SHOTGUN METAGENOMICS AND MICROSCOPY INDICATE DIVERSE CYANOPHYTES, OTHER BACTERIA, AND MICROEUKARYOTES IN THE EPIMICROBIOTA OF A NORTHERN CHILEAN WETLAND *NOSTOC* (CYANOBACTERIA)[1]

Anchittha Satjarak

Plants of Thailand Research Unit, Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok 103330, Thailand

Linda E. Graham [ID],[2] *Michael J. Piotrowski, Marie T. Trest, Lee W. Wilcox*

Department of Botany, University of Wisconsin, Madison, Wisconsin 53706, USA

Martha E. Cook

School of Biological Sciences, Illinois State University, Normal, Illinois 61790, USA

Jennifer J. Knack

Department of Biology, University of Minnesota, Duluth, Minnesota 55812, USA

and Patricia Arancibia-Avila

Department of Basic Sciences, University of Bío-Bío, Chillan, Chile

Prokaryotic *Nostoc*, one of the world's most conspicuous and widespread algal genera (similar to eukaryotic algae, plants, and animals) is known to support a microbiome that influences host ecological roles. Past taxonomic characterizations of surface microbiota (epimicrobiota) of free-living *Nostoc* sampled from freshwater systems employed 16S rRNA genes, typically amplicons. We compared taxa identified from 16S, 18S, 23S, and 28S rRNA gene sequences filtered from shotgun metagenomic sequence and used microscopy to illuminate epimicrobiota diversity for *Nostoc* sampled from a wetland in the northern Chilean Altiplano. Phylogenetic analysis and rRNA gene sequence abundance estimates indicated that the host was related to *Nostoc punctiforme* PCC 73102. Epimicrobiota were inferred to include 18 epicyanobacterial genera or uncultured taxa, six epieukaryotic algal genera, and 66 anoxygenic bacterial genera, all having average genomic coverage ≥90X. The epicyanobacteria *Geitlerinemia, Oscillatoria, Phormidium,* and an uncultured taxon were detected only by 16S rRNA gene; *Gloeobacter* and *Pseudanabaena* were detected using 16S and 23S; and *Phormididesmis, Neosynechococcus, Symphothece, Aphanizomenon, Nodularia, Spirulina, Nodosilinea, Synechococcus, Cyanobium,* and *Anabaena* (the latter corroborated by microscopy), plus two uncultured cyanobacterial taxa (JSC12, O77) were detected only by 23S rRNA gene sequences. Three chlamydomonad and two heterotrophic stramenopiles genera were inferred from 18S; the streptophyte green alga *Chaetosphaeridium globosum* was detected by microscopy and 28S rRNA genes, but not 18S rRNA genes. Overall, >60% of epimicrobial taxa were detected by markers other than 16S rRNA genes. Some algal taxa observed microscopically were not detected from sequence data. Results indicate that multiple taxonomic markers derived from metagenomic sequence data and microscopy increase epimicrobiota detection.

Key index words: Chile; epimicrobiota; *Nostoc*; shotgun metagenomics; wetlands

Abbreviations: DIC, differential interference contrast

The filamentous, heterocyte-bearing cyanobacterial genus *Nostoc* (Order Nostocales, Phylum Cyanobacteria [alternatively, Cyanobacteriota; Soo et al. 2019])–is one of the world's most conspicuous and widespread algal representatives. Though single filaments can occur, this genus is known for forming macroscopic ball-, cylinder-, sheet-, or ribbon-shaped colonies of bent or kinked filaments held within a firm mucilaginous matrix often colored yellow, brown, or black by the sunscreen pigment scytonemin (Graham et al. 2016). Though inferred to occur in tropical seawater samples analyzed with the Nanopore MinION™ sequencing platform (Curren et al. 2019), this genus has most often been associated with freshwater-terrestrial habitats, including extreme environments of the Arctic and Antarctic,

and warm deserts (Dodds et al. 1995). Some *Nostoc* species are known for exceptional desiccation resistance (Inoue-Sakamoto et al. 2018) that fosters survival in arid environments. For example, environmental genomics revealed the presence of *Nostoc* among the cyanobacteria detected from soils or stones in the northern Chilean Atacama Desert, one of Earth's driest places (Jung et al. 2019). Our previous metagenomic study of a *Nostoc commune* from Patagonia (Graham et al. 2014) revealed the presence of a *wspA* gene associated with water stress (Arima et al. 2012).

Nostoc's resistance to high irradiance exposure has been linked to occurrence in the colonial matrix or sheath of brown-colored scytonemin, a UVA sunscreen pigment whose biosynthesis-related genes have been used as deep-time evolutionary markers (Garcia-Pichel et al. 2019). The use of mutants and confocal autofluorescence microscopy indicated that the *ebo* gene cluster functions in the export of scytonemin monomers to the periplasm for final oxidative dimerization (Klicki et al. 2018). A study of desiccation tolerance indicated that *N. punctiforme* strain M-15 tolerated desiccation and freeze-thawing (Inoue-Sakamoto et al. 2018), another advantageous trait in extreme environments. In reviewing the ecological significance of *Nostoc*, Dodds et al. (1995) particularly noted its relevance to global nitrogen fixation. The magnitude of global nitrogen fixation (and other biogeochemical impacts) associated with modern freshwater-terrestrial *Nostoc* has recently been estimated at 2.7×10^{10} kg annually, and as much as 3.9×10^{19} kg over the past 1.45 billion y (Graham et al. 2018a).

We had previously used metagenomic methods and microscopy to characterize microbiota and some functional genes of a lacustrine *Nostoc* sampled from Chilean Patagonia. In that study, 16S rRNA genes filtered from shotgun metagenomic long-read (Roche 454) sequences indicated host taxonomic classification as *N. commune* and suggested presence of diverse prokaryotic epibionts, including nitrate-reducing bacteria that foster carbonate production by reducing pH, and genes specific to nitrate reduction and scytonemin biosynthesis, but not a significant diversity of epicyanobacteria (Graham et al. 2014). Subsequently, epimicrobiota of *Nostoc* sampled from New Zealand wetlands were inferred from 16S rRNA gene amplicons and light microscopy (Secker et al. 2016), and such amplicons were also employed to characterize host taxonomic relationships and bacterial communities associated with lake and stream *Nostoc* sampled from the Altiplano of northern Chile (Aguilar et al. 2019).

Reviews of previous studies of microbial diversity in various Altiplano wetland systems (Albarracín et al. 2015, Orellana et al. 2018) indicate that the microbiota of *Nostoc* sampled from this region have not previously been studied using metagenomic methods, which offer advantages such as availability of multiple marker sequences (Graham et al. 2015). We hypothesized that employing 23S, 18S, and 28S rRNA gene markers filtered from metagenomic data might reveal greater taxonomic diversity of the *Nostoc* epimicrobiota than would 16S rRNA genes alone, particularly evidence for algal epiphytes. To test this concept, we compared numbers and types of epimicrobiota taxa inferred at $\geq 90X$ average genomic coverage when 16S rRNA genes were employed as a taxonomic marker to those inferred using 23S, 18S, and 28S rRNA genes, focusing on cyanobacteria and eukaryotic algae. We also used light and SEM to gain insight into epimicrobial morphotype diversity and to compare results obtained from microscopy with molecular assessments.

MATERIALS AND METHODS

Sampling. *Nostoc* samples for metagenomic and microscopic analysis were collected from the Bofedales de Parinacota, a wetland in Lauca National Park in the Altiplano of the Arica and Parinacota Region, Chile, during January 2016. Lauca National Park, which includes Chungará Lake and is known for distinctive puna steppe shrublands and scenic volcanic peaks, has been designated a UNESCO International Biosphere Reserve. Like Chungará Lake, the Bofedales de Parinacota is drained by the Rio Lauca, from which water is diverted for hydroelectricity generation and agriculture in northern Chile. Rio Lauca then flows eastward into Bolivia, and so has been the subject of international water rights issues. Like other inland wetlands of the world, bofedales are considered key to wildlife conservation, regional hydrology, and carbon storage (Rundel and Palma 2000). Throughout its extent, the river basin is characterized by high salinity (electrical conductivity > 2 mS $\cdot$ cm^{-1}) and high concentrations of heavy metals (Copaja and Muñoz 2018).

Sampling for this study occurred at GPS coordinates 18.219112 S, 69.281515 W, at an elevation of 4,495.8 m, near the edge of Highway A181 between the intersection with Highway 11 and the village of Parinacota. Within sight were bird flocks and grazing herds of camelids such as guanaco (*Lama guanaco*), llama (*Lama glama*), vicuña (*Vicugna vicugna*), and alpaca (*Vicugna pacos*). From pools of standing water occurring between cushion plant hummocks, multiple randomly chosen pieces, a few square cm in size, of macroscopic dark-brown sheets inferred to represent colonies of the filamentous cyanobacterium *Nostoc* were collected from just beneath the water surface. In this report, we refer to such samples as Parinacota *Nostoc*, to facilitate comparisons with *Nostoc* samples taken by other investigators from northern Chilean Altiplano lake or stream locales in the same year (Aguilar et al. 2019). Park rangers were present during the collection of Parinacota *Nostoc* samples, which did not include obvious evidence for organisms other than microbes. No other biological materials were sampled at this locale. Water pH was 7.7, water temperature 22.3°C, air temperature near the water surface 22.5°C, and photosynthetically active radiation at mid-day was 1,915 µmol photons·m^{-2} · s^{-1}, determined using a Licor LI-250A meter equipped with LI-190R quantum sensor (LI-COR Biosciences, Lincoln, NE, USA).

Three technical replicates were aseptically collected by a gloved operator into sterile plastic screw-cap conical tubes containing reagent-grade 90% ethanol for later DNA extraction for metagenomic analysis. The ethanol and other useful materials were graciously provided by Professor José Delatorre-Herrera (Universidad Arturo Prat, Iquique, Chile).

Sealing screw caps with a strip of Parafilm (Bemis, Inc. Neenah, WI, USA) proved essential to preventing ethanol evaporation during transport.

Small additional *Nostoc* pieces a few centimeters long were collected into a sterile conical tube containing bofedales water, and shortly thereafter fixed in a buffered 2% glutaraldehyde solution freshly prepared from 70% electron microscopy-grade glutaraldehyde, for later microscopic analysis. The concentrated glutaraldehyde had been shipped directly to Chile by supplier Ted Pella, Inc., Redding, CA, USA. Fresh glutaraldehyde solution was prepared in the field, just prior to use, by diluting 0.2 mL of 70% glutaraldehyde in 6.8 mL of 0.1M pH 7 phosphate buffer. For transport to the site, the 70% glutaraldehyde was stored free of air in a 5 mL syringe, which was wrapped with Parafilm and double-bagged in plastic to retard leakage or air entry. Unused glutaraldehyde was left in Chile with a local researcher (P. Arancibia-Avila, University of Bio-Bio, Chillan, Chile). Fixed materials were washed three times in phosphate buffer prior to air-transport from Chile. No living materials were transported from Chile. A formal document from USDA APHIS, declaring that these sample-transport procedures had been reviewed and were not under jurisdiction, accompanied DNA and fixed samples during import into the United States.

Imaging. Preliminary microscopic examinations of fresh materials were conducted in the field using a Zeiss Primo Star compound microscope equipped with sub-stage illumination and adaptors for international current (B&H Photo Video, New York, NY, USA). Glutaraldehyde-fixed material was later examined with a variety of light microscope optical techniques, including dark-field and differential interference contrast (DIC), and images were obtained using an Olympus BX50 microscope and Nikon D810 camera. For SEM, fixed material was dehydrated in an ethanol series, then critical point dried, and coated with iridium before examination with a Hitachi S-4800 ultra-high-resolution cold cathode field emission SEM operated at 5 kV.

DNA extraction and sequencing. Three days after collection, DNA extraction was performed in Professor Luis Sobrevia's molecular biology laboratory at the Pontificia Universidad Católica de Chile in Santiago. Samples were carefully washed to remove EtOH, soil, and any other loosely associated materials by swirling in three changes of sterile water in sterile petri dishes, before transfer to bead beater tubes. Three technical replicates were generated. To improve the proportion of epibiont DNA to host DNA, we purposely did not employ grinding. These processing methods mirrored those we had previously used to prepare other field-sampled aquatic algal or wetland plant species for metagenomic analyses, to facilitate comparative analyses (e.g., Knack et al. 2015, Graham et al. 2017).

To reduce the potential for introduction of exogenous microbes during the DNA extraction process, transfers of *Nostoc* sheets were accomplished using forceps that had been dipped into 95% EtOH and then flamed, by operators using new latex gloves. A MoBio PowerSoil kit (Qiagen, Hilden, Germany) was used for DNA extraction, and we employed a modified lysis procedure designed to reduce DNA shearing, suggested by Qiagen technical personnel: Samples were vortexed for only a few seconds prior to heating at 70°C for 5 min, a process that was then repeated. Extracted DNA was transported to the US in freezer tubes at ambient temperature, which included moderate temperatures during flights, occurring immediately after the DNA was extracted, to the northern hemisphere destination in its cold season. Tubes were kept frozen at −80°C for a few days prior to delivery to the University of Wisconsin-Madison Biotechnology Center for quantification, library production, and sequencing of the pooled technical replicates. Shotgun metagenomic sequence was obtained using the Illumina HiSeq 2500 platform (Table S1 in Supporting Information). The obtained reads were paired-end, with read length equal to 101 bp. The total number of reads was 110,578,102.

Microbiota classifications. Informatic analyses were performed in 2019, taking advantage of the incorporation into classification pipelines of relatively recent changes in bacterial classification. Because no raw reads were flagged for poor quality (Table S2 in the Supporting Information), raw reads were trimmed using Trimmomatic version 0.39 (Bolger et al. 2014), then MEGAHIT version 1.2.6, designed for handling complex metagenomics data (Li et al. 2015), was employed for assembling contigs from raw sequences.

For classifications using rRNA markers, we processed contigs using the SILVA incremental aligner (SINA SINA version 1.2.10 for ARB SVN, revision 21008; Pruesse et al. 2012) applied against the SILVA small subunit (SSU) and large subunit (LSU) rRNA SEED and quality controlled (Quast et al. 2012). Reads shorter than 50 aligned bases and those with 12% ambiguities or homopolymers were excluded before further processing, as were likely contaminants and artifacts and those having low alignment quality scores. Identical reads were identified and unique reads clustered using cd-hit-est software (ver. 3.1.2; Li and Godzik 2006) running in accurate mode, ignoring overhangs, and applying identity criteria of 1.00 and 0.98 as operational taxonomic units (OTUs) for classification performed by local nucleotide BLAST search against the non-redundant version of the SILVA SSU and LSU reference datasets (release 132), using blastn, version 2.2.30+, with standard settings (Camacho et al. 2009; Table S2). Classification of each OTU reference read was mapped onto all reads assigned to the respective OTU, performed in July, 2019. Reads for which there were no or weak (% sequence identity + % alignment coverage)/2 = <93) BLAST hits remained unclassified and were labeled "no relative" in SILVAngs fingerprint and Krona charts (Ondov et al. 2011). These methods were initially used by Klindworth et al. (2013) and Ionescu et al. (2012). The above paragraph has been paraphrased from SILVAngs project reports, as recommended. SILVAngs output employed the term "Oxyphotobacteria" as a synonym for phylum Cyanobacteria.

Host phylogenetic relationships. The *Nostoc* and *Anabaena* 23S rRNA gene sequences employed for phylogenetic analysis were obtained from GenBank, accessed on June 14, 2020. *Anabaena variabilis* CP000117 was used as an outgroup. Nucleotide sequences were aligned using MAFFT v7.402 (Katoh and Standley 2013) and the substitution model computed using ModelTest-NG v0.1.5 (Darriba et al. 2020). Maximum likelihood (ML) analysis was performed using RAxML v8.2.12 (Stamatakis 2014) on the XCEDE Portal for CIPRES (Miller et al. 2010) using GTR+I+G substitution model. Bayesian analysis was performed with MrBayes v3.2.6 (Ronquist and Huelsenbeck 2003). Four independent chains were run for 100,000 cycles, and consensus topologies were calculated after 25,000 burn-in cycles.

Marker sequence genomic coverage estimation. Genomic coverage level was estimated for taxa as a means of gaining insight into relative abundance and to develop a reporting cut-off standard. Quality control results indicating that none of the raw sequences was flagged for poor quality (Table S1) supports this approach, which utilizes trimmed raw sequence read data as input. Because sequence data were obtained by shotgun metagenomic methods, amplification bias was not an issue. Only those taxa inferred from 16S, 18S, 23S, or 28S rRNA gene sequences at an estimated average sequence depth level ≥ 90X were reported. We note that genomic coverage is likely a better indicator of relative abundance for

prokaryotic taxa than for eukaryotes, which may be multicellular or whose genomes may have duplication histories. Even so, some cyanobacteria are known to have multiple genome copies per cell (Griese et al. 2011); for such cyanobacteria, genomic coverage might likewise overestimate relative population abundance.

To estimate genomic coverage, trimmed raw reads were aligned to annotated rRNA gene SILVA references using BWA (Li and Durbin 2009) version 0.7.17-r11880.7.4 non-model species alignment. Mapped reads were filtered and converted to fastq format using SAMtools version 1.7 and the mean, maximum, and minimum coverage of each annotated rRNA gene sequence were obtained using BEDtools (Quinlan and Hall 2010) version 2.26.0. Reads were then mapped to their corresponding SILVA reference sequence using Geneious version 9.1.3 (https://www.geneious.com). Mean, maximum, and average coverage were calculated using the Geneious function "statistics." We employed coverage for the gene region of greatest abundance, indicated by use of the CyVerse Discovery Environment (https://de.cyverse.org/de/). We reported the average coverage for each classified taxon, and when average coverage data were available for more than one SILVAngs reference contig for a particular taxon, we reported the largest estimate (Table S3 in the Supporting Information).

Data accessibility. A dry piece of glutaraldehyde-fixed/buffer-washed *Nostoc* sheet has been deposited in the University of Wisconsin-Madison (WIS) herbarium under barcode number WIS-A-0000002. Because assembler pipelines evolve, and investigators may prefer particular assemblers, to facilitate use by others, we archived raw (rather than assembled) metagenomic sequence in the NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra) within accession SAMN12326833 in BioProject PRJNA555713.

RESULTS

Host and epimicrobiota imaging. The host primarily occurred as submerged brown macroscopic sheets of morphology consistent with that of genus *Nostoc* (Fig. 1A). Microscopic examination revealed the presence of small ball-shaped *Nostoc* colonies having vegetative cells of the same dimensions (~5 μm diameter) at sheet edges. Freehand sections of sheets provided evidence for presence of more abundant epibiota on one surface of sheets than the other (Fig. 1, B and C). *Nostoc* colonies consisted of curving unbranched filaments of bead-shaped cells held within a firm, layered golden-brown mucilage, with interspersed heterocytes (Fig. 1D).

Microscopic examination indicated that a tapered filamentous cyanobacterium consistent with identification as genus *Calothrix*, having one or more (when multiples present, not heteromorphic) terminal heterocytes at the broader end, and with trichomes enclosed by a firm dark-brown sheath, was common on *Nostoc* surfaces, as were narrower filaments and other types of microbial structures (Fig. 1, E and F). Among these were filaments of prokaryotic cells having more elongate shapes than the round, bead-like cells of *Nostoc*, consistent with generic identification as *Anabaena* (Fig. 1G). Microscopy also indicated presence of eukaryotic algae having distinctive

features that allowed generic determination (e.g., the chlorophyte green algal genera *Oedogonium* and *Bulbochaete* having characteristic wall rings, and streptophyte green algal genera *Zygnema* with distinctive stellate plastids and *Chaetosphaeridium globosum* with distinctive hair cells). Other distinctive structures interpreted as eukaryotic algae that were visible with microscopy included a centric diatom, pennate diatoms, a thecate dinoflagellate, a loricate protist containing particles that might be endosymbiotic algal cells, and fungal spores (Fig. S1 in the Supporting Information). The brown sheath color of the *Nostoc* host obscured some epimicrobiota, making them difficult to characterize by means of microscopic structure.

SEM indicated that *Nostoc* surface microbiota included diverse morphotypes of bacteria-sized particles (Fig. 2A) and filamentous structures of narrow widths consistent with prokaryotic taxa. SEM also revealed filaments of greater widths consistent with filamentous cyanobacteria and eukaryotic algae, and other structures having sizes and shapes consistent with eukaryotic cells or cellular aggregations (Fig. 2, B–D).

Host and epimicrobiota rRNA gene-based classifications. 11,168 Mb of sequence was obtained, and no raw sequences were flagged as poor quality (Table S1). Of the sequences passing further quality control checks, 9,784 contained ribosomal RNA genes, 4,065,551 contained predicted proteins with known functions, and 3,080,299 sequences (43.05%) contained predicted proteins of unknown function. Average contig length employed for classifications was 760 bp, and maximum length was 1,363,648 bp (Table S2).

Relative sequence abundance inferred from estimated average genomic coverage levels for cyanobacterial components (Table 1) and phylogenetic assessment (Fig. 3) indicated that *Nostoc punctiforme* PCC-73102 (=ATCC 29133) was likely the database taxon most closely related to the host species. Among epicyanobacterial genera, *Geitlerinemia*, *Oscillatoria*, *Phormidium*, and an uncultured taxon were detected only by 16S rRNA gene; *Gloeobacter* and *Pseudanabaena* were detected using 16S and 23S; and *Phormidesmis*, *Neosynechococcus*, *Symphothece*, *Aphanizomenon*, *Nodularia*, *Spirulina*, *Nodosilinea*, *Synechococcus*, *Cyanobium*, and *Anabaena*, plus two uncultured cyanobacterial taxa (JSC12, O77) were detected only by 23S rRNA gene sequences (Table 1). *Phormidium* CYN63 represented ~3% of SSU (16S rRNA gene + 18S rRNA gene) sequences, indicated by a Krona chart included in the SILVAngs report (Fig. S2 in the Supporting Information).

Epimicrobial taxa that were inferred using 18S and 23S rRNA gene markers at ≥100X average sequence coverage estimate are listed in Table 2. Three chlamydomonad genera (*Chlamydomonas*, *Rhysamphichloris*, and *Vitreochlamys*), and two
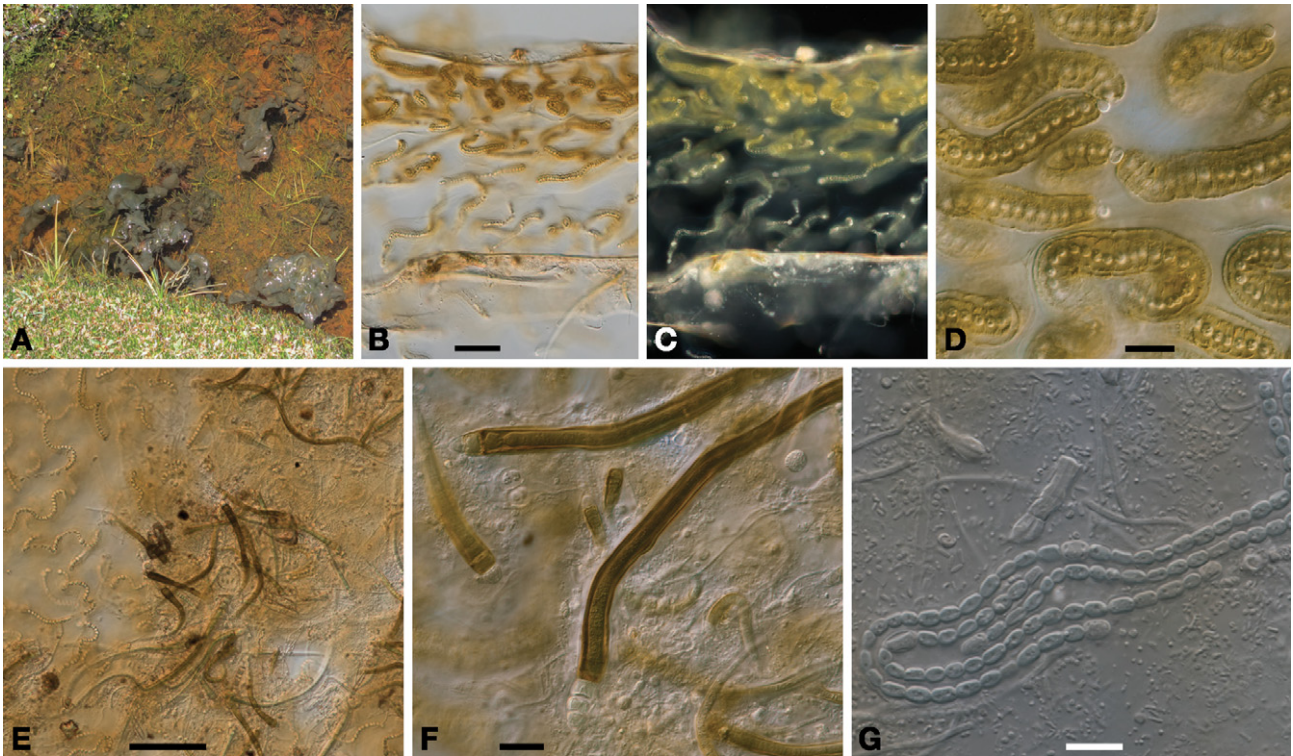
Fɪɢ. 1. *Nostoc punctiforme* of the Bofedales de Parinacota, and representative epimicrobiota. Scale bars: B, C, and E 100 µm; D and F 20 µm; G 10 µm. (A) Habit view of *Nostoc* sheets in water pools occurring among higher plant hummocks. (B) Differential interference contrast (DIC) microscopy view of a sectioned *Nostoc* sheet showing that one surface (below) was coated with more epimicrobiota than was the other (top). This orientation does not necessarily represent top-bottom orientation of *Nostoc* sheets in the natural environment. (C) Dark-field microscopy view of the same section shown in (B), showing bright-walled *Nostoc* heterocytes. (D) Higher-magnification view of representative contorted *Nostoc* filaments that, except for heterocytes, are enclosed by layered, brown-colored sheaths. (E) DIC view of *Nostoc* sheet surface, illustrating a diverse epimicrobial community, which includes an abundant cyanobacterial morphotype characterized by tapered filaments with terminal heterocytes, enclosed by a dark-brown-colored sheath. (F) Higher magnification view of the tapered, dark-brown filamentous morphotype, showing one-several terminal heterocytes. Diverse additional epimicrobial morphotypes are also illustrated. (G) Filamentous morphotype having cells more elongate than those of *Nostoc*, surrounded by narrower filamentous and unicellular morphotypes.

heterotrophic chrysophyte stramenopiles, *Spumella,* and *Pedospumella,* were inferred from 18S rRNA gene sequences. 18S rRNA gene sequences also indicated presence of the ciliate *Phascolodon*, and the rhizarian *Heteromita* was detected by both 18S and 28S rRNA genes. The streptophyte green alga *Chaetosphaeridium globosum* was inferred from 28S rRNA gene (and, as noted, was observed using microscopy), but was not inferred from 18S rRNA gene sequence data (Table 2). Epibiontic Oedogoniales (*Bulbochaete* and *Oedogonium*) that seemed common in microscopic view (Fig. S1) were not detected from sequence data.

A total of 66 non-oxygenic bacterial genera having ≥90X average sequence coverage were inferred to occur in the Parinacota *Nostoc* epimicrobiota. Of these, 38 were inferred with the use of 16S rRNA genes; 19 of these were inferred only from 16S rRNA gene sequences and 19 were inferred from both 16S and 23S rRNA gene sequences. 47 bacterial genera were inferred only from 23S rRNA genes at the same detection limit (Table S3). Additional uncultured bacterial taxa (including cyanobacterial representatives) were detected using the ≥90X average sequence coverage criterion (Table S4 in the Supporting Information). Figure 4 indicates that >60% of epimicrobial taxa inferred to occur on the Parinacota *Nostoc* (at estimated average genomic coverage levels ≥90X) were detected only by 18S, 23S, or 28S markers. Many rRNA gene contigs could not be classified based on the content of current taxonomic databases (see e.g., a Krona chart from the SILVAngs report shown as Fig. S2).

DISCUSSION

In overview, results of the current study of Parinacota *Nostoc* epimicrobiota supported the hypothesis that use of microscopy and 16S, 18S, 23S, and 28S ribosomal marker genes derived from metagenomic sequence data can reveal greater diversity than might be apparent from the use of 16S rRNA gene sequences alone. Employing a minimal abundance criterion of ≥90X average genomic coverage, more
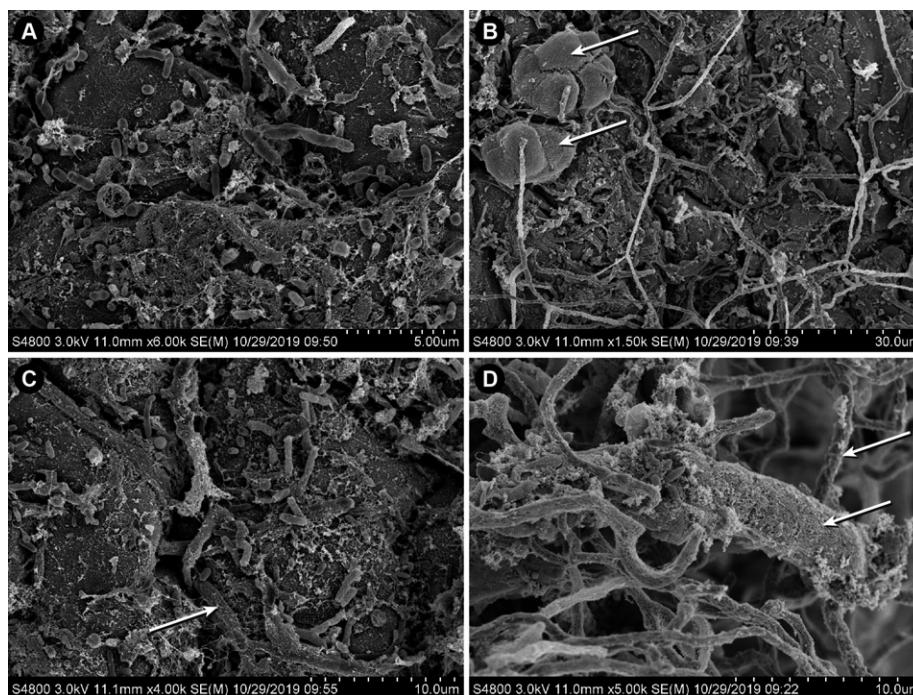
FIG. 2. Scanning electron microscopic views of the surface of Parinacota *Nostoc*, showing diverse microbial morphotypes. (A) High-magnification view showing small-diameter cells of sizes consistent with bacteria. (B) Lower-magnification view showing putative eukaryotic cell clusters (arrows) and numerous smaller-diameter threads. (C), (D) Higher magnification views of filamentous microbial epibiota (arrows).

than half of reported taxa in our study were observed only from 18S, 23S, or 28S rRNA gene sequences. With respect to algal components of the microbiota, a study focus, we inferred the presence of 10 more epicyanobacterial genera from 23S rRNA gene sequences than from 16S rRNA gene sequences employed alone, and the use of 18S and 28S rRNA gene sequences revealed six genera of eukaryotic algal epimicrobiota, most of which were not obvious from microscopic examination. Likewise, 47 more anoxygenic epibacterial genera were detected (at ≥90X average genomic coverage estimate) by 23S rRNA gene sequences than with the use of 16S rRNA gene sequences. We hypothesize that some of our sequences matched database 16S (or 18S) rRNA gene sequences more closely and others matched database 23S (or 28S) rRNA gene sequences better, possibly explaining why not all taxa were indicated by both markers. Light and scanning microscopy also provided evidence for diverse morphotypes of bacteria and eukaryotes in the Parinacota *Nostoc* epimicrobiota and in some cases corroborated molecular detections. Our study also yielded new information about diversity of *Nostoc* species in the northern Chilean Altiplano and indicated epibacterial genera that might play key ecological roles, including toxin production.

Comparisons of Parinacota Nostoc *and its epimicrobiota with those of other Chilean* Nostoc *studied with high-throughput sequencing methods.* Consistent with microscopic observations, the relative abundance of

16S and 23S rRNA gene sequences that we obtained from shotgun metagenomic sequencing of presumptive Parinacota *Nostoc,* and phylogenetic analysis of ribosomal sequence data, indicated that the sheet-forming host was *Nostoc,* likely a form closely related to *N. punctiforme* (PCC 73102 = ATCC 29133). By contrast, using partial 16S rRNA gene amplicons, Aguilar et al. (2019) tentatively assigned spherical colonies collected from northern Chilean Altiplano Chungará Lake to *N. commune,* and identified smaller structures collected from nearby Culco stream as *N. flagelliforme,* neither of which was closely related to a *Nostoc* sp. sampled from a northern Chilean food market (Galetovic et al. 2017). Together, our findings and other studies indicate that aquatic environments of the northern Chilean Altiplano harbor several aquatic *Nostoc* genotypes that host microbiota.

Our previous metagenomic study of the microbiome of a ball-forming *Nostoc* sampled from an alkaline Patagonian lake had revealed that all but one of the relatively long (Roche 454) 16S rRNA gene reads classified as cyanobacteria clustered with *N. commune* (Graham et al. 2014). As noted, *Nostoc* balls from northern Chilean Altiplano Chungará Lake were also tentatively identified as *N. commune* based on 16S rRNA gene amplicons (Aguilar et al. 2019). However, the Patagonian *Nostoc* was associated with the formation of distinctive clotted microbialites likely arising from carbonate encrustation of the small ball-shaped cyanobacterial colonies,

TABLE 1. Cyanobacteria inferred for the Parinacota *Nostoc* epimicrobiota from 16S and/or 23S rRNA gene having average genomic coverage estimates ≥ 90X.

	Genus/Strain	Average 16S rRNA gene genomic coverage	Average 23S rRNA gene genomic coverage
Oxyphotobacteria (=phylum Cyanobacteria)	*Gloeobacter* PCC-7421	311X	158X
	Geitlerinema LD9	361X	
	Phormidesmis ANT.L52.6		171X
	Neosynechococcus		155X
	Symphothece PCC-700		129X
	Anabaena PCC-7108		136X
	Aphanizomenon NIES81		106X
	Nodularia PCC-9350		112X
	Spirulina PCC-6313		180X
	Nodosilinea PCC-7104		151X
	Pseudanabaena PCC-7429 or PCC-6802	303X	273X
	Synechococcus PCC-7502		134X
	Cyanobium PCC-6307		141X
	Oscillatoria SAG 1459-8	313X	
	Phormidium CYN64	419X	
	Phormidium SAG 37.90	325X	

consistent with inferred presence in epimicrobiota of sulfate-reducing bacteria (*Desulfomicrobium* and *Sulfospirillum*) and genes associated with sulfate reduction, a process hypothesized to foster carbonate deposition by raising local pH (Graham et al. 2014). By contrast, carbonate encrustation was not reported for spherical colonies of Chungará Lake *Nostoc* (Aguilar et al. 2019) or observed for Parinacota *Nostoc* in this study, both sampled from alkaline waters.

Two of the bacterial genera inferred for the Parinacota *Nostoc* epimicrobiota by 16S or 23S rRNA gene sequences (employing a ≥90X average genomic coverage as a detection criterion), namely *Hydrogenophaga* and *Flavobacterium*, had also been reported for both the Patagonian *Nostoc* studied by 16S rRNA gene sequences filtered from metagenomic data (Graham et al. 2014) and northern Chilean aquatic *Nostoc* epibacteria inferred from 16S rRNA gene amplicons (Aguilar et al. 2019). Some of the other bacterial genera we had inferred from 16S rRNA gene sequences to associate with Patagonian *N. commune* (Graham et al. 2014) were also detected using 16S or 23S rRNA gene sequences in the current study of Parinacota *Nostoc* epimicrobiota,

namely *Roseomonas, Devosia, Rhodobacter, Blastomonas,* and *Methylibium.* Although 16S rRNA gene sequence evidence filtered from metagenomic data for a microbiome described for *Nostoc commune* from Patagonia had indicated some evidence for presence of Archaea (Graham et al. 2014), Archaea were not detected in the present study of Parinacota *Nostoc.*

The Aguilar et al. (2019) 16S rRNA gene amplicon study of northern Chilean aquatic *Nostoc* was based on sampling performed in 2016, as was the current metagenomic study of Parinacota *Nostoc,* aiding microbiota comparison. For such a comparison, we did not include bacterial taxa that Aguilar et al. (2019) had inferred only from water samples. We considered only bacterial taxa reported from their samples labeled Chungará (lake *Nostoc*) or Culco (stream *Nostoc*) in their figure 3. We considered all bacterial taxa these authors had inferred to be closely associated with *Nostoc,* whether localized to the inner gelatinous matrix or on the outer surface of the colony, even though our microscopic observations, like those of New Zealand wetland *Nostoc* colonies (Secker et al. 2016), did not reveal many bacteria in the gelatinous matrix. Our 16S rRNA gene sequences for Parinacota *Nostoc* indicated several bacterial genera in common with those inferred for northern Chilean Altiplano lake or stream *Nostoc* by Aguilar et al. (2019): *Tabrizicola, Pseudorhodobacter, Hydrogenophaga, Gemmatimonas, Flavobacterium, Polymorphobacter,* and *Pirellula.* Our 23S rRNA gene sequences for Parinacota *Nostoc* indicated additional bacterial genera in common with *Nostoc* epibacteria reported by Aguilar et al. (2019), namely *Terrimicrobium, Porphyrobacter,* and *Novosphingobium* that we had not also inferred from 16S rRNA gene sequences. Pending further study, these bacterial genera might represent components of a core microbiota more generally associated with aquatic *Nostoc.* Some of the listed taxa (*Flavobacterium, Hydrogenophaga*) had previously been hypothesized to represent core microbiota for *Nostoc* sampled from aquatic or terrestrial habitats (Graham et al. 2018).

Making quantitative comparisons to epimicrobiota inferred from other studies of Chilean Altiplano *Nostoc* sampled in the same year (2016) would be complicated by the use of differing methods and numbers and types of samples. However, we note that some of the bacterial genera inferred for the Parinacota *Nostoc* at ≥1% in the Krona chart shown in Figure S2, namely *Hydrogenobacter, Tabrizicola, Rhizorhapis,* and *Polymorphobacter,* were also reported to occur at the >1% level by Aguilar et al. (2019), suggesting that these four bacterial genera may play some important role in the ecology of aquatic *Nostoc* in the northern Chilean Altiplano. Future studies might indicate these roles.

Cyanobacterial epibionts of ecological interest. Parinacota *Nostoc* cyanobacterial epibionts of particular interest included an abundant morphotype resembling the widespread genus *Calothrix* in having
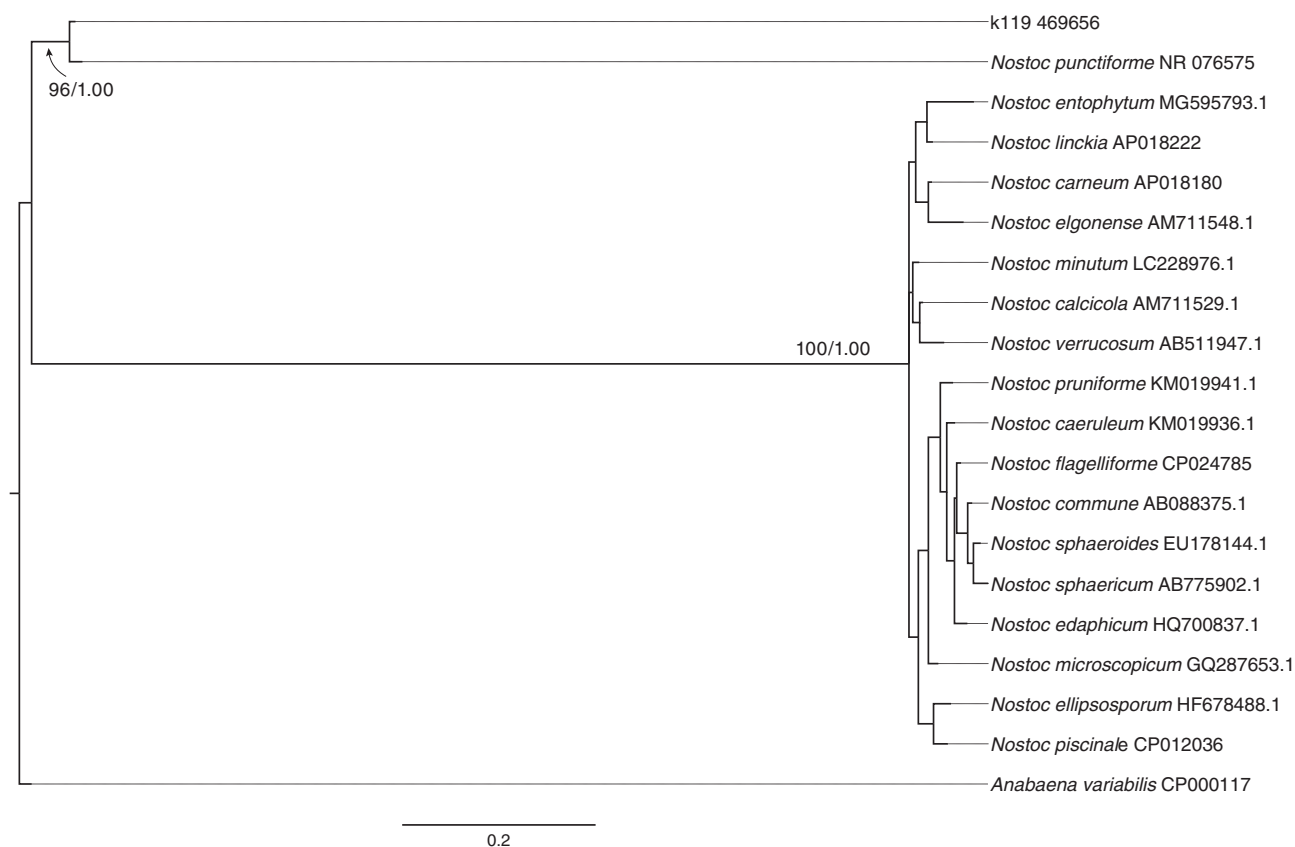
FIG. 3. Phylogenetic relationships of the Parinacota *Nostoc* host. At nodes indicating relationship of Parinacota *Nostoc* to *N. punctiforme*, maximum likelihood bootstrap values are shown first, followed by Bayesian values. Values for other nodes (relationships among other database *Nostoc* accessions) are not shown because relationship patterns were not congruent between ML and Bayesian analyses.

tapered filaments with one or more terminal heterocytes, demonstrated by microscopy. In trichome dimensions and dark sheath color, the Parinacota *Nostoc* epibiont resembled *C. fusca* commonly present in neutral-alkaline U.S. rivers and streams (Rinkel and Manoylov 2014). The *Calothrix*-like epibiont observed on Parinacota *Nostoc* sometimes displayed more than one terminal heterocyte, which has previously been reported for some *Calothrix* species; however, when more than one heterocyte was present, these were not heteromorphic, as is the case for particular *Calothrix*-related taxa recently segregated as *Macrochaete* (Gómez et al. 2016). 23S rRNA gene sequences (present at 112X average genomic coverage) that classified as *Nodularia* PCC-9350 in our SILVAngs analysis were also listed as possible *Calothrix* sp. PCC 7507. This finding might reflect views that *Calothrix* may be a polyphyletic genus (Will et al. 2018).

A *Cyanobium* relative, inferred for the Parinacota *Nostoc* epimicrobiota from 23S rRNA gene sequence data, has been used as a model system to assess proteomic changes associated with marine-freshwater transition (Cabello-Yeves and Rodriguez-Valera, 2019). Inferred presence of *Cyanobium* may reflect the relatively high salinity conditions of the river

system that feeds the Parinacota bofedales (Copaja and Muñoz 2018).

Some of the cyanobacterial taxa we inferred from metagenomic sequence are commonly associated with toxin production and thus possibly relevant to animal food safety. The Parinacota *Nostoc* was classified as a relative of *N. punctiforme* PCC 73102, which is considered to be a prolific producer of secondary metabolites, including the depsipeptides nostopeptolide (Liaimer et al. 2015) and the cyclic peptides nostamide A and anabaenopeptin, plus unusual forms of the tricyclic peptide microviridin indicated by expression of gene clusters assessed using RNA-seq and fluorescence reporters (Dehm et al. 2019). *Nostoc punctiforme* PCC 73102 is known to produce sesquiterpene synthases (Agger et al. 2008), a feature that may aid defense. *Nostoc punctiforme* PCC 73102 is also considered to possess the genetic capacity to produce the bad-tasting compound geosmin, as do some other cyanobacteria and non-oxygenic myxobacteria and streptomycetes (Wang et al. 2019).

Cytotoxic secondary metabolites such as microcystin and nodularin, in addition to aeruginosins, nosperin, nostocyclopeptides, and many other peptides not as yet identified (Liaimer et al. 2016) are

TABLE 2. Eukaryotic taxa inferred for the Parinacota *Nostoc* epimicrobiota from 18S and/or 28S rRNA gene sequences having average genomic coverage estimates ≥ 90X.

	Genus/species	Average 18S rRNA gene genomic coverage	Average 28S rRNA gene genomic coverage
Embryophyta liverworts	*Nardia compressa?*	353X	235X
Streptophyte Green algae	*Chaetosphaeridium globosum*		280X
Chlorophyte Green algae	Unclassified chlorophyte	432X	
	Unclassified chlorophyte	619X	
	Chlamydomonas	442X	
	Rhizamphichloris	342X	
	Vitreochlamys	469X	
Alveolate ciliates	*Phascolodon*	327X	
	Unclassified ciliate		137X
	Unclassified ciliate		154X
Rhizarian cercozoans	*Heteromita*	415X	233X
Stramenopile chrysophytes	*Spumella*	413X	145X
	Pedospumella	408X	
	Unclassified chrysophyte		157X
	Unclassified chrysophyte		176X
	Unclassified chrysophyte		140X
	Unclassified chrysophyte		202X
	Unclassified chrysophyte		203X
	Unclassified chrysophyte		101X
Fungi, Helotiales		139X	

FIG. 4. Chart showing relative numbers of Parinacota *Nostoc* epimicrobial taxa having average genomic coverage estimates ≥ 90X that were inferred using different rRNA gene sequence markers, labeled rDNA for simplicity. Numbers refer to all classified taxa, including bacterial genera (Table S3) plus uncultured bacterial taxa currently classifiable only at higher taxonomic levels (Table S4). More than 60% of the epimicrobial taxa were detected by ribosomal markers other than 16S rRNA gene sequences.

commonly produced by well-studied cases of terrestrial *Nostoc* occurring in symbioses, illustrated by the glomalean *Geosiphon pyriforme* (Kluge 2015), cyanolichens (Graham et al. 2018b), and land plants. Plant symbioses with *Nostoc* or related cyanobacteria that have been recently studied with the use of high-throughput sequencing methods include the angiosperm *Gunnera* (Fernández-Martinez et al. 2013), cycads such as *Dioon* (Suárez-Moo et al. 2019), water-ferns that contribute to rice paddy fertility (Kumar et al. 2019), and bryophytes such as the liverwort *Blasia* (Liaimer et al. 2016). By contrast, assessments of the potential for aquatic *Nostoc* to produce toxins have generally focused on testing market samples of wild material harvested for food use, and then, only for a few of the hundreds of types of toxins that cyanobacteria are known to produce (Du et al. 2019, Janssen 2019).

In the northern Chilean altiplano, and some other areas of the world, *Nostoc* has traditionally been harvested for sale in local markets for use as a foo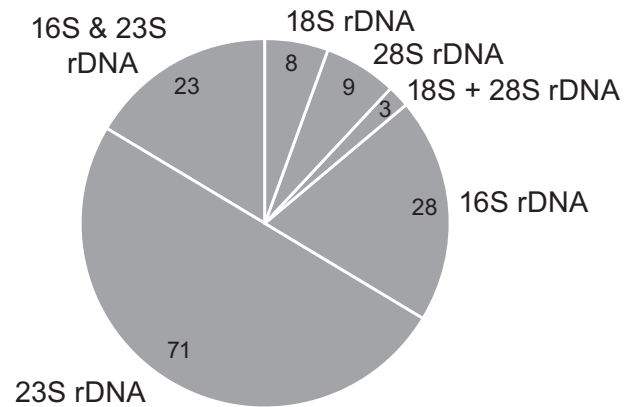d, with some experts recommending revitalization of this practice (Rivera et al. 2018). However, the documented capacity of genus *Nostoc* to produce diverse secondary compounds has generated concerns regarding food safety and stimulated chemical testing for potential toxins. For example, samples of field-collected *Nostoc commune* purchased in Peruvian markets tested using multiple methods commonly contained the neurotoxic amino acid BMAA (beta-N-methylamino-L-alanine; Johnson et al. 2008). Other investigators, assessing the nutritional features of *Nostoc* (locally known as llayta) purchased in markets in northern Chile or southern Peru, also tested market samples for presence of a gene in the microcystin-biosynthesis pathway to assess degree of safety for human consumption (Galetovic et al. 2017); although these authors did not find genetic marker evidence for microcystin production, they noted that the samples were not tested for presence of other potentially harmful compounds, or genetic capacity to synthesize other toxins. Modern chemical and molecular studies have revealed that as a group, cyanobacteria can produce hundreds, perhaps thousands, of types of secondary metabolites, many of known animal health concern or of unknown health impact, to the extent that chemical testing for all potentially harmful toxins is challenging (Du et al. 2019, Janssen 2019).

Some (notably *Anabaena, Aphanizomenon, Nodularia/Calothrix*) of the 15 genera of epicyanobacteria we inferred to occur in the microbiota of Parinacota *Nostoc* have been linked to toxin production, indicating that even if the *Nostoc* host did not itself produce harmful cyanotoxins, the epimicrobiota might. For example, *Calothrix* is reported to possess toxin-

related biosynthesis gene clusters (Khumalo et al. 2020). Modern high-throughput molecular methods could be used to infer degree of risk by indicating presence in microbiota of cyanobacterial taxa known to commonly produce harmful compounds. Metagenomic data also offer the potential for finding key toxin biosynthesis pathway genes. Cataloguing such genes and pathways is beyond the scope of the current study focused on epimicrobiota diversity. In view of the vast numbers of potential cyanobacterial toxins and pathways genes known to exist (Du et al. 2019, Janssen 2019), such investigations would likely need to focus on particular toxin groups or key pathway genes.

In view of its diverse microbiota, potential for the *Nostoc* holobiont to produce toxins, and modeling indications that freshwater *Nostoc* plays an important role in global N-fixation now and over deep time (Graham et al. 2018a), to foster conservation goals articulated by Rundel and Palma (2000), perhaps Altiplano *Nostoc* should not be extensively harvested, but generally left in place to generate fixed N that may foster the bofedales vegetation upon which valued, high-diversity wildlife communities depend.

Comparisons of Parinacota Nostoc *eukaryotic epimicrobiota with those of other freshwater* Nostoc *studied with high-throughput sequencing methods.* Previous studies of freshwater *Nostoc* epimicrobiota were accomplished by means of 16S rRNA gene amplicons (Secker et al. 2016, Aguilar et al. 2019) or 16S rRNA gene sequences filtered from shotgun metagenomic sequence (Graham et al. 2014) and therefore did not report 18S or 28S rRNA gene sequence related to eukaryotes. However, wetland *Nostoc*-associated 16S rRNA gene sequences similar to plastid sequences of green algae such as the chlorophyte *Nannochloropsis* and the streptophyte *Coleochaete* and some diatoms were reported, in addition to microscopic detection of additional eukaryotic autotrophs and heterotrophs (Secker et al. 2016). Diatoms and a few other types of eukaryotic associates were observed with the use of light and scanning electron microscopy on surfaces of *Nostoc* commune sampled from a Patagonian lake (Graham et al. 2014). By contrast, in the present study of Parinacota *Nostoc*, the epimicrobiota was inferred by ≥90X average sequence coverage of 18S or 28S rRNA gene sequences to include a liverwort (putatively *Nardia compressa*, possibly in the form of spores) and eight protist genera, including six eukaryotic algal genera. Additional algal taxa were indicated by sequences not currently classifiable to the generic level and the use of light microscopy.

Our finding that *Chaetosphaeridium* was associated with Parinacota *Nostoc*, indicated by both molecular sequence evidence and microscopy, together with molecular evidence for *Coleochaete* in the epimicrobiota of another (New Zealand) wetland *Nostoc* (Secker et al. 2016), suggests that Coleochaetales might occur more widely on surfaces of *Nostoc*.

Among the chlorophyte epibionts detected by rRNA gene sequences, the chlorophyte green alga *Vitreochlamys* sp. SAG 12.93 (MB-B) is known to undergo binary cell division that leaves two daughter cells connected by a protoplasmic link, and so illustrates possible evolutionary transition to coloniality in the volvocalean lineage Tetrabaenaceae–Goniaceae–Volvocaceae (Nakada et al. 2019). Among the chrysophyte eukaryotic components of the Parinacota *Nostoc* epimicrobiota, heterotrophic *Spumella* is considered to be an important bactivore in aquatic systems (Grossmann et al. 2016). Additional tough-walled eukaryotic algal genera identified by distinctive microscopic features, but not by molecular sequencing (e.g., Oedogoniales) may have evaded DNA extraction, which had been conducted without grinding, to foster extraction of longer, less-sheared DNA pieces. These observations indicate that the use of multiple ribosomal taxonomic markers and microscopy together provide a broader view of *Nostoc* epimicrobiota, particularly the algal components, than does the use of 16S rRNA gene sequences alone.

CONFLICT OF INTEREST

The authors declare that they do not have a conflict of interest.

Agger, S. A., Lopez-Gallego, F., Hoye, T. R. & Schmidt-Dannert, C. 2008. Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120. *J. Bacteriol.* 190:6084–96.

Aguilar, P., Dorador, C., Vila, I. & Sommaruga, R. 2019. Bacterial communities associated with spherical *Nostoc* macrocolonies. *Front. Microbiol.* 10:483.

Albarracín, V. H., Kurth, D., Ordoñez, O. R., Belfiore, C., Luccini, E., Salum, G. M., Placentini, R. D. & Farías, M. E. 2015. High-up: A remove reservoir of microbial extremophiles in Central Andean wetlands. *Front. Microbiol.* 6:1404.

Arima, H., Horiguchi, N., Takaichi, S., Kofuji, R., Ishida, K.-I., Wada, K. & Sakamoto, T. 2012. Molecular genetic and chemotaxonomic characterization of the terrestrial cyanobacterium *Nostoc commune* and its neighboring species. *FEMS Microbiol. Ecol.* 79:34–45.

Bolger, A. M., Lohse, M. & Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–20.

Cabello-Yeves, P. J. & Rodriguez-Valera, F. 2019. Marine-freshwater prokaryotic transitions required extensive changes in the predicted proteome. *Microbiome* 7:117.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

Copaja, S. V. & Muñoz, F. J. 2018. Heavy metals concentration in sediment of Lluta River Basin. *J. Chil. Chem. Soc.* 63:3878–83.

Curren, E., Yoshida, T., Kuwahara, V. S. & Leong, S. C. Y. 2019. Rapid profiling of tropical marine cyanobacterial communities. *Reg. Stud. Mar. Sci.* 25:100485.

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B. & Flouri, T. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37:291–4.

Dehm, D., Krumbholz, J., Baunach, M., Wiebach, V., Hinrichs, K., Guljamow, A., Tabuchi, T., Jenke-Kodama, H., Süssmuth, R. D. & Dittmann, E. K. 2019. Unlocking the spatial control of secondary metabolism uncovers hidden natural product diversity in *Nostoc punctiforme*. *ACS Chem. Biol.* 14:1271–9.

Dodds, W. K., Gudder, D. A. & Mollenhauer, D. 1995. The ecology of *Nostoc*. *J. Phycol.* 31:2–18.

Du, X., Liu, H., Yuan, L., Wang, Y., Ma, Ya, Wang, Rui, Chen, X., Losiewicz, M. D., Guo, H. & Zhang, H. 2019. The diversity of cyanobacterial toxins on structural characterization, distribution and identification: a systematic review. *Toxins* 11:530.

Fernández-Martinez, M. A., de los Ríos, A., Sancho, L. G. & Pérez-Ortega, S. 2013. Diversity of endosymbiotic *Nostoc* in *Gunnera magellanica* (L) from Tierra del Fuego. *Chile. Microb. Ecol.* 66:335–50.

Galetovic, A., Araya, J. & Gómez-Silva, B. 2017. Biochemical composition and toxicity of edible colonies of the cyanobacterium *Nostoc* sp. *Llayta*. *Rev. Chil. Nutr.* 44:360–70.

Garcia-Pichel, F., Lombard, J., Soule, T., Dunaj, S., Wu, S. H. & Wojciechowski, M. F. 2019. Timing the evolutionary advent of cyanobacteria and the later great oxidation event using gene phylogenies of a sunscreen. *mBio* 10:e00561-19.

Gómez, E. B., Johansen, J. R., Kastovsky, J., Bohunická, M. & Capková, K. 2016. *Macrochaete* gen nov. (Nostocales, Cyanobacteria), a taxon morphologically and molecularly distinct from *Calothrix*. *J. Phycol.* 52:638–55.

Graham, L. E., Graham, J. M., Knack, J. J., Trest, M. T., Piotrowski, M. J. & Arancibia-Avila, P. 2017. A sub-Antarctic peat moss metagenome indicates microbiome resilience to stress and the biogeochemical functions of early Paleozoic terrestrial ecosystems. *Int. J. Plant Sci.* 178:618–28.

Graham, L. E., Graham, J. M., Wilcox, L. W. & Cook, M. E. 2016. *Algae*. LJLM Press, Madison, WI, 720 pp.

Graham, L. E., Graham, J. M., Wilcox, L. W., Cook, M. E., Arancibia-Avila, P. & Knack, J. J. 2018a. Evolutionary roots of plant microbiomes and biogeochemical impacts of nonvascular autotroph-microbiome systems over deep time. *Int. J. Plant Sci.* 179:505–22.

Graham, L. E., Knack, J. J., Piotrowski, M. J., Wilcox, L. W., Cook, M. E., Wellman, C. H., Taylor, W., Lewis, L. & Arancibia-Avila, P. 2014. Lacustrine *Nostoc* (Nostocales) and associated microbiome generate a new type of modern clotted microbialite. *J. Phycol.* 50:280–91.

Graham, L. E., Trest, M. T., Will-Wolf, S., Miicke, N. S., Atonio, L. M., Piotrowski, M. J. & Knack, J. J. 2018b. Microscopic and metagenomic analyses of *Peltigera ponojensis* (Peltigerales, Ascomycota). *Int. J. Plant Sci.* 179:241–55.

Graham, L. E., Wilcox, L. W. & Knack, J. J. 2015. Why we need more algal metagenomes. *J. Phycol.* 51:1029–36.

Griese, M., Lange, C. & Soppa, J. 2011. Ploidy in Cyanobacteria. *FEMS Microbiol. Lett.* 323:124–31.

Grossmann, L., Bock, C., Schweikert, M. & Boenigk, J. 2016. Small but manifold–hidden diversity in "*Spumella*-like flagellates". *J. Euk. Micro.* 63:419–39.

Inoue-Sakamoto, K., Tanji, Y., Yamaba, M., Natsume, T., Masaura, T., Asano, T., Nishiuchi, T. & Sakamoto, T. 2018. Characterization of extracellular matrix components from the desiccation-tolerant cyanobacterium *Nostoc commune*. *J. Gen. Appl. Microbiol.* 64:15–25.

Ionescu, D., Siebert, C., Polerecky, L., Munwes, Y. Y., Lott, C., Hausler, S., Bozoc-Ionescu, M. et al. 2012. Microbial and chemical characterization of underwater fresh water springs in the Dead Sea. *PLoS ONE* 7:e38319.

Janssen, E. M. L. 2019. Cyanobacterial peptides beyond microcystins – A review on co-occurrence, toxicity, and challenges for risk assessment. *Water Res.* 151:488–99.

Johnson, H. E., King, S. R., Banack, S. A., Webster, C., Callanaupa, W. J. & Cox, P. A. 2008. Cyanobacteria (*Nostoc commune*) used as a dietary item in the Peruvian highlands produce the neurotoxic amino acid BMAA. *J. Ethnopharmacol.* 118:159–65.

Jung, P., Schemer, M., Briegel-Williams, L., Baumann, K., Leinweber, P., Karsten, U., Lehnert, L., Achilles, S., Bendix, J. & Büdel, B. 2019. Water availability shaped edaphic and lithic cyanobacterial communities in the Atacama Desert. *J. Phycol.* 55:1306–18.

Katoh, K. & Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–80.

Khumalo, M. J., Nzuza, N., Padayachee, T., Chen, W., Yu, J. H., Nelson, D. R. & Syed, K. 2020. Comprehensive analyses of cytochrome P450 monooxygenases and secondary metabolite biosynthetic gene clusters in Cyanobacteria. *Int. J. Mol. Sci.* 21:656.

Klicki, K., Ferreira, D., Hamill, D., Dirks, B., Mitchell, N. & Garcia-Pichel, F. 2018. The widely-conserved ebo cluster is involved in precursor transport to the periplasm during scytonemin synthesis by *Nostoc punctiforme*. *mBio* 9:e02266-18.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. & Glöckner, F. O. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1.

Kluge, M. 2015. Fascinated by a curiosity: Dieter Mollenhauer and *Geosiphon pyriformis*. *Algol. Stud.* 148:5–14.

Knack, J. J., Wilcox, L. W., Delaux, P. -M., Ané, J. M., Piotrowski, M. J., Cook, M. E., Graham, J. M. & Graham, L. E. 2015. Microbiomes of streptophyte algae and bryophytes suggest that a functional suite of microbiota fostered plant colonization of land. *Int. J. Plant Sci.* 176:405–20.

Kumar, U., Nayak, A. K., Panneerselvam, A. K., Mohanty, S., Shahid, M., Sahoo, A., Kaviraj, M. et al. 2019. Cyanobiont diversity in six *Azolla* spp. and relation to *Azolla*-nutrient profiling. *Planta* 249:1435–47.

Li, D., Liu, C. M., Sadakane, K. & Lam, T. W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 15:1674–6.

Li, W. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–60.

Li, W. & Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–9.

Liaimer, A., Helfrich, E. J. N., Hinrichs, K., Gulijamow, A., Ishida, K., Hertweck, C. & Dittmann, E. 2015. Nostopeptolide plays a governing role during cell differentiation of the symbiotic cyanobacterium *Nostoc punctiforme*. *Proc. Natl. Acad. Sci. USA* 112:1862–7.

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

Miller, M. A., Pfeiffer, W. & Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *In 2010 gateway computing environments workshop (GCE)*. IEEE, pp. 1–8.

Nakada, T., Tsuchida, Y. & Tomita, M. 2019. Improved taxon sampling and multigene phylogeny of unicellular chlamydomonads closely related to the colonial volvocalean lineage Tetrabaenaceae-Goniaceae-Volvocaceae (Volvocales, Chlorophyceae). *Mol. Phylo. Evol.* 130:1–8.

Ondov, B., Bergman, N. & Phillippy, A. 2011. Interactive metagenomic visualization in a web browser. *BMC Bioinformatics* 12:385.

Orellana, R., Macaya, C., Bravo, G., Dorochesl, F., Cumsille, A., Valencia, R., Rojas, C. & Seeger, M. 2018. Living at the frontiers of life: Extremophiles in Chile and their potential for bioremediation. *Front. Microbiol.* 9:2309.

Pruesse, E., Peplies, J. & Glöckner, F. O. 2012. SINA: accurate high throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–9.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41:D590–6.

Quinlan, A. R. & Hall, I. M. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–2.

Rinkel, B. & Manoylov, K. 2014. *Calothrix*–An evaluation of fresh water species in United States rivers and streams, their distribution and preliminary ecological findings. *Proc. Acad. Nat. Sci. Phila.* 163:43–59.

Rivera, M., Galetovic, A., Licuime, R. & Gómez-Silva, B. 2018. A microethnographic and ethnobotanical approach to llayta consumption among Andes feeding practices. *Foods* 7:202.

Ronquist, F. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–4.

Rundel, P. W. & Palma, B. 2000. Preserving the unique puna ecosystems of the Andean Altiplano. A descriptive account of Lauca National Park, Chile. *Mountain Res. Devel.* 20:262–71.

Secker, N. H., Chua, J. P. S., Laurie, R. E., McNoe, L., Guy, P. L., Orlovich, D. A. & Summerfield, T. C. 2016. Characterization of the cyanobacteria and associated bacterial community from an ephemeral wetland in New Zealand. *J. Phycol.* 52:761–73.

Soo, R. M., Hemp, J. & Hugenholtz, P. 2019. Evolution of photosynthesis and aerobic respiration in the cyanobacteria. *Free Radical Biol. Med.* 140:200–5.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–3.

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

Wang, Z., Song, G., Li, Y., Yu, G., Hou, X., Gan, Z. & Li, R. 2019. The diversity, origin, and evolutionary analysis of geomin synthase gene in cyanobacteria. *Sci. Total Environ.* 689:789–96.

Will, S. E., Henke, P., Boedeker, C., Huang, S., Brinkmann, H., Rohde, M., Jarek, M. et al. 2018. Day and night: Metabolic profiles and evolutionary relationships of six axenic non-marine cyanobacteria. *Genome Biol. Evol.* 11:270–94.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Figure S1. Eukaryotic members of the Parinacota *Nostoc* epimicrobiota detected by light microscopy. (A), (B) A loricate ciliate that may include algal endosymbionts. Scale bar = 20 μm. (C), (D) Darkly-pigmented *Alternaria*-like muriform fungal spores with associated hyphae. Scale bar = 20 μm. (E) Theca of an armored dinoflagellate. Scale bar = 20 μm. (F) Frustules of a centric *Melosira*-like diatom. Scale bar = 10 μm. (G) Frustule of pinnate diatom. Scale bar = 10 μm. (H) Frustule of pinnate diatom. Scale bar = 10 μm. (I) Frustule of pinnate diatom. Scale bar = 5 μm. (J) *Oedogonium*, an unbranched oedogonialean chlorophyte alga. Scale bar = 50 μm. (K) *Oedogonium* septum showing wall rings characteristic of Oedogoniales. Scale bar = 10 μm. (L) *Bulbochaete*, an oedogonialean chlorophyte alga. Scale bar = 20 μm. (M) Putative chlorophycean green alga. Scale bar = 10 μm. (N) Coccoid green alga. Scale bar = 5 μm. (O) Filamentous alga. Scale bar = 10 μm. (P) *Chaetosphaeridium globosum*, a streptophyte coleochaetalean green alga. Scale bar = 20 μm. (Q) *Zygnema*, a streptophyte zygnematalean green alga. Scale bar = 20 μm. (R) Cell wall remains resembling desmidialean green algal semi-cell. Scale bar = 20 μm. (S) Cell wall remains resembling desmidialean green algal semi-cell. Scale bar = 10 μm. (T) Cell wall remains resembling desmidialean green alga. Scale bar = 10 μm Taxonomic affinities of phototrophic eukaryotes were ascertained using structural characters described and illustrated in Graham et al. (2016).

Figure S2. SILVAngs-generated Krona chart indicating percentages of microbial taxa in the Parinacota *Nostoc* epimicrobiota indicated by the ribosomal small-subunit markers 16S rRNA and 18S rRNA gene sequences.

Table S1. Sequencing information.

Table S2. SILVAngs report details July, 2019.

Table S3. Prokaryotic genera inferred for the Parinacota *Nostoc* epimicrobiota from 16S and/or 23S rRNA gene sequences having average genomic coverage estimates ≥ 90X.

Table S4. Uncultured bacterial taxa detected in the Parinacota *Nostoc* epimicrobiota.