Received: 14 December 2020 | Accepted: 17 May 2021

DOI: 10.1002/ajb2.1729

RESEARCH ARTICLE

American Journal of Botany

The diversity and community structure of symbiotic cyanobacteria in hornworts inferred from long-read amplicon sequencing

Jessica M. Nelson[1] | **Duncan A. Hauser**[1] | **Fay-Wei Li**[1,2]

[1]Boyce Thompson Institute, Ithaca, NY, USA

[2]Plant Biology Section, Cornell University, Ithaca, NY, USA

Correspondence

Fay-Wei Li, Boyce Thompson Institute, Ithaca, NY, USA; Plant Biology Section, Cornell University, Ithaca, NY, USA.
Email: fl329@cornell.edu

Abstract

Premise: Nitrogen-fixing endosymbioses with cyanobacteria have evolved independently in five very different plant lineages. Expanding knowledge of these symbioses promises to improve the understanding of symbiosis evolution and broaden the toolkit for agricultural engineering to reduce artificial fertilizer use. Here we focused on hornworts, a bryophyte lineage in which all members host cyanobacteria, and investigated factors shaping the diversity of their cyanobiont communities.

Methods: We sampled hornworts and adjacent soils in upstate New York throughout the hornwort growing season. We included all three sympatric hornwort species in the area, allowing us to directly compare partner selectivity. To profile cyanobacteria communities, we established a metabarcoding protocol targeting *rbcL-X* with PacBio long reads.

Results: The hornwort cyanobionts detected were phylogenetically diverse, including clades that do not contain other known plant symbionts. We found significant overlap between hornwort cyanobionts and soil cyanobacteria, a pattern not previously reported in other plant–cyanobacteria symbioses. Cyanobiont communities differed between host plants only centimeters apart, but we did not detect an effect of sampling time or host species on the cyanobacterial community structure.

Conclusions: This study expands the phylogenetic diversity of known symbiotic cyanobacteria. Our analyses suggest that hornwort cyanobionts have a tight connection to the soil background, and we found no evidence that time within growing season, host species, or distance at the scale of meters strongly govern cyanobacteria community assembly. This study provides a critical foundation for further study of the ecology, evolution, and interaction dynamics of plant–cyanobacteria symbiosis.

KEYWORDS

Anthoceros, cyanobacteria, endophytes, hornworts, *Nostoc*, *Notothylas*, PacBio, *Phaeoceros*, symbiosis

Nitrogen is a critical plant macronutrient, but only certain prokaryotes can access the abundant nitrogen in the atmosphere. Most plants therefore need to obtain combined nitrogen from the soil that has been fixed by such microbes or released from decaying organisms. However, some plants expedite this process by hosting nitrogen-fixing bacteria on their surfaces (e.g., boreal feather mosses; Holland-Moritz et al., 2018; Stuart et al., 2020) or inside their tissues (Santi et al., 2013). A few groups of plants have evolved specialized structures that can harbor the nitrogen-fixing bacteria to optimize their nitrogen fixation and transfer (Mus et al., 2016). The most well-known and studied examples are the rhizobia bacteria that live in the root nodules of legume plants. However, five other plant groups have independently evolved nitrogen-fixing symbioses in which they build cavities for cyanobacteria: the angiosperm genus *Gunnera*, cycads (gymnosperms), the water fern genus *Azolla*, a small family of liverworts (Blasiaceae), and hornworts (Meeks, 1998; Rai et al., 2000). In all these lineages, the symbiotic cyanobacteria (hereafter cyanobionts) are hosted inside specialized plant structures such as slime cavities in hornworts and coralloid roots in cycads that, unlike legume

root nodules, develop independently of contact with the symbiotic bacteria (Rai et al., 2000; Santi et al., 2013).

While the natural history and genetics of the root nodule symbiosis are currently known in some detail (Oldroyd et al., 2011; Wang et al., 2012; Andrews and Andrews, 2017), the diversity and interaction mechanisms of cyanobacterial partners are poorly known. Filling in knowledge of these cyanobacterial symbioses would allow comparative evolutionary studies between independent origins of nitrogen-fixing associations. We would also improve our understanding of the evolution of symbioses in general and expand the toolkit for researchers interested in engineering agricultural systems that are less dependent on artificial fertilizers.

Of the five plant groups that host cyanobionts internally, hornworts are well suited to investigate the ecological and evolutionary dynamics of cyanobiont communities. They are broadly distributed globally and encompass deep evolutionary splits (Villarreal and Renzaglia, 2015; Frangedakis et al., 2020), allowing comparative studies across biomes and clades. Hornworts also present a number of advantages as laboratory models: they are compact, they can be grown in axenic culture for symbiosis reconstitution experiments, identical clones are simple to produce as experimental replicates, and establishment of the symbiosis is easy to confirm in intact plants since the cyanobacteria form discrete dark colonies in the translucent plant thalli. Indeed, the hornwort *Anthoceros punctatus* and cyanobacterial strain *Nostoc punctiforme* PCC 73102 have long been used to investigate symbiotic physiology (Meeks, 2009). In addition, several high-quality hornwort and cyanobiont genomes have been recently published, accelerating research on the genetic mechanisms of the symbiosis (Nelson et al., 2019; Li et al., 2020; Zhang et al., 2020).

Determining the factors controlling cyanobiont community assembly in hornworts would allow comparison to other N-fixing systems and future experiments investigating mechanistic details of partner choice and regulation in this symbiosis. Therefore, research is needed on the ecology of cyanobacterial interactions with hornworts in the wild. Few studies have attempted to sample this diversity (West and Adams, 1997; Costa et al., 2001; Bouchard et al., 2020), covering only a handful of the approximately 200 known hornwort species.

Soil cyanobacteria are usually assumed to be the source of cyanobionts for terrestrial plant hosts (Rasmussen and Johansson, 2002). However, only a few studies have attempted to survey the cyanobacterial diversity of soils near plant hosts (hornworts: West and Adams, 1997; liverworts: Liaimer et al., 2016; Rikkinen and Virtanen, 2008; cycads: Cuddy et al., 2012 and Suarez-Moo et al., 2019) and have found differing amounts of overlap between soil and symbiotic communities. Only one of these studies has used the current methods of metabarcoding with next generation sequencing (Suárez-Moo et al., 2019). Furthermore, terrestrial cyanobacteria communities are not well characterized beyond soil crusts in certain extreme environments (for example Rippin et al., 2018). Therefore, information on the diversity of cyanobacteria available to temperate hornworts is scarce. Without knowing the background cyanobacteria compositions, it is difficult to infer and compare selectivity and specificity across species, or to understand the factors structuring cyanobiont communities.

In addition to selection from the soil background, cyanobiont communities, like other microbiomes, are likely to be structured across space and through time (O'Malley, 2008; Rout and Callaway, 2012; Ranjard et al., 2013; Peršoh, 2015; Goldmann et al., 2020). In lichen symbioses, symbiont specificity has even been found to differ depending on the geographic scale investigated (Rikkinen, 2013). Previous studies sampling plant cyanobionts from distant sites indicate that, while some bacterial strains can be shared across wide distances, the symbiont communities can also vary considerably between plants collected at the same site (Nilsson et al., 2000; Costa et al., 2001; Bouchard et al., 2020). The spatial scale of variation for this symbiosis is therefore unclear, and more studies at various spatial scales are needed to clarify these patterns of diversity. Plant microbial communities can also exhibit complex patterns of microbial succession over seasons or years (Shade et al., 2013; Wagner et al., 2016; Materatski et al., 2018; Goldmann et al., 2020). The structures that plants produce to house cyanobionts are usually continuously developed and colonized by cyanobacteria (Rai et al., 2000; Adams and Duggan, 2008), implying that cyanobiont compositions may fluctuate over time, but no study to date has investigated the temporal dynamics of cyanobiont populations.

In any symbiosis, the degree of specificity and preference between the partners is an important characteristic because partner quality may vary. Legume–rhizobia symbioses involve complex signaling pathways to determine compatibility between partners (Oldroyd et al., 2011; Wang et al., 2012), but whether such a mechanism exists for plant–cyanobacterial symbioses is still an open question. Surveys and symbiosis reconstitution experiments on plant cyanobionts so far have not provided evidence of strong partner preferences (Johansson and Bergman, 1994; West and Adams, 1997; Nilsson et al., 2000; Rasmussen and Johansson, 2002; Gehringer et al., 2010; Suárez-Moo et al., 2019), aside from the obligate, vertically transmitted cyanobionts found in *Azolla* (Papaefthimiou et al., 2008; Li et al., 2018). While this lack of preference may be the reality for plant–cyanobacteria relationships, it could also be due in part to host species and location often being confounded or to problems with cyanobacterial taxonomy. Most plant cyanobionts have been commonly referred to as *Nostoc* (Rai et al., 2000), but this genus is notoriously polyphyletic and contains multiple distinct genetic clades (Rajaniemi et al., 2005; Shih et al., 2013; Komárek et al., 2014). The taxonomic confusion for *Nostoc* likely conceals wider symbiont diversity and patterns of compatibility and preference between partners.

Until recently, research on the diversity of plant cyanobionts relied on culturing the cyanobacteria or limited methods of genetic differentiation such as RFLPs

(Leizerovich et al., 1990; West and Adams, 1997; Costa et al., 1999, 2001; Nilsson et al., 2000; Guevara et al., 2002; Zheng et al., 2002; Papaefthimiou et al., 2008; Rikkinen and Virtanen, 2008; Thajuddin et al., 2010; Yamada et al., 2012; Fernández-Martínez et al., 2013; Liaimer et al., 2016). While metabarcoding and metagenomic methods using next-generation-sequencing platforms have begun to be applied, covering *Azolla*, cycads, and one hornwort (Dijkhuizen et al., 2018; Zheng et al., 2018; Gutiérrez-García et al., 2019; Suárez-Moo et al., 2019; Zheng and Gong, 2019; Bell-Doyon et al., 2020; Bouchard et al., 2020), we are still far from fully capturing the diversity of plant cyanobionts and any patterns of host preferences. The conventional 16 S amplicon approach for barcoding bacteria poses several challenges in characterizing cyanobiont communities. First, the standard 16 S primers are nonspecific and will amplify sequences from the entire bacterial community as well as plant organelles, meaning much of the pool of sequences will contain taxa other than cyanobacteria. This lack of specificity is a particular challenge when targeting cyanobacteria since they share ancestry with plastids. Second, the short 16 S amplicons obtained with Illumina sequencing do not provide much phylogenetic signal, which is problematic because the classification of cyanobacteria—from order to species levels—has been in flux (Komárek et al., 2014). Third, cyanobacterial genomes typically contain multiple rRNA operons, which skews read counts, and in some cases, the operons can be highly divergent in the same genome (Johansen et al., 2017), which makes rRNA genes inappropriate species markers in this case.

In this study, we established a new protocol to circumvent the problems with 16 S barcoding of cyanobionts by targeting the cyanobacteria-specific *rbcL-X* region (~800 bp). We built our approach based on the recent developments of using the PacBio circular consensus sequencing (CCS) method for plant phylogenomics (Rothfels et al., 2016) and full-length 16 S and ITS metabarcoding studies (Schlaeppi et al., 2016; Tedersoo et al., 2018; Callahan et al., 2019). Using this technique, we aimed to address what factors are important for shaping hornwort cyanobiont communities and forming a basis for future studies of cyanobiont community dynamics and partner choice. We investigated the diversity of cyanobionts in three temperate hornwort species and looked for effects of soil cyanobacterial communities, host species, time of collection (on the scale of months), and locality (on the scale of centimeters and meters). Specifically, we asked: (1) What is the phylogenetic diversity of cyanobionts in these hornworts? (2) To what extent do the symbiotic and soil cyanobacteria communities overlap? (3) How stable are cyanobiont communities over the time of the hornwort growing season? (4) How do cyanobiont communities differ between hornworts sampled a few centimeters or meters apart? And (5) do sympatric hornwort species exhibit different preferences in cyanobiont selection?

MATERIALS AND METHODS

Sample collection

We sampled hornworts and adjacent soil at two sites in Tompkins County, New York about 20 km apart: one in the Cornell Botanic Gardens on the bank of Grossman Pond and one on a roadside bank in Potato Hill State Forest (Figure 1A, B). The Grossman Pond site had a population of *Notothylas orbicularis* (Schwein.) Sull. while the Potato Hill

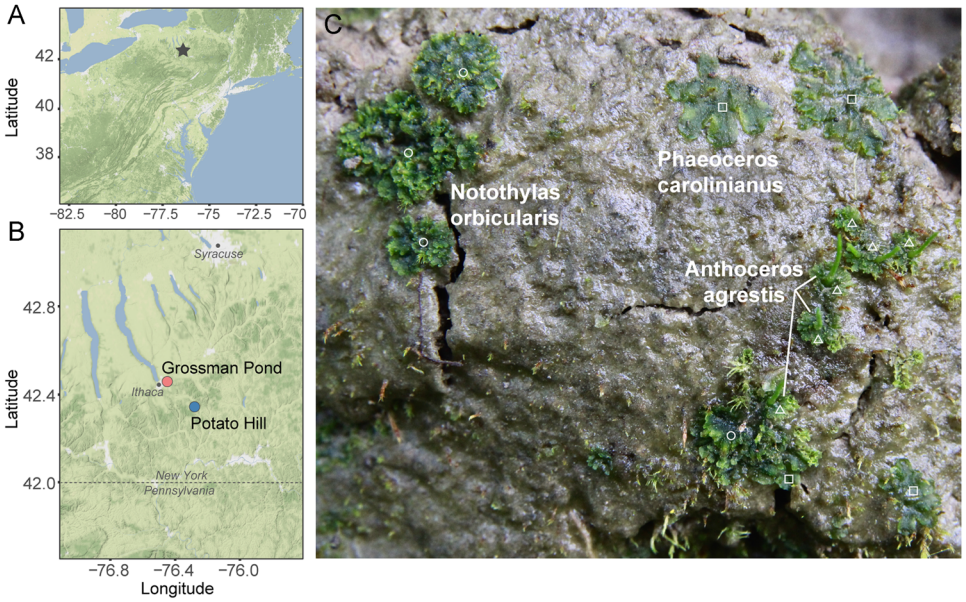

FIGURE 1 Hornwort population samplings. (A) Grossman Pond and Potato Hill sites in upstate New York (star). (B) Map showing the locations of the two sites. (C) Three sympatric hornwort species can be found growing together at the Potato Hill site. Circle: *Notothylas orbicularis*. Square: *Phaeoceros carolinianus*. Triangle: *Anthoceros agrestis*

site had a mixed population of *N. orbicularis, Phaeoceros carolinanus* (Michx.) Prosk., and *Anthoceros agrestis* Paton (Figure 1C). These three species are the most common, if not the only, hornworts in the northeastern United States (Schuster, 1992).

In the study area, hornworts are annual with a growing season typically starting in early- to mid-summer and ending in fall. To capture variation over this growing season, we collected samples every 3 weeks from 23 August to 30 October in 2018, for a total of four sampling times at each site.

We used quadrats of 0.25 m^2 as the sampling unit to delineate patches of hornworts to sample and marked the locations with pin flags so the same ground area was used each time. We sampled two quadrat locations (about 20 m apart) at Grossman Pond and three at Potato Hill (about 20 and 7 m apart). Within each quadrat, we collected four soil samples along the diagonals using sterile 15-mL Falcon tubes. We took four plugs of soil approximately 1 cm deep for each sample, moving from the center of the quadrat toward one corner (Appendix S1). We collected four plants (entire gametophytes, sometimes with attached sporophytes) of each present hornwort species within the area of the quadrat using an ethanol-cleaned knife. *Notothylas* declined toward the end of the season earlier than the other species, resulting in two quadrats that lacked this species at the final sampling time. Because individual thalli are small and destroyed for DNA extraction, it was not possible to resample the same plant each time. We stored soil and plant samples at 4°C until processing (usually about 24 h, but no longer than 48 h).

Soil sample preparation

We collected a total of 76 soil samples. We passed soils through a clean 2-mm-mesh sieve to even out particle sizes, remove large debris or organic materials, and mix the sample together (four plugs into one, see Appendix S1). From this mixture, we put a small lump of soil (about 150–250 mg) directly into an E.Z.N.A. soil DNA kit (Omega Bio-Tek, Norcross, GA, USA) disruptor tube with glass beads. Between samples, we washed adhering soil off the sieve and cleaned the sieve with 70% ethanol. We extracted DNA from soils using the E.Z.N.A. soil DNA kit with two cHTR reagent washes.

Hornwort sample preparation

We collected a total of 168 hornwort samples. We cleaned fresh hornworts in sterile water to remove all adhering soil. First, we rinsed them in multiple changes of sterile water using vortexing and forceps to dislodge soil particles. Once most of the soil was removed, we sonicated the plants in sterile water at 1% amplitude for 15 s in intervals of 3 s with 2-s rests in between. We determined this protocol by testing for the strongest sonication that would not break the hornwort cells. We then blotted the plants dry on clean Kimwipes and transferred them to homogenization tubes with 2-mm zirconia beads. We froze the tubes containing the plants with liquid nitrogen and ground the plants for 2 min at 1300 strokes/min on a 1600 MiniG tissue homogenizer (SPEX SamplePrep, Metuchen, NJ, USA) in a metal box pre-chilled in liquid nitrogen. We extracted DNA from plant tissues using the E.Z.N.A. Plant DNA Kit (Omega Bio-Tek).

Amplicon library preparation and sequencing

We amplified the *rbcL-X* region by PCR from all DNA samples using barcoded versions of the primers CW and CX (Rudi et al., 1998). We used eight forward primer barcodes and 12 reverse primer barcodes to make 96 dual barcode combinations (Appendix S2). We amplified samples with the appropriate barcoded primers using Phusion high-fidelity polymerase (New England Biolabs, Ipswich, MA, USA). The PCR recipe and thermocycler program can be found in Appendix S3.

We used a mock community of five cyanobacteria as a positive sequencing control. We extracted DNA from pure laboratory cultures of cyanobacteria with a standard CTAB protocol and grinding with copper beads, as previously described (Nelson et al., 2019). Then we amplified *rbcL-X* from each using the same PCR program as above with unbarcoded primers. We measured the concentrations of these PCR products with the Qubit HS DNA kit (ThermoFisher, Waltham, MA, USA) and mixed in equal quantity. We then added a unique barcode to this mix by PCR in the same way as all other samples.

We measured the amplicon concentrations with the Qubit HS DNA kit and pooled the samples in equal quantity into three libraries. In each library, we included one mock community sample. We cleaned the libraries using ProNex beads (Promega, Madison, WI, USA) with a 1.5:1 ratio of beads to sample. We sent the samples to Duke University for PacBio circular consensus sequencing on the Sequel platform with v3.0 chemistry. We deposited the sequencing reads at NCBI SRA under the accession PRJNA632853.

Data processing

We used the PacBio ccs (v4.0.0; https://github.com/PacificBiosciences/ccs) package to generate circular consensus sequences from the raw PacBio reads and only kept those with at least five complete passes and a predicted consensus accuracy over 0.999 (–min-passes=5 –min-rq=0.999). We used lima (v1.10.0; https://github.com/PacificBiosciences/barcoding) for demultiplexing with the minimum barcode score set to 26 (–min-score 26), then removed primers and reoriented reads with the removePrimers function of the dada2 package (Callahan et al., 2016). We further filtered the data set using the dada2 filterAndTrim function to remove sequences longer

than 1200 bp, shorter than 400 bp, or having a quality score lower than 3 (minQ=3, minLen=400, maxLen=1200). We used the dada2 (v.1.14.1) PacBio pipeline (Callahan et al., 2019) for dereplication, error learning, amplicon sequence variant (ASV) inference, and chimera detection (Callahan et al., 2019). In contrast to the traditional clustering-based method, which uses an arbitrary identity cutoff to group sequences into operational taxonomic units (OTUs), the ASV approach incorporates models of sequencing error to derive exact sequence variants. To benchmark the ASV approach against traditional OTU clustering, we used VSEARCH (Rognes et al., 2016) to carry out dereplication, chimera removal, and clustering at 97% and 95% sequence identities. We analyzed the resulting ASV and OTU tables using the R package phyloseq v1.30.0 (McMurdie and Holmes, 2013).

Phylogenetic analysis

We inferred phylogenies using two data sets. The first data set included ASVs from this study and a large collection of cyanobacterial *rbcL-X* sequences, so that we could place our ASVs into a broader phylogenetic context. To build an *rbcL-X* database, we (1) downloaded all the available cyanobacterial genomes and extracted their *rbcL-X* region, (2) queried all the ASVs by BLASTn against the NCBI database and retrieved the hits (with an e-value threshold of 0.001), and (3) incorporated *rbcL-X* sequences (done by Sanger sequencing) from some of our own cyanobacterial isolates. We removed redundant and identical sequences. Combining ASVs from the present study and previously sequenced references resulted in a total of 1748 sequences. We aligned these using PASTA (Mirarab et al., 2015) and inferred the phylogeny using RAxML-HPC2 v8.2.12 (Stamatakis, 2014) on CIPRES (Miller et al., 2010), with 1000 replicates of rapid bootstrapping to assess branch supports.

To better visualize and delineate the hornwort cyanobiont clades, we analyzed a second data set including only the ASVs. For the plant samples in this data set, we filtered out low abundance ASVs (<3% per sample) because they likely represented epiphytic or soil bacteria, or PCR/sequencing errors. In other words, if an ASV never reached at least 3% abundance in any plant sample, then we excluded this ASV from the data matrix. We selected this threshold based on the mock community results (see below). We did the alignment with PASTA and tree inference with RAxML as before. We plotted the number of times an ASV could be found in each sample category (soil, *Anthoceros*, *Notothylas*, and *Phaeoceros*) using ggtree (Yu et al., 2018). To be conservative, we defined cyanobiont clades on the phylogeny as those containing ASVs present in at least three plant samples and with bootstrap support over 95. To assign taxonomic affinities to nodes, we cross-referenced this phylogeny with (1) the one described above with other known cyanobacterial sequences and (2) trees from Shih et al. (2013) based on a large sampling of cyanobacteria genomes, and from Otálora et al. (2010) and Magain et al. (2017, 2018) on lichen photobionts. Due to the relatively poor taxonomic understanding of cyanobacteria, we could not confidently assign families to nodes. Instead, we listed any strains/isolates that were present in the same clade as each ASV group.

Analysis of cyanobacteria diversity and community composition

We used the R package phyloseq (v1.30.0) (McMurdie and Holmes, 2013) to transform ASV counts per sample into relative abundance, calculate Unifrac and weighted Unifrac distance between samples, and carry out a series of principal coordinate analyses (PCoA). To determine whether cyanobacteria communities varied significantly by sampling location, time, and host species, we conducted PERMANOVA tests with 10,000 permutations using the adonis function in the vegan package (Dixon, 2003). To identify specific ASVs that showed differential abundance among host taxa, we conducted an analysis of composition of microbiomes (ANCOM) (Mandal et al., 2015) in the R package ANCOM (v2.1; https://github.com/FrederickHuangLin/ANCOM). We carried out the PERMANOVA and ANCOM analyses with low abundance ASVs (3%) removed.

We calculated Chao, Shannon, and Simpson alpha diversity indices with phyloseq, and Faith's phylogenetic diversity and mean pairwise distance indices with the R package picante (Kembel et al., 2010) using the ASV-only phylogeny from above. We used the UpSetR package to visualize the overlaps of ASVs among quadrats, sample types, and times (Conway et al., 2017). For these comparisons, we excluded ASVs that only appeared in one sample since these would artificially inflate the list of ASVs unique to individual categories. We did an unpaired *t*-test comparing the number of ASVs from soil versus plant samples in R. The scripts, commands, and data matrices from this study can be found at https://github.com/fayweili/hornwort_cyano_interaction.

Based on the community compositions, we assessed specificity and selectivity of hornwort–cyanobiont combinations. We followed the definitions of Bubrick et al. (1985) and defined specificity as partner compatibility and selectivity as partner preference.

RESULTS

Amplicon sequencing

We sequenced a total of 288 samples on three PacBio Sequel runs, 244 of which (168 hornwort, 76 soil) were from the focal Grossman Pond and Potato Hill samplings (the rest were controls and other samples not part of this study). Over the three PacBio runs, we obtained 1,573,867 raw sequences, which yielded 669,780 CCS reads passing filters on number of passes, accuracy, length, and correct barcodes and primers; of these, 598,442 reads were from the focal samples.

Our benchmark on mock communities clearly demonstrated that the ASV approach had a much higher accuracy and sensitivity than traditional OTU clustering (Figure 2). In two of the three mock samples, we recovered exactly five ASVs, which had sequences identical to those from the five input strains (Figure 2D). One mock sample returned one extra ASV (ASV25), which was not part of the mock community. However, this ASV had a low relative abundance (3%; Figure 2E). In mock 1 and 3, the ASV counts significantly deviated from equal abundance ($\chi^2 P = 0.002$ and 0.047, respectively; Figure 2E), but this deviation could be due to an extra PCR step that was used when making the mock, potentially introducing more biases. We detected no significant count difference in mock 2 ($P = 0.815$; Figure 2E). In contrast, the OTU method recovered many more taxa than were actually included (12 and 7 OTUs on average per mock when clustered at 97% and 95%, respectively, after removing singletons; Figure 2F, G).

In total, dada2 inferred 382 ASVs from the soil and plant samples (Figure 2A–C). Across all samples, the correlation between read count and ASV number was very weak ($R^2 = 0.0145$, $P = 0.0317$; Appendix S4), suggesting sufficient sequencing coverage. As expected with our sampling design of mixing multiple soil plugs together, soil samples usually harbored significantly more ASVs than plants (average ASVs/sample: 17.4 vs 10.3, $P < 10^{-14}$ unpaired t-test; Figure 2B). After filtering out ASVs with lower than 3% relative abundance, a hornwort sample on average had about 4.4 cyanobiont ASVs (Figure 2C) (5.8 ASVs for soil samples), and over 67% of the hornwort samples contained an ASV whose relative abundance reached at least 50% (Appendix S5).

Phylogenetic diversity of hornwort cyanobionts

Because of the unresolved nature of cyanobacteria taxonomy, we used the "subclade" scheme from Shih et al.

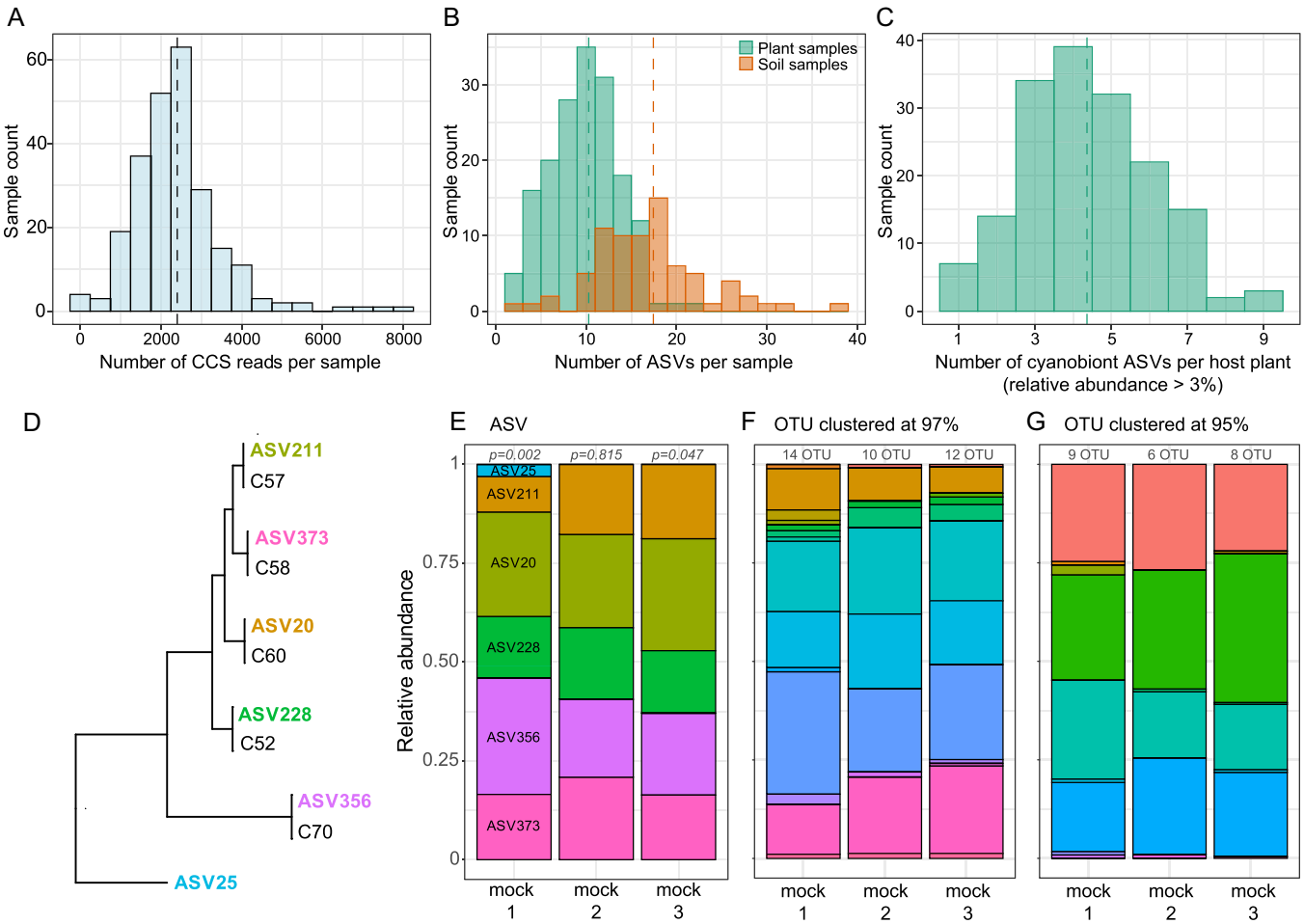

FIGURE 2 Summary of sequencing results. Distributions of (A) final filtered CCS reads, (B) all ASVs, and (C) ASVs from plants with relative abundance over 3%. Dashed lines are the group means. (D) The mock community consisted of five isolates and was sequenced in three separate PacBio runs (mock 1–3). A total of 6 ASVs were recovered from mock 1–3, five of which have identical sequences with the original isolates. The only exception is ASV25, which was not part of the mock community and only found in mock 1 and in low abundance. (E) Relative abundance of ASVs in mock 1–3. The italic numbers above bars are $\chi^2 p$-values tested against equal abundance. OTU clustering-based approach at (F) 97% and (G) 95% performed poorly in our PacBio data sets and cannot accurately recover the mock strains. Note that the colors of (F) and (G) do not correspond to the same OTUs

(2013) to provide a higher level classification for our ASVs. The soil cyanobacteria from our study spanned a wide phylogenetic range, including members from Subclade A (e.g., *Oscillatoria*), Subclade F (e.g., *Pseudanabaena*), and a few that could not be mapped onto the Shih et al. (2013) phylogeny (Figure 3; Appendix S6). All the hornwort cyanobiont ASVs fell under Subclade B1, roughly corresponding to Subsection IV (Nostocales) plus Subsection V (Stigonematales) from Bergey's Manual (Garrity et al., 2001). Subsections IV and V were circumscribed solely

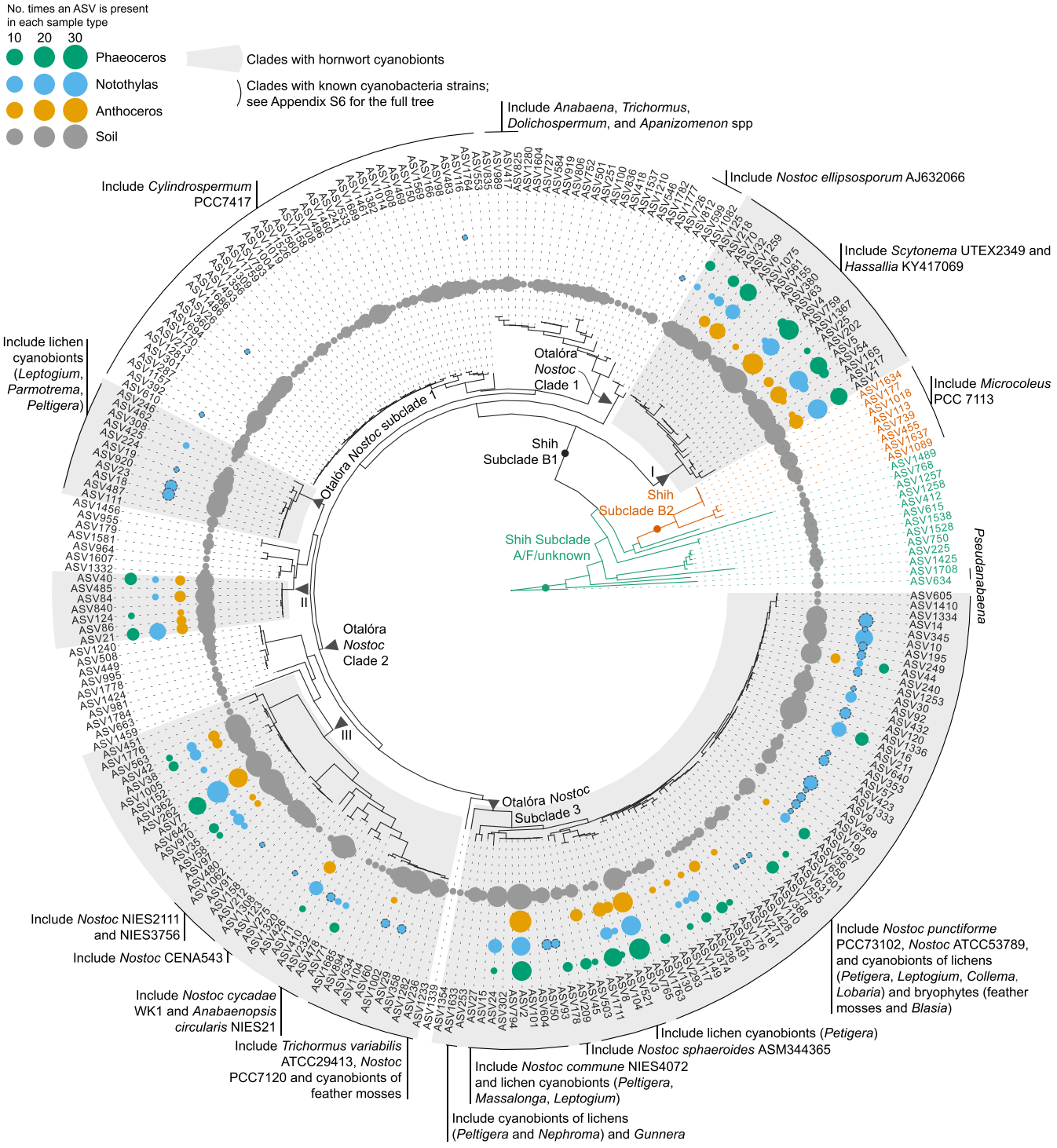

FIGURE 3 Phylogeny of cyanobacteria ASVs. The designations of Shih Subclades (denoted by filled circles) were based on Shih et al. (2013), and Otalóra *Nostoc* Clades and Otalóra *Nostoc* Subclades (denoted by arrowheads) were based on Otalóra et al. (2010) and Magain et al. (2017, 2018). The roman numerals (I–III) mark the hornwort cyanobiont groups that are not part of the Otalóra subclades. Blue circles with dashed outlines are *Notothylas* cyanobionts exclusively from Grossman Pond

based on morphology and were not recovered as monophyletic groups in our *rbcL-X* phylogeny.

The majority of the available sequences in Subclade B1 were annotated as "*Nostoc*", although this genus is clearly a polyphyletic assemblage. We mapped the *Nostoc* clade and subclade designations made by previous cyanolichen studies (Otálora et al., 2010, 2013; O'Brien et al., 2013; Magain et al., 2017, 2018) to our phylogeny (arrowheads in Figure 3), providing useful anchors for comparing across studies. The hornwort cyanobionts fell into five monophyletic groups, two of which overlapped with lichen photobionts ("Otálora *Nostoc* subclade 1" and "Otálora *Nostoc* subclade 3"), and one of which included cycad cyanobionts and cyanobacteria epiphytically associated with feather mosses (group III). The remaining two clades had no other known symbiotic members (groups I and II).

Overlap of symbiotic and soil cyanobacteria

We found a substantial overlap between the soil and the cyanobiont communities, with over half (53.8%) of the ASVs that appeared in more than one sample being found in both. This connection to soil can also be seen in Figure 3. Very few ASVs were unique to soils at Potato Hill (5.26%, Figure 4C), while 32.6% of the ASVs at Grossman Pond were unique to soil. An opposite but less extreme pattern appeared for ASVs unique to plants: 40.3% at Potato Hill and 20.9% at Grossman Pond.

Despite this strong overlap, the soil samples largely clustered by sites on the PCoA using weighted unifrac distance, with one cluster exclusively of Potato Hill samples and the other mostly Grossman Pond ones (Figure 4A), indicating phylogenetically distinct communities in the soil of the two sites. However, the separation of plants samples by site is weaker, and they showed much more variation, clustering most densely in a different corner of the ordination space from either of the soil clusters (Figure 4A).

Variation in hornwort and soil cyanobacteria communities over time and space

We found that sampling time explained the least of the community variance (Grossman Pond: $R^2 = 0.022$, $P = 0.559$; Potato Hill: $R^2 = 0.061$, $P < 0.001$) compared to location and sample type (Appendix S7). When we analyzed each quadrat separately, we also did not see samples clustered by

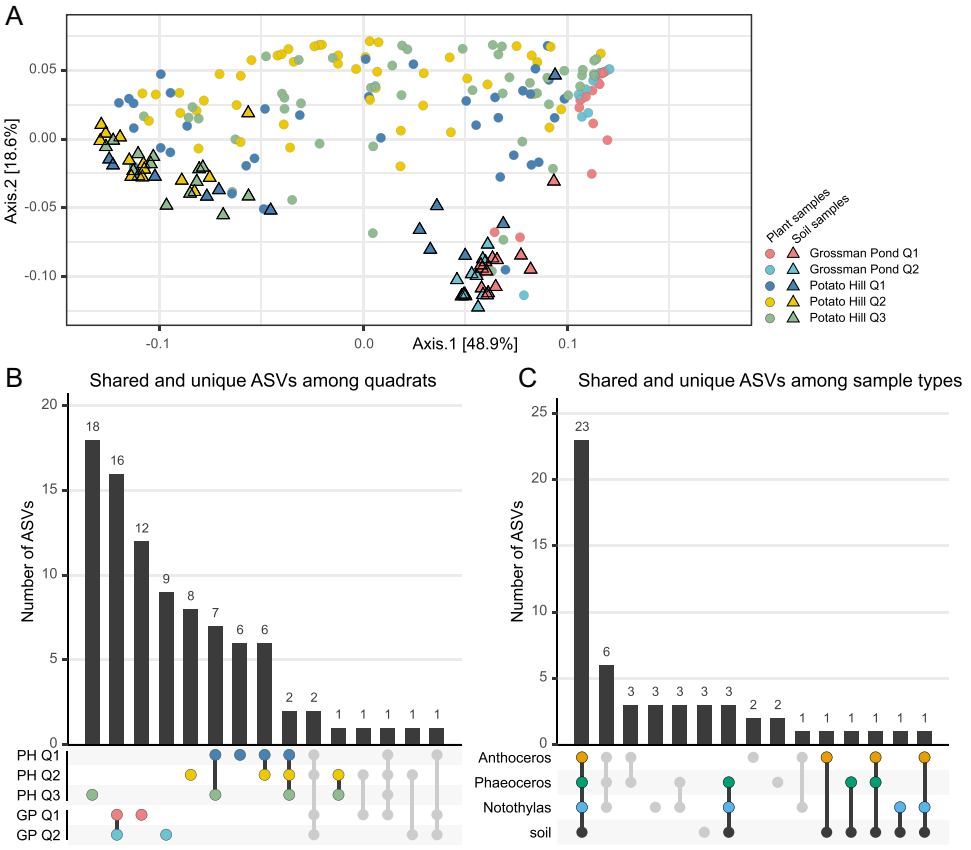

FIGURE 4 Comparison of soil and cyanobiont community composition. (A) PCoA based on weighted unifrac distance of all samples. (B) Shared and unique ASVs among the five quadrats. All the top categories are either unique or shared ASVs within the same sampling sites (highlighted in color). (C) Shared and unique ASVs among the sample types at Potato Hill site. Many cyanobiont ASVs can also be found in soil samples (highlighted in color). For (B) and (C) overlaps with zero count were not plotted, and only ASVs present in more than one sample were included. GP, Grossman Pond; PH, Potato Hill. Data from all time points were plotted together

time points on the PCoA plots (Appendix S8), and PERMANOVA tests found no significant association. The only exception was quadrat 1 at Potato Hill ($P < 10^{-4}$), which was largely driven by the *Notothylas* samples that did show a time-structured pattern (Appendix S8). Our ANCOM analysis could not identify any individual ASV that consistently changed across the sampling time points within each quadrat. ASVs that appeared in all time points or only one time point were the most common category for each quadrat (Appendix S9). Additionally, cyanobacterial communities did not appear to consistently become more similar or dissimilar (measured as pairwise weighted unifrac distance) through time (Appendix S10).

In terms of spatial variation, few ASVs were present (at more than 3% of sample reads) in a high percentage of samples, showing a limited core cyanobacterial community with substantial variation between nearby hosts, even on the scale of a few centimeters between collection points. Across the whole data set, the most common ASV (ASV7, in group III) only appeared in 35.25% of samples and most ASVs appeared in less than 10% (Appendix S11). When looking among the four replicate plant samples (collected at each time point/quadrat/hornwort species), very few ASVs were shared (Appendix S12). Additionally, all these shared ASVs had highly variable relative abundances across the four replicates (Appendix S12). While the two sites (~20 km apart) differed somewhat in the PCoA, quadrats did not form separate clusters (Figure 4A), indicating no overarching differences between collections separated by a few meters. However, separate ordinations of soil and plant samples from Potato Hill showed some separation between quadrats in soils but not in plants (Appendix S7). In terms of ASV overlap, we found that most ASVs were either unique to one quadrat or shared between the quadrats at the same site (Figure 4B), consistent with the ordination results (Figure 4A; Appendix S13). The two quadrats at Grossman Pond showed more overlap than the three at Potato Hill (Figure 4B).

Cyanobiont communities among three sympatric hornwort species

A large number of phylogenetically diverse ASVs were shared among the three sympatric host species (Figures 3, 4C), and alpha diversity indices were similar across the hosts (Appendix S14). Although Figure 3 showed a few distinct ASVs that only came from *Notothylas*, they were restricted to the Grossman Pond site (dashed blue circles in Figure 3) and absent in the soil samples elsewhere.

We did not observe clustering by host identity in ordinations of all samples or only Potato Hill plants (Figure 4A; Appendix S13), but did see some signal when looking at individual quadrats at Potato Hill. We found a by-host association in quadrat 3 (PERMANOVA $P < 10^{-4}$, $R^2 = 0.352$; Figure 5C) with *Notothylas* and *Anthoceros* samples mostly separating along the first PCoA axis (Figure 5E). A difference also appeared in quadrat 2 based on PERMANOVA ($P = 0.023$, $R^2 = 0.126$; Figure 5C, F), but we found no difference in partner preference in quadrat 1 ($P = 0.230$).

For quadrats 2 and 3, we further searched for the cyanobacteria ASVs that were differentially recruited by the three host species. Our ANCOM analyses identified ASV1 and ASV8 in quadrat 3, but none in quadrat 2, consistent with the much weaker by-host interaction in quadrat 2. ASV1 and ASV8 belonged to two distantly related cyanobacteria clades (group I and *Nostoc* Subclade 3, respectively; Figure 3) and exhibited different host association patterns: ASV1 tended to have higher abundances in *Notothylas* samples (Figure 5D), whereas ASV8 was more frequent in *Anthoceros* and *Phaeoceros* (Figure 5E). These two ASVs were either absent or in very low abundance in quadrats 1 and 2 soil samples (Figure 5D, E).

DISCUSSION

A high phylogenetic diversity of cyanobionts

The hornworts we sampled hosted a phylogenetically diverse community of cyanobacteria within the order Nostocales. Many of the hornwort cyanobionts were closely related to those from cyanolichens (e.g., *Peltigera*) or other plant associations. The overlaps between lichen and plant cyanobionts suggests that diverse fungal and plant lineages have independently used the same cyanobacteria lineages for symbiotic partners, consistent with earlier studies based on fewer samples (Costa et al., 2001; O'Brien et al., 2005). No hornwort cyanobionts were matched to two known lichen photobiont lineages, "Otálora *Nostoc* Subclade 2" and *Rhizonema* in this study. However, this result could simply reflect our geographic focus in a small part of temperate North America. "Otálora *Nostoc* Subclade 2" is mostly found in polar and boreal regions (Magain et al., 2017) and *Rhizonema* in the tropics (Lücking et al., 2009).

On the other hand, in two cases hornwort cyanobionts were either the only members of a clade (group II), or they made up the bulk of a clade with a handful of free-living strains (group I; Figure 3), indicating detection of novel cyanobiont diversity. Group I is particularly interesting. This clade diverged very deeply within Subclade B1 (Figure 3) and consists of cyanobionts from all three hornwort species from Potato Hill (absent in Grossman Pond), as well as free-living cyanobacteria strains of "*Scytonema hofmanni*" UTEX 2349 (from Watkins Glen, New York) and "*Hassallia* sp." (from Mojave National Preserve, California). These two free-living strains each share an identical or near-identical *rbcL-X* sequence with at least another cyanobiont ASV. While some lichen-forming basidiomycetes were thought to harbor *Scytonema*, Lücking et al. (2009) showed that their photobionts are phylogenetically distinct from *Scytonema* and belong to a separate monophyletic genus *Rhizonema*. Group I is neither related to *Scytonema* sensu stricto nor *Rhizonema* (Appendix S6). In addition, the species identity of "*S. hofmanni*" UTEX 2349 is questionable, as it

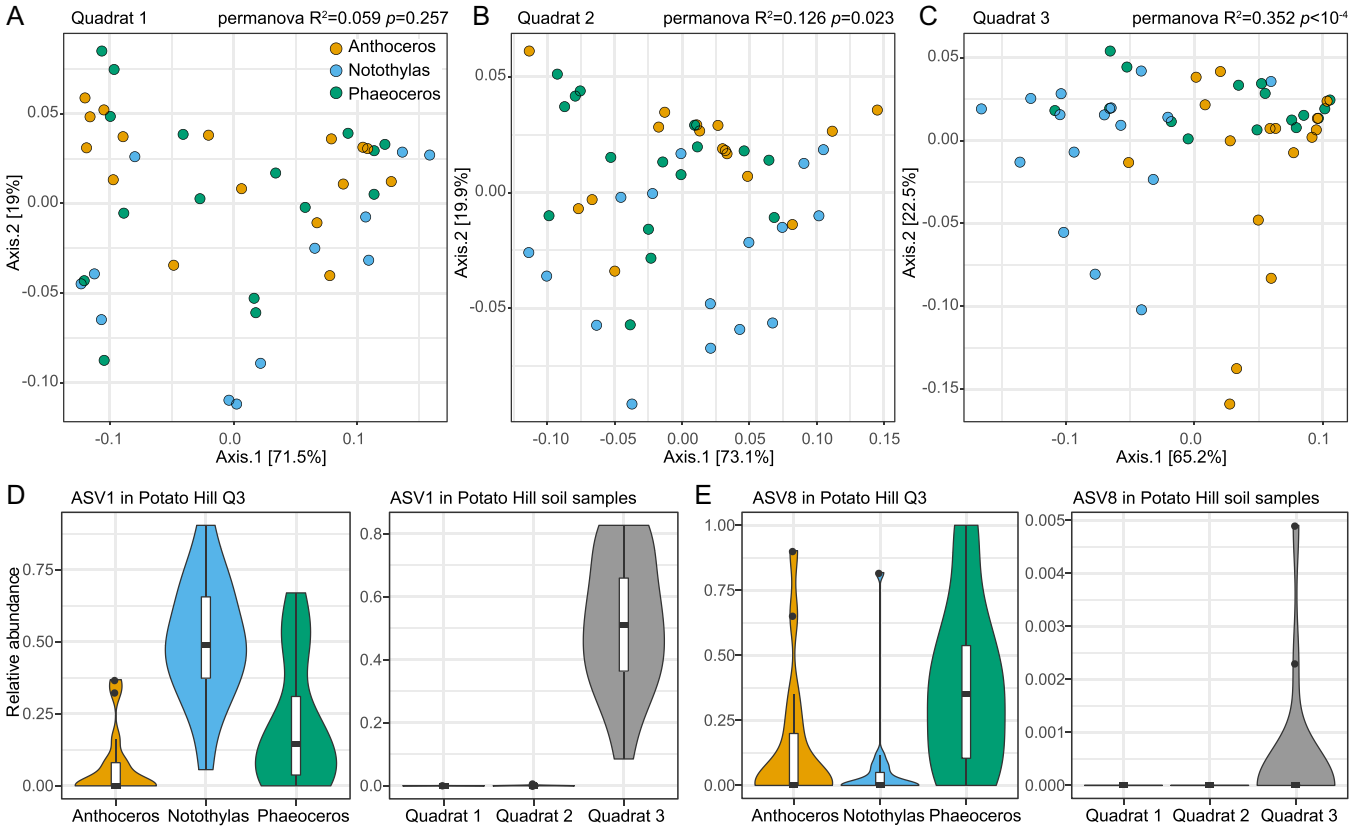

FIGURE 5 Cyanobiont communities among three sympatric hornwort species. PCoA plots (based on weighted unifrac distance) of (A) Quadrat 1, (B) Quadrat 2, and (C) Quadrat 3 in the Potato Hill site. The cyanobiont communities are color-coded by the host species, and PERMANOVA tests were done with host species as the grouping term. In Quadrat 3, cyanobiont communities are significantly organized by host species, with two ASVs that significantly vary in abundance among the three host species: (D) ASV1 and (E) ASV8 (the corresponding abundances in soil are shown on the right in each panel). Data from all time points were plotted together

did not group with the *Scytonema* species nor with other *S. hofmanni* isolates. The other known strain in this clade was attributed to *Hassallia*, which has only one *rbcL-X* sequence available on GenBank, and more data are needed to resolve its taxonomic status. This result suggests that further sampling of cyanobacteria from hornworts would be useful to describe not only broader diversity of symbioses than previously recorded but also new cyanobacterial taxa.

While these results suggest greater plant cyanobiont diversity than previously thought, these relationships must be tested in experimental studies to confirm the symbiotic status of these taxa. Some of the taxa may have been epi- or endophytes not associated with N-fixing symbiosis in the slime cavities. This study therefore provides targets for future culture-based experiments on symbiotic capabilities and for further field sampling.

A strong soil connection

Our study is the first to comprehensively profile soil cyanobacteria communities from around hornworts and gives a rare view of temperate soil cyanobacterial communities. We observed substantial overlap between cyanobiont and soil communities, but plant samples did not clearly group with the associated soil samples (Figure 4). The wider spread of plant samples observed might indicate that plants recruited a small but diverse subset of soil cyanobacteria. If two plant samples had cyanobionts from disparate parts of the phylogeny, the weighted unifrac distance between them would be high and hence show a more dispersed distribution in PCoA. In contrast to the homogenized soil samples, the plant ones would therefore have more variable positions on the ordination. This hypothesis may also explain why soils clustered by quadrat but plant samples did not. The overlap between hornworts and soils, coupled with the relatively low amount of ASVs unique to soils is not consistent with hornworts picking a narrow subset of soil cyanobacteria as symbionts. Thus, it is likely that hornworts are not very selective in symbionts or that hornwort samples were frequently contaminated with external soil cyanobacteria. However, if the strong overlap were mostly from contamination of plant samples, we would expect to see the shared ASVs with consistently low abundances in plants, and this is not the case. Future culture-based laboratory experiments could target taxonomically similar cyanobacteria isolated from soils to test whether they indeed form functional symbioses with hornworts. If it is in fact the case that the hornwort symbiosis is not tightly selective, investigating the

molecular mechanisms behind it may provide new, simpler avenues for engineering nitrogen-fixing symbioses into economically important plants.

The substantial overlap that we found between soil and hornwort cyanobacterial communities contrasts with the rather minor overlap found in previous work with cycads (Cuddy et al., 2012; Suárez-Moo et al., 2019) and bryophytes (West and Adams, 1997; Rikkinen and Virtanen, 2008; Liaimer et al., 2016). This difference could be due to the fact that most earlier studies were culture-based and hence could only catalog a small portion of the soil diversity. More recently, Suárez-Moo et al. (2019) used 16 S amplicon-seq to profile the microbiomes of coralloid roots of six Mexican cycad species (*Dioon* spp.), as well as the immediate rhizosphere and bulk soils. Of the 12 cyanobiont OTUs recovered, only five were also found in rhizosphere and/or soil samples. A similar study was done by Zheng and Gong (2019) on a Chinese *Cycas* species; however, very few cyanobacteria reads were recovered from their root and soil samples, making it difficult to compare the results. It should be noted that in our study, we did not sample the soil immediately below hornworts (but instead as multiple plugs across the quadrats), which means that the physical distance between soil and plant samples might be greater here than in that of Suárez-Moo et al. (2019) on cycads. Even so, our study still found that hornworts might have a tighter soil connection than other plant lineages, although more amplicon studies are needed to validate this pattern. This difference could arise from the thin, flat structure of hornworts that results in a much higher percent of their surface being in direct contact with the substrate and fewer physical barriers to the bacteria than is the case for larger, non-thalloid plants.

Our findings on hornwort–soil relations somewhat resemble what was reported in cyanolichens. Zúñiga et al (2017) characterized the substrate cyanobacteria diversity underneath *Peltigera* lichens, and reported a similarly large species overlap with cyanobionts from the same site. One important difference, however, is that in their study, lichens actually harbored a greater cyanobacteria diversity than the substrate (Zúñiga et al., 2017), which is opposite from what we found in hornworts. Aside from the differences in sequencing technologies and sampling strategies, one possible explanation is that lichens can disperse through asexual propagules that include the cyanobionts, thereby adding diversity to the symbiotic communities but not necessarily to the substrate. Hornwort gemmae, on the other hand, are not known to carry cyanobionts, and none of the hornwort species included here produce gemmae.

Taken together, our data imply that the contrast between symbiotic and substrate cyanobacteria might vary considerably among different symbiotic systems, potentially based on differences in structure and growth form, but the available data on this area are currently limited. More surveys and taxonomic studies of terrestrial cyanobacterial diversity would support investigations of these possible distinctions between host plants in soil connection.

No clear trend of cyanobiont community succession over the growing season

Our sampling did not detect clear patterns of cyanobacterial community change over the hornwort growing season. Based on the fact that the many of the ASVs were found at all sampling times, it seems that some parts of a site's cyanobacterial community are quite stable over the growing season. Our data do not provide evidence of cyanobacterial community succession over the time period studied. This lack of temporal pattern in the plant samples could suggest that colonization events are very stochastic or simply that the scale of the patterns present is greater than that covered by our sampling.

It is possible that the high spatial variation we observed could mask a relatively small temporal pattern. This caveat could be exacerbated by the small numbers of taxa present in each plant sample and the small number of replicate plants taken per time. In addition, the hornwort growing season in our sampling sites (August–October) might be too short to observe significant cyanobiont turnover. It would be therefore interesting to conduct similar time-course samplings in regions with less pronounced seasonality and where hornwort populations are present year-round. Finally, in the temperate climate with long winters where we surveyed, hornworts likely regenerate from spore banks each year, and the cyanobiont communities would therefore be determined by soil conditions at germination time. In other words, interannual variation should be of particular interest for future studies.

Cyanobiont communities vary between physically close hosts

We found that hornwort plants growing only a few centimeters apart can have very different sets of strains. Our study is the first to apply next-generation sequencing on hornworts that form discrete cyanobacterial colonies, confirming the within- and between-thallus variability of cyanobionts detected with culturing previously (West and Adams, 1997; Costa et al., 2001). Our data are consistent with separate slime cavities being independently colonized as they form (Adams and Duggan, 2008) and also show that individual hornworts are often dominated by one or two cyanobionts (Appendix S5). However, these dominant taxa differ between nearby plants (Appendix S12), suggesting stochasticity in the colonization process. Future laboratory culture experiments could determine how much this variability may be due to competition between potential partners, priority effects, or random chance. Based on the high variability at small spatial scales, it will also be important to test in future work if the 0.25-m^2 quadrat size can meaningfully represent hornwort cyanobiont communities.

Similar partner preferences among sympatric species

Sympatric species offer a unique opportunity to compare selectivity because the hosts are exposed to the same pool of symbionts (and vice versa). We did not observe host species to be a strong organizing factor, which is in line with previous work on cycad-cyanobacteria interactions (Gehringer et al., 2010; Suárez-Moo et al., 2019) and other work indicating that plant–cyanobiont relationships outside the vertically transmitted *Azolla* symbiosis are not strongly specific (Rasmussen and Johansson, 2002). Since our data used more precise methods of genetic differentiation than many previous studies, we can confirm that this lack of strong specificity or selectivity is not simply an artifact of low-resolution symbiont identification.

Nevertheless, we detected some possible differential selectivity of the three hornwort host species. Because the two differentially abundant ASVs detected in one quadrat were absent in the soils of other quadrats, this selectivity may be dependent on the background cyanobacteria pool. However, in this study, we only had three quadrats with sympatric species and observed high levels of variation between individual plants, so some selectivity may not have been detectable. Future field sampling with more sites and replicates should allow the effects of host selectivity versus other environmental variation to be clarified. Laboratory experiments are also needed to determine the direction of selection—whether hornworts select cyanobacteria or cyanobacteria select hornworts—and to explore whether competition for partners (and not selection per se) can also result in the by-host association.

CONCLUSIONS

This study opens new opportunities for understanding the ecology and natural history of plant-cyanobacteria symbiosis. We demonstrated the efficacy of the *rbcL-X* PacBio metabarcoding approach to profile cyanobacteria communities. We found that cyanobiont communities strongly overlap with adjacent soils but are highly variable between individual hornworts in the same habitat. These communities are not clearly determined by distance, time, or host species. Our results highlight the importance of sampling soil and sympatric species with greater replication to tackle the apparently complex drivers of symbiotic community composition. This study marks an important step for using hornworts to explore the dynamics of nitrogen-fixing symbioses and defines many future directions for laboratory and field studies on this topic.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation grant (DEB-1831428) to F.-W. Li. We thank Norm Trigoboff for the information on hornwort distribution in Potato Hill State Forest and Juan Carlos Villarreal, two anonymous reviewers, and an associate editor for invaluable comments.

AUTHOR CONTRIBUTIONS

J.M.N. and F.-W.L. conceived the project, carried out the sampling and data analyses, and wrote the manuscript. J.M.N. and D.A.H. carried out the molecular work.

DATA AVAILABILITY STATEMENT

The sequencing reads were deposited at NCBI SRA under the accession PRJNA632853.

The scripts, commands, and data matrices from this study can be found at https://github.com/fayweili/hornwort_cyano_interaction.

ORCID

Fay-Wei Li http://orcid.org/0000-0002-0076-0152

REFERENCES

Adams, D. G., and P. S. Duggan. 2008. Cyanobacteria–bryophyte symbioses. *Journal of Experimental Botany* 59: 1047–1058.

Andrews, M., and M. E. Andrews. 2017. Specificity in legume–rhizobia symbioses. *International Journal of Molecular Sciences* 18: 705.

Bell-Doyon, P., J. Laroche, K. Saltonstall, and J. C. Villarreal. 2020. Specialized bacteriome uncovered in the coralloid roots of the epiphytic gymnosperm, Zamia pseudoparasitica. *Environmental DNA* 2: 418–428.

Bouchard, R., G. Peñaloza-Bojacá, S. Toupin, Y. Guadalupe, J. Gudiño, N. S. Allen, F. W. Li, and J. C. Villarreal. 2020. Contrasting bacteriome of the hornwort Leiosporoceros dussii in two nearby sites with emphasis on the hornwort-cyanobacterial symbiosis. *Symbiosis* 81: 39–52.

Bubrick, P., A. Frensdorff, and M. Galun. 1985. Selectivity in the lichen symbiosis. *In* D. H. Brown [ed.], Lichen physiology and cell biology, 319–334. Springer, Basel, Switzerland.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13: 581–583.

Callahan, B. J., J. Wong, C. Heiner, S. Oh, C. M. Theriot, A. S. Gulati, S. K. McGill, and M. K. Dougherty. 2019. High-throughput amplicon sequencing of the full-length 16 S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research* 47: e103.

Conway, J. R., A. Lex, and N. Gehlenborg. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33: 2938–2940.

Costa, J. L., P. Paulsrud, and P. Lindblad. 1999. Cyanobiont diversity within coralloid roots of selected cycad species. *FEMS Microbiology Ecology* 28: 85–91.

Costa, J.-L., P. Paulsrud, J. Rikkinen, and P. Lindblad. 2001. Genetic diversity of Nostoc symbionts endophytically associated with two bryophyte species. *Applied and Environmental Microbiology* 67: 4393–4396.

Cuddy, W. S., B. A. Neilan, and M. M. Gehringer. 2012. Comparative analysis of cyanobacteria in the rhizosphere and as endosymbionts of cycads in drought-affected soils. *FEMS Microbiology Ecology* 80: 204–215.

Dijkhuizen, L. W., P. Brouwer, H. Bolhuis, G.-J. Reichart, N. Koppers, B. Huettel, A. M. Bolger, et al. 2018. Is there foul play in the leaf pocket? The metagenome of floating fern Azolla reveals endophytes that do not fix N_2 but may denitrify. *New Phytologist* 217: 453–466.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* 14: 927–930.

Fernández-Martínez, M. A., A. de los Ríos, L. G. Sancho, and S. Pérez-Ortega. 2013. Diversity of endosymbiotic Nostoc in Gunnera magellanica from Tierra del Fuego, Chile. *Microbial Ecology* 66: 335–350.

Frangedakis, E., M. Shimamura, J. C. Villarreal, F.-W. Li, M. Tomaselli, M. Waller, K. Sakakibara, et al. 2020. The hornworts: morphology, evolution and development. *New Phytologist* 229: 735–754.

Garrity, G. M., D. R. Boone, and R. W. Castenholz [eds.]. 2001. Bergey's manual of systematic bacteriology, 2nd ed. Springer-Verlag, NY, NY, USA.

Gehringer, M. M., J. J. L. Pengelly, W. S. Cuddy, C. Fieker, P. I. Forster, and B. A. Neilan. 2010. Host selection of symbiotic cyanobacteria in 31 species of the Australian cycad genus: Macrozamia (Zamiaceae). *Molecular Plant-Microbe Interactions* 23: 811–822.

Goldmann, K., R. S. Boeddinghaus, S. Klemmer, K. M. Regan, A. Heintz-Buschart, M. Fischer, D. Prati, et al. 2020. Unraveling spatiotemporal variability of arbuscular mycorrhizal fungi in a temperate grassland plot. *Environmental Microbiology* 22: 873–888.

Guevara, R., J. J. Armesto, and M. Caru. 2002. Genetic diversity of Nostoc microsymbionts from Gunnera tinctoria revealed by PCR-STRR fingerprinting. *Microbial Ecology* 44: 127–136.

Gutiérrez-García, K., E. D. Bustos-Díaz, J. A. Corona-Gómez, H. E. Ramos-Aboites, N. Sélem-Mojica, P. Cruz-Morales, M. A. Pérez-Farrera, et al. 2019. Cycad coralloid roots contain bacterial communities including cyanobacteria and Caulobacter spp. that encode niche-specific biosynthetic gene clusters. *Genome Biology and Evolution* 11: 319–334.

Holland-Moritz, H., J. Stuart, L. R. Lewis, S. Miller, M. C. Mack, S. F. McDaniel, and N. Fierer. 2018. Novel bacterial lineages associated with boreal moss species. *Environmental Microbiology* 20: 2625–2638.

Johansen, J. R., J. Mareš, N. Pietrasiak, M. Bohunická, J. Zima, L. Štenclová, and T. Hauer. 2017. Highly divergent 16 S rRNA sequences in ribosomal operons of Scytonema hyalinum (Cyanobacteria). *PLoS One* 12: e0186393.

Johansson, C., and B. Bergman. 1994. Reconstitution of the symbiosis of Gunnera manicata Linden: cyanobacterial specificity. *New Phytologist* 126: 643–652.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463–1464.

Komárek, J., J. Kaštovský, J. Mareš, and J. R. Johansen. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86: 295–335.

Leizerovich, I., N. Kardish, and M. Galun. 1990. Comparison between eight symbiotic, cultured Nostoc isolates and a free-living Nostoc by recombinant DNA. *Symbiosis* 8: 75–85.

Li, F.-W., P. Brouwer, L. Carretero-Paulet, S. Cheng, J. Vries, P.-M. Delaux, A. Eily, et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* 4: 1–16.

Li, F.-W., T. Nishiyama, M. Waller, E. Frangedakis, J. Keller, Z. Li, N. Fernandez-Pozo, et al. 2020. Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature Plants* 6: 259–272.

Liaimer, A., J. B. Jensen, and E. Dittmann. 2016. A genetic and chemical perspective on symbiotic recruitment of cyanobacteria of the genus Nostoc into the host plant Blasia pusilla L. *Frontiers in Microbiology* 7: 1693.

Lücking, R., J. D. Lawrey, M. Sikaroodi, P. M. Gillevet, J. L. Chaves, H. J. M. Sipman, and F. Bungartz. 2009. Do lichens domesticate photobionts like farmers domesticate crops? Evidence from a previously unrecognized lineage of filamentous cyanobacteria. *American Journal of Botany* 96: 1409–1418.

Magain, N., J. Miadlikowska, B. Goffinet, E. Sérusiaux, and F. Lutzoni. 2017. Macroevolution of specificity in cyanolichens of the genus Peltigera section Polydactylon (Lecanoromycetes, Ascomycota). *Systematic Biology* 66: 74–99.

Magain, N., C. Tniong, T. Goward, D. Niu, B. Goffinet, E. Sérusiaux, O. Vitikainen, et al. 2018. Species delimitation at a global scale reveals high species richness with complex biogeography and patterns of symbiont association in Peltigera section Peltigera (lichenized Ascomycota: Lecanoromycetes). *Taxon* 67: 836–870.

Mandal, S., W. V. Treuren, R. A. White, M. Eggesbø, R. Knight, and S. D. Peddada. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease* 26: 27663.

Materatski, P., C. Varanda, T. Carvalho, A. Dias, M. D. Campos, F. Rei, and M. doR. Félix. 2018. Spatial and temporal variation of fungal endophytic richness and diversity associated to the phyllosphere of olive cultivars. *Fungal Biology* 123: 66–76.

McMurdie, P. J., and S. Holmes. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8: e61217.

Meeks, J. C. 1998. Symbiosis between nitrogen-fixing cyanobacteria and plants. *Bioscience* 48: 266–276.

Meeks, J. C. 2009. Physiological adaptations in nitrogen-fixing Nostoc–plant symbiotic associations. *Microbiological Monograph* 8: 181–205.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE), 1–8.

Mirarab, S., N. Nguyen, S. Guo, L.-S. Wang, J. Kim, and T. Warnow. 2015. PASTA: Ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *Journal of Computational Biology* 22: 377–386.

Mus, F., M. B. Crook, K. Garcia, A. G. Costas, B. A. Geddes, E. D. Kouri, P. Paramasivan, et al. 2016. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Applied and Environmental Microbiology* 82: 3698–3710.

Nelson, J., D. Hauser, J. A. Gudiño, Y. A. Guadalupe, J. C. Meeks, N. Salazar, J. C. Villarreal, and F.-W. Li. 2019. Complete genomes of symbiotic cyanobacteria clarify the evolution of vanadium-nitrogenase. *Genome Biology and Evolution* 11: 1959–1964.

Nilsson, M., B. Bergman, and U. Rasmussen. 2000. Cyanobacterial diversity in geographically related and distant host plants of the genus Gunnera. *Archives of Microbiology* 173: 97–102.

O'Brien, H. E., J. Miadlikowska, and F. Lutzoni. 2005. Assessing host specialization in symbiotic cyanobacteria associated with four closely related species of the lichen fungus Peltigera. *European Journal of Phycology* 40: 363–378.

O'Brien, H. E., J. Miadlikowska, and F. Lutzoni. 2013. Assessing population structure and host specialization in lichenized cyanobacteria. *New Phytologist* 198: 557–566.

Oldroyd, G. E. D., J. D. Murray, P. S. Poole, and J. A. Downie. 2011. The rules of engagement in the legume–rhizobial symbiosis. *Annual Review of Genetics* 45: 119–144.

O'Malley, M. A. 2008. 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Studies in History and Philosophy of Science Part C* 39: 314–325.

Otálora, M. A. G., I. Martínez, H. O'Brien, M. C. Molina, G. Aragón, and F. Lutzoni. 2010. Multiple origins of high reciprocal symbiotic specificity at an intercontinental spatial scale among gelatinous lichens (Collemataceae, Lecanoromycetes). *Molecular Phylogenetics and Evolution* 56: 1089–1095.

Otálora, M. A. G., C. Salvador, I. Martínez, and G. Aragón. 2013. Does the reproductive strategy affect the transmission and genetic diversity of bionts in cyanolichens? A case study using two closely related species. *Microbial Ecology* 65: 517–530.

Papaefthimiou, D., C. V. Hove, A. Lejeune, U. Rasmussen, and A. Wilmotte. 2008. Diversity and host specificity of Azolla cyanobionts. *Journal of Phycology* 44: 60–70.

Peršoh, D. 2015. Plant-associated fungal communities in the light of meta'omics. *Fungal Diversity* 75: 1–25.

Rai, A. N., E. Soderback, and B. Bergman. 2000. Cyanobacterium–plant symbioses. *New Phytologist* 147: 449–481.

Rajaniemi, P., P. Hrouzek, K. Kaštovská, R. Willame, A. Rantala, L. Hoffmann, J. Komárek, and K. Sivonen. 2005. Phylogenetic and morphological evaluation of the genera Anabaena, Aphanizomenon, Trichormus and Nostoc (Nostocales, Cyanobacteria). *International Journal of Systematic and Evolutionary Microbiology* 55: 11–26.

Ranjard, L., S. Dequiedt, N. C. Prévost-Bouré, J. Thioulouse, N. P. A. Saby, M. Lelievre, P. A. Maron, et al. 2013. Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nature Communications* 4: 1434.

Rasmussen, U., and C. Johansson. 2002. Diversity and specificity in cyanobacterial symbioses. *Biology & Environment: Proceedings of the Royal Irish Academy* 102: 53–56.

Rikkinen, J. 2013. Molecular studies on cyanobacterial diversity in lichen symbioses. *MycoKeys* 6: 3–32.

Rikkinen, J., and V. Virtanen. 2008. Genetic diversity in cyanobacterial symbionts of thalloid bryophytes. *Journal of Experimental Botany* 59: 1013–1021.

Rippin, M., S. Lange, N. Sausen, and B. Becker. 2018. Biodiversity of biological soil crusts from the Polar Regions revealed by metabarcoding. *FEMS Microbiology Ecology* 94: fiy036.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4: e2584.

Rothfels, C. J., K. M. Pryer, and F. Li. 2016. Next-generation polyploid phylogenetics: rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing. *New Phytologist* 213: 413–429.

Rout, M. E., and R. M. Callaway. 2012. Interactions between exotic invasive plants and soil microbes in the rhizosphere suggest that 'everything is not everywhere.' *Annals of Botany* 110: 213–222.

Rudi, K., O. M. Skulberg, and K. S. Jakobsen. 1998. Evolution of cyanobacteria by exchange of genetic material among phyletically related strains. *Journal of Bacteriology* 180: 3453–3461.

Santi, C., D. Bogusz, and C. Franche. 2013. Biological nitrogen fixation in non-legume plants. *Annals of Botany* 111: 743–767.

Schlaeppi, K., S. F. Bender, F. Mascher, G. Russo, A. Patrignani, T. Camenzind, S. Hempel, et al. 2016. High-resolution community profiling of arbuscular mycorrhizal fungi. *New Phytologist* 212: 780–791.

Schuster, R. M. 1992. The Hepaticae and Anthocerotae of North America East of the hundredth meridian. Field Museum of Natural History, Chicago, IL, USA.

Shade, A., J. G. Caporaso, J. Handelsman, R. Knight, and N. Fierer. 2013. A meta-analysis of changes in bacterial and archaeal communities with time. *ISME Journal* 7: 1493–1506.

Shih, P. M., D. Wu, A. Latifi, S. D. Axen, D. P. Fewer, E. Talla, A. Calteau, et al. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences, 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, R. K., E. R. A. Pederson, P. D. Weyman, P. K. Weber, U. Rassmussen, and C. L. Dupont. 2020. Bidirectional C and N transfer and a potential role for sulfur in an epiphytic diazotrophic mutualism. *ISME Journal* 14: 3068–3078.

Suárez-Moo, P. de J., A. P. Vovides, M. P. Griffith, F. Barona-Gómez, and A. Cibrián-Jaramillo. 2019. Unlocking a high bacterial diversity in the coralloid root microbiome from the cycad genus Dioon. *PLoS One* 14: e0211271.

Tedersoo, L., A. Tooming-Klunderud, and S. Anslan. 2018. PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytologist* 217: 1370–1385.

Thajuddin, N., G. Muralitharan, M. Sundaramoorthy, R. Ramamoorthy, S. Ramachandran, M. A. Akbarsha, and M. Gunasekaran. 2010. Morphological and genetic diversity of symbiotic cyanobacteria from cycads. *Journal of Basic Microbiology* 50: 254–265.

Villarreal, J. C., and K. S. Renzaglia. 2015. The hornworts: important advancements in early land plant evolution. *Journal of Bryology* 37: 157–170.

Wagner, M. R., D. S. Lundberg, T. G. del Rio, S. G. Tringe, J. L. Dangl, and T. Mitchell-Olds. 2016. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications* 7: 12151.

Wang, D., S. Yang, F. Tang, and H. Zhu. 2012. Symbiosis specificity in the legume-rhizobial mutualism. *Cellular Microbiology* 14: 334–342.

West, N. J., and D. G. Adams. 1997. Phenotypic and genotypic comparison of symbiotic and free-living cyanobacteria from a single field site. *Applied and Environmental Microbiology* 63: 4479–4484.

Yamada, S., S. Ohkubo, H. Miyashita, and H. Setoguchi. 2012. Genetic diversity of symbiotic cyanobacteria in Cycas revoluta (Cycadaceae). *FEMS Microbiology Ecology* 81: 696–706.

Yu, G., T. T.-Y. Lam, H. Zhu, and Y. Guan. 2018. Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Molecular Biology and Evolution* 35: 3041–3043.

Zhang, J., X.-X. Fu, R.-Q. Li, X. Zhao, Y. Liu, M.-H. Li, A. Zwaenepoel, et al. 2020. The hornwort genome and early land plant evolution. *Nature Plants* 6: 107–118.

Zheng, W., T. Song, X. Bao, B. Bergman, and U. Rasmussen. 2002. High cyanobacterial diversity in coralloid roots of cycads revealed by PCR fingerprinting. *FEMS Microbiology Ecology* 40: 215–222.

Zheng, Y., T.-Y. Chiang, C.-L. Huang, and X. Gong. 2018. Highly diverse endophytes in roots of Cycas bifida (Cycadaceae), an ancient but endangered gymnosperm. *Journal of Microbiology* 56: 337–345.

Zheng, Y., and X. Gong. 2019. Niche differentiation rather than biogeography shapes the diversity and composition of microbiome of Cycas panzhihuaensis. *Microbiome* 7: 152.

Zúñiga, C., D. Leiva, M. Carú, and J. Orlando. 2017. Substrates of Peltigera lichens as a potential source of cyanobionts. *Microbial Ecology* 74: 561–569.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Appendix S1. Soil sampling scheme.

Appendix S2. The barcode sequences used in PCR primers.

Appendix S3. PCR recipe.

Appendix S4. Assessment of sufficient sequencing coverage.

Appendix S5. Abundance distribution of the top ASV per sample.

Appendix S6. Phylogeny of ASVs from this study and other *rbcL-X* sequences.

Appendix S7. PERMANOVA table.

Appendix S8. Cyanobacteria community through time.

Appendix S9. Shared and unique ASVs across the time points (T1–T4).

Appendix S10. Changes of pairwise weighted unifrac distance through time.

Appendix S11. Statistics on the core cyanobacteria in different datasets.

Appendix S12. High variation of ASV abundance among replicate plants.

Appendix S13. Cyanobacteria community in the Potato Hill site.

Appendix S14. Comparison of cyanobacteria alpha diversity indices among the sampled species.

How to cite this article: Nelson, J. M., D. A. Hauser, and F.-W. Li. 2021. The diversity and community structure of symbiotic cyanobacteria in hornworts inferred from long-read amplicon sequencing. *American Journal of Botany* 108(9): 1731–1744. https://doi.org/10.1002/ajb2.1729