation of gene tree topologies with the species tree topology for 28 gene families shared by all symbiotic *Nostoc* strains but missing in *N.* CALU 996. HGTs were detected in 25 gene families and involved transfers between all clades of symbiotic *Nostoc* strains (fig. 4). We found 126 transfer events for the 25 gene families and many of the HGT events occur between crown groups (fig. 4), which indicates a high prevalence of HGT within the Nostocales. This evolutionary mechanism appears to have been frequent for gene families shared between plant symbiotic *Nostoc* strains, as reported for cyanobacteria in general (Zhaxybayeva et al. 2006; Yerrapragada et al. 2009; Szollosi et al. 2012). Considering that most of the transfers were coming from the Intra/Extra clade (fig. 4), our results suggest that the plant symbiotic gene set

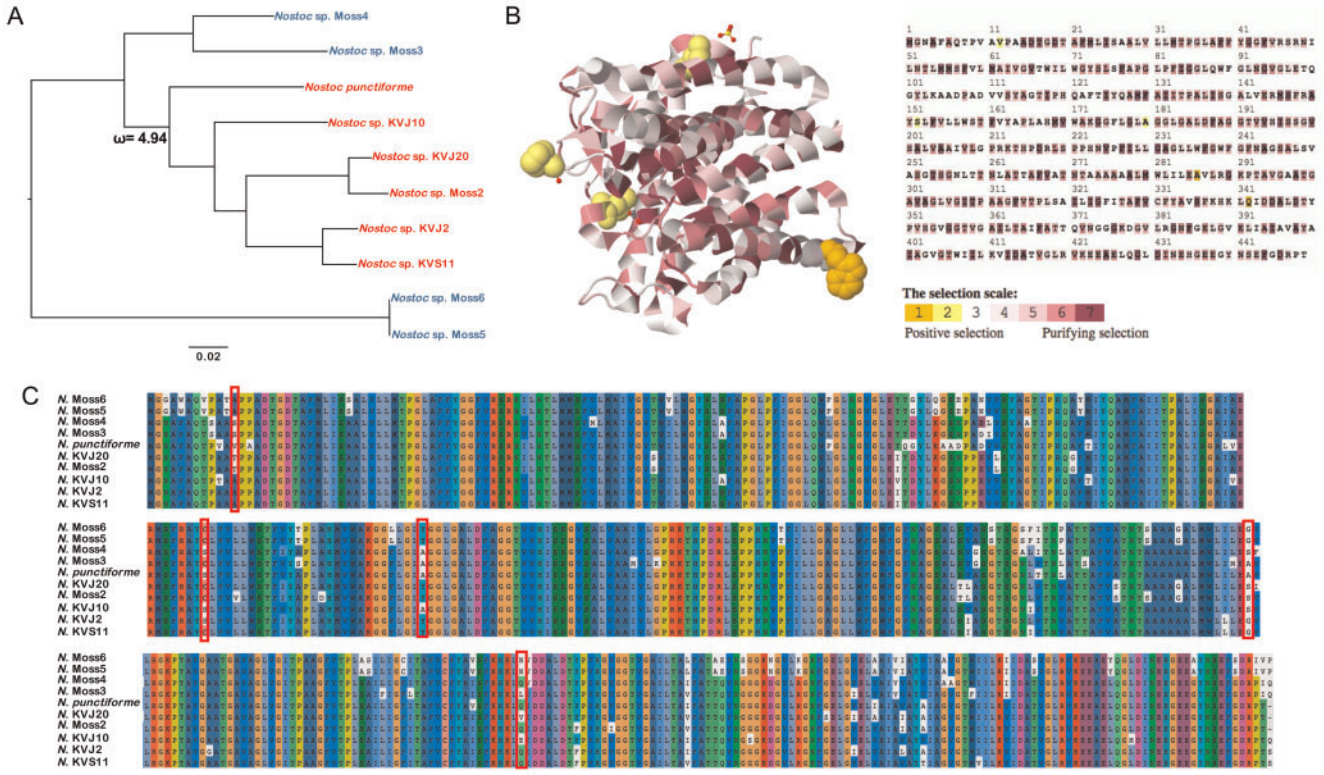

FIG. 5. Positive selection analysis of the *amtB* in extracellular-restricted and intracellular competent *Nostoc*. (*A*) Maximum likelihood gene tree based on the *amtB* amino acid sequences, the tree constructed by RAxML using GTR + GAMMA model. A foreground branch was specified as the separation between extracellular restricted (in blue) and intracellular competent *Nostoc* (in red). Ka/Ks substitution rates (ω value) is specified next to the branch. (*B*) Sites under positive selection mapped onto the *amtB* structure. Sites under positive selection are indicated in yellow, and the one under purifying selection are indicated in purple. The amino acid sequence of *N. punctiforme* was mapped with a color code. (*C*) Alignments of 10 AmtB proteins retrieved in the different facultative symbiotic *Nostoc* strains. Sites under positive selection are indicated by a red box.

was at least partially horizontally transferred from this clade to *N. azollae* and the Extra I and II clade, which might explain the polyphyletic origin found for those symbiotic cyanobacteria (fig. 2). It is notable that HGTs seem to be prevalent between members of the Intra/Extra clade (fig. 4), indicating that some adaptations of the intracellular symbiotic lifestyle could have been gained through HGT. Besides, late and early transfers involving the obligate symbiont *N. azollae* were detected, which implies that substantial gene flow might have occurred before its host restriction, at least 90 Ma (Metzgar et al. 2007), and that this strain might have been a symbiont of other host plants.

Shared Putative Symbiosis-Related Genes under Positive Selection

A total of 74 gene families shared by the members of the Intra/Extra clade and the two Extra clades represented single-copy gene families (again, *N. azollae* was excluded from this analysis due to its distinct evolutionary history; supplementary table S2, Supplementary Material online). For these single-copy gene families, the branch-site model was used to detect genes with signals of positive and purifying selection. Twenty-seven gene families under purifying selection were identified in the Intra/Extra clade ($\omega < 1$; $P < 0.01$) and ten gene families showed evidence of positive selection ($\omega > 1$; $P < 0.01$) (supplementary table S2, Supplementary Material online). Purifying selection generally allows maintenance of the ancestral function of a gene during evolution (Yang and Bielawski 2000), and these 27 gene families under selection were enriched in functions related to "Carbohydrate transport and metabolism," "Cell wall/membrane/envelope biogenesis," "Defense mechanisms," "Inorganic ion transport and metabolism," "Nucleotide transport and metabolism," and "Signal transduction mechanisms" (supplementary table S2, Supplementary Material online).

Genes under positive selection have usually undergone adaptive divergence (Yang and Bielawski 2000), and the ten gene families under positive selection were related to specific adaptive traits, such as "Cell wall/membrane/envelope biogenesis," "Defense mechanisms," "Energy production and conversion," and "Inorganic ion transport and metabolism" (supplementary table S2, Supplementary Material online). It is notable that the *amtB* gene encoding an ammonium transporter showed positive selection in the members of the Intra/Extra clade, but not in the members of the two Extra clades (fig. 5A). Five amino acid residues at position 12, 152, 180, 286, and 342 within the signal peptide, and the transmembrane helices three, four, seven, and nine, respectively, were significantly detected under positive selection (fig. 5B and C). In the

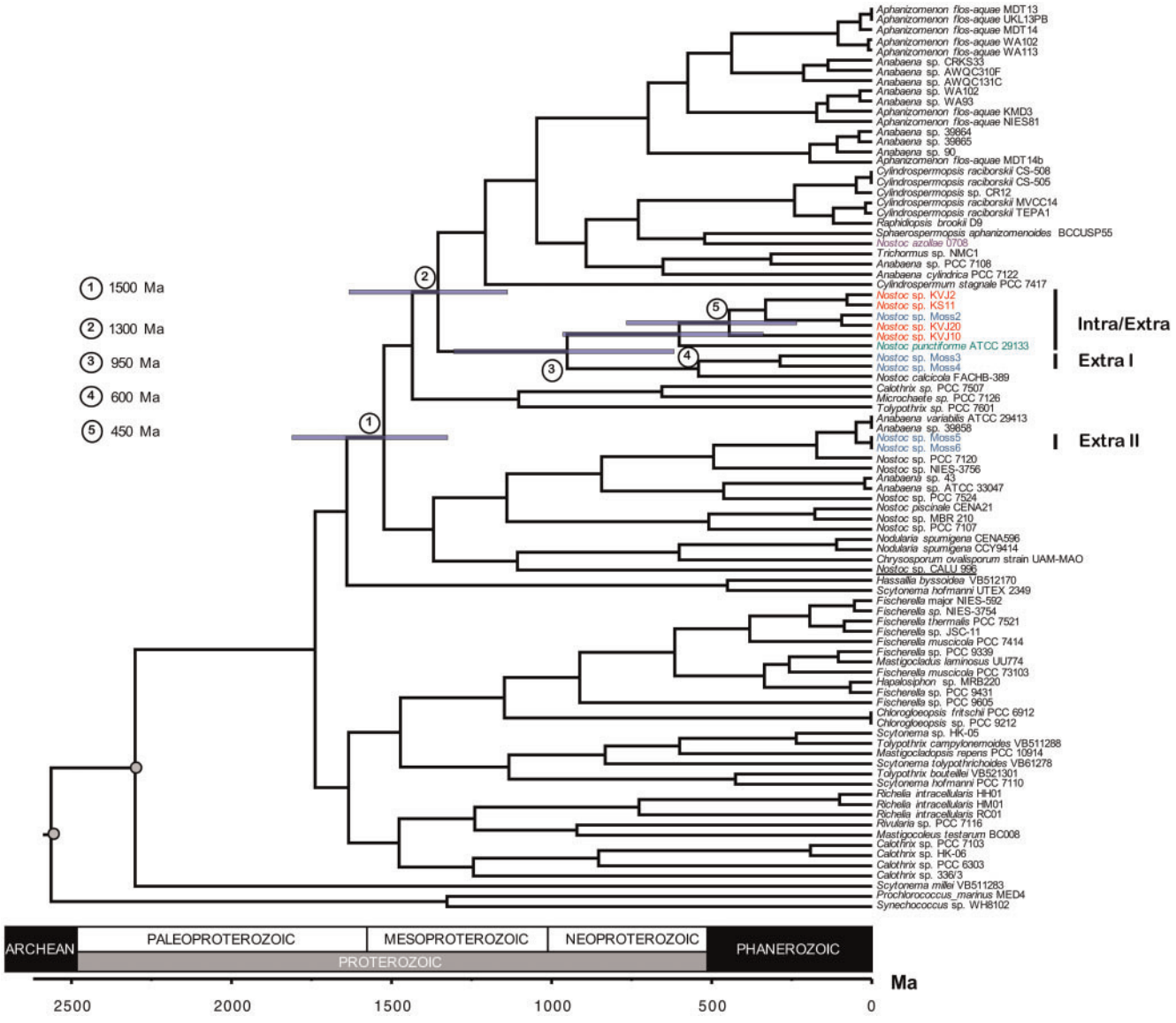

Fig. 6. The origin and diversification of symbiotic *Nostoc*. Bayesian relaxed molecular clock analyses were implemented using MCMCtree to estimate divergence times. *Nostoc* strains isolated from feathermosses are highlighted in blue, in red for *Blasia pusilla* isolates, in green for *N. punctiforme* isolated from cycad and in purple for *Nostoc azollae*, the obligate symbiont of *Azolla filiculoides*. The symbiotic incompetent *N.* CALU 996 is underlined. Two calibrations (gray circles) were used for the tree. The root of the tree was set between 2,320 and 2,700 Ma. Numbered nodes 1–5 indicate divergence times for the symbiotic *Nostoc*. Blue bars represent the posterior 95% confidence intervals for the node ages. Clades of extracellular-exclusive *Nostoc* isolates (Extra I and II) and the clade formed by intracellular competent strains (Intra/Extra) are shown.

hornwort/liverwort, *Azolla* and *Gunnera* symbioses the N-source for the host is ammonium produced from cyanobacterial N_2-fixation (Meeks 2009); the only exceptions are the cycad- and probably the feathermoss symbioses (Costa and Lindblad 2002; Warshan et al. 2017). This is achieved partly by a modulation of glutamine synthetase activity, with a reduction of enzyme activity, to 70% or 15% during symbiosis with *Gunnera* sp. and *Anthoceros* sp., respectively (Meeks 2009). The positive selection of *amtB* in the members of the Intra/Extra clade identified in this study suggests a specific adaptation in this clade. For rhizobial symbioses, expression of the gene encoding the bacterial AmtB importer is repressed in bacteroids, ensuring that ammonium originating from symbiont N_2-fixation is not recovered by the bacterium but rather provided to the host plant (Mus et al. 2016). Although no particular molecular function of the identified amino acid residues has been proven, they might be associated with the repression of the activity of the ammonium importer *amtB* in symbiotic *Nostoc* belonging to the Intra/Extra clade which would be beneficial for the host plant, with the exception of cycads and feathermosses.

An Evolutionary Timeline of Cyanobacterial-Plant Symbiosis

By implementing phylogenomic and Bayesian relaxed molecular clock analyses, we estimated the phylogeny and age divergences of symbiotic *Nostoc* strains (fig. 6). We identified

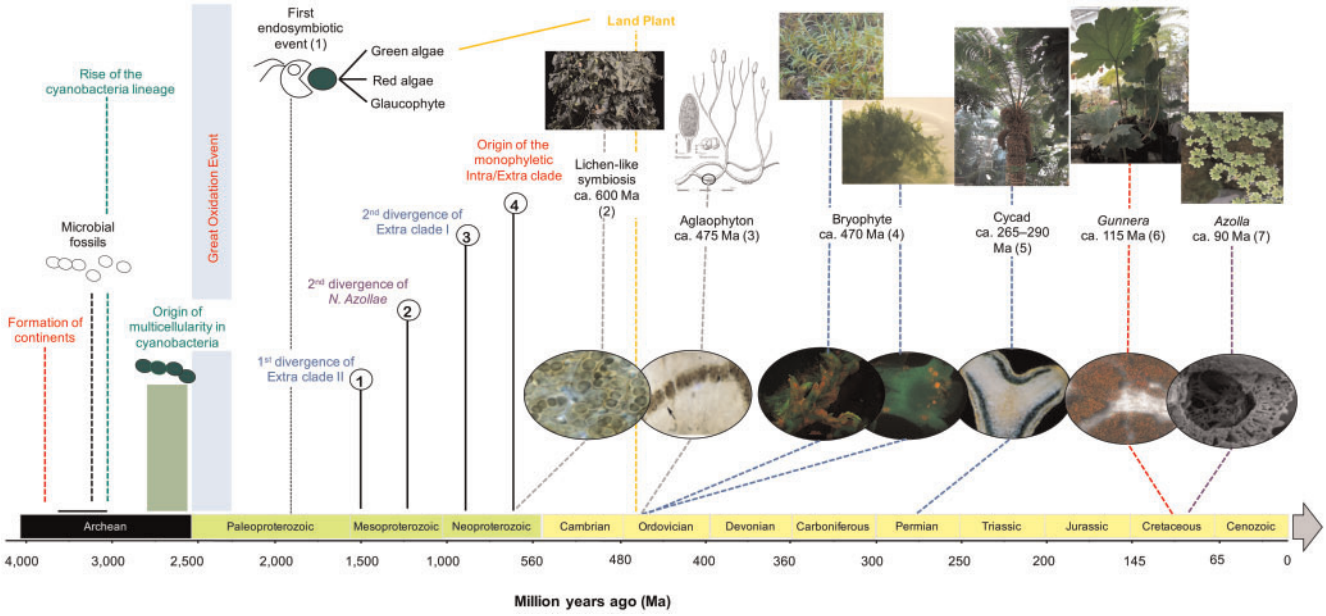

Fig. 7. A possible scenario of the evolution of *Nostoc* in symbiosis with plants. Image shows the different extant terrestrial symbioses involving *Nostoc* sp., and the cyanobacterial symbiont of the early land plant *Aglaophyton major* from the Lower Devonian Rhynie Chert (Krings et al. 2009). Photos illustrating the lichen-like symbiosis represent the bipartite cyanolichen *Leptogium* sp. and its *Nostoc* symbiont (Rikkinen 2017). Divergence time and origin of symbiotic *Nostoc* were estimated from bayesian relaxed molecular clock analyses using MCMCtree (see fig. 5). (1) Based on the scenario and timing of primary endosymbiosis by Sánchez-Baracaldo et al. (2017), (2) Yuan et al. (2005), (3) Krings et al. (2009), (4) Edwards et al. (2014); Brown et al. (2015), (5) Brenner et al. (2003); Condamine et al. (2015), (6) Vekemans et al. (2012), and (7) Metzgar et al. (2007).

the Mesoproterozoic and the beginning of Neoproterozoic as the interval when the ancestors of present day symbiotic *Nostoc* strains diverged and diversified. The first divergence was estimated to have taken place ca. 1,500 Ma, when the cyanobacterial feathermoss isolates *N.* Moss5 and Moss6 (Extra II clade), as well as the nonsymbiotic *N.* CALU 996 diverged from the other clades of symbiotic *Nostoc* strains. The second divergence between symbiotic *Nostoc* strains was estimated to ca. 1,300 Ma, when *N. azollae* diverged from the Intra/Extra clade and the Extra I clade. Finally, the last divergence between the Intra/Extra clade and the Extra I clade was dated to ca. 950 Ma. We also identified the origin of the Intra/Extra clade, dated to ca. 600 Ma, as well as within this clade, the more recent divergence between the cycad isolate *N. punctiforme* and the bryophyte *Nostoc* isolates at ca. 450 Ma. Surprisingly, the emergence of the Intra/Extra clade predated the origin of land plants in the Mid-Ordovician period ca. 471 Ma, as well as the emergence of their host genus *Gunnera* which was dated to ca. 115 Ma (fig. 7; Vekemans et al. 2012; Magallón et al. 2013; Edwards et al. 2014). The earliest direct fossil evidence of an intracellular cyanobacterial symbiosis is the observation of cyanobacterial filaments in prostrate axes of *Aglaophyton major*, an early nonvascular land plant, which have been dated to ca. 400 Ma (fig. 7; Krings et al. 2009). Due to the lack of observation of heterocyst or akinete structure in these fossils it was hypothesized that Oscillatoriales and not Nostocales were the cyanobacteria associated with *A. major* (Krings et al. 2009). Our results suggest that the intracellular lifestyle of *Nostoc* first manifested with another host than a member of the Gunnerales. A metagenomic study of the epiphytic microbiome associated with streptophyte algae (green algae most closely related to embryophytes), revealed the presence of *Nostoc* in the microbial community and comparison of the microbiomes of streptophyte algae and bryophytes suggest that symbiosis with microbial community might have been inherited from algal ancestors to early land plants (Knack et al. 2015). Their results together with our study might indicate that the origin of *Nostoc* in symbiosis with plants could have predated the origin of embryophytes and this lifestyle could have been initiated with streptophyte algae. Another intracellular symbiosis involving *N. punctiforme* is the one with *Geosiphon pyriformis*, a glomeromycotinan fungus which is hosting *N. punctiforme* filaments in the bladder, a structure formed upon contact with the cyanobacterium (Schüßler et al. 1994; Mollenhauer et al. 1996; Schüßler 2012). This interaction might represent the only remnant of a once widespread fungal-cyanobacterial symbiosis. Thus, it is possible that the LCA of the Intra/Extra clade was forming intracellular symbioses with streptophyte algae, early land plants and/or fungi, and that capacity was lost during evolution with the exception of *Geosiphon* and/or *Gunnera* sp.

Conclusions

In this study, we provide evidence that symbiotic *Nostoc* strains with a broad host-range, including both intracellular and extracellular physical localizations, form a monophyletic clade indicating that the members of this clade have a common evolutionary history. On the contrary, a polyphyletic origin was found for *Nostoc* strains only able to enter extracellular symbioses, suggesting that this trait was most likely

gained or lost several times in the evolution of the Nostocales. Consistent with this assumption, we found that only a small set of genes was retained in the genomes of all facultative symbiotic *Nostoc* strains and that transfer events of this set between symbiotic strains was frequent; so the trait of capability for symbiosis could have been transferred several times in the evolution of the Nostocales. It is important to note that only a single genome of nonsymbiotic *Nostoc* was used for comparative genomic analysis and the characterization of more nonsymbiotic strains is a future requirement to consolidate our findings. When comparing the members of the two Extra clades of *Nostoc* to the LCA of the members of the Intra/Extra clade, we found that a specific set of genes was gained during evolution that might have enabled the Intra/Extra clade to form an intracellular symbiosis with *Gunnera* sp. The reconstruction of the metabolic and functional capabilities of the LCA of the Intra/Extra clade and of the two Extra clades provided new insights into the genomic changes associated with the evolution of *Nostoc* symbioses, and highlighted interesting parallels with rhizobial-plant symbioses. Some of the genomic changes implied nutrient exchange between partners, host recognition and adaptation to the host environment, and argue for specific adaptations to a mutualistic type of symbiosis. We found that genes families, specific to facultative symbiotic *Nostoc* strains, were under positive selection in the lineage of Intra/Extra strains, thus potentially had undergone adaptive divergence compared with the other symbiotic *Nostoc* strains. Molecular clock analysis suggested that the intracellular competent clade emerged ca. 600 Ma, which predates the origin of land plants, and the origin of Gunnerales. Even though these results are not based on the direct evidence of early land plant fossils being intracellularly colonized by *Nostoc*, our findings suggest that the Intra/Extra clade evolved before the genus *Gunnera*, and that intracellular cyanobacterial symbioses even may have predated the emergence of extant terrestrial plants. We hypothesize that LCA that originated ca. 600 Ma likely adapted the ability to form associations with extinct land plants, or perhaps even with streptophyte algae and fungi.

Materials and Methods

Growth Conditions for Cyanobacteria, Feathermosses, *Blasia pusilla*, and *Gunnera manicata*

Nostoc strains from feathermosses (*P. schreberi* and *H. splendens*) and *B. pusilla* were isolated as previously described by Warshan et al. (2017) and West and Adams (1997), respectively. All *Nostoc* isolates (table 2) were grown in liquid $BG11_0$ medium at 20°C with constant shaking and constant light (35 μmol photons $m^{-2}s^{-1}$). *Blasia pusilla* was grown in 20% strength $BG11_0$ medium under the same conditions. Feathermosses were maintained on a thin layer of vermiculite in the greenhouse at 10°C and constant light at 35 μmol photons $m^{-2}s^{-1}$ and misted with deionized water once per week. *Gunnera manicata* propagation and maintenance was done according to Liaimer et al. (2015).

Symbiosis Reconstitution Experiments

To reconstitute the cyanobacterial-feathermoss symbiosis, experiments with *Nostoc* isolates and *P. schreberi* were performed as described by Bay et al. (2013). Infections of *B. pusilla* and *G. manicata* were performed as described by Liaimer et al. (2015). Microscopy was used to verify the symbiotic competency of the different *Nostoc* isolates. Plant tissues from different hosts were observed under an epifluorescence microscope (Axiovert, Carl Zeiss GmbH, Jena, Germany) and a stereomicroscope (SteReo Lumar.V12, Carl Zeiss) equipped with a green excitation filter (510–560 nm). Confocal microscopy (Zeiss LSM 780, Carl Zeiss) utilizing the argon laser (488 nm) and the Plan-Apo 20×/0.8 lens was utilized for the verification of the internal colonization of *G. manicata* via *N.* Moss2. The photos were taken and extracted with the ZEN blue software (Carl Zeiss).

Genome Sequencing and Assembly

Genomic DNA of the feathermoss isolates was extracted as explained in Warshan et al. (2017) and sequencing was performed using PacBio technology with subsequent assembly at the DOE JGI (Joint Genome Institute, Walnut Creek, CA). Genomic DNA of *B. pusilla* isolates was extracted as explained in Meeks et al. (2001) and sequencing was done using Illumina MiSeq platform at the UiT (The Arctic University of Norway). More specifically for *B. pusilla* isolates, the resulting reads were quality checked using FastQC. A draft de novo assembly was created using the MIRA 4.0.2 software with default settings (Chevreux et al. 1999). Initial analyses of the genomes of *B. pusilla* isolates KVJ2, KVJ10, KVJ20, and KVS11 revealed some contamination with bacterial DNA. Thus, assemblies for these individual genomes were binned separately using t-SNE ordination in VizBin (Laczny et al. 2015) which identified cyanobacterial chromosome and plasmid assemblies as well as bacterial contaminants as separate bins. The genome characteristics of the symbiotic *Nostoc* strains are described in supplementary table S1, Supplementary Material online.

Determining Orthologous Gene Families

For phylogenomic analyses, we determined orthologous gene families between the 11 genomes of symbiotic Nostocales (table 2), the nonsymbiotic strain *Nostoc* sp. CALU 996 (*N.* CALU 996) and 74 genomes of Nostocales for which no symbiosis with plants is known. An all-against-all BLASTP analysis of the proteomes of these 86 Nostocales led to the identification of candidate orthologs. BLAST hits with an e-value $>10^{-5}$ and for which the query and the hit sequence had <50% overlap of their gene length, were excluded. Clusters of orthologous gene families were created using OrthoMCL (Li et al. 2003) with the recommended settings resulting in a total of 22,839 gene families.

Inferring Phylogenetic Trees

Genome-wide phylogenies were inferred from 65 single-copy orthologous gene families based on orthologous gene clustering. The partition model involves estimating independent evolutionary models for different genes or subsets of genes,

using the PartitionFinder software (Lanfear et al. 2012). The genomes of symbiotic *Nostoc* strains were used to build a concatenated core gene phylogeny across 89 Nostocales genomes. The genomes of symbiotic and nonsymbiotic Nostocales were added to the alignment as concatenated core protein strings, with a reconstruction of the whole genome phylogeny using RaxML (Stamatakis 2014).

Molecular dating was performed using Bayesian relaxed molecular clock analyses by implementing MCMCtree (Yang 2007). As molecular dating requires a phylogenomic tree with fossil calibrations, the minimum and maximum ages for the cyanobacterial root were set to 2,320 Ma (Bekker et al. 2004) and 2,700 Ma (Brocks et al. 2003; Farquhar et al. 2011), respectively. For the Nostocales, a minimum age of 1,650 and a maximum age of 2,320 Ma were used (Tomitani et al. 2006; Butterfield 2015; Uyeda et al. 2016).

Inferring Gene Gain and Gene Loss

To study the evolution of life style strategies of symbiotic and nonsymbiotic cyanobacteria, we used orthologous gene clustering to compile a set of 9,573 orthologous gene families covering the 11 genomes of symbiotic *Nostoc* strains, the genome of the nonsymbiotic strain *N.* CALU 996 and one outgroup genome. Ancestral family sizes were inferred using COUNT (Csurös 2010) by assuming a probabilistic framework involving a phylogenetic birth-and-death model (Nei and Rooney 2005) along the rooted phylogeny. The model is described by lineage- and family-specific gene loss and duplication rates, coupled with a family gain process accounting for arrival by horizontal gene transfer. Selected sets of predicted gene gain, loss, expansion, and contraction occurrences were categorized using the COG, KEGG, and BioCyc databases.

Detecting HGTs in Gene Trees Given Candidate Species Trees

We used ecceTERA (Jacox et al. 2016) to compute the most parsimonious number of gene duplications, losses, and transfers to reconciliate a given gene tree with the species tree. All gene families missing in *N.* CALU 996 but retained in all symbiotic *Nostoc* strains were screened for such conflict. We did not consider transfer events to or from unsampled species, and to account for uncertainty regarding parsimony costs for events, we used Pareto optimality according to To et al. (2015). A dated subtree of symbiotic *Nostoc* strains based on the whole genome phylogeny was supplied as the species tree. For each gene family, amino acid alignments were generated in MUSCLE (Edgar 2004) and sites filtered with trimAl (Capella-Gutiérrez et al. 2009). RAxML (Stamatakis 2014) was used to infer gene trees from amino acid alignment for each gene family with automatic model selection and 1,000 bootstraps using the rapid bootstrapping algorithm.

Positive Selection Detection to Identify Genes Putatively Involved in Symbiosis

Positive selection can be inferred from a higher proportion of nonsynonymous (Ka) over synonymous substitutions (Ks) per site ($Ka/Ks > 1$). In this analysis, only single-copy orthologous gene families retained in all facultative symbiotic *Nostoc* strains and missing in the nonsymbiotic *N.* CALU 996 were used. To calculate the nonsynonymous and synonymous substitution rates for each single-copy ortholog, amino acid alignment for all gene families was generated in MUSCLE (Edgar 2004) and sites were filtered with trimAl (Capella-Gutiérrez et al. 2009) using default settings. Then the resulting protein alignments were reverse-translated to codon-based nucleotide alignments with PAL2NAL (Suyama et al. 2006). For each alignment, a gene tree was constructed by RAxML software (Stamatakis 2014) using maximum likelihood criteria under the GTR + GAMMA model. Using each gene tree topology, we applied the improved branch-site model (Zhang et al. 2005) implemented in codeml from PAML 4 (Yang 2007) to estimate the Ka/Ks substitution rates (ω value) for each gene family. A foreground branch was specified as the separation between *Nostoc* strains able to enter extracellular symbioses, and those able to enter both extra- and intracellular symbioses. A significant likelihood ratio test was conducted to determine whether positive selection was operating in the foreground branch. We set the threshold for positively selected sites at a posterior probability >95% based on empirical Bayes analysis (Yang et al. 2005).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Nostoc sp. CALU 996 was kindly provided by Professor AV Pinevich, St Petersburg State University, St Petersburg, Russia. A portion of this research was performed under the JGI-EMSL Collaborative Science Initiative and used resources at the DOE Joint Genome Institute, which is DOE Office of Science User Facilities. The DOE Joint Genome Institute facilities is sponsored by the Office of Biological and Environmental Research and operated under contract no. DE-AC02-05CH11231 (JGI). C.L.D. was supported by the NASA Astrobiology Institute Alternative Earths.

References

Archibald JM. 2009. The puzzle of plastid evolution. *Curr Biol.* 19(2):R81–R88.

Archibald JM. 2015a. Endosymbiosis and eukaryotic cell evolution. *Curr Biol.* 25(19):R911–R921.

Archibald JM. 2015b. Genomic perspectives on the birth and spread of plastids. *Proc Natl Acad Sci U S A.* 112(33):10147–10153.

Barron AR, Wurzburger N, Bellenger JP, Wright SJ, Kraepiel AML, Hedin LO. 2009. Molybdenum limitation of asymbiotic nitrogen fixation in tropical forest soils. *Nat Geosci.* 2(1):42–45.

Bay G, Nahar N, Oubre M, Whitehouse MJ, Wardle DA, Zackrisson O, Nilsson MC, Rasmussen U. 2013. Boreal feather mosses secrete chemical signals to gain nitrogen. *New Phytol.* 200(1):54–60.

Bekker A, Holland HD, Wang P-L, Rumble D, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ. 2004. Dating the rise of atmospheric oxygen. *Nature* 427(6970):117–120.

Bergman B. 2002. The *Nostoc-Gunnera* Symbiosis. In: Rai AN, Bergman B, Rasmussen U, editors. Cyanobacteria in symbiosis. Dordrecht: Springer Netherlands. p. 207–232.

Brenner ED, Stevenson DW, Twigg RW. 2003. Cycads: evolutionary innovations and the role of plant-derived neurotoxins. *Trends Plant Sci*. 8(9):446–452.

Brocks JJ, Buick R, Summons RE, Logan GA. 2003. A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochim Cosmochim Acta* 67(22):4321–4335.

Brown RC, Lemmon BE, Shimamura M, Villarreal JC, Renzaglia KS. 2015. Spores of relictual bryophytes: diverse adaptations to life on land. *Rev Palaeobot Palynol*. 216:1–17.

Burki F. 2017. The convoluted evolution of eukaryotes with complex plastids. *Adv Bot Res*. 84:1–30.

Butterfield NJ. 2015. Proterozoic photosynthesis – a critical review. *Palaeontology* 58(6):953–972.

Campbell EL, Hagen KD, Chen R, Risser DD, Ferreira DP, Meeks JC. 2015. Genetic analysis reveals the identity of the photoreceptor for phototaxis in hormogonium filaments of *Nostoc punctiforme*. *J Bacteriol*. 197(4):782–791.

Campbell EL, Meeks JC. 1989. Characteristics of hormogonia formation by symbiotic nostoc spp. in response to the presence of *Anthoceros punctatus* or its extracellular products. *Appl Environ Microbiol*. 55(1):125–131.

Campbell GRO, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, Walker GC. 2006. *Sinorhizobium meliloti bluB* is necessary for production of 5, 6-dimethylbenzimidazole, the lower ligand of B12. *Proc Natl Acad Sci U S A*. 103(12):4634–4639.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.

Carpenter EJ, Foster RA. 2002. Marine cyanobacterial symbioses. In: Rai AN, Bergman B, Rasmussen U, editors. Cyanobacteria in symbiosis. Dordrecht: Springer Netherlands. p. 11–17.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB), 99:45–56.

Condamine FL, Nagalingum NS, Marshall CR, Morlon H. 2015. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evol Biol*. 15:65.

Costa J-L, Lindblad P. 2002. Cyanobacteria in symbiosis with cycads. In: Rai AN, Bergman B, Rasmussen U, editors. Cyanobacteria in symbiosis. Dordrecht: Springer Netherlands. p. 195–205.

Csurös M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26(15):1910–1912.

Dodd D, Cann IKO. 2009. Enzymatic deconstruction of xylan for biofuel production. *GCB Bioenergy* 1(1):2–17.

Duggan PS, Thiel T, Adams DG. 2013. Symbiosis between the cyanobacterium *Nostoc* and the liverwort *Blasia* requires a CheR-type MCP methyltransferase. *Symbiosis* 59(2):111–120.

Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, Dearth SP, Van Mooy BAS, Campagna SR, Kujawinski EB. 2015. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc Natl Acad Sci U S A*. 112(2):453–457.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.

Edwards D, Morris JL, Richardson JB, Kenrick P. 2014. Cryptospores and cryptophytes reveal hidden diversity in early land floras. *New Phytol*. 202(1):50–78.

Farquhar J, Zerkle AL, Bekker A. 2011. Geological constraints on the origin of oxygenic photosynthesis. *Photosynth Res*. 107(1):11–36.

Foster RA, Carpenter EJ, Bergman B. 2006. Unicellular cyanobionts in open ocean dinoflagellates, radiolarians, and tintinnids: ultrastructural characterization and immuno-localization of phycoerythrin and nitrogenase. *J Phycol*. 42(2):453–463.

Foster RA, Kuypers MMM, Vagner T, Paerl RW, Musat N, Zehr JP. 2011. Nitrogen fixation and transfer in open ocean diatom–cyanobacterial symbioses. *ISME J*. 5(9):1484–1493.

Gehringer MM, Pengelly JJL, Cuddy WS, Fieker C, Forster PI, Neilan BA. 2010. Host selection of symbiotic cyanobacteria in 31 species of the australian cycad genus: *macrozamia* (Zamiaceae). *Mol Plant-Microbe Interact*. 23(6):811–822.

Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. *ISME J*. 8(8):1553–1565.

Gould SB, Waller RF, McFadden GI. 2008. Plastid evolution. *Annu Rev Plant Biol*. 59:491–517.

Ito N, Itakura M, Eda S, Saeki K, Oomori H, Yokoyama T, Kaneko T, Tabata S, Ohwada T, Tajima S. 2006. Global gene expression in *Bradyrhizobium japonicum* cultured with vanillin, vanillate, 4-hydroxybenzoate and protocatechuate. *Microbes Environ*. 21(4):240–250.

Jacox E, Chauve C, Szöllősi GJ, Ponty Y, Scornavacca C. 2016. EcceTERA: comprehensive gene tree-species tree reconciliation using parsimony. *Bioinformatics* 32(13):2056–2058.

Johansson C, Bergman B. 1994. Reconstitution of the symbiosis of *Gunnera manicata* Linden: cyanobacterial specificity. *New Phytol*. 126(4):643–652.

Keeling PJ. 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu Rev Plant Biol*. 64:583–607.

Knack JJ, Wilcox LW, Delaux P-M, Ané J-M, Piotrowski MJ, Cook ME, Graham JM, Graham LE. 2015. Microbiomes of streptophyte algae and bryophytes suggest that a functional suite of microbiota fostered plant colonization of land. *Int J Plant Sci*. 176(5):405–420.

Krings M, Hass H, Kerp H, Taylor TN, Agerer R, Dotzler N. 2009. Endophytic cyanobacteria in a 400-million-yr-old land plant: a scenario for the origin of a symbiosis? *Rev Palaeobot Palynol*. 153(1–2):62–69.

Kulkarni N, Shendye A, Rao M. 1999. Molecular and biotechnological aspects of xylanases. *FEMS Microbiol Rev*. 23(4):411–456.

Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian H, Coronado S, der Maaten L, Vlassis N, Wilmes P. 2015. VizBin – an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* 3(1):1.

Laenen B, Shaw B, Schneider H, Goffinet B, Paradis E, Désamoré A, Heinrichs J, Villarreal JC, Gradstein SR, McDaniel SF, et al. 2014. Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nat Commun*. 5:5134.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*. 29(6):1695–1701.

Larsson J, Nylander JA, Bergman B. 2011. Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evol Biol*. 11:187.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 13(9):2178–2189.

Li ZH, Wang Q, Ruan X, Pan CD, Jiang DA. 2010. Phenolics and plant allelopathy. *Molecules* 15(12):8933–8952.

Liaimer A, Helfrich EJN, Hinrichs K, Guljamow A, Ishida K, Hertweck C, Dittmann E. 2015. Nostopeptolide plays a governing role during cellular differentiation of the symbiotic cyanobacterium *Nostoc punctiforme*. *Proc Natl Acad Sci U S A*. 112(6):1862–1867.

Liaimer A, Jenke-Kodama H, Ishida K, Hinrichs K, Stangeland J, Hertweck C, Dittmann E. 2011. A polyketide interferes with cellular differentiation in the symbiotic cyanobacterium *Nostoc punctiforme*. *Environ Microbiol Rep*. 3(5):550–558.

Liaimer A, Jensen JB, Dittmann E. 2016. A genetic and chemical perspective on symbiotic recruitment of cyanobacteria of the genus *Nostoc* into the host plant *Blasia pusilla* L. *Front Microbiol*. 7:1693.

MacLean AM, MacPherson G, Aneja P, Finan TM. 2006. Characterization of the beta-ketoadipate pathway in *Sinorhizobium meliloti*. *Appl Environ Microbiol*. 72(8):5403–5413.

Magallón S, Hilu KW, Quandt D. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am J Bot*. 100(3):556–573.

Meeks JC. 2009. Physiological adaptations in nitrogen-fixing *Nostoc*–plant symbiotic associations. In: Pawlowski K, editor. Prokaryotic symbionts in plants. Heidelberg: Springer Berlin. p. 181–205.

Meeks JC, Elhai J, Thiel T, Potts M, Larimer F, Lamerdin J, Predki P, Atlas R. 2001. An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. *Photosynth Res*. 70(1):85–106.

Metzgar JS, Schneider H, Pryer KM. 2007. Phylogeny and divergence time estimates for the fern genus *Azolla* (Salviniaceae). *Int J Plant Sci*. 168(7):1045–1053.

Mollenhauer D, Mollenhauer R, Kluge M. 1996. Studies on initiation and development of the partner association in *Geosiphon pyriforme* (Kütz.) v. Wettstein, a unique endocytobiotic system of a fungus (Glomales) and the cyanobacterium *Nostoc punctiforme* (Kütz.) Hariot. *Protoplasma* 193:3–9.

Moller I, Sørensen I, Bernal AJ, Blaukopf C, Lee K, Øbro J, Pettolino F, Roberts A, Mikkelsen JD, Knox JP, et al. 2007. High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. *Plant J*. 50(6):1118–1128.

Mus F, Crook MB, Garcia K, Costas AG, Geddes BA, Kouri ED, Paramasivan P, Ryu MH, Oldroyd GED, Poole PS, et al. 2016. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl Environ Microbiol*. 82(13):3698–3710.

Nardi S, Pizzeghello D, Bragazza L, Gerdol R. 2003. Low-molecular-weight organic acids and hormone-like activity of dissolved organic matter in two forest soils in N Italy. *J Chem Ecol*. 29(7):1549–1564.

Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*. 39:121–152.

Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci U S A*. 108(33):13624–13629.

Paz-Yepes J, Herrero A, Flores E. 2007. The NtcA-regulated *amtB* gene is necessary for full methylammonium uptake activity in the cyanobacterium *Synechococcus elongatus*. *J Bacteriol*. 189(21):7791–7798.

Paz-Yepes J, Merino-Puerto V, Herrero A, Flores E. 2008. The *amt* gene cluster of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol*. 190(19):6534–6539.

Prell J, Bourdès A, Kumar S, Lodwig E, Hosie A, Kinghorn S, White J, Poole P. 2010. Role of symbiotic auxotrophy in the *Rhizobium*-legume symbioses. *PLoS One* 5(11):e13933.

Rai AN, Söderbäck E, Bergman B. 2000. Tansley Review No. 116. Cyanobacterium-plant symbioses. *New Phytol*. 147(3):449–481.

Ran L, Larsson J, Vigil-Stenman T, Nylander JAA, Ininbergs K, Zheng WW, Lapidus A, Lowry S, Haselkorn R, Bergman B. 2010. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. *PLoS One* 5(7):e11486.

Rikkinen J. 2017. Cyanobacteria in terrestrial symbiotic systems. In: Hallenbeck PC, editor. Modern topics in the phototrophic prokaryotes. Cham: Springer International Publishing. p. 243–294.

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

Rousk K, Degboe J, Michelsen A, Bradley R, Bellenger JP. 2017. Molybdenum and phosphorus limitation of moss-associated nitrogen fixation in boreal ecosystems. *New Phytol*. 214(1):97–107.

Saha BC. 2003. Hemicellulose bioconversion. *J Ind Microbiol Biotechnol*. 30(5):279–291.

Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH. 2017. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc Natl Acad Sci U S A*. 201620089.

Santi C, Bogusz D, Franche C. 2013. Biological nitrogen fixation in nonlegume plants. *Ann Bot*. 111(5):743–767.

Schüßler A. 2012. 5 The *Geosiphon–Nostoc* endosymbiosis and its role as a model for arbuscular mycorrhiza research. In: Hock B, editor. Fungal associations. Heidelberg (Germany): Springer Berlin. p. 77–91.

Schüßler A, Mollenhauer D, Schnepf E, Kluge M. 1994. *Geosiphon pyriforme*, an endosymbiotic association of fungus and cyanobacteria: the spore structure resembles that of arbuscular mycorrhizal (AM) fungi. *Bot Acta* 107(1):36–45.

Shimizu M, Fujii T, Masuo S, Takaya N. 2010. Mechanism of de novo branched-chain amino acid synthesis as an alternative electron sink in hypoxic *Aspergillus nidulans* cells. *Appl Environ Microbiol*. 76(5):1507–1515.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Stucken K, John U, Cembella A, Murillo AA, Soto-Liebe K, Fuentes-Valdés JJ, Friedel M, Plominsky AM, Vásquez M, Glöckner G. 2010. The smallest known genomes of multicellular and toxic cyanobacteria: comparison, minimal gene sets for linked traits and the evolutionary implications. *PLoS One* 5(2):e9235.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(Web Server issue):W609–W612.

Svenning MM, Eriksson T, Rasmussen U. 2005. Phylogeny of symbiotic cyanobacteria within the genus *Nostoc* based on 16S rDNA sequence analyses. *Arch Microbiol*. 183(1):19–26.

Szollosi GJ, Boussau B, Abby SS, Tannier E, Daubin V. 2012. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proc Natl Acad Sci U S A*. 109(43):17513–17518.

Taga ME, Larsen NA, Howard-Jones AR, Walsh CT, Walker GC. 2007. BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature* 446(7134):449–453.

Taga ME, Walker GC. 2010. *Sinorhizobium meliloti* requires a cobalamin-dependent ribonucleotide reductase for symbiosis with its plant host. *Mol Plant-Microbe Interact*. 23(12):1643–1654.

To TH, Jacox E, Ranwez V, Scornavacca C. 2015. A fast method for calculating reliable event supports in tree reconciliations via Pareto optimality. *BMC Bioinformatics* 16:384.

Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proc Natl Acad Sci U S A*. 103(14):5442–5447.

Usher KM, Bergman B, Raven JA. 2007. Exploring cyanobacterial mutualisms. *Annu Rev Ecol Evol Syst*. 38(1):255–273.

Uyeda JC, Harmon LJ, Blank CE. 2016. A comprehensive study of cyanobacterial morphological and ecological evolutionary dynamics through deep geologic time. *PLoS One* 11(9):e0162539.

Varman AM, He L, Follenfant R, Wu W, Wemmer S, Wrobel SA, Tang YJ, Singh S. 2016. Decoding how a soil bacterium extracts building blocks and metabolic energy from ligninolysis provides road map for lignin valorization. *Proc Natl Acad Sci U S A*. 113(40):E5802–E5811.

Vekemans D, Proost S, Vanneste K, Coenen H, Viaene T, Ruelens P, Maere S, Van De Peer Y, Geuten K. 2012. Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-BOX gene and species diversification. *Mol Biol Evol*. 29(12):3793–3806.

Vitousek PM, Menge DN, Reed SC, Cleveland CC. 2013. Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. *Philos Trans R Soc L B Biol Sci*. 368:20130119.

Warshan D, Bay G, Nahar N, Wardle DA, Nilsson M-C, Rasmussen U. 2016. Seasonal variation in *nifH* abundance and expression of cyanobacterial communities associated with boreal feather mosses. *ISME J*. 10(9):2198–2208.

Warshan D, Espinoza JL, Stuart RK, Richter RA, Kim SY, Shapiro N, Woyke T, C Kyrpides N, Barry K, Singan V, et al. 2017. Feathermoss and epiphytic *Nostoc* cooperate differently: expanding the spectrum of plant-cyanobacteria symbiosis. *ISME J*. 11(12):2821–2833.

West NJ, Adams DG. 1997. Phenotypic and genotypic comparison of symbiotic and free-living cyanobacteria from a single field site. *Appl Environ Microbiol*. 63(11):4479–4484.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.

Yang Z, Bielawski JR. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*. 15(12):496–503.

Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 22(4):1107–1118.

Yerrapragada S, Siefert JL, Fox GE. 2009. Horizontal gene transfer in cyanobacterial signature genes. *Methods Mol Biol*. 532:339–366.

Yuan X, Xiao S, Taylor TN. 2005. Lichen-like symbiosis 600 million years ago. *Science* 308(5724):1017–1020.

Zehr JP, Shilova IN, Farnelid HM, Muñoz-MarínCarmen M. d C, Turk-Kubo KA. 2016. Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A. *Nat Microbiol*. 2(1):16214.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22(12):2472–2479.

Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res*. 16(9):1099–1108.

Zimorski V, Ku C, Martin WF, Gould SB. 2014. Endosymbiotic theory for organelle origins. *Curr Opin Microbiol*. 22:38–48.