

# Complete Genomes of Symbiotic Cyanobacteria Clarify the Evolution of Vanadium-Nitrogenase

Jessica M. Nelson<sup>1,2,†</sup>, Duncan A. Hauser<sup>1,2,†</sup>, José A. Gudiño<sup>3</sup>, Yessenia A. Guadalupe<sup>3</sup>, John C. Meeks<sup>4</sup>, Noris Salazar Allen<sup>3</sup>, Juan Carlos Villarreal<sup>3,5</sup>, and Fay-Wei Li<sup>1,2,\*</sup>

<sup>1</sup>Boyce Thompson Institute, Ithaca, New York

<sup>2</sup>Plant Biology Section, Cornell University, Ithaca, New York

<sup>3</sup>Smithsonian Tropical Research Institute, Panama City, Panama

<sup>4</sup>Department of Microbiology and Molecular Genetics, University of California, Davis, California

<sup>5</sup>Department of Biology, Laval University, Quebec City, Quebec, Canada

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding author: E-mail: fl329@cornell.edu.

Accepted: June 24, 2019

Data deposition: This project has been deposited at NCBI BioProject under the accession PRJNA534312.

## Abstract

Plant endosymbiosis with nitrogen-fixing cyanobacteria has independently evolved in diverse plant lineages, offering a unique window to study the evolution and genetics of plant–microbe interaction. However, very few complete genomes exist for plant cyanobionts, and therefore little is known about their genomic and functional diversity. Here, we present four complete genomes of cyanobacteria isolated from bryophytes. Nanopore long-read sequencing allowed us to obtain circular contigs for all the main chromosomes and most of the plasmids. We found that despite having a low 16S rRNA sequence divergence, the four isolates exhibit considerable genome reorganizations and variation in gene content. Furthermore, three of the four isolates possess genes encoding vanadium (V)-nitrogenase (*vnf*), which is uncommon among diazotrophs and has not been previously reported in plant cyanobionts. In two cases, the *vnf* genes were found on plasmids, implying possible plasmid-mediated horizontal gene transfers. Comparative genomic analysis of *vnf*-containing cyanobacteria further identified a conserved gene cluster. Many genes in this cluster have not been functionally characterized and would be promising candidates for future studies to elucidate V-nitrogenase function and regulation.

**Key words:** *Nostoc*, *Blasia*, hornworts, liverworts, phylogenomics, nanopore.

## Introduction

Symbiotic interactions with microbes play a pivotal role in eukaryote evolution. In plants, microbial symbionts produce growth-promoting substances, boost tolerance to environmental stress, and can greatly expand access to essential nutrients (Lugtenberg and Kamilova 2009). Partnerships with nitrogen-fixing bacteria are of particular importance because they relieve plants from nitrogen limitation. The best-known example is the symbiosis between legumes and rhizobia, which is an area of active investigation aiming to transfer the symbiosis to nonlegume crop plants (Mus et al. 2016).

There is another and often overlooked type of N<sub>2</sub>-fixing symbiosis in which certain plants partner with heterocyst-forming cyanobacteria (supplementary fig. 1,

Supplementary Material online). Compared with rhizobia, they are less dependent on the plant partner—they have their own specialized cells (i.e., heterocysts) for N<sub>2</sub>-fixation and do not require hosts to create a low-oxygen environment—and therefore hold a promising translational potential for installing N<sub>2</sub>-fixation into crop plants. In addition, unlike the legume-rhizobia symbiosis, the plant-cyanobacteria endosymbiotic partnership has evolved independently multiple times—in two lineages of bryophytes (liverworts and hornworts; supplementary fig. 1, Supplementary Material online), ferns (*Azolla*), gymnosperms (cycads), and a genus of angiosperms (*Gunnera*) (Adams et al. 2006), thereby offering a unique opportunity to study convergent evolution of plant–microbe symbiosis.

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

The key function of cyanobacteria in plant symbiosis is nitrogen fixation, which is carried out by nitrogenase. Nitrogenase is a two-component protein complex consisting of dinitrogenase and dinitrogenase reductase. The most common nitrogenase uses molybdenum (Mo) as the cofactor; however, certain cyanobacteria and other diazotrophs possess an “alternative” nitrogenase that uses vanadium (V) instead of Mo (Mus et al. 2018). Although structurally similar, Mo- and V-nitrogenases are encoded by separate gene clusters and have different catalytic properties; for example, Mo-nitrogenase is more efficient above 16 °C whereas V-nitrogenase performs better at a lower temperature (Miller and Eady 1988). Additionally, Mo is more scarce than V in the terrestrial environment (Wedepohl 1995) and can be a limiting factor for biological nitrogen fixation (Rousk et al. 2017). It is therefore likely that having both nitrogenases could be advantageous for diazotrophs to fine-tune nitrogen fixation in response to mineral availability and temperature. However, although the existence of V-nitrogenase in cyanobacteria is known, its phylogenetic distribution is unclear, and its role in plant symbiosis has not been explored.

As part of our ongoing effort to probe the diversity of plant symbiotic cyanobacteria, we assembled four complete circular genomes of cyanobacteria isolated from hornworts and a liverwort. We report here the discovery of V-nitrogenase in plant cyanobionts, and present comparative genomic analyses to shed new light on the distribution and evolution of V-nitrogenase.

## Materials and Methods

### Isolation and Culture of Cyanobionts

Symbiotic cyanobacteria were cultured from the hornworts *Phaeoceros carolinianus* and *Leiosporoceros dussii*, as well as the liverwort *Blasia pusilla* (table 1). For *Blasia pusilla* and *P. carolinianus*, gametophyte thalli were dissected to obtain pieces with one cyanobacterial colony each. These fragments were wetted in 1% Tween 20, surface sterilized with 5% bleach solution (that has 6% sodium hypochlorite) for 2 min, and rinsed three times in sterile water. For *Leiosporoceros dussii*, cleaned thalli were sterilized in a 10% bleach solution for 1 min and washed three times with sterile water (1 min each). One-millimeter (in thickness) sections from the center of thalli were cut and placed in microporous capsules that were partially submerged in sterile water. Transverse sections of thalli were sterilized for the second time, by submerging the microporous capsules in a 10% bleach solution for 30 s and washing them in three changes of sterile water, 1 min each.

Surface-sterilized thallus fragments were put on combined nitrogen-free BG11 medium (BG11<sub>0</sub>; Rippka et al. 1979), solidified with 1% (w/v) agar and incubated at 22 °C under a 12-h day/night cycle. After growth was detected on the

BG11<sub>0</sub> plates, cultures were transferred to liquid BG11<sub>0</sub> and grown with continuous shaking under the same incubation conditions. Cultures were checked for contamination by fungi and heterotrophic bacteria by streaking on 1.5% malt extract agar and lysogeny broth agar, respectively, and incubating at room temperature in the dark. No sign of contamination was found for these four cultures after at least 10 days of incubation.

### DNA Extraction

Aliquots of liquid cultures of cyanobacteria were centrifuged at 17,000 × g for 2 min to pellet the cells, liquid medium was removed, and one copper bead was added to each pellet. The tubes holding the cell pellets were frozen in liquid nitrogen and ground for 1 min at 1,300 strokes/min on a SPEX SamplePrep 1600 MiniG tissue homogenizer. The metal rack holding the tubes was also submerged in liquid nitrogen before grinding to ensure the samples stayed cold during disruption. Pellets were resuspended in prewarmed 2× cetyl trimethylammonium bromide (CTAB) with 1% β-mercaptoethanol and incubated at 65 °C for 1 h, with gentle mixing every 15 min. Samples were extracted with an equal volume of 24:1 chloroform:isoamyl alcohol twice, each consisting of 5 min gentle mixing on a Labnet Mini Labroller and 5 min centrifugation at 17,000 × g. Wide bore pipette tips were used to transfer the aqueous layer in order to maintain DNA integrity. DNA was precipitated with an equal volume of isopropanol and pelleted for 30 min at 17,000 × g and 4 °C. The DNA pellets were washed twice with 70% ethanol, air dried in a sterile air hood, resuspended in Tris-HCl pH 7.5, and treated with RNase for 1 h at 37 °C.

### Nanopore and Illumina Sequencing

Nanopore libraries were prepared by using the Ligation Sequencing Kit (SQK-LSK109) in conjunction with the Native Barcoding Expansion Kit (1-12). No size selection was done prior to the library preps. We sequenced the four libraries on two MinION R9 flowcells (strains C52 + TCL26-01 on one and C57 + TCL240-02 on the other). Sequencing duration was set to 60 h. We basecalled the raw data using the “flip-flop” algorithm in guppy (version 2.3.7).

For Illumina library construction, we used the SparQ DNA Frag & Library Prep kit and Adapter Barcode Set A following the manufacturer’s protocol. Approximately 100 ng DNA from each strain was used as the input, and fragmented for 5 min in the Fragmentation mix. Libraries were amplified with eight polymerase chain reaction cycles, followed by size selection at 300–700 bp using sparQ PureMag Beads. The four cyanobacteria libraries were pooled with eight other samples in equal quantity and sequenced together on an Illumina NexSeq500 mid-output flowcell at Cornell Institute of Biotechnology.

**Table 1**

Cyanobacteria Strains, Nanopore Sequencing, and Assembly Statistics

Strain	Taxonomic Assignment	Host Species	Locality	Raw Reads (Mb)	Raw Read Length N50 (kb)	Filtered Reads (Mb)	Filtered Read Length N50 (kb)	Assembled Chromosome Size (Mb)	Assembled Plasmids (Mb)
C52	<i>Nostoc</i> sp.	<i>Blasia pusilla</i>	USA, New York, Adirondacks	8,050	4,380	847	12,147	7.83 <sup>c</sup>	0.54 <sup>c</sup> , 0.34 <sup>c</sup> , 0.30 <sup>c</sup> , 0.28 <sup>c</sup> , 0.23 <sup>c</sup> , 0.22 <sup>c</sup> , 0.10 <sup>l</sup> , 0.04 <sup>l</sup>
C57	<i>Nostoc</i> sp.	<i>Phaeoceros carolinianus</i>	USA, California, Mendocino Co.	253	3,662	204	5,098	7.50 <sup>c</sup>	0.24 <sup>c</sup> , 0.16 <sup>c</sup> , 0.06 <sup>c</sup> , 0.06 <sup>c</sup> , 0.06 <sup>c</sup> , 0.04 <sup>c</sup> , 0.33 <sup>l</sup> , 0.32 <sup>l</sup> , 0.16 <sup>l</sup> , 0.13 <sup>l</sup> , 0.08 <sup>l</sup> , 0.03 <sup>l</sup> , 0.03 <sup>l</sup> , 0.02 <sup>l</sup> , 0.02 <sup>l</sup>
TCL240-02	<i>Nostoc</i> sp.	<i>Leiosporoceros dussii</i>	Panamá, Río Guayabo, El Valle de Antón.	960	7,764	587	14,348	7.89 <sup>c</sup>	NA
TCL026-01	<i>Nostoc</i> sp.	<i>Leiosporoceros dussii</i>	Panamá, Río Guayabo, El Valle de Antón.	3,640	2,420	421	15,235	6.77 <sup>c</sup>	0.22 <sup>c</sup> , 0.14 <sup>c</sup> , 0.11 <sup>c</sup> , 0.05 <sup>c</sup> , 0.04 <sup>c</sup> , 0.04 <sup>c</sup> , 0.03 <sup>c</sup>

Note.—<sup>c</sup>Circular assembly and <sup>l</sup>linear assembly.

### Genome Assembly and Annotation

Basecalled nanopore reads were first filtered by length: a 10-kb cutoff was applied to strains C52 and TCL026-01, 5 kb to TCL240-02, and 1 kb to C57. We used Flye (Kolmogorov et al. 2019) for genome assembly with five iterations of built-in polishing, followed by four rounds of minimap2-Racon (Li 2016; Vaser et al. 2017) and one round of medaka (<https://github.com/nanoporetech/medaka>; last accessed April 2019) error correction steps. For strain C57, because we were unable to obtain a circular chromosome after Flye assembler, circlator (Hunt et al. 2015; with the option “--assembler spades --split\_all\_reads --merge\_min\_id 85 --merge\_breaklen 1000”) was used to merge and circularize contigs. The final nanopore assemblies were further polished by Illumina reads with four iterations of bwa-Pilon (Li and Durbin 2009; Walker et al. 2014). Genome annotations were done using prokka (Seemann 2014), with a custom *Nostoc* database compiled from GenBank.

### Phylogenetic Analyses

We first reconstructed a preliminary 16S phylogeny including our sequenced strains and all the cyanobacteria genomes available from GenBank. From this phylogeny, we then selected 96 genomes that are related to our cyanobiont isolates or have *vnf* genes (supplementary table 1, Supplementary Material online). To build a phylogenomic data set, we used BUSCO (Benchmarking Universal Single-Copy Orthologs; Simão et al. 2015) to extract conserved, single-copy genes. A total of 834 cyanobacterial BUSCO loci (“cyanobacteria\_odb9”) were searched against the genomes, and sequences annotated as “fragment” or “duplicated” were excluded from the data set. Each BUSCO locus was aligned separately by MAFFT (Kato et al. 2002). The final concatenated alignment contains 326,259 amino acid sites

(available as supplementary materials, Supplementary Material online). To infer a phylogenetic relationship, a maximum likelihood analysis was carried out by RAxML with the JTT substitution model (Stamatakis 2014), and a total of 1,000 rapid bootstraps were done to assess the branch support.

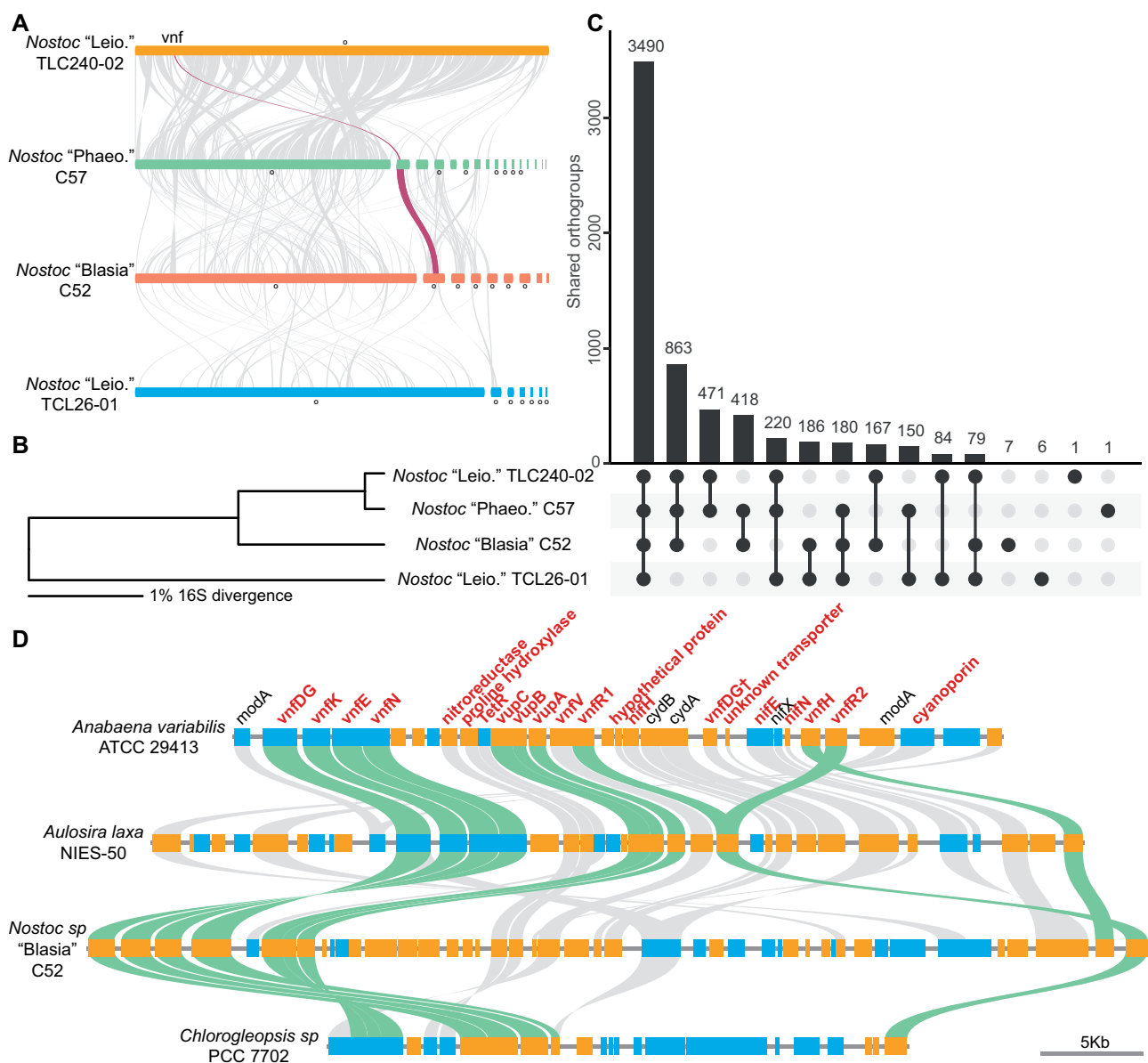
### Comparative Genomic Analyses

To infer large-scale genomic rearrangements among the four sequenced strains, we used the *jcvi.compara.synteny* function in the package MCscan (Tang et al. 2008; with the option “screen --minspan = 15”) to identify and visualize syntenic blocks. For the finer-scale syntenic relationship around the *vnf* genes, the same MCscan function was used but with “mcscan --iter = 1.” To infer the degree of gene content conservation among the four genomes, Orthofinder2 (Emms and Kelly 2018) was used to classify gene orthogroups. The numbers of shared orthogroups were then plotted by UpSetR (Conway et al. 2017).

## Results and Discussion

### Cyanobiont Genome Assemblies

Although genomes have been published for several plant cyanobionts (Meeks et al. 2001; Ran et al. 2010; Li et al. 2018; Kanesaki et al. 2018; Warshan et al. 2018; Gutiérrez-García et al. 2019), only *Nostoc punctiforme* PCC 73102 (from a cycad; Meeks et al. 2001) and *Nostoc azollae* (from *Azolla*; Ran et al. 2010) have genomes assembled into complete circular chromosomes. All others are highly fragmented with hundreds to thousands of contigs, making comparative genomic studies difficult. In this study, even with nanopore coverage as low as ~20× (table 1), we were able to obtain circular contigs for all four main chromosomes and for most of the plasmids (fig. 1a; table 1). However, we found that our



**Fig. 1.**—Comparative analyses of cyanobacteria genomes. (A) Considerable differences in genome organization were found among the four cyanobiont genomes sequenced in this study. Ribbons connect the syntenic blocks, with red ribbons marking the location of the *vnf* gene cluster. Note the contig lengths are not to scale. (B) The four cyanobiont strains show very little 16S sequence divergence. (C) Although the majority of the orthogroups are shared among the four genomes, gains and losses appear to be common. "Leio.": strain isolated from *Leiosporoceros dussii*; "Phaeo.": from *Phaeoceros carolinianus*; "Blasia": from *Blasia pusilla*. "o" indicates circular chromosomes or plasmids. (D) A conserved *vnf*-associated gene cluster was found in cyanobacteria. Genes shared among at least three genomes are labeled in red. Green ribbons connect genes that have been functionally characterized in *Anabaena variabilis* ATCC 29413. Blue and orange gene blocks distinguish the transcription directions. "+" indicates a truncated gene in *Anabaena variabilis*.

nanopore assemblies all have a low nucleotide accuracy. After multiple rounds of Racon and medaka correction, over half of the BUSCO genes were still missing or fragmented (supplementary fig. 2, Supplementary Material online), often due to frameshifts introduced by inaccurate homopolymer resolution. Further polishing using Illumina reads was required to bring BUSCO genes to >99% completion (supplementary fig. 2, Supplementary Material online). This indicates that

the error profile of nanopore reads is not random, and further development is needed so that one could bypass the Illumina data generation step.

#### Variation in Genome Organization and Gene Content

Comparing the four newly sequenced genomes, we found that the level of 16S sequence divergence is low, ranging from 6.33% to 0.34% (fig. 1b). This is a stark contrast to the large-

scale genomic rearrangements identified by synteny analysis (fig. 1a), as well as the relatively low number of shared orthogroups (fig. 1c). For example, even though 16S sequences of TCL240-02 and C57 are 99.66% identical, the two genomes have very different genome organizations and plasmid numbers (fig. 1a). Furthermore, over 17% of the orthogroups are present in only one of two genomes (fig. 1c). Although 16S is a commonly used marker to profile microbial communities, the results shown here imply that 16S could seriously underestimate the diversity at the genomic (and presumably functional) level, especially when a less stringent cutoff (e.g., 97%) is applied for clustering Operational Taxonomic Units.

### Phylogenetic Relationship of Symbiotic Cyanobacteria

Overall, our phylogeny shows that a large number of cyanobacteria genera are polyphyletic with respect to morphology-based taxonomic identity (supplementary fig. 3, Supplementary Material online), which is consistent with several recent studies (Herdman and Rippka 2017; Gagunashvili and Andrésón 2018; Warshan et al. 2018; Will et al. 2019). When placed under a broader phylogenetic context, our sequenced strains fell into two clades (supplementary fig. 3, Supplementary Material online). Clade A is mostly comprised of cyanobacterial endosymbionts of lichens, bryophytes, and a cycad, as well as those found as epiphytes on boreal feather mosses (Warshan et al. 2018). This clade corresponds to the “Intracellular/Extracellular” group in a recent phylogenomic study by Warshan et al (2018). Clade B on the other hand, has fewer strains isolated from plants or lichens, although symbiotic competency has not been tested for most of the strains in this clade. Warshan (2018)’s “Extracellular II” group is nested within this clade.

### Presence of Vanadium-Type Nitrogenase in Plant Cyanobionts

Three of the four cyanobacteria genomes we assembled contain the genes encoding vanadium (V)-nitrogenase (*vnf*), in addition to the conventional Mo-type nitrogenase. Although *vnf* genes have been previously identified in cyanobacteria associated with lichens (Hodkinson et al. 2014; Darnajoux et al. 2017; Gagunashvili and Andrésón 2018), this is the first report, to our knowledge, of them in plant cyanobionts. Interestingly, in C57 and C52, *vnf* genes are located in plasmids (fig. 1a), implying there is a potential for plasmid-mediated horizontal gene transfers. To better understand the distribution of *vnf* in cyanobacteria, we mined the NCBI database and identified four separate cyanobacteria lineages having *vnf* (Clades A–D; see the red taxa in supplementary fig. 3, Supplementary Material online). None of these clades are exclusively composed of *vnf*-containing strains, and it appears that *vnf* losses and/or gains are frequent among

cyanobacteria. For strains having complete circular genomes, only strains C57 and C52 have *vnf* in plasmids.

### A Conserved *vnf* Cluster in Cyanobacteria

V-nitrogenase was originally discovered in *Azotobacter* (Robson et al. 1986) and subsequently identified and characterized in the cyanobacterium *Anabaena variabilis* ATCC 29413 (Thiel 1993; Thiel et al. 2014). In *Anabaena variabilis* ATCC 29413, *vnf* genes form a cluster consisting of *vnfD* (nitrogenase structural proteins; Thiel 1993), *vnfH* (dinitrogenase reductase; Pratte et al. 2006), and *vnfEN* (for cofactor synthesis; Thiel 1996). Additionally, genes encoding vanadate transporters (*vupABC*; Pratte and Thiel 2006) and *vnf* transcriptional regulators (*vnfR1* and *R2*; Pratte et al. 2013) have also been identified in a close proximity to *vnf* genes. To determine if this gene cluster is conserved across the four *vnf*-containing clades (Clades A–D; supplementary fig. 3, Supplementary Material online), we carried out a detailed microsynteny analysis with one representative genome from each clade. We were able to delineate a clear conserved gene cluster associated with V-nitrogenase (fig. 1d). This cluster includes not only genes with known functions such as *vnfR2* and *vupABC* but also a large number of uncharacterized genes (fig. 1d). Whether these conserved genes are involved in V-nitrogenase activity is unresolved, but our finding here provides a list of promising candidates to analyze.

### Supplementary Material

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

### Acknowledgments

This research was financially supported by grant 429440 from the Simons Foundation to the Smithsonian Tropical Research Institute and National Science Foundation Dimensions of Biodiversity grant (1831428) to F.-W.L. and J.C.M. We thank Denis Warshan for providing the annotated genomes from their study, and Li lab members for discussions.

### Literature Cited

- Adams DG, Bergman B, Nierzwicki-Bauer SA, Rai AN, Schüßler A. 2006. Cyanobacterial-plant symbioses. In: *The prokaryotes*. New York: Springer. p. 331–363.
- Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33(18):2938–2940.
- Darnajoux R, et al. 2017. Biological nitrogen fixation by alternative nitrogenases in boreal cyanolichens: importance of molybdenum availability and implications for current biological nitrogen fixation estimates. *New Phytol.* 213(2):680–689.
- Emms DM, Kelly S. 2018. OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *bioRxiv*. 13:466201. doi:10.1101/466201.

- Gagunashvili AN, Andr sson  S. 2018. Distinctive characters of *Nostoc* genomes in cyanolichens. *BMC Genomics*. 19(1):434.
- Guti rrez-Garc a K, et al. 2019. Cycad coralloid roots contain bacterial communities including cyanobacteria and *Caulobacter* spp. that encode niche-specific biosynthetic gene clusters. *Genome Biol Evol*. 11(1):319–334.
- Herdman M, Rippka R. 2017. Cyanobacterial phylogeny. Available from: <http://cyanobact.000webhostapp.com/>
- Hodkinson BP, et al. 2014. Lichen-symbiotic cyanobacteria associated with *Peltigera* have an alternative vanadium-dependent nitrogen fixation system. *Eur J Phycol*. 49(1):11–19.
- Hunt M, et al. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol*. 16(1):294.
- Kanesaki Y, et al. 2018. Draft genome sequence of the nitrogen-fixing and hormogonia-inducing cyanobacterium *Nostoc cycadae* strain WK-1, isolated from the coralloid roots of *Cycas revoluta*. *Genome Announc*. 6(7):e00021-18.
- Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 37(5):540–546.
- Li F-W, et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* 4(7):460–416.
- Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32(14):2103–2110.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Lugtenberg B, Kamilova F. 2009. Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol*. 63(1):541–556.
- Meeks JC, et al. 2001. An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. *Photosyn Res*. 70(1):85–106.
- Miller RW, Eady RR. 1988. Molybdenum and vanadium nitrogenases of *Azotobacter chroococcum*. Low temperature favours N<sub>2</sub> reduction by vanadium nitrogenase. *Biochem J*. 256(2):429–432.
- Mus F, Alleman AB, Pence N, Seefeldt LC, Peters JW. 2018. Exploring the alternatives of biological nitrogen fixation. *Metallomics* 10(4):523–538.
- Mus F, et al. 2016. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl Environ Microbiol*. 82(13):3698–3710.
- Pratte BS, Eplin K, Thiel T. 2006. Cross-functionality of nitrogenase components NifH1 and VnfH in *Anabaena variabilis*. *J Bacteriol*. 188(16):5806–5811.
- Pratte BS, Sheridan R, James JA, Thiel T. 2013. Regulation of V-nitrogenase genes in *Anabaena variabilis* by RNA processing and by dual repressors. *Mol Microbiol*. 88(2):413–424.
- Pratte BS, Thiel T. 2006. High-affinity vanadate transport system in the cyanobacterium *Anabaena variabilis* ATCC 29413. *J Bacteriol*. 188(2):464–468.
- Ran L, et al. 2010. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. *PLoS One* 5(7):e11486.
- 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.
- Robson RL, et al. 1986. The alternative nitrogenase of *Azotobacter chroococcum* is a vanadium enzyme. *Nature* 322(6077):388–390.
- Rousk K, Degboe J, Michelsen A, Bradley R, Bellenger J-P. 2017. Molybdenum and phosphorus limitation of moss-associated nitrogen fixation in boreal ecosystems. *New Phytol*. 214(1):97–107.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069.
- Sim o FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Tang H, et al. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res*. 18(12):1944–1954.
- Thiel T. 1993. Characterization of genes for an alternative nitrogenase in the cyanobacterium *Anabaena variabilis*. *J Bacteriol*. 175(19):6276–6286.
- Thiel T. 1996. Isolation and characterization of the *vnfEN* genes of the cyanobacterium *Anabaena variabilis*. *J Bacteriol*. 178(15):4493–4499.
- Thiel T, et al. 2014. Complete genome sequence of *Anabaena variabilis* ATCC 29413. *Stand Genomic Sci*. 9(3):562–573.
- Vaser R, Sovi c I, Nagarajan N,  iki c M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 27(5):737–746.
- Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
- Warshan D, et al. 2018. Genomic changes associated with the evolutionary transitions of *Nostoc* to a plant symbiont. *Mol Biol Evol*. 35(5):1160–1175.
- Wedepohl KH. 1995. The composition of the continental crust. *Geochim Cosmochim Acta* 59:1217–1232.
- Will SE, et al. 2019. Day and night: metabolic profiles and evolutionary relationships of six axenic non-marine cyanobacteria. *Genome Biol Evol*. 11(1):270–294.

Associate editor: Howard Ochman