

## ORIGINAL ARTICLE

# Feathermoss and epiphytic *Nostoc* cooperate differently: expanding the spectrum of plant–cyanobacteria symbiosis

Denis Warshan<sup>1</sup>, Josh L Espinoza<sup>2</sup>, Rhona K Stuart<sup>3</sup>, R Alexander Richter<sup>2</sup>, Sea-Yong Kim<sup>1</sup>, Nicole Shapiro<sup>4</sup>, Tanja Woyke<sup>4</sup>, Nikos C Kyrpides<sup>4</sup>, Kerrie Barry<sup>4</sup>, Vasanth Singan<sup>4</sup>, Erika Lindquist<sup>4</sup>, Charles Ansong<sup>5</sup>, Samuel O Purvine<sup>6</sup>, Heather M Brewer<sup>6</sup>, Philip D Weyman<sup>2</sup>, Christopher L Dupont<sup>2</sup> and Ulla Rasmussen<sup>1</sup>

<sup>1</sup>Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden; <sup>2</sup>J Craig Venter Institute, La Jolla, CA, USA; <sup>3</sup>Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA; <sup>4</sup>US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA; <sup>5</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA and <sup>6</sup>Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA

**Dinitrogen (N<sub>2</sub>)-fixation by cyanobacteria in symbiosis with feathermosses is the primary pathway of biological nitrogen (N) input into boreal forests. Despite its significance, little is known about the cyanobacterial gene repertoire and regulatory rewiring needed for the establishment and maintenance of the symbiosis. To determine gene acquisitions and regulatory changes allowing cyanobacteria to form and maintain this symbiosis, we compared genomically closely related symbiotic-competent and -incompetent *Nostoc* strains using a proteogenomics approach and an experimental set up allowing for controlled chemical and physical contact between partners. Thirty-two gene families were found only in the genomes of symbiotic strains, including some never before associated with cyanobacterial symbiosis. We identified conserved orthologs that were differentially expressed in symbiotic strains, including protein families involved in chemotaxis and motility, NO regulation, sulfate/phosphate transport, and glycosyl-modifying and oxidative stress-mediating exoenzymes. The physical moss–cyanobacteria epiphytic symbiosis is distinct from other cyanobacteria–plant symbioses, with *Nostoc* retaining motility, and lacking modulation of N<sub>2</sub>-fixation, photosynthesis, GS-GOGAT cycle and heterocyst formation. The results expand our knowledge base of plant–cyanobacterial symbioses, provide a model of information and material exchange in this ecologically significant symbiosis, and suggest new currencies, namely nitric oxide and aliphatic sulfonates, may be involved in establishing and maintaining the cyanobacteria–feathermoss symbiosis.**

*The ISME Journal* (2017) 11, 2821–2833; doi:10.1038/ismej.2017.134; published online 11 August 2017

## Introduction

Dinitrogen (N<sub>2</sub>)-fixation by filamentous heterocystous cyanobacteria living in symbiosis with pleurocarpous feathermosses (for example, *Pleurozium schreberi* and *Hylocomium splendens*) and acrocarpous mosses (for example, *Sphagnum fuscum*) serves as the primary input of nitrogen (N) into boreal forests and can account for 50% of total N input (DeLuca *et al.*, 2002, 2008; Turetsky *et al.*, 2012; Rousk *et al.*, 2015). The boreal forest ecosystem occupies 11% of the Earth's terrestrial surface (Bonan and Shugart, 1989),

where feathermoss often account for >65% of the forest floor (DeLuca *et al.*, 2008). This partnership has a vital role in primary productivity of the boreal forest (Turetsky, 2003; Lindo and Gonzalez, 2010). High-resolution secondary ion mass spectrometry verified transfer of fixed N from the cyanobacteria to the moss host, demonstrating that the symbiosis governs the main N entrance to boreal forests (Bay *et al.*, 2013). Despite the significance of the symbiosis, knowledge of the interaction between the two partners is rudimentary and, in particular, little is known about the gene repertoire and regulation needed to form the symbiosis. No studies have examined global gene or protein expression of the moss or the cyanobacterial partner.

Current knowledge on colonization steps and maintenance of cyanobacterial–plant symbioses is primarily based on studies of endophytes,

Correspondence: U Rasmussen, Department of Ecology, Environment and Plant Sciences, Stockholm University, Svante Arrheniusvägen 20A, Stockholm 114 18, Sweden.  
E-mail: ulla.rasmussen@su.se

Received 28 April 2017; revised 20 June 2017; accepted 7 July 2017; published online 11 August 2017

specifically the liverwort *Blasia*, and hornworts *Anthoceros* and *Phaeoceros*, where the cyanobacteria, mainly *Nostoc* sp., are extracellular but enclosed in a specialized cavity, and with the angiosperm *Gunnera*, where the cyanobacteria are located intracellularly (reviewed in Santi *et al.*, 2013). During the establishment of these symbioses, there are two phases of interaction, an early phase, which includes chemical signaling between partners, followed by a later phase where the cyanobacteria are physically associated with the host.

In the early phase, the cyanobacteria differentiate from vegetative filaments into a transient motile stage termed hormogonia. The differentiation is induced by a variety of environmental stimuli and host-produced chemical signals of unknown structure, known as hormogonia-inducing factor, HIF (Rasmussen *et al.*, 1994; Meeks and Elhai, 2002; Meeks, 2003; Adams and Duggan, 2008). The feathermoss secretes HIF and possibly other chemoattractants to recruit and direct the hormogonia toward the host, according to the only study addressing feathermoss colonization (Bay *et al.*, 2013). Cyanobacterial genes responding to plant HIF have been identified, and many appear to be involved in signal transduction and transcriptional activation (Campbell *et al.*, 2007; Duggan *et al.*, 2013; Risser and Meeks, 2013; Campbell *et al.*, 2015). A recent study has identified *che*-family genes as important for both motility and symbiotic competence (Duggan *et al.*, 2013), whereas genes involved in phototaxis, chemotaxis and motility (*ptx*, *hmp*, *hps* and *pil* genes) are thought to be implicated in the colonization process (Campbell *et al.*, 2003, 2008; Risser and Meeks, 2013; Risser *et al.*, 2014).

The initial signaling and motility stage is followed by a physical interaction phase between host and cyanobacterium. In the symbiosis with hosts such as *Gunnera*, *Blasia* and *Anthoceros*, the cyanobacteria exhibit shared morphological and physical changes (reviewed in Adams and Duggan, 2012). Specific changes include repression of hormogonia development regulated by the *hrm* locus, increases in heterocyst frequency (*hetR* gene) and N<sub>2</sub>-fixation activity (*nif* genes), repression of GS-GOGAT activity, and reduced photosynthetic CO<sub>2</sub> fixation (reviewed in Adams and Duggan, 2012). These changes imply that the cyanobacterium shifts from photo-autotrophy to heterotrophic carbon metabolism, with organic C provided by host to the cyanobacteria to sustain N<sub>2</sub>-fixation (Meeks and Elhai, 2002). The carbohydrate exported from the host is not identified, but in *Gunnera* it was suggested to be glucose and/or fructose (Khamar *et al.*, 2010). In exchange, the cyanobacteria release 80–90% of their fixed N, mainly as ammonia for *Gunnera*, liverworts and hornworts symbioses (Meeks, 2009).

The aim of this study is to determine what gene acquisitions and regulatory rewiring allow cyanobacteria to form symbioses with feathermosses. To identify gene acquisitions, we sequenced and

compared the genomes of novel *Nostoc* isolates to target putative symbiotic-competent gene clusters. To target regulatory rewiring, we then selected closely related symbiotic-competent and -incompetent *Nostoc* strains to investigate gene, protein and exoprotein expression during contact with the feathermoss. To distinguish effects of different phases of colonization on gene and protein expression, we used a novel experimental system to isolate physical versus chemical contact conditions with the host (Bay *et al.*, 2013). We mined the gene expression of shared orthologs using random forest analysis, a supervised machine learning analysis based on decision tree predictors to identify expression signatures of genes that best discriminate symbiotic cyanobacteria from the non-symbiotic. The cellular and extracellular proteomes were also examined for patterns associated with cyanobacterial adaptation to the moss contact. We hypothesized that (1) sets of genes are absent from the genome of symbiotic-incompetent *Nostoc* and that these genes are of importance for initiating and maintaining a symbiotic association with feathermoss, (2) symbiotic-competent *Nostoc* strains exhibit a distinct expression profile in response to contact with the moss, as compared with incompetent strains, (3) competent *Nostoc* foster a distinct extracellular environment when in chemical contact with the host and (4) the epiphytic location of the symbiosis on feathermoss involves different physiological and metabolic processes during colonization relative to the endophytic symbioses in the liverwort/hornwort and *Gunnera*.

## Materials and methods

### *Sampling, cyanobacteria isolation and growth conditions*

Two feathermosses, *P. schreberi* and *H. splendens*, were randomly collected in June and September 2013 at the site of Ruttjeheden in Reivo boreal forest nature reserve in northern Sweden (65°80'N–19°10'E). Mosses were maintained on a thin layer of vermiculite in the greenhouse at 10 °C and light 35 μmol photons m<sup>-2</sup> s<sup>-1</sup> (see Supplementary Methods for description of the isolation procedure). Individual cyanobacteria isolates were grown in liquid BG11<sub>o</sub> media (Rippka *et al.*, 1979) with constant shaking under continuous light at 35 μmol photons m<sup>-2</sup> s<sup>-1</sup> and 20 °C. *Nostoc punctiforme* ATCC 29133 and *Nostoc* sp. CALU 996 (*N.* 996), a free-living symbiotic-incompetent soil isolate (culture collection of St Petersburg State University, St Petersburg, Russia), which is phylogenetically similar to symbiotic *Nostoc* (Papaefthimiou *et al.*, 2008) were also grown as liquid cultures in BG11<sub>o</sub>.

### *DNA extraction, genome sequencing and analysis*

DNA was extracted from *N.* 996 and five *Nostoc* strains isolated from mosses (Supplementary Table S1, Supplementary Methods for description)

and sequenced using PacBio technology at the Joint Genome Institute (JGI) (see Supplementary Methods and Supplementary Figures S1 and S2 for description of genome sequencing, annotation, assembly and genome closure of *N. 996*). After kmer binning to remove low abundance contaminant contigs (Laczný *et al.*, 2015), the chromosome and plasmid assemblies for the newly sequenced cyanobacteria were used to build a concatenated core gene phylogeny across all cyanobacteria. The existing phylogeny from the cyanoGEBA project (Shih *et al.*, 2013) was constructed using a concatenated alignment set of proteins from 126 cyanobacterial genomes (Wu and Eisen, 2008) and provided a validated starting alignment. The existing cyanoGEBA alignment was used to construct HMMs for each core protein (HMMER3) (Eddy, 2011), with subsequent searches of the new genomes and retrieval of the correct proteins. The new genomes were added to the alignment as concatenated core protein strings, with a reconstruction of the whole genome phylogeny (Dupont *et al.*, 2012) using RaxML (Stamatakis, 2014). All alignments and subsequent trees were manually examined with subsequent re-generation. The HMMER-based protein family annotations were used to compare the presence, absence and relative abundance of protein domains between all of the cyanobacterial genomes. To define orthologous protein families (OPFs) shared between *Nostoc* sp. Moss2 (*N. Moss2*), *N. punctiforme* and *N. 996* (OPF<sub>all</sub>) and between symbiotic strains (OPF<sub>symb</sub>), we used reciprocal best BLAST hit clustering with predicted amino-acid protein sequences (see Supplementary Methods).

#### *Experimental set up for chemical and physical contact*

To mimic the chemical and physical contact of the cyanobacteria–moss symbiosis, experiments between *P. schreberi* and *N. Moss2*, *N. 996* or *N. punctiforme* were performed as described by Bay *et al.* (2013) with slight modifications. For each cyanobacteria strain, triplicate cultures were set up with 1 and 8 µm membrane pore size filters (Corning, Inc., Corning, NY, USA) positioned above the cyanobacteria (for chemical and physical contact, respectively). In addition, triplicates were set up for moss alone (without cyanobacteria) and for each of the cyanobacteria (no mosses) (see Supplementary Methods). The chemical contact experiment was repeated to gain enough material for exoproteomic analyses, with a merging of media from both experiments before analyses. After 24 h, cyanobacteria and filtered spent media from the 1 µm experiment were harvested by centrifugation at 4 °C and promptly frozen in liquid N, and stored at –80 °C until RNA and protein extractions. The mosses from the 8 µm experiment were harvested when colonization was noticed (4–5 weeks), frozen in liquid N and stored at –80 °C until RNA extraction

(see Supplementary Methods for RNA and protein extraction procedure).

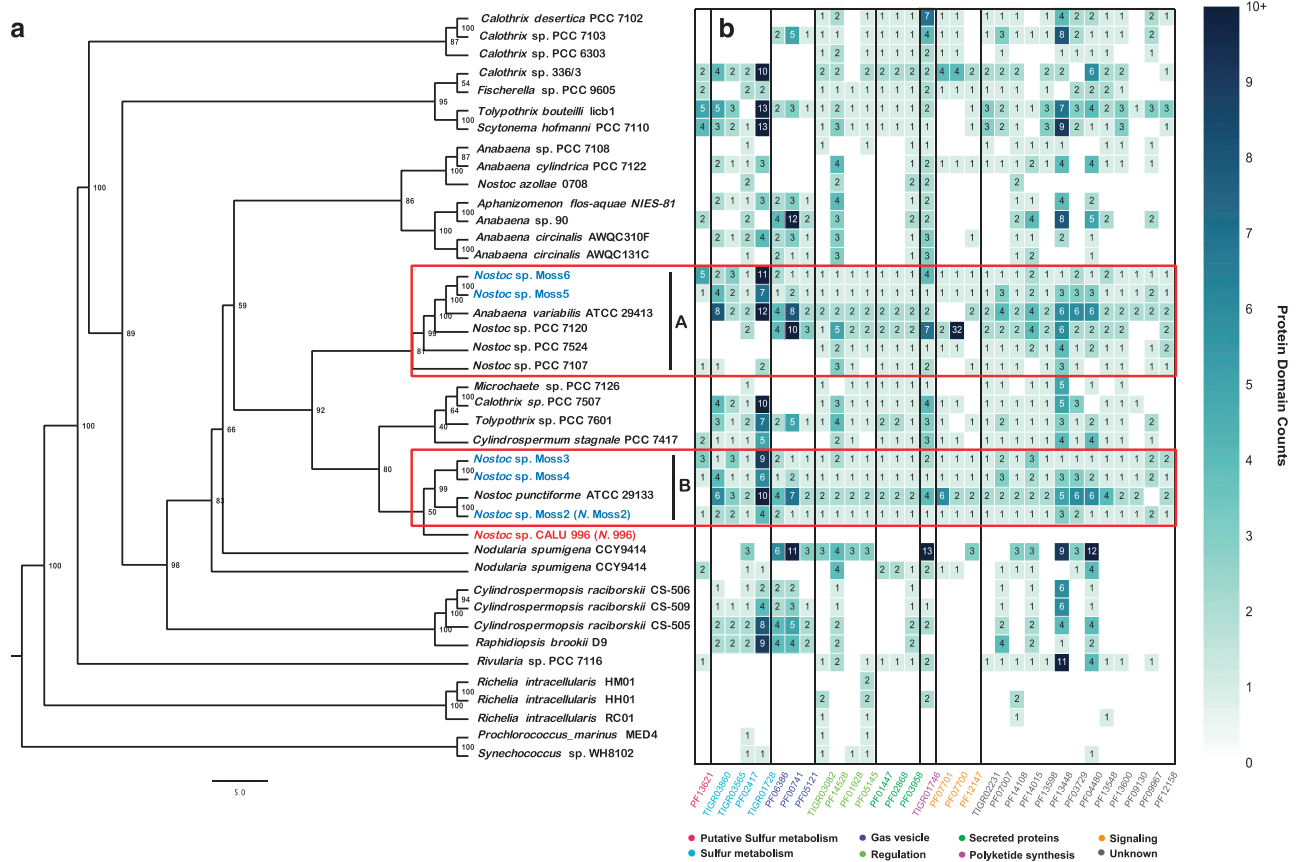
#### *Transcriptome analysis*

The moss–cyanobacteria physical contact transcriptome was treated as a mini-metatranscriptome (Supplementary Methods for description of transcriptomes sequencing, annotation and assembly). For all *Nostoc*-specific mapping reads, raw transcriptomic read mappings were normalized using TMM (Robinson and Oshlack, 2010) and analyzed for differentially expressed genes using the Bioconductor package EdgeR (Robinson *et al.*, 2010). For symbiotic-competent cyanobacteria, there were two pairwise comparisons: (i) cyanobacteria in isolation versus chemical contact with *P. schreberi* and (ii) cyanobacteria in isolation versus physical contact with *P. schreberi*. For *N. 996*, the only possible pairwise comparison was the cyanobacteria in isolation versus chemical contact with *P. schreberi*.

A random forest analysis of the gene expression data was also used to identify putative gene expression patterns statistically associated with chemical or physical contact (see Supplementary Methods). The differential gene expression of profile-identified putative symbiotic genes defined by the random forest classification, hereafter referred to as PIPS genes, was visualized using AutoSOME software and co-expression modules were identified as described by Newman and Cooper (2010). Genomic locations of OPF<sub>all+symb</sub> orthologs were manually curated and spatially colocalized regulons (SCRs) were defined as sets of orthologs genomically localized with a given PIPS gene; we examined the transcription profile of flanking genes in either direction in the genome of *N. punctiforme*, *N. Moss2* and *N. 996*, cut-off distance of a flanking region was decided based on manual curation. SCRs were compared between the three strains, and spatially conserved and co-regulated SCRs (based on transcriptomics and/or proteomics) in the symbiotic cyanobacteria were selected for further analysis.

#### *Proteome and exoproteome analysis*

Extracted proteins from cyanobacteria alone (that is, *N. 996*, *N. punctiforme*, *N. Moss2*) and cyanobacteria–moss in chemical contact (that is, *N. 996*, *N. punctiforme*, *N. Moss2*) were prepared for proteome analyses as described in Supplementary Methods. Raw mass spectrometry data were converted to peak lists (DTA files) using DeconMSn (Mayampurath *et al.*, 2008) and searched with MSGF+ (Kim and Pevzner, 2014) against predicted proteins from the JGI-annotated genome for each species as the search database, except for *N. punctiforme*, which utilized the Uniprot *N. punctiforme* ATCC 29133 genome. The identified spectra were filtered based on their MSGF+ scores



**Figure 1** Nostocales species tree and distribution of HMM occurrences missing in *Nostoc* sp. CALU 996 (*N. 996*). (a) Maximum-likelihood phylogeny of Nostocales based on 15 conserved protein comparisons of the five moss symbiotic-competent strains sequenced in our study (in blue), the one non-symbiotic competent (in red) and 31 other Nostocales sequenced during the cyanoGEBa project (Shih *et al.*, 2013). The genomes of *Prochlorococcus marinus* MED4 and *Synechococcus* sp. WH8102 were included as outgroups. Bootstrap values (out of 100) are shown at each node. Red boxes show subclades A and B, which include the new symbiotic-competent strains. (b) Distribution of the HMM occurrences missing in *N. 996*. Columns in the heat map represent individual gene families (Pfam or TIGRfams) grouped into eight functional protein categories. Increasing representation of each gene family in a given cyanobacteria strain is shown by increasing protein domain counts and depth color.

and only the proteins with two proteo-specific peptides were conserved resulting in a protein and peptide false discovery rate < 1%. For the quantitative analysis, the tandem mass tag reporter ion intensities were extracted with MASIC (Monroe *et al.*, 2008). Tandem mass tag reporter intensities were summed from the different peptides belonging to the same proteins.

For exoprotein extraction, 1.5–2 ml spent media from five replicate incubations of moss alone and moss–cyanobacteria in chemical contact for 24 h were analyzed for each *N. Moss2*, *N. 996* and *N. punctiforme* (30 samples total). Scaffold (Proteome Software Inc., Portland, OR, USA) was used to validate peptide and protein identified by tandem mass spectrometry (MS/MS) and to calculate normalized weighted spectra values (NSAF). NSAF values were then averaged between replicates and log<sub>2</sub> fold change values calculated between the two treatments. A Student’s *t*-test was used to determine differences between the treatments ( $P < 0.05$ ). (For a detailed description of extraction methods, search

database and peptide identification thresholds see Supplementary Methods).

## Results

### Comparative genomic analysis

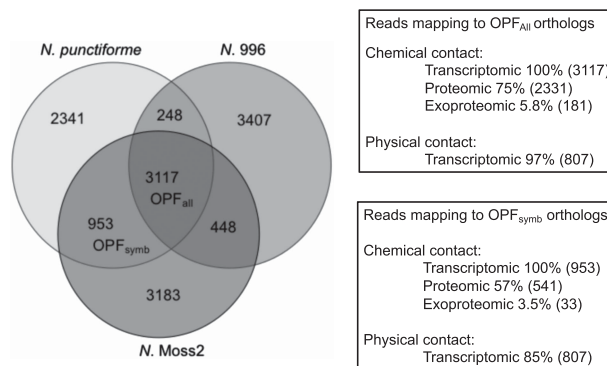
To determine if symbiotic-competent strains have unique genome contents, we sequenced five *Nostoc* strains that were isolated from mosses. For comparative genomics, we also sequenced one free-living *Nostoc* strain (*N. 996*), which we have experimentally verified as lacking symbiotic competence with *P. schreberi*, *Blasia pusilla* and *Gunnera manicata*. To establish the evolutionary relatedness of the six newly sequenced cyanobacteria, the sequences were included with 31 other Nostocales genomes in a phylogenomic inference built from the concatenated alignments of 15 conserved proteins (Figure 1a). The sequenced cyanobacteria belong to one clade, which consists of 12 strains currently classified as Nostocaceae, two as Microchaetaceae (*Microchaete* sp.

PCC 71269 and *Tolypothrix* sp. PCC 7601) and one as Rivulariaceae (*Calothrix* sp. PCC 7507). The five moss-derived isolates were divided into two subclades within the Nostocales. The first subclade (A on tree) contained two of the isolates, *Nostoc* sp. Moss6 and *Nostoc* sp. Moss5 and were closely related to the free-living cyanobacteria *Anabaena variabilis* ATCC 29413 and *Nostoc* sp. PCC 7120. A second subclade (B) contained the other three isolates, *Nostoc* sp. Moss3, *Nostoc* sp. Moss4 and *N. Moss2*, which grouped together with the symbiotic strain *N. punctiforme* ATCC 29133. The symbiotic-incompetent strain *N. 996* is a weakly supported (50%) basal outgroup for subclade B (Figure 1a).

We performed protein family profiling to identify gene families that were retained in the symbiotic strains but absent from the symbiotic-incompetent *N. 996*. We identified 32 gene families that lacked an HMM hit in *N. 996* but were found in all feathermoss symbiotic isolates (Figure 1b, Supplementary Table S2). Many of the 32 gene families have non-redundant functional roles in gas vesicle formation, gene regulation and signaling, and sulfur metabolism. Three of these gene families have predicted functions in gas vesicle formation (PF00741, PF05121 and PF06386). Among the seven regulatory domains are a heme-nitric oxide (NO; PF07700) binding protein, and a heme NO-binding associated protein (PF07701). Domains involved in sulfur metabolism include putative aliphatic sulfonate transporter substrate binding protein (TIGR01728), and a putative aliphatic sulfonate monooxygenase (TIGR03565). The aliphatic sulfonate transporter substrate binding protein was the most abundant differentially retained gene family found in symbiotic cyanobacteria, ranging from four protein domain counts in the *N. Moss2* genome to 11 counts in *Nostoc* sp. Moss6 (Figure 1b, Supplementary Table S2).

#### Transcriptional and proteomic regulation by *Nostoc* in physical or chemical contact with *P. schreberi*

To better understand the changes in gene regulation occurring during different phases of the cyanobacterial association with the moss, we performed transcriptomic, proteomic and exoproteomic analyses on triplicate cultures of *N. Moss2*, *N. punctiforme* and *N. 996* under three conditions: cyanobacteria in isolation, in chemical contact with the moss and in physical contact with the moss (Supplementary Dataset S1 and Supplementary Table S3). To compare gene and protein expression between different strains and genomes, it is necessary to know the orthology of the different genome contents. Therefore, the predicted proteins from the genomes of *N. punctiforme*, *N. Moss2* and *N. 996* were clustered into orthologous groups (Supplementary Table S4). OPFs shared between all three strains (OPF<sub>all</sub>) consisted of 3117 orthologs, and 953 orthologs were only shared between *N.*

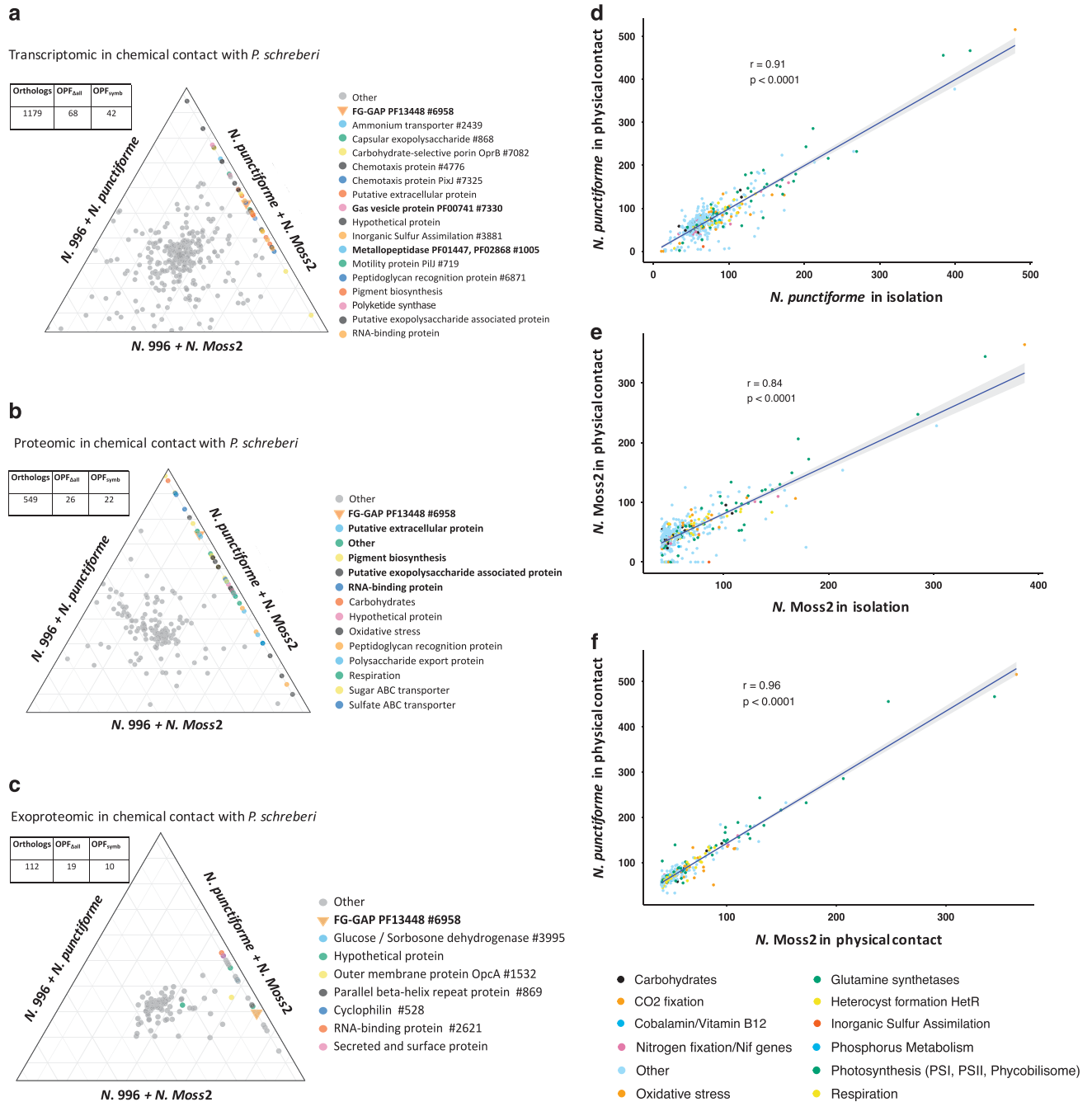


**Figure 2** Venn diagram showing shared and unique orthologs between *N. Moss2*, *N. punctiforme* and *N. 996*. Orthologs shared between the three *Nostoc* were classified as OPF<sub>all</sub> and orthologs shared only between symbiotic strains as OPF<sub>symb</sub>. The percentage of each OPF category detected in the different datasets is shown in boxes.

*punctiforme* and *N. Moss2* (OPF<sub>symb</sub>) (Figure 2, Supplementary Table S4). For the transcriptomics in chemical contact, all of the OPF<sub>all</sub> and OPF<sub>symb</sub> orthologs were detected. Despite the overwhelming representation of moss transcripts in the physical contact treatment data set, we identified reads mapping to 97–85% of OPF<sub>all</sub> and OPF<sub>symb</sub> orthologs during physical contact with *P. schreberi* (Figure 2). Proteomics identified 75% and 57% of OPF<sub>all</sub> and OPF<sub>symb</sub>, respectively, and 5.8% and 3.5% of the orthologs were detected in the exoproteome (Figure 2). The depth of transcriptomic and proteomic coverage allows for maximal data input to analyses.

#### Most abundant orthologous transcripts and proteins

We identified the 10% highest expressed OPF<sub>all</sub> and OPF<sub>symb</sub> at the transcript, protein and exoprotein levels for each strain, based on average normalized counts in chemical and physical contact with the moss. This included 1179 transcripts across all cyanobacteria in chemical contact, 549 proteins and 112 exoproteins (Figures 3a–c). Figures 3a–c show the orthologs that are present in OPF<sub>all</sub> and OPF<sub>symb</sub>. *N. Moss2* and *N. punctiforme* shared a unique complement of abundant transcripts and proteins belonging to OPF<sub>symb</sub> and also a set included in OPF<sub>all</sub> that are more highly expressed in symbiotic-competent cyanobacteria than *N. 996* (Figures 3a–c). The OPF<sub>symb</sub> orthologs are involved in pigment biosynthesis, exopolysaccharide production, extracellular proteins and RNA-binding proteins (Figures 3a and b). Three specific gene families missing from *N. 996* were highly expressed, a protein of unknown function (PF13448), a secreted protein (metallopeptidase PF01447, PF02868), and one gas vesicle protein (PF00741) (Figure 3a). Only the PF13448 containing protein was found as highly produced in the proteome and exoproteome of symbiotic cyanobacteria (Figures 3b and c).



**Figure 3** Global comparison of gene and protein expression between the three *Nostoc* strains. The triplots visually present the OPF<sub>all+symb</sub> orthologs that were among the 10% most abundant orthologs in each strain, based on average normalized counts. (a) Transcriptomes, (b) proteomes and (c) exoproteomes. For each triplot, the number of abundant orthologs, OPF<sub>all</sub> orthologs not represented in *N. 996* and OPF<sub>symb</sub> orthologs is shown in the box. The colors correspond to the ortholog group functions/annotation and marked in bold if identified as missing protein family in *N. 996*; the closer the symbol is to the node of the triplot the more abundant the transcript or protein is in the particular cyanobacterium. Functions in the middle are shared between the three strains and those on the edge are shared between two cyanobacteria. ‘#’ indicate the ortholog group identifier for unique function/annotations. Parts (d-f) show correlation of the 10% most abundant OPF<sub>all+symb</sub> orthologs expressed in the transcriptomic data when in isolation and in physical contact for *N. punctiforme* (d) for *N. Moss2* (e) and the two symbiotic strains post-colonization (f). Axes represent the average normalized counts. The colors correspond to selected ortholog group functions.

To investigate how cyanobacteria adapt their lifestyle post-colonization, we compared the 10% most abundant OPF<sub>all</sub> and OPF<sub>symb</sub> orthologs when in isolation (that is, free-living N-fixing state) with the most abundant orthologs identified in physical

contact (that is, symbiotic N-fixing state). In both symbiotic strains, most of the highly expressed orthologs in isolation are also significantly highly expressed in physical contact (Figures 3d-f). OPF<sub>all+symb</sub> orthologs highly expressed in both phases

were involved in CO<sub>2</sub> fixation (*rbcLXS*), photosynthesis (photosystem I *psaABCDEFGHIJKLM*, photosystem II *psbABCDE*) and cytochrome b6-f complex (*petABCD*), GS-GOGAT (the glutamine synthetase *glnA* and the glutamate synthase *gltB*), phosphorus assimilation (*ptsS*), oxidative stress (peroxiredoxin and glutaredoxin synthesis), heterocyst differentiation (*hetR*), and N<sub>2</sub>-fixation (*nifBHKU*). This represents a major departure from endophytic cyanobacterial–plant symbioses.

#### Symbiotic-competent exoprotein composition

Of the 214 orthologs detected in the exoproteomes of symbiotic-competent strains, 33 are unique to OPF<sub>symb</sub> (Figure 2, Supplementary Dataset S1, Supplementary Table S5). OPF<sub>symb</sub> orthologs included a putative pectate lyase (polysaccharide lyase family 9 CAZy domain, cluster #869), and hemolysin type Ca-binding protein (cluster #7058; also identified as PIPS gene, see below). Both proteins were upregulated in the exoproteome when in contact with the host, and were in the top 10% abundance in the cellular proteome. Of the OPF<sub>all</sub> exoproteins, several involved in oxidative stress response were differentially expressed in the competent strain exoproteomes, including a catalase protein (cluster #2351; PIPS gene), upregulated under chemical contact in the *N. punctiforme* exoproteome and not detected in *N. 996*. This catalase was also transcriptionally upregulated in both *N. Moss2* and *N. punctiforme* under physical contact (Supplementary Dataset S1, Supplementary Table S5). In addition, a ferritin protein (cluster #5454; PIPS gene) and a flavorubredoxin (cluster #2363) were detected in all three exoproteomes, but only upregulated in the presence of the moss in the competent strains.

#### Machine learning identification of symbiosis-specific expression patterns

A supervised machine learning approach, random forest analysis with decision trees, was used to

identify expression patterns of the subset of OPF<sub>all+symb</sub> orthologs that best differentiated the symbiotic cyanobacteria *N. Moss2* and *N. punctiforme* (4070 orthologs, Supplementary Dataset S1) from the non-symbiotic *N. 996* transcriptome across physical and chemical contact conditions. We identified 405 PIPS genes (282 OPF<sub>all</sub> and 123 OPF<sub>symb</sub> orthologs) that together predicted the cyanobacterial symbiotic state with 98.3% confidence (Supplementary Table S6).

We conducted unsupervised clustering of the expression profile of PIPS genes using self-organizing maps (Newman and Cooper, 2010), which revealed 21 colonization phase-dependent gene co-expression modules (Supplementary Figure 3). The second large module (35 OPF<sub>all</sub> and 70 OPF<sub>symb</sub>) was upregulated PIPS genes during chemical contact with the moss, and was particularly enriched in genes involved in gas vesicle organization and chemotaxis (Supplementary Figure 3a). The largest module (164 OPF<sub>all</sub> and 17 OPF<sub>symb</sub> orthologs) comprised PIPS upregulated genes during the physical phase and was enriched with GO terms involved in reactive oxygen species homeostasis, photosynthesis, c-di-GMP biosynthetic process and sugar, phosphorus and sulfur metabolisms (Supplementary Figure 3c). The other 19 modules were composed of genes similarly regulated during both phases and merged into one cluster, which contained 83 OPF<sub>all</sub> and 36 OPF<sub>symb</sub> orthologs with genes enriched in GO terms associated with chemotaxis, type II secretion system and pili genes (Supplementary Figure 3b).

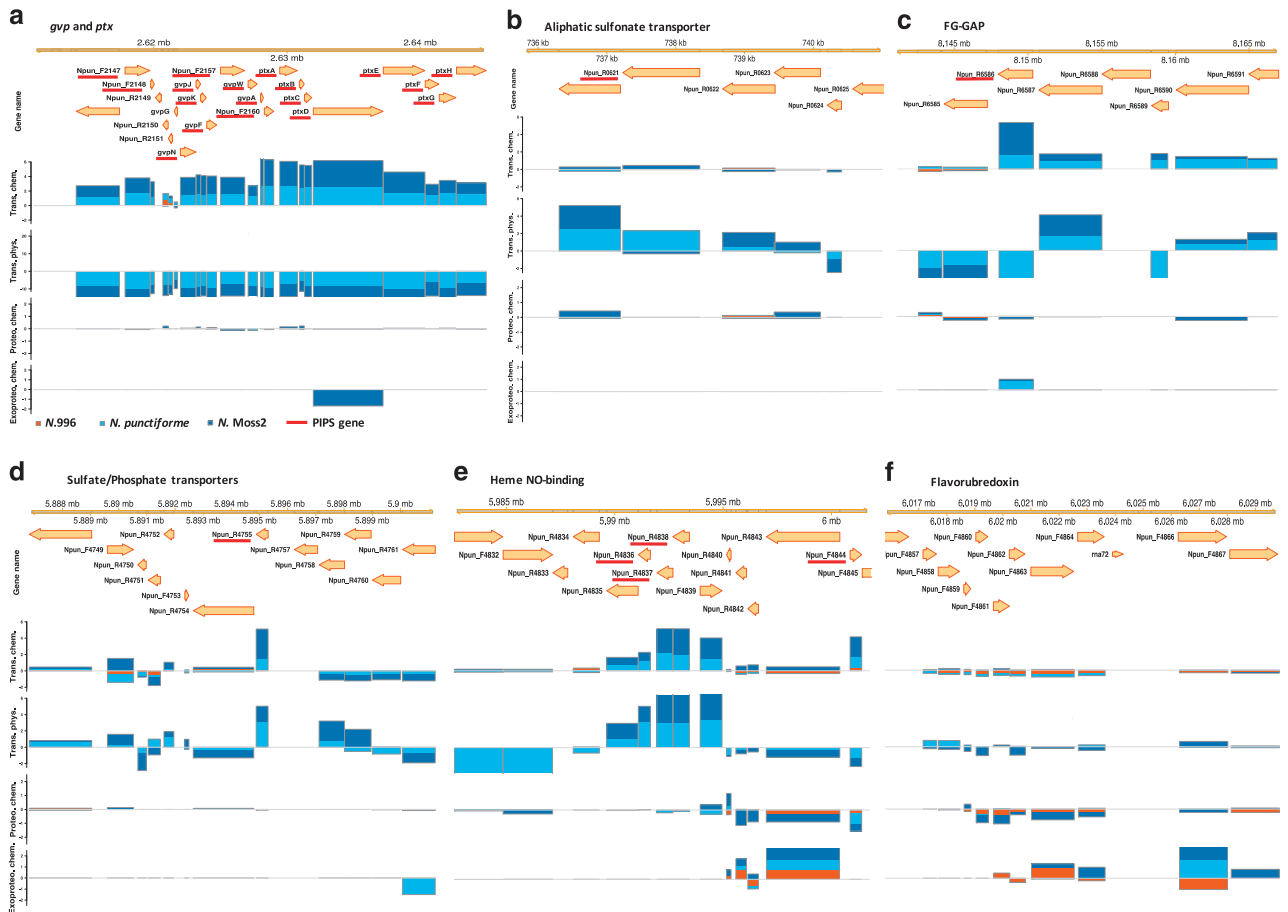
#### SCRs shared between the strains

We examined the genomic spatial organization of the 4070 OPF<sub>all+symb</sub> orthologs and combined it with the differential gene expression analyses to identify spatially colocalized and co-regulated sets of genes (SCRs) shared in *N. Moss2* and *N. punctiforme* and genes differentially expressed between the symbiotic-competent and -incompetent strains (Table 1, Figure 4, Supplementary Table S7). For a

**Table 1** SCRs identified in *N. Moss2*, *N. punctiforme* and *N. 996*

SCR function	Interaction phase	SCR name	SCR statistics
Chemotaxis and motility	Chemical contact	<i>gvp</i> and <i>ptx</i>	22 genes; 20 OPF <sub>symb</sub> ; 18 PIPS genes
	Chemical and physical contact	<i>hrm</i>	6 genes; 4 OPF <sub>symb</sub> ; 2 PIPS genes
		<i>pix</i>	13 genes; 6 OPF <sub>symb</sub> ; 6 PIPS genes
		<i>hmp</i>	8 genes; 7 OPF <sub>all</sub> ; 1 PIPS gene
		<i>hps</i>	15 genes; 15 OPF <sub>all</sub> ; 9 PIPS genes
		<i>pil</i> I	15 genes; 13 OPF <sub>all</sub> ; 4 PIPS genes
Sensing and signal transduction Biosynthetic	Chemical and physical contact	<i>pil</i> II	31 genes; 30 OPF <sub>all</sub> ; 8 PIPS genes
		Heme NO-binding	12 genes; 6 OPF <sub>symb</sub> ; 4 PIPS genes
	Chemical and physical contact	FG-GAP	6 genes; 5 OPF <sub>symb</sub> ; 1 PIPS gene
		Flavorubredoxin	10 genes; 10 OPF <sub>all</sub> ; 0 PIPS genes
		Glycosyltransferase	11 genes; 11 OPF <sub>all</sub> ; 2 PIPS genes
		Fructose/mannose metabolism	13 genes; 13 OPF <sub>all</sub> ; 2 PIPS genes
Transport	Physical contact	Aliphatic sulfonate transporter	5 genes; 4 OPF <sub>symb</sub> ; 1 PIPS gene
	Physical contact	Sulfate/phosphate transport	12 genes; 8 OPF <sub>all</sub> ; 1 PIPS gene

Abbreviations: FG-GAP, FG-GAP repeat-containing protein; NO, nitric oxide; OPF, orthologous protein family; PIPS, profile-identified putative symbiotic; SCRs, spatially colocalized regulons.



**Figure 4** Gene and protein expression of select conserved genomic regions. Genomic regions correspond to *N. punctiforme* genomic loci of the *gvp* and *ptx* (a), aliphatic sulfonate (b), FG-GAP repeat-containing protein (c), sulfate/phosphate transporters (d), Heme NO-binding (e) and flavorubredoxin SCRs (f). Note that the sulfate/phosphate transporters (d), Heme NO-binding (e) and flavorubredoxin (f) SCRs are also colocalized with each other. Log fold changes presented are the regulation compared with cyanobacteria in isolation for *N. Moss2* (blue), *N. punctiforme* (light blue) and *N. 996* (red). PIPS genes for each SCR are underlined in red. Chem., chemical contact; Exoproteo., exoproteomic; Phys., physical contact; Proteo., proteomic; Trans., transcriptomic.

detailed description of the SCRs, see Supplementary Results. Overall, we identified 14 SCRs, with two that were regulated in chemical contact with *P. schreberi*, nine SCRs differentially expressed during both chemical and physical contact and three differentially expressed only during physical contact (Table 1, Supplementary Table S7). The SCRs were mostly mono-functional, as expected given the genomic colocalization, and separately involved in chemotaxis and motility, sensing and signal transduction and macromolecule biosynthesis and transport. We found four SCRs composed of OPF<sub>all</sub> orthologs, which were differentially expressed in *N. 996* compared with symbiotic strains. Those SCRs were associated with hormogonia motility (*hmp*, *hps*, *pil I*, *pil II* and glycosyltransferase) (Table 1, Supplementary Table S7). Six SCRs were composed of gene families missing from *N. 996*, involved in chemotaxis (*gvp* and *ptx*, and *pix*), motility (*pil II*), sensing of NO (heme NO-binding) and other functions (FG-GAP repeat-containing protein and aliphatic sulfonate transport) (Figure 4, Table 1, Supplementary Table S7).

## Discussion

### Overview and hypotheses restatement

With any data-rich study, experimental design and explicit hypotheses focus a robust analysis. Here, the genomes of closely related *Nostoc* strains, five competent and one incompetent for symbioses with feathermosses, were sequenced and compared. From our phylogenetic analysis, the feathermoss *Nostoc* isolates are divided in two subclades suggesting the taxonomy of *Nostoc* and related genus need further revisions (Komárek, 2010). The feathermoss isolate most closely related to the model symbiotic strain, *N. punctiforme*, and the closely related symbiotic-incompetent strain were chosen for further global gene and protein expression experiments. The samples for these analyses originated from a unique culturing framework allowing for controlled and replicated sampling of cyanobacteria under three conditions: in isolation, in chemical contact with *P. schreberi* (early colonization phase), and in physical contact with *P. schreberi* (late colonization phase). Our hypotheses framing the data analyses



were that (1) symbiotic-competent *Nostoc* strains contain symbiosis-related gene families not found in incompetent strains, (2) symbiotic-competent *Nostoc* have a shared expression profile in response to contact with *P. schreberi*, distinct from incompetent strains, (3) symbiotic-competent strains foster a different extracellular environment when in chemical contact with the mosses and (4) the epiphytic nature of the cyanobacteria–moss symbiosis is unique because of novel regulation of metabolism and phenotype, as well as molecular currency exchange. The first hypothesis entails fundamental evolutionary adaptations shaped by gene acquisition or selective gene retention, whereas the second and third hypotheses require regulatory reprogramming and likely involve host-based signaling. The final hypothesis challenges singularity in cyanobacterial host adaptations, thereby refining existing theory to create a gradient of molecular adaptations based upon defined criteria of physical contact, namely endophyte, both intra and extracellular, and epiphyte. Each of these hypotheses will be addressed with regards to the experimental results below, and based on interpretations of integrated results, we will detail a model of new molecular currency exchanges between host and symbiont.

#### *Gene families specific to symbiotic strains*

We found that symbiotic-competent cyanobacteria genomes contain 32 gene families not found in closely related strains including the symbiotic-incompetent *N. 996*. These families were significantly expressed during the early colonization phase and are enriched in functional attributes such as chemotaxis and motility, signaling, and aliphatic sulfonate transport. This includes known molecular participants in cyanobacteria–plant symbioses. For example, seven upregulated gas vesicle genes (*gvp*), here predicted as PIPS genes, previously showed transcriptional upregulation in *N. punctiforme* upon addition of HIF, indicating a role in cyanobacterial–plant symbioses establishment (Campbell *et al.*, 2008; Christman *et al.*, 2011; Risser *et al.*, 2014). Deletion of *gvpA* showed that this gene is essential to *N. punctiforme* motility on a solid substrate (Risser *et al.*, 2014). The upregulation of *gvp* and *ptx* genes specifically during the early phase indicate that *P. schreberi* activates a signal transduction pathway for hormogonia motility unique to this phase. By contrast, another taxis locus *pix*, which is missing in *N. 996*, was upregulated during both colonization phases, suggesting phase-dependent signaling pathways. A FG-GAP repeat-containing protein (Npun\_R6586) was also only present in the genomes of the symbiotic-competent strains, and found among the 10% most abundant transcripts, cellular proteins, and exoproteins during chemical contact with *P. schreberi*. The FG-GAP repeat-containing protein may have a role in the establishment of symbiosis, and is clearly secreted, given its

abundance in the exoproteome. Two more gene families missing in *N. 996* genome are involved in NO signaling. The heme-NO binding and a NO-associated (annotated as a histidine kinase) genes were identified as PIPS genes and were upregulated during chemical and physical contact, suggesting a role of NO in signaling during establishment and maintenance of the symbiosis. Increased NO concentration may lead to upregulation of the heme-NO binding gene and c-di-GMP synthesis genes inducing biofilm and adhesion of cyanobacteria to the host. Rao *et al.* (2015) showed that *Silicibacter* sp. TrichCH4B and the marine cyanobacterium *Trichodesmium erythraeum* have a NO-mediated mechanism for symbiosis, in which *T. erythraeum* produces a chemical signal, inducing NO formation in the symbiont. NO activates the heme-NO binding protein signaling pathway of *Silicibacter* sp. TrichCH4B, leading to higher levels of cellular c-di-GMP controlling adhesion and biofilm formation. As NO synthase genes were not identified in our *Nostoc* strains, the host would be the likely source of NO. There is evidence that *P. schreberi* NO synthase gene was upregulated when in physical contact with *N. Moss2* and *N. punctiforme* and accumulation of NO was observed by fluorescence probing with 4,5-diaminofluorescein-2 diacetate (unpublished work). Biofilm formation is often critical to host–symbiont communication, especially when it is mediated by diffusible signaling molecules (Norsworthy and Visick, 2013), so biofilm formation may favor host–symbiont NO-mediated communication and enhance currency exchange. NO-mediated signaling may also allow the host to guide cyanobacteria migration and colonization of new plant tissue as NO is highly produced in protonema cells during the growth of the moss *Physcomitrella patens* (Medina-Andres *et al.*, 2015).

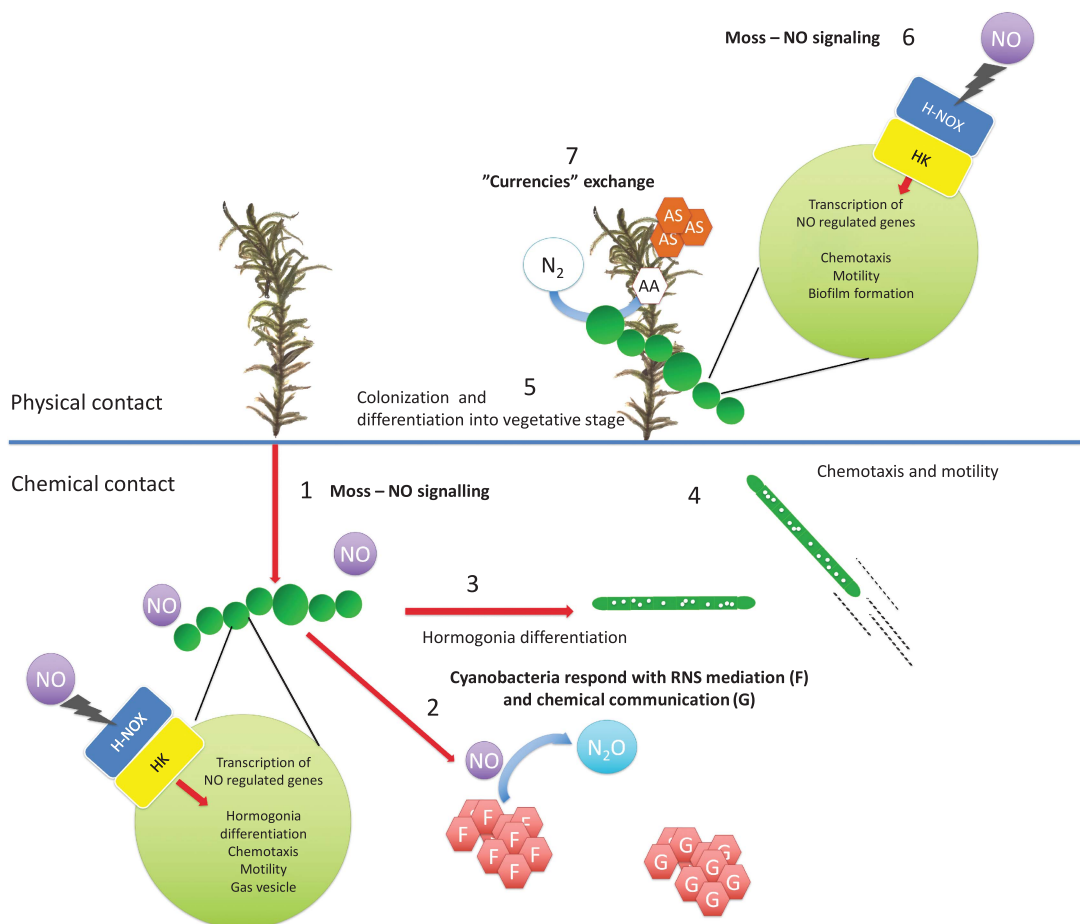
In support of our second hypothesis, we found that OPFs shared by both symbiotic-competent and -incompetent cyanobacteria are expressed differently in contact with the moss. Our random forest analysis shows that more than half of the 405 PIPS genes, those with expression profiles diagnostic of the phenotype, are OPF<sub>all</sub> orthologs. These are 282 genes present in the symbiotic-incompetent strain that have been transcriptionally reprogrammed. In particular, genes involved in hormogonia motility (*hmp*, *hps*, *pil I*, *pil II* and glycosyltransferase), further classified into SCRs, were upregulated in the two competent strains but not in *N. 996*. *N. punctiforme* motility involves co-ordinated expression of the *pil* genes encoding a type IV pilus that acts as a nanomotor, driving motility on a secreted polysaccharide layer encoded by the *hps* genes, with hormogonium gliding regulated by the *hmp* locus (Duggan *et al.*, 2007; Risser and Meeks, 2013; Risser *et al.*, 2014; Khayatan *et al.*, 2015). The glycosyltransferase SRC may also be involved in the motility process by producing an excreted polysaccharide. The motility loci were also

upregulated during physical contact with *P. schreberi*, which we propose are because of the cyanobacteria being epiphytic, retaining motility and chemotaxis allowing for colonization of new parts of the moss. Along with upregulation of the motility loci, the *hrm* locus, which controls hormogonia repression (Cohen and Meeks, 1997; Meeks, 1998; Campbell *et al.*, 2003), was downregulated or not detected in physical contact with *P. schreberi*. Our results suggest that although *N. 996* shares many of the genes coding for motility, chemotaxis and cellular differentiation, these are insensitive to the chemical signaling derived from *P. schreberi*. This differential expression of a shared gene set is a concrete indication of moss-derived chemical signaling of regulatory reprogramming in the symbiotic-competent strains. Such organism-specific information exchange is one hallmark of symbiotic interactions.

Exoprotein composition provides support for our third hypothesis, which the symbiotic-competent strains foster a distinct extracellular environment when in chemical contact with their host, specifically through glycosyl modifications and oxidative stress mediation. In cyanobacteria, hydrolytic exoproteins in biofilms support heterotrophic acquisition of complex polymers, whereas in a free-living cyanobacterium diverse exoenzymes, such as hemolysin, virulent factor-like metalloproteases and chitinases, may indicate competitive interactions dominate (Christie-Oleza *et al.*, 2015; Stuart *et al.*, 2016). Our data provide evidence for specific exoenzymes produced by the symbiotic-competent cyanobacteria in response to the moss. For example, microbial pectate lyases have been implicated in plant pathogenesis (reviewed in Payasi *et al.*, 2009), and legume-produced pectate lyases allow symbiont-rhizobia infection (Xie *et al.*, 2012). Our symbiont-specific putative pectate lyase exoprotein may perform a similar function, assisting in colonization of the feathermoss surface through cell wall modifications. We also found that *hpsJ*, involved in polysaccharide synthesis, is downregulated in the exoproteome of *N. punctiforme* providing evidence for a modification of the exopolysaccharide composition during interaction with the moss. Deletion of *hpsJ* affects polysaccharide composition, indicating the role of this gene in defining the nature of hormogonium polysaccharide (Risser *et al.*, 2014; Khayatan *et al.*, 2015). Among the 10% most abundant exoproteins expressed in chemical contact, we also found several linked to oxidative stress. Specifically, we found two flavorubredoxins upregulated in the exoproteome of both symbiotic-competent cyanobacteria and downregulated in *N. 996*. There were also several symbiont-upregulated exoproteins that may be involved in oxidative stress mediation, including a catalase and ferritin, supporting the importance of mediating reactive nitrogen species and reactive oxygen species in establishing and maintaining the symbiosis.

In support to our last hypothesis, we found that the cyanobacteria in symbiosis with feathermoss have a distinct morphological and metabolic regulation relative to endophytic cyanobacteria–plant symbioses. In endophytic systems, cyanobacterial heterocyst frequency and the N<sub>2</sub>-fixation activity increases when in contact with the host (Adams and Duggan, 2012). We did not observe any statistically significant upregulation of *nifH* or *hetR* genes when associated with *P. schreberi* relative to N-fixing cells in isolation, indicating that the cyanobacteria do not increase N<sub>2</sub>-fixation activity when on the moss. Unequivocally, cyanobacterial fixed N<sub>2</sub> is transferred to the moss (Bay *et al.*, 2013), and the continued expression of *nifH* here is consistent with that. Despite lack of differential expression when comparing isolation and physical contact treatments, both *hetR* and *nif* genes were among the 10% most abundant transcripts when on the moss. In other cyanobacterial symbioses, except the cycad association, ammonium is released by *Nostoc*, achieved partly by a modulation of glutamine synthetase activity, with a reduction from 70% to 15% of *in vitro* activity of the enzyme during symbiosis with *Gunnera* and *Anthoceros*, respectively (reviewed in Meeks, 2009). In our study, *glnA* (glutamine synthetase, GS) and *gltB* (glutamate synthase, GOGAT) transcripts are highly abundant post-colonization and statistically indistinguishable from the free-living condition. The results could indicate that the cyanobacteria may not transfer ammonium to the moss, but rather amino acids, as in the cycad–cyanobacteria symbiosis (Costa and Lindblad, 2002). Finally, in symbiosis with *Gunnera* and *Blasia*, cyanobacterial light-dependent CO<sub>2</sub> fixation is reduced by 90–100%, in return for combined N, the cyanobacteria is compensated by organic C from their host (Meeks and Elhai, 2002; Khamar *et al.*, 2010). By contrast, our data indicate that light-dependent CO<sub>2</sub> and photosynthesis-related genes are highly and equally abundant in association with *P. schreberi* as in isolation, suggesting that cyanobacteria remain photo-autotrophic when in symbiosis with the moss.

Nitrogenase activity of feathermoss associated cyanobacteria is limited by phosphorus availability (Rousk *et al.*, 2016), and sulfur is an essential component of abundant Fe-S clusters in a mature nitrogenase. Neither seems logically available in sufficient quantities on a plant surface. During physical contact, we observed transcriptional upregulation of phosphate and sulfate transporters in the symbiotic-competent strains. Interestingly, the protein families of aliphatic sulfonate transporters and monooxygenases in the genomes of moss symbiotic-competent cyanobacteria are highly upregulated during physical contact. The alkane sulfonate sulfur-specific catabolism pathways were identified using phylogenetic profiling and hypothesized to catalyze the anabolic incorporation of sulfonate-derived sulfur into homocysteine, a direct input to



**Figure 5** Conceptual model for establishment and maintenance of cyanobacteria on feathermoss. A putative time-line includes the following steps: (1) moss releases NO into the environment, which is sensed by the heme NO-binding protein (H-NOX) fused with associated histidine kinase (HK), (2) cyanobacteria respond with reactive nitrogen species mediation by the secretion of flavorubredoxin (F) and FG-GAP repeat-containing protein (G) as chemical cues to the host and other cyanobacteria, (3) NO signaling pathway triggers hormogonia differentiation (*hrm* locus) and gas vesicle formation (*gvp* locus), (4) chemotaxis and motility (*ptx*, *pix* and *hmp* loci and *pil* and *hsp* loci, respectively), (5) cyanobacteria colonize the moss and differentiate into vegetative stage, (6) the host signals to the cyanobacteria through NO for spatial localization and re-direction of the hormogonia, (7) increased NO signaling induces biofilm formation to favor host–symbiont NO-mediated communication and enhance currency exchange, (8) production of aliphatic sulfonate (AS) compound and transfer from the host to the cyanobacteria. The cyanobacteria provide N as amino acid (AA) to the moss. Newly described loci/function are in bold.

cysteine synthesis. We propose aliphatic sulfonates may act as a ‘currency’ in the symbiosis, which has been suggested in marine diatom–bacteria interactions (Durham *et al.*, 2015). In diatom–bacteria interactions, the bacteria utilizes carbon from the sulfonate, whereas in our study the genomically colocalized genes for sulfur anabolism suggest the sulfur is the desired component of the sulfonate. These findings indicate that cyanobacteria epiphytically located on the moss do not obligately depend on the host for fixed carbon, and instead the host may ‘reward’ their associated symbiotic cyanobacteria with an aliphatic sulfonate compound.

Our study defined a repertoire of important functional genes for the moss–cyanobacteria symbiosis composed of previously described OPFs as well as OPFs never before associated with cyanobacteria–plant symbioses. The latter include genes involved in gas vesicles formation, NO-sensing and aliphatic sulfonate transport. We also show that the

competent strains differentially express a unique suite of genes related to motility, NO regulation, sulfate/phosphate transport compared with the non-competent strain. Based on these findings, we propose a model for the establishment and maintenance of the symbiosis, schematically presented in Figure 5. We demonstrate that the *Nostoc* symbiosis with feathermoss is distinct from that with other plants, most likely due to its epiphytic nature. The differences include the retention of motility and chemotaxis post-colonization, as well as constitutive regulation of N<sub>2</sub>-fixation, photosynthesis, GS-GOGAT cycle, and heterocyst formation. The epiphytic cyanobacteria is spatially mobile and chemotactic, therefore predictably redeployed, while still exchanging needed metabolites with the host. Along with describing essential genes to the *Nostoc* feathermoss symbioses, and newly described pathways that either shared among cyanobacterial symbionts, or specific to epiphytic symbioses, our conceptual

model and newly described gene candidates provide targets for verification by gene knockout and testing on feathermosses and other hosts.

#### Data availability

The data have been deposited in <https://gold.jgi.doe.gov/studies?id=Gs0110198>.

#### Conflict of Interest

The authors declare no conflict of interest.

#### Acknowledgements

*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 and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under contract nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). CLD, JE and RAR were supported by NSF grants IOS-1354423 and OCE-1259994 to CLD. RKS was supported by Laboratory Directed Research and Development (LDRD) funding (16-LW-030), and that portion of the work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. We thank Michelle Salemi and Brett Phinney for exoproteomic analysis at the UC Davis Proteomics Core Facility.

#### References

- Adams DG, Duggan PS. (2008). Cyanobacteria-bryophyte symbioses. *J Exp Bot* **59**: 1047–1058.
- Adams DG, Duggan PS. (2012). Signalling in cyanobacteria–plant symbioses. In: Perotto S, Baluška F (eds). *Signaling and Communication in Plant Symbiosis*. Springer: Berlin, Heidelberg, pp 93–121.
- Bay G, Nahar N, Oubre M, Whitehouse MJ, Wardle DA, Zackrisson O *et al*. (2013). Boreal feather mosses secrete chemical signals to gain nitrogen. *New Phytol* **200**: 54–60.
- Bonan GB, Shugart HH. (1989). Environmental-factors and ecological processes in boreal forests. *Annu Rev Ecol Syst* **20**: 1–28.
- Campbell EL, Wong FCY, Meeks JC. (2003). DNA binding properties of the HrmR protein of *Nostoc punctiforme* responsible for transcriptional regulation of genes involved in the differentiation of hormogonia. *Mol Microbiol* **47**: 573–582.
- Campbell EL, Summers ML, Christman H, Martin ME, Meeks JC. (2007). Global gene expression patterns of *Nostoc punctiforme* in steady-state dinitrogen-grown heterocyst-containing cultures and at single time points during the differentiation of akinetes and hormogonia. *J Bacteriol* **189**: 5247–5256.
- Campbell EL, Christman H, Meeks JC. (2008). DNA microarray comparisons of plant factor- and nitrogen deprivation-induced hormogonia reveal decision-making transcriptional regulation patterns in *Nostoc punctiforme*. *J Bacteriol* **190**: 7382–7391.
- 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**: 782–791.
- Christie-Oleza JA, Armengaud J, Guerin P, Scanlan DJ. (2015). Functional distinctness in the exoproteomes of marine *Synechococcus*. *Environ Microbiol* **17**: 3781–3794.
- Christman HD, Campbell EL, Meeks JC. (2011). Global transcription profiles of the nitrogen stress response Resulting in heterocyst or hormogonium development in *Nostoc punctiforme*. *J Bacteriol* **193**: 6874–6886.
- Cohen MF, Meeks JC. (1997). 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 In* **10**: 280–289.
- Costa JL, Lindblad P. (2002). Cyanobacteria in symbiosis with cycads. In: Rai AN, Bergman B, Rasmussen U (eds). *Cyanobacteria in Symbiosis*. Kluwer Academic Publishers: Dordrecht, pp 195–205.
- DeLuca TH, Zackrisson O, Nilsson MC, Sellstedt A. (2002). Quantifying nitrogen-fixation in feather moss carpets of boreal forests. *Nature* **419**: 917–920.
- DeLuca TH, Zackrisson O, Gundale MJ, Nilsson MC. (2008). Ecosystem feedbacks and nitrogen fixation in boreal forests. *Science* **320**: 1181.
- Duggan PS, Gottardello P, Adams DG. (2007). Molecular analysis of genes in *Nostoc punctiforme* involved in pilus biogenesis and plant infection. *J Bacteriol* **189**: 4547–4551.
- Duggan PS, Thiel T, Adams DG. (2013). Symbiosis between the cyanobacterium *Nostoc* and the liverwort *Blasia* requires a CheR-type MCP methyltransferase. *Symbiosis* **59**: 111–120.
- Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Richter RA, Valas R *et al*. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* **6**: 1186–1199.
- Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ *et al*. (2015). Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc Natl Acad Sci USA* **112**: 453–457.
- Eddy SR. (2011). Accelerated profile HMM searches. *PLoS Comput Biol* **7**: e1002195.
- Khamar HJ, Breathwaite EK, Prasse CE, Fraley ER, Secor CR, Chibane FL *et al*. (2010). Multiple roles of soluble sugars in the establishment of *Gunnera*-*Nostoc* endosymbiosis. *Plant Physiol* **154**: 1381–1389.
- Khayatan B, Meeks JC, Risser DD. (2015). Evidence that a modified type IV pilus-like system powers gliding motility and polysaccharide secretion in filamentous cyanobacteria. *Mol Microbiol* **98**: 1021–1036.
- Kim S, Pevzner PA. (2014). MS-GF plus makes progress towards a universal database search tool for proteomics. *Nat Commun* **5**: 5277.
- Komárek J. (2010). Modern taxonomic revision of planktic nostocacean cyanobacteria: a short review of genera. *Hydrobiologia* **639**: 231–243.
- Laczný CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH *et al*. (2015). VizBin - an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3**: 1.

- Lindo Z, Gonzalez A. (2010). The bryosphere: an integral and influential component of the Earth's biosphere. *Ecosystems* **13**: 612–627.
- Mayampurath AM, Jaitly N, Purvine SO, Monroe ME, Auberry KJ, Adkins JN *et al.* (2008). DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **24**: 1021–1023.
- Medina-Andres R, Solano-Peralta A, Saucedo-Vazquez JP, Napsucialy-Mendivil S, Pimentel-Cabrera JA, Sosa-Torres ME *et al.* (2015). The nitric oxide production in the moss *Physcomitrella patens* is mediated by nitrate reductase. *PLoS ONE* **10**: e0119400.
- Meeks JC. (1998). Symbiosis between nitrogen-fixing cyanobacteria and plants - the establishment of symbiosis causes dramatic morphological and physiological changes in the cyanobacterium. *Bioscience* **48**: 266–276.
- Meeks JC, Elhai J. (2002). Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiol Mol Biol R* **66**: 94–121.
- Meeks JC. (2003). Symbiotic interactions between *Nostoc punctiforme*, a multicellular cyanobacterium, and the hornwort *Anthoceros punctatus*. *Symbiosis* **35**: 55–71.
- Meeks JC. (2009). Physiological adaptations in nitrogen-fixing *Nostoc*–plant symbiotic associations. In: Pawlowski K (ed). *Prokaryotic Symbionts in Plants*. Springer-Verlag: Berlin Heidelberg, pp 181–205.
- Monroe ME, Shaw JL, Daly DS, Adkins JN, Smith RD. (2008). MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput Biol Chem* **32**: 215–217.
- Newman AM, Cooper JB. (2010). AutoSOME: a clustering method for identifying gene expression modules without prior knowledge of cluster number. *BMC Bioinform* **11**: 117.
- Norsworthy AN, Visick KL. (2013). Gimme shelter: how *Vibrio fischeri* successfully navigates an animal's multiple environments. *Front Microbiol* **4**: 356.
- Papaefthimiou D, Hrouzek P, Mugnai MA, Lukesova A, Turicchia S, Rasmussen U *et al.* (2008). Differential patterns of evolution and distribution of the symbiotic behaviour in nostocacean cyanobacteria. *Int J Syst Evol Microbiol* **58**: 553–564.
- Payasi A, Sanwal R, Sanwal GG. (2009). Microbial pectate lyases: characterization and enzymological properties. *World J Microb Biot* **25**: 1–14.
- Rao M, Smith BC, Marletta MA. (2015). Nitric oxide mediates biofilm formation and symbiosis in *Silicibacter* sp. strain TrichCH4B. *Mbio* **6**: e00206–e00215.
- Rasmussen U, Johansson C, Bergman B. (1994). Early communication in the *Gunnnera*–*Nostoc* symbiosis - plant-induced cell-differentiation and protein-synthesis in the *Cyanobacterium*. *Mol Plant Microbe In* **7**: 696–702.
- Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. (1979). Generic assignments, strain histories and properties of pure cultures of *Cyanobacteria*. *J Gen Microbiol* **111**: 1–61.
- Risser DD, Meeks JC. (2013). Comparative transcriptomics with a motility-deficient mutant leads to identification of a novel polysaccharide secretion system in *Nostoc punctiforme*. *Mol Microbiol* **87**: 884–893.
- Risser DD, Chew WG, Meeks JC. (2014). Genetic characterization of the hmp locus, a chemotaxis-like gene cluster that regulates hormogonium development and motility in *Nostoc punctiforme*. *Mol Microbiol* **92**: 222–233.
- Robinson MD, McCarthy DJ, Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Robinson MD, Oshlack A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25.
- Rousk K, Sorensen PL, Lett S, Michelsen A. (2015). Across-habitat comparison of diazotroph activity in the subarctic. *Microb Ecol* **69**: 778–787.
- Rousk K, Degboe J, Michelsen A, Bradley R, Bellenger JP. (2016). Molybdenum and phosphorus limitation of moss-associated nitrogen fixation in boreal ecosystems. *New Phytol*. **214**: 97–107.
- Santi C, Bogusz D, Franche C. (2013). Biological nitrogen fixation in non-legume plants. *Ann Bot* **111**: 743–767.
- Shih PM, Wu DY, Latifi A, Axen SD, Fewer DP, Talla E *et al.* (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA* **110**: 1053–1058.
- Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Stuart RK, Mayali X, Lee JZ, Craig Everroad R, Hwang M, Bebout BM *et al.* (2016). Cyanobacterial reuse of extracellular organic carbon in microbial mats. *ISME J* **10**: 1240–1251.
- Turetsky MR. (2003). The role of bryophytes in carbon and nitrogen cycling. *Bryologist* **106**: 395–409.
- Turetsky MR, Bond-Lamberty B, Euskirchen E, Talbot J, Froelking S, McGuire AD *et al.* (2012). The resilience and functional role of moss in boreal and arctic ecosystems. *New Phytol* **196**: 49–67.
- Wu M, Eisen JA. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* **9**: R151.
- Xie F, Murray JD, Kim J, Heckmann AB, Edwards A, Oldroyd GE *et al.* (2012). Legume pectate lyase required for root infection by rhizobia. *Proc Natl Acad Sci USA* **109**: 633–638.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2017

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)