Cycad Coralloid Roots Contain Bacterial Communities Including Cyanobacteria and *Caulobacter* spp. That Encode Niche-Specific Biosynthetic Gene Clusters

Karina Gutiérrez-García[1,2], Edder D. Bustos-Díaz[1], José Antonio Corona-Gómez[2], Hilda E. Ramos-Aboites[1], Nelly Sélem-Mojica[1], Pablo Cruz-Morales[1,2], Miguel A. Pérez-Farrera[3], Francisco Barona-Gómez[1,*], and Angélica Cibrián-Jaramillo[2,*]

[1]Evolution of Metabolic Diversity Laboratory, Unidad de Genómica Acanzada (Langebio), Irapuato, Guanajuato, México

[2]Ecological and Evolutionary Genomics Laboratory, Unidad de Genómica Avanzada (Langebio), Irapuato, Guanajuato, México

[3]Herbario Eizi Matuda, Laboratorio de Ecología Evolutiva, Instituto de Ciencias Biológicas, Universidad de Ciencias y Artes del Estado de Chiapas, Tuxtla Gutiérrez, Chiapas, México

*Corresponding authors: E-mails: angelica.cibrian@cinvestav.mx; francisco.barona@cinvestav.mx.

Accepted: December 10, 2018

Data deposition: This project has been deposited at GenBank under de accesion numbers as follows:

SUBID	BioProject	BioSample	Accession	Microorganism
SUB2297132	PRJNA360300	SAMN06208854	MTAV00000000	*Nostoc* sp. T09
SUB2299096	PRJNA360305	SAMN06208961	MTAW00000000	*Nostoc* sp. 106C
SUB2299173	PRJNA360315	SAMN06209042	MTAX00000000	*Nostoc* sp. RF31YmG
SUB4062599	PRJNA472998	SAMN09257966	QHKA00000000	*Nostoc* sp. 3335mG
SUB4062599	PRJNA472998	SAMN09257968	QHJY00000000	*Caulobacter* sp. D5
SUB4062599	PRJNA472998	SAMN09257967	QHJZ00000000	*Caulobacter* sp. D4A

Metagenomes are available at: *Dioon merolae* coralloid roots: https://www.ncbi.nlm.nih.gov/bioproject/369422. *Dioon edule* coralloid roots: http://www.ncbi.nlm.nih.gov/bioproject/472628. Cyanobacterial genomic sequences: https://www.ncbi.nlm.nih.gov/sra/SRP149992.

Abstract

Cycads are the only early seed plants that have evolved a specialized root to host endophytic bacteria that fix nitrogen. To provide evolutionary and functional insights into this million-year old symbiosis, we investigate endophytic bacterial sub-communities isolated from coralloid roots of species from *Dioon* (Zamiaceae) sampled from their natural habitats. We employed a sub-community co-culture experimental strategy to reveal both predominant and rare bacteria, which were characterized using phylogenomics and detailed metabolic annotation. Diazotrophic plant endophytes, including *Bradyrhizobium*, *Burkholderia*, *Mesorhizobium*, *Rhizobium*, and *Nostoc* species, dominated the epiphyte-free sub-communities. Draft genomes of six cyanobacteria species were obtained after shotgun metagenomics of selected sub-communities. These data were used for whole-genome inferences that suggest two *Dioon*-specific monophyletic groups, and a level of specialization characteristic of co-evolved symbiotic relationships. Furthermore, the genomes of these cyanobacteria were found to encode unique biosynthetic gene clusters, predicted to direct the synthesis of specialized metabolites, mainly involving peptides. After combining genome mining with detection of pigment emissions using multiphoton excitation fluorescence microscopy, we also show that *Caulobacter* species co-exist with cyanobacteria, and may interact with them by means of a novel indigoidine-like specialized metabolite. We provide an unprecedented view of the composition of the cycad coralloid root, including phylogenetic and functional patterns mediated by specialized metabolites that may be important for the evolution of ancient symbiotic adaptations.

Key words: cycads, coralloid roots, *Dioon*, cyanobacteria, *Caulobacter*, EcoMining.

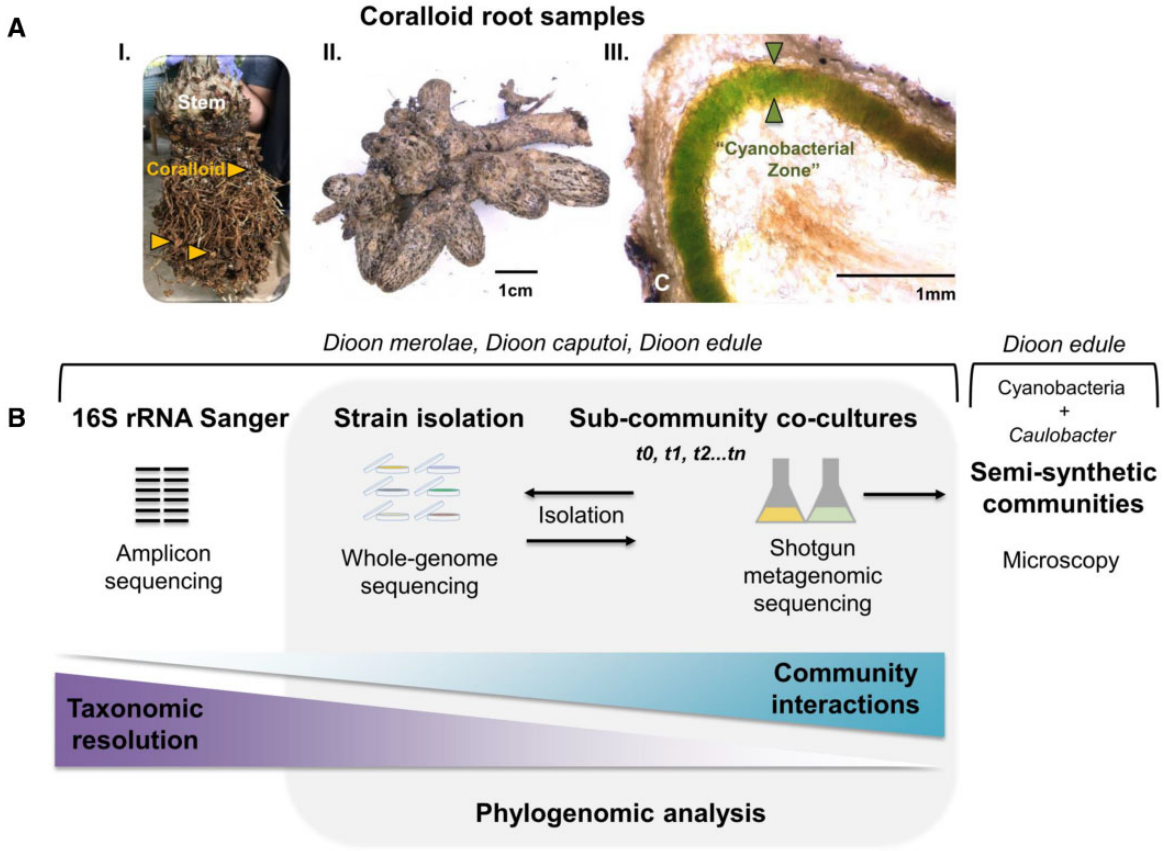

Fig. 1.—Diagram of experimental strategy to capture bacterial diversity (EcoMining). (*A*) Coralloid roots from *Dioon merolae*, *Dioon caputoi*, and *Dioon edule* were sampled. The primary root is shown with coralloids marked by an arrow (I); photograph of a coralloid root (II); cross-section of a coralloid root showing the "cyanobacterial zone" (III). (*B*) Endophytes from the macerated root were isolated, following two strategies: directly from the sample (*t0*) and after enrichment of sub-community co-cultures in BG-11_0 (*t1*, *t2*). In both cases, cultivable bacteria were obtained using an array of six different media. co-cultures were characterized using shotgun metagenomics, and the resulting data were used to select representative genomes from the endophyte culture collection that we mined for functional and evolutionary information. All the information obtained was used to construct a semi-synthetic community (Cyanobacteria plus *Caulobacter*) that could be mediated by specialized metabolites.

Introduction

Cycads (Cycadales) are ancient gymnosperms of Permian origin, currently mostly distributed in endangered low-density or isolated populations throughout New World tropical and subtropical regions (Gutiérrez-Ortega et al. 2017). Cycads are the only gymnosperms and among the few plant lineages that evolved a specialized root organ to host symbiotic bacteria. Their lateral roots can develop into coralloid roots, a dichotomous and coral-like small cluster of roots, typically growing above ground, that acquire symbiotic bacteria (Norstog and Nicholls 1997). Its main function is nitrogen fixation for the plant (Bergersen et al. 1965; Grobbelaar et al. 1987), similar to the adaptive functions in legume nodules, but having appeared millions of years before them. In natural habitats coralloid roots appear in early life stages of the plant (Halliday and Pate 1976); and in adults mainly in habitats with poor or inaccessible nutrients, such as sand dunes, sclerophyll forests, steep rock outcrops with high exposure to salt, and lowland forests with recurrent fires (Grove et al. 1980; our observations). It is possible that coralloid roots and their bacteria are a key early trait that enabled cycads to thrive and adapt to changing environments during millions of years.

Coralloid root endophytes have been studied since the 19th century (Grilli Caiola 1980 and references therein). However, most studies have focused on resolving the taxonomy of cultured cyanobacteria, which can be visually detected in the root forming a green "cyanobacterial ring" under differentiated cortical cells lying beneath the epidermis (Storey 1968; fig. 1*A*). These studies have been carried out with samples collected from either botanic garden collections in greenhouses (Zimmerman and Rosen 1992; Costa et al. 1999; Thajuddin et al. 2010), natural populations (Yamada et al. 2012; Cuddy et al. 2012), or a mixture of both (Costa et al. 2004; Gehringer et al. 2010; Zheng et al. 2018). Studies testing for the specificity of cyanobacteria report contrasting

results regarding the specialization of coralloid root symbionts. In wild populations, a single *Nostoc* strain was found in different coralloid roots (Costa et al. 2004). Other studies report a significant degree of biodiversity and compatibility with the plant (Grilli Caiola 1980; Ow et al. 1999; Yamada et al. 2012). There is also inconclusive information on the origin and transmission of cycad bacterial endophytes (Lobakova et al. 2003a; Cuddy et al. 2012).

Anatomical studies of the cyanobacteria ring have shown the presence of mucilaginous or protein-rich material, which may host other associated bacteria in addition to cyanobacteria (Grilli Caiola 1980; Baulina and Lobakova 2003; Lobakova et al. 2003b). The taxonomic associations of these bacteria, however, have been mostly suggested on the basis of generally low-resolution markers (Grobbelaar et al. 1987; Huang and Grobbelaar 1989; Thajuddin et al. 2010; Zvyagintsev et al. 2010; Zheng et al. 2018), and doubts about their endophytic nature remain to be dissipated. In addition to nitrogen fixation there have been suggestions of other unknown roles for the coralloid root (Grilli Caiola 1980), but there is no clear evidence of a broader function to date. Likewise, various chemical, physical, and physiological processes appear to regulate the cycad–bacteria interaction (Lobakova et al. 2003a; Meeks 2009), although genes coding for these specialized mechanisms of the symbiosis have been not identified. The presence of unique specialized metabolites in the cycad coralloid root bacterial endophytes is of interest to us because they may be a result of co-evolution between the cycad host and the endophyte symbiotic bacterial community (Liaimer et al. 2016; Warshan et al. 2018).

Certain soil-dwelling and aquatic bacteria, including cyanobacteria, have genomic plasticity (Warshan et al. 2018) and the capacity to synthesize specialized metabolites with overwhelming chemical diversity, believed to be produced to cope with biotic and abiotic pressures. These bacterial traits are concomitant with large genomes that code for specialized metabolites in genetic loci of ~25–40 kb, called biosynthetic gene clusters (BGCs) (Tan 2007). Specialized metabolites play many important roles in competition and collaboration in microbe–microbe and host–microbe interactions (Massalha et al. 2017; Scherlach and Hertweck 2018). Hence, the term of "specialized metabolite" rather than secondary metabolite or natural product provides a more accurate description of the evolution and biology related to these compounds.

Our goal in this study is to investigate the endophytic bacterial community of the coralloid roots of species of *Dioon*, which are long-lived, perennial, entomophilous, dioecious, and arborescent cycads endemic to Mexico (De Luca et al. 1981; Lázaro-Zermeño et al. 2012). *Dioon* is the only genus that spans all the Neotropical biogeographic provinces in Mexico, distributed in a wide range of environmental conditions often associated with very steep, rocky slopes and cliffs with low nutrient availability (Gutiérrez-Ortega et al. 2017). The subject species of this study, *Dioon caputoi*, *Dioon edule*, and *Dioon merolae*, are smaller species with shorter trunks, fronds and cones (Norstog and Nicholls 1997), and they belong to two closely related phylogenetic clades (Gutiérrez-Ortega et al. 2017). Coralloid root samples were collected from natural populations from pine-oak forests, deciduous tropical forests, and xeric shrubland from their natural range, currently distributed in moderate population sizes of a few hundreds of individuals throughout the States of Queretaro, Puebla, Oaxaca, and Chiapas, in Central and South of Mexico (supplementary fig. s1, Supplementary Material online).

Here, we provide evidence of the existence of a diverse endophytic bacterial community inside the coralloid root of *Dioon* species beyond cyanobacteria, which consists of a core of diazotrophic bacteria, and the remarkably small Alphaproteobacteria *Caulobacter*. Unique BGCs in monophyletic Nostocales and *Caulobacter* endophytes suggest functional specificity throughout the evolutionary history of the cycad–cyanobacteria coralloid root system, and that some of these metabolites may mediate bacterial interactions within this symbiotic niche.

Materials and Methods

Overall Strategy

To overcome technical difficulties in characterizing the breadth of microbial diversity in these environmental samples, we used EcoMining, a combined sub-community co-culture, metagenomics and phylogenomics strategy to detect and measure taxonomic diversity and phylogenetic relationships in the endophytes of the cycad coralloid root (Cibrián-Jaramillo and Barona-Gómez 2016) (fig. 1). We grew and isolated bacteria from environmental samples using a diverse set of semiselective media that aim to capture as much of the cultured bacterial diversity as possible (*t0*). Simultaneously, we enriched the same samples for specific sub-communities in co-cultures grown under oligotrophic conditions specific for cyanobacteria, namely, BG-11_0 medium. This second approach aims to capture other bacterial groups that interact with, and depend upon, the metabolic capabilities of cyanobacteria grown under oligotrophic conditions. The sub-community co-cultures were grown over time and sampled after one month (*t1*) and at the end of one year (*t2*), leading to bacterial isolates, their genomes and metagenomes, with the exception of bacterial isolates for *t2*. Specifically, six cyanobacterial draft genomes from sub-community co-cultures (strains JP106B, JP106C, TVP09, RF15115, 35k25, and 33k59), and two draft genomes reconstructed from metagenomic data (RF31YmG and 3335mG) were obtained. These genomic sequences were mined for BGCs potentially directing the synthesis of specialized metabolites. Also, two *Caulobacter* strains, D5 and D4A, associated with cyanobionts from *D. edule* were isolated, and their genomes sequenced and mined for niche-specific metabolites. Both cyanobacteria

and *Caulobacter* strains were characterized after pigment emission using multiphoton excitation fluorescence microscopy. Specific methods, materials and analytical tools are described in the following subsections.

Field Collections

We sampled coralloid roots from wild populations for each of the three *Dioon* species based on their previously reported distribution (Gutiérrez-Ortega et al. 2017). We sampled from two *D. merolae* populations in deciduous tropical forests, at JP Oaxaca Mexico (JP) at 560 m above mean sea level, with an average annual precipitation of 320 mm and average annual temperature of 18 °C; and at RF, Oaxaca, Mexico (RF) at 900 m above mean sea level, with 2,500 mm and 25 °C annual average precipitation and temperature, respectively. We also sampled one population for *D. caputoi* in Xeric shrubland, Tehuacán Valley Puebla, in Puebla, Mexico (TVP), at 1,600 m above mean sea level with an average annual precipitation of 474 mm and average annual temperature of 18 °C. We sampled one pine oak forest *D. edule* population in SJD, Queretaro, Mexico (SJD) at 1,300 m above mean sea level, with an average annual precipitation of 966 mm and average annual temperature of 24 °C (supplementary fig. s1, Supplementary Material online). Exact coordinates and names of locations can be shared upon request to preserve the integrity of these endangered populations.

In some cycads coralloid roots were easily visible above ground, whereas in others we dug to ~20–30 cm around the main root until coralloid roots were found, especially when they were adult plants. We generally found 6–12 plants with coralloid roots, almost exclusively in seedlings, or in adult plants living on rock outcrops. A total of 20 coralloid apogeotropic roots from all species were removed from plants that appeared healthy, to minimize the negative impact on the plant. Roots were placed in 15 ml sterile Falcon tubes (Beckton Dickinson) and transported immediately to the laboratory at room temperature.

Coralloid Root Processing

We focused our sequencing effort on 13 samples: 1) two *D. merolae* samples from two plants from JP, named JP2 and JP6, and two *D. merolae* samples from two individuals from RF, named RF1 and RF3; 2) one sample from *D. caputoi* (TVP09); and 3) eight samples from two individuals from SJD, named SJD1–SJD8 (fig. 1*A*). The remaining coralloid root samples were stored at –80 °C for subsequent studies. We treated the coralloid root in a laminar flow hood (Nuaire Model Nu-126-400) with a series of washes to remove exogenous bacteria from the rhizosphere or epiphytic contamination. Each root was introduced in 50 ml sterile Falcon tubes containing 10 ml of each of the following solutions, and gently stirred for 3 min in hydrogen peroxide (H_2O_2), 7 min in 70% ethanol, 30 s in sterile dd-MilliQ water, 4 min in 6% sodium hypochlorite (NaClO), and three 1-min washes in sterile dd-MilliQ water. After this procedure, we plated out water from the last wash in two media described below for as long as the plate dry out (between 3 and 5 weeks). Lack of growth in the last wash was considered a negative control. Only samples complying with this criterion were used for endophyte isolation (i.e., *Caulobacter*). We undertook two approaches to bacterial isolation (fig. 1*B*): sampling without enrichment directly from field samples after treatment for removal of epiphytes (*t0*), and sampling from the enriched co-cultures (*t1*).

Bacterial Growth and Isolation in Selective Media

To isolate bacteria from the original epiphyte-free field samples (*t0*) and after enrichment (*t1*), coralloid roots were macerated and the macerate or co-culture broth were used as inoculant. In all cases, roots were macerated in 10 ml of sterile water using a pestle and mortar until plant material was completely disintegrated. After enrichment (*t1*) we recover bacteria expected to be initially present in low abundances and required time to grow, and that did so as a response to the sub-community nutritional interactions (e.g., amino acids derived from the process of fixing nitrogen). We used 100 µl from the root macerate to directly isolate bacteria in Petri dishes containing a total of six different media to recover bacterial diversity as much as possible. Four selective media were chosen to target bacterial groups that are known to be either plant endophytes or rhizosphere bacteria, and included: 1) *Caulobacter* medium (glucose: 1 g/l; peptone: 1 g/l; yeast extract: 1.5 g/l; trace metals solution: 1 ml/l; and 10 g/l of agar for solid medium); 2) *Rhizobium* medium (mannitol: 10 g/l; dipotassium phosphate: 0.5 g/l; magnesium sulfate: 0.2 g/l; yeast extract: 1 g/l; sodium chloride: 0.1 g/l; final pH 6.8; and of agar for solid medium; 3) ISP4 and $ISP4^{N-}$, for isolation of actinomycetes (starch: 10.0 g/l; dipotassium phosphate: 1 g/l; magnesium sulfate: 1 g/l; sodium chloride: 1 g/l; ammonium sulfate: 2 g/l or none in the case of $ISP4^{N-}$; calcium carbonate: 2 g/l; ferrous sulfate: 1 mg/l; magnesium chloride: 1 mg/l; zinc sulfate: 1 mg/l; final pH 7.0; and 20 g/l for solid media); 4) $BG\text{-}11_0$, a cyanobacteria medium (sodium nitrate: 1.5 g/l; dipotassium phosphate: 0.04 g/l; magnesium sulfate: 0.075 g/l; calcium chloride: 0.036 g/l; citric acid: 0.006 g/l; ferric ammonium citrate: 0.006 g/l; EDTA [disodium salt]: 0.001 g/l; sodium carbonate: 0.02 g/l; final pH 7.1; and agar solid media 10.0 g/l). Moreover, two nonselective rich media were included: 5) Nutrient Broth (BD Bioxon, Mexico); and 6) As in *Caulobacter* medium, but supplemented with mannitol (*Caulobacter* + mannitol medium): 1 g/l, with aim of providing a carbon source closer to that hypothetically encountered inside the cycad root.

Bacterial Growth and Cultivation in Sub-communities

Hundred microliters of the macerated roots free from epiphytes from the 13 samples (two *D. merolae* samples from

JP, two *D. merolae* samples from RF, one from *D. caputoi*, and eight *D. edule* samples from SJD), which passed the negative growth control were used to inoculate 100 ml of media in 250 ml flasks. The remaining macerated roots not used for fresh cultures were kept as frozen stocks for future studies (–80 °C in 50% glycerol), although community viability after freezing was found to diminish over time. We used selective oligotrophic medium, that is, BG-11_0 (medium No. 4) for these experiments. BG-11_0 cyanobacteria-centric co-cultures were grown for up to one year with constant stirring, with cycles of 16/8 h of light/darkness. Sampling of the oligotrophic sub-communities was carried out at two different times, between 30 and 45 days (*t1*) and one year (*t2*): *D. merolae* oligotrophic sub-communities were sampled after 30 days (*t1*) and one year (*t2*), and treated independently; whereas *D. edule* co-cultures were sampled after 45 days (*t1*) and *D. caputoi* after 35 days (*t1*), although the latter did not include bacterial isolation. Bacterial isolates were only obtained from *t0* and *t1*, whereas shotgun metagenomics were obtained for *t1* (all three species) and *t2* (solely *D. merolae*).

Genomics and Shotgun Metagenomics

The genomic sequence of five cyanobacteria isolates, JP106B, JP106C, RF15115, 33k59, and 35k25, were obtained from samples JP2, JP6, and RF1 (*D. merolae*) plus SJD1 and SJD2 (*D. edule*), respectively (*t1*). The genome sequence of the cyanobacterial isolate TPV09 was obtained from coralloid roots of *D. caputoi*. After many replates in both solid and liquid BG-11_0 medium, all the microorganisms were grown on BG-11_0 plates before DNA extraction, which was done using a CTAB-phenol chloroform standard protocol. Metagenomic DNA of all enriched sub-community co-cultures (200 ml), coming from JP, RF, and SJD samples, were extracted using the same protocol, with the exception that biomass was obtained from liquid cultures by high-speed centrifugation during 15 min. Additionally, two *Caulobacter* strains, D5 and D4A, were isolated from 35k25 sample (*t1*) using ISP4 medium (medium No. 3). The axenic microorganisms were grown on liquid *Caulobacte*r medium (medium No. 1) before DNA extraction, carried out using the protocol described previously.

DNA samples were processed with the truseq nano kit and sequenced at Langebio, Cinvestav (Irapuato, Mexico) using the MiSeq Illumina platform in the 2X250 paired end reads format (TVP09), and the NextSeq mid output 2X150 paired-end read format for all other samples. The reads for each library were filtered with fastQ and trimmed using Trimommatic version 0.32 (Bolger et al. 2014). Assembly of JP106B, JP106C, 33k59, 35k25, RF15115, TVP09, D4A, and D5 sequences was done de novo, whereas the RF31YmG and 3335mG genomic sequences were obtained from metagenomic-assembled reads of the sub-community coculture RF31Y of *D. merolae* (*t2*), and 33k59 plus 35k25 of *D. edule* (*t1*), respectively. For these assemblies, the newly obtained sequence of strain JP106C and *Nostoc punctiforme* PCC 73102 were used as references. The metagenomic reads were filtered by mapping them against the assembly of each reference genomes with BWA (Li and Durbin 2009). In all cases the resulting reads were assembled with Velvet (Zerbino and Birney 2008) using different *k*-mers: the assemblies with the largest length and the smaller number of contigs were selected and annotated using RAST (Aziz et al. 2008). The genomic sequences that were released as independent metagenomic-assembled genomes (from 33k50 plus 35k25, RF31Y, TVP09, and JP6106C, respectively, fig. 2) are 3335mG (GenBank PRJNA472998), RF31YmG (GenBank PRJNA360315), T09 (GenBank PRJNA360300), and 106C (GenBank PRJNA360305).

To confirm the presence of diazotrophic microorganisms in the sub-community co-cultures, we searched for the NifH protein sequence, an enzyme involved in nitrogen fixation and relevant in symbiotic relationships (Warshan et al 2018). We used the NifH protein from *N. punctiforme* PCC 73102 (B2IV87), *Bradyrhizobium japonicum* OX = 375 (I4DDG5), and *Mesorhizobium japonicum* LMG 29417 (Q98AP7) from the NCBI databases (July 2018), as query sequences to perform standard protein BLAST searches with the default parameters.

Taxonomic Diversity

We estimated taxonomic diversity in single isolates using their 16S rRNA gene sequences obtained after PCR amplification and Sanger sequencing of our entire bacterial endophyte collection (*t0* and *t1*). PCR products of 1.4 kb in length, obtained using the F27 and R1492 primers (Lane 1991), were cloned and sequenced using the Sanger method. Taxonomic identification was made using BlastN with an initial cut-off e-value of 1e–5 against the SILVA database (Quast et al. 2012) and >98% sequence identity. All the sequences were aligned and trimmed to 1.1 kb, using MUSCLE v3.8.31 with default parameters (Edgar 2004). ModelTest v3.7 (Posada and Crandall 1998). Kimura 2 parameters was found as the best substitution model. A multiple alignment matrix was used for phylogenetic reconstruction using MrBayes v3.2 (Ronquist et al. 2012) with a gamma distribution type range with 1,000,000 generations. The resulting tree was edited and sorted with Environment for Tree Exploration Toolkit v3.0.0b35 (Huerta-Cepas et al. 2010) in Python v2.7.6. We only report genus-level taxonomy as species determination can be biased by reference 16S rRNA databases.

We used Kraken (http://ccb.jhu.edu/software/kraken/; last accessed December 2018) to assign taxonomic labels to metagenomic DNA sequences based on exact alignment of *k*-mers (Wood and Salzberg 2014). Kraken was used to exclude sequence contaminants from the draft assembly, allowing us to generate a symbiotic cyanobacteria marker database as reference for future classification. We made a standard Kraken

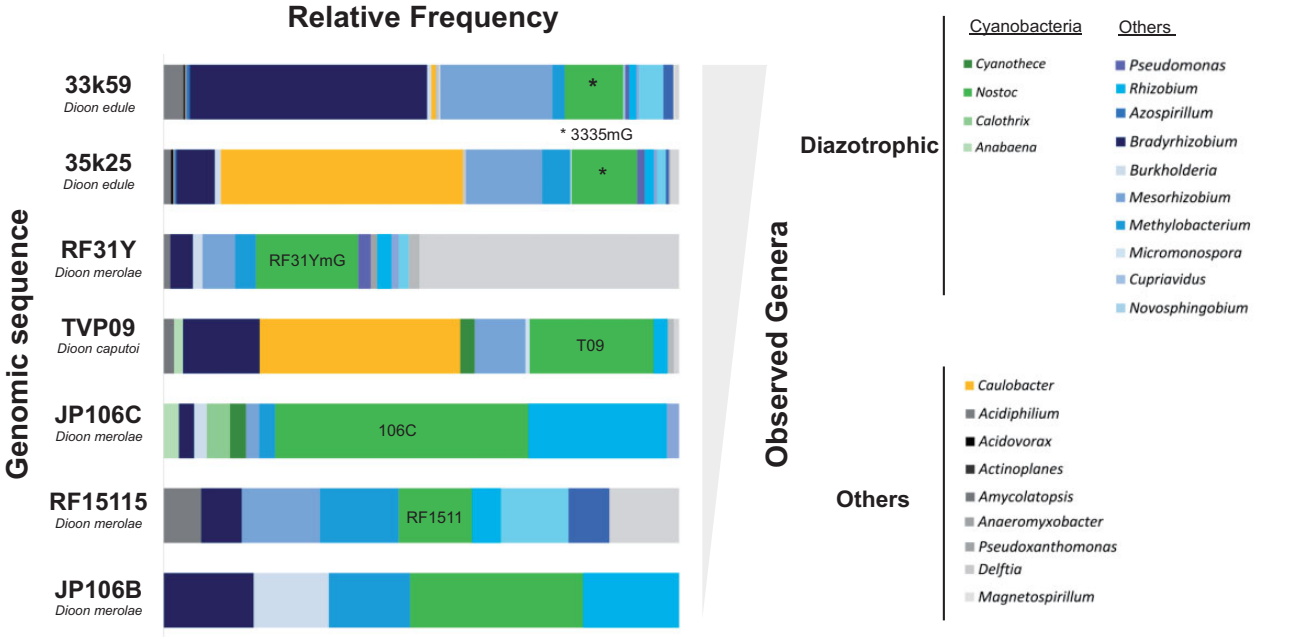

Fig. 2.—Taxonomic diversity of selected endophytic sub-community co-cultures characterized with shotgun metagenomics. Seven endophytic genomic sequences obtained from sub-community co-cultures were organized according to the number of genera observed. Diazotrophic microorganisms are prevalent: cyanobacterial species (marked in green) and *Bradyrhizobium*, *Burkholderia*, *Mesorhizobium*, and *Rhizobium* species (marked in blue). In some samples *Caulobacter* was also found (marked in yellow). 3335mG, RF31YmG, T09, and 106C assemblies were obtained from 33k50 plus 35k25 (*), RF31Y, TVP09, and JP6106C, respectively. See Materials and Methods for more details of the sequences nomenclature.

database using NCBI taxonomic information for all bacteria, as well as the bacterial, archaeal, and viral complete genomes in RefSeq (October 2016). This database contains a mapping of every *k*-mer in Kraken's genomic library to the lowest common ancestor in a taxonomic tree of all genomes that contain that *k*-mer. We summarized the results in genera-level tables for each metagenome and filtered taxonomy hits that had 50 or more reads assigned directly to a taxon. To visualize this diversity, we selected the best genomic sequences based in two criteria: 1) coverage over 25×; and 2) assemblies with <20,000 contigs. From these sequences, we highlight only taxa with 500 reads or more in figure 2.

Reconstruction of Phylogenomic Relationships

To reconstruct a complete phylogeny of the cyanobacteria and *Caulobacter* strains using their whole genomes, we adopted a phylogenomics approach that we termed OrthoCores based on our in-house scripts (https://github.com/nselem/orthocore/wiki; last accessed December, 2018, dockerhub: nselem/orthocores). We assume that single gene or even housekeeping multilocus phylogenies are often not enough to resolve the evolutionary history of certain taxonomic groups, and thus total evidence tree evidence approaches must be used (Narechania et al. 2012). We first obtained a set of orthologous groups that are shared among all high-quality genomes, or the "core genome," excluding paralogs and signatures of expansions (not all genes from a gene family are part of the Bidirectional Best Hits, or BBH set). We define orthologs from the core genome as hits that satisfy that all gene members are BBHs, all versus all, to cyanobacteria and *Caulobacter*. To place our samples within the cyanobacteria and *Caulobacter* phylogeny, we used OrthoCores to generate the tree. For the cyanobacterial phylogenetic tree, we selected closed and well-annotated genomes, and for Nostocales all the publicly available genomes. We included representatives from public Chroococcidiopsidales, Oscillatoriales, Synechococcales, Chroococcales, Pleurocapsales, Gloeobacterales, and Nostocales and the eight genomic sequences obtained in this study. The *Caulobacter* phylogenetic tree was constructed with 22 publicly available genomes including the two *Caulobacter* genomes D5 (GenBank PRJNA472998) and D4A (GenBank PRJNA472998) and *Hyphomonas jannaschiana* VP2, used as outgroup.

The OrthoCores cut-off e-value was 1e–6, although most genes had a 1e–50 score given our all versus all criterion. Set of sequences were concatenated and aligned using MUSCLE v3.8.31 with default parameters (Edgar 2004), and trimmed by Gblocks (Talavera and Castresana 2007) with five positions as minimum block length, 10 as a maximum number of contiguous nonconserved positions, and only considering positions with a gap in <50% of the sequences in the final alignment. The final cyanobacterial matrix included 81 taxa

and 198 core protein sequences with 45,475 amino acids that represent a set of 71 cyanobacterial genomes including the six cyanobacterial genomic sequences (supplementary table s1, Supplementary Material online), plus the RF31YmG and 3335mG metagenomic assemblies, and *Gloeobacter kilaueensis* JS1 and *Gloeobacter violaceus* PCC 7421 as outgroups. To build each cyanobacteria subsystems tree, the core proteome was used and proteins were classified using RAST (Aziz et al. 2008) and were used to construct a concatenated matrix. Each concatenated matrix was used to construct a phylogenetic tree using the methodology described above. The final *Caulobacter* matrix included 82 core protein sequences with 20902 amino acids (supplementary table s2, Supplementary Material online). The resulting matrices were used to reconstruct the trees with MrBayes, using a mixed substitution model based on posterior probabilities (aamodel[Wag]1.000) for proteins for one million generations. Both trees were visualized with FigTree http://tree.bio.ed.ac.uk/software/figtree/; last accessed December, 2018).

Genome Mining for BGCs

To identify BGCs potentially directing the synthesis of specialized metabolites among selected cyanobacteria (JP106C, TVP09, JP106B, RF31YmG, RF15115, 3335mG, 35k25, and 33k59) and *Caulobacter* (D5 and D4A), we annotated their genomic sequences, as well as selected genomes from the database, using antiSMASH (Weber et al. 2015). The quality of the predicted BGCs in both cyanobacteria and *Caulobacter* strains (248 and 7 potential BGCs, respectively) was confirmed (only 49 cyanobacterial antiSMASH outputs did not pass this filter), leading to 77 and 7 nonredundant BGCs used as references for subsequent searches in 41 Nostocales and 21 *Caulobacter* selected genomes, respectively. This process allowed us to identify both conserved and specific BGCs throughout our databases generated from publicly available genomes on NCBI (supplementary tables s3 and s4, Supplementary Material online). We considered that a BGC hit was found whenever any given enzyme, plus any other enzyme gene from the cognate BGC, were found in a vicinity of 20 genes, using a BLAST e-value cut-off of 0.001.

The prediction of the chemical structures of the putative specialized metabolites associated with some BGCs, namely, those *Dioon*-specific present in the cyanobacterial endophytes genomic sequences (BGCs 1, 5, and 12), and a BGC specific within *Caulobacter* endophytes strains, were done after domain identification and specificity prediction of adenylation and acyl transfer domains, using the nonribosomal peptide synthetase (NRPS) and polyketide synthases (PKS) NRPS–PKS server (Bachmann and Ravel 2009), PRISM (Skinnider et al. 2015), and antiSMASH (Weber et al. 2015). The functional inferences of selected BGCs present in these strains, in parallel with their structural chemical prediction, was carried out after in-depth functional annotation of all genes belonging to these BGCs, using standard databases and sequence similarity approaches, together with a detailed revision of the literature.

Pigment Emissions Using Multiphotonic Microscopy

To probe if the genus *Caulobacter* co-exists with the cyanobacterial strains isolated from cycads, we exploited the occurrence of specialized metabolites produced by these microorganisms. We obtained pigment emissions of both isolated microbial cultures using intensity, time, and wavelength-resolved multiphoton excitation fluorescence microscopy, with a LSM880-NLO Multiphoton Microscope (Zeiss, Germany) coupled to an Infrared laser Ti: Sapphire (Chameleon vision II, COHERENT, Scotland). To perform in vitro assays, we used the *Caulobacter* endophyte D5, and the cyanobacteria isolate 35k25. Each isolate was grown on different plates containing BG-11_0 medium. We also cultured both isolates together on plates containing BG-11_0 and BG-11 media. All the samples were visualized through multiphoton excitation fluorescence microscopy, adding sterile water as a propagation medium. The observation areas were made using the objective 60x/1.3 NA ∞-0.17 (Zeiss Plan-Neofluar, Germany). The samples were excited at different excitation energies (700–850 nm). Three emissions were detected: 1) Multialkali detector (Ch1), used to collect photons between 371 and 471 nm; 2) Detector GAsP (Ch2), used to collect photons between 501 and 601 nm, and 3) Detector Multialkali (Ch3), used to collect photons between 616 and 688 nm. Scans were performed in the image acquisition mode and in the lambda mode to obtain the spectral fingerprint of each analyzed area. The spectral signature is specific for every molecule and determined for each pixel of scanned image. These data were used for the subsequent digital separation of fluorescent molecules.

Results

Dioon Coralloid Roots Show Diazotrophic Endophyte Diversity beyond and within Cyanobacteria

The 16S rRNA Sanger sequences that we obtained yielded 247 isolates distributed in 17 families and 10 bacterial orders, with 27 genera in total, representing most of the known bacterial groups (supplementary table s5, fig. s2, Supplementary Material online). Eighty six percent of the taxa identified can be taxonomically classified as diazotrophic plant endophytes, including species belonging to the genera *Bradyrhizobium*, *Burkholderia*, *Mesorhizobium*, *Rhizobium*, and *Nostoc* (supplementary table s5, fig. s2, Supplementary Material online). We also isolated *Bacillus*, which was previously reported as associated with the outside of the coralloid root (Lobakova et al. 2003a); *Streptomyces*, isolated as an epiphyte (Zvyagintsev 2010); and four strains identified as *Nostoc* spp., this genus previously identified in several cycads.

However, as it will be further discussed in detail in the subsection dealing with the genomic analysis, these isolates turned out to be a microcommunity. Moreover, we continuously isolated *Caulobacter*, a genus that has been associated with cyanobacteria from *Azolla* ferns (Newton and Herman 1979) and aquatic cyanobacteria (Bunt 1961; Berg et al. 2009), but to the best of our knowledge this is the first report of *Caulobacter* as a cycad endophyte.

Metagenomic sequencing of endophytic sub-communities was successful for JP2, JP6, RF1, RF3, and SJD1 to SJD8 grown until *t1*, and RF31Y and JP61Y, grown until *t2*. Basic indicators of metagenome quality are provided in supplementary table s6, Supplementary Material online. Taxonomic analysis of this data was used to identify the most abundant genera, starting at 50 reads (supplementary table s7, Supplementary Material online). Dominant taxa defined as those with >500 reads were identified and these are shown in figure 2. As expected from visual inspection, *Nostoc* was prevalent in both *t1* and *t2* samples. To represent the taxonomic diversity of the endophytic sub-communities, shown as solid-medium co-cultures in supplementary figure s3, Supplementary Material online, we selected only the genomic sequences that passed the abovementioned criteria, namely, 33k59, 35k25, RF31Y, TVP09, JP106C, RF15115, and JP106B.

The endophytic sub-communities are enriched with diazotrophic microorganisms, including cyanobacteria belonging to the Nostocales order, such as those from the genera *Nostoc* (Thajuddin et al. 2010; Yamada et al. 2012) and *Calothrix*, previously reported in *Encephalartos* (Grobbelaar et al. 1987; Huang and Grobbelaar 1989; Gehringer et al. 2010) and in *Cycas revoluta* (Thajuddin et al. 2010). Other diazotrophic microorganisms, such as *Bradyrhizobium*, *Burkholderia*, *Mesorhizobium*, and *Rhizobium* species, were also found. Moreover, *Caulobacter* species were found in 33k59, 35k25, and TVP09 samples (fig. 2). The presence of diazotrophic microorganisms in the sub-community cultures was confirmed by BLAST-positive hits of the NifH enzyme to selected genera, such as *Bradyrhizobium* and *Mesorhizobium* species, with 99% sequence identity.

Overall, the 16S rRNA Sanger results and the taxonomic analysis of our metagenomic data agrees with the bacterial diversity reported using single-amplicon metagenomics of environmental samples (Suarez-Moo et al. 2018). Among the most abundant, we confirm the occurrence of *Amycolaptopsis*, *Burkholderia*, *Caulobacter*, *Microbacterium*, *Pseudomonas*, *Ralstonia*, *Rhizobium*, *Serratia*, *Sphingobium*, *Sphingomonas*, *Streptomyces*, and *Xanthomonas*, in addition to the Nostocales genera, as *Dioon* coralloid root endophytes.

Cyanobacteria Endophytes Are a Monophyletic Group within the Nostocales

Given that cyanobacteria phylogenetic history is likely reticulated (Zhaxybayeva et al. 2006), we adopted a whole-genome inference approach to explore the phylogenetic position of endophytes genomic sequences, including JP106B, JP106C, RF15115, 33k59, 35k25, and TVP09, and two draft genomes reconstructed from metagenomic data of co-culture sub-communities, RF31YmG and 3335mG. We reconstructed a phylogeny from 198 proteins identified as the core proteome encoded in 75 selected genomes including our own sequences, which provide a broad taxonomic coverage and represent good quality sequences (supplementary tables s1, s3, Supplementary Material online). With the exception of the Nostocales, we found lack of monophyly of the cyanobacteria groups in our phylogeny (supplementary fig. s4, Supplementary Material online). However, the genomic sequences TVP09 from *D. caputoi*, plus JP106B, JP106C and RF31YmG from *D. merolae*, form a monophyletic clade, and group with *Tolypothrix* sp. PCC 7601. As can be seen in figure 3, their clade is sister to *Calothrix* sp. PCC 7507 and *Fortiea contorta* PCC 7126, both reported as aquatic microorganisms. These microorganisms have not been tested for symbiotic competence, although a related species, *Calothrix rhizosoleniae*, is a symbiont of the diatom *Chaetoceros* (Foster et al. 2010). Likewise, RF15115 from *D. merolae*, with 33k59, 35k25, and 3335mG from *D. edule*, form a monophyletic clade with *N. punctiforme* PCC 73102 from the cycad *Macrozamia* and *Nostoc* sp. KVJ20 from the liverwort *Blasia pusilla* (Warshan et al. 2018).

In agreement with our phylogenetic reconstruction, the genomic sequences from endophytes of the coralloid roots were notably syntenic with the genome of KVJ20, indicating that these microorganisms may share adaptive traits within equivalent ecological niches, that is, specialized organs devoted to nitrogen fixation. It should be noted that aside from *N. punctiforme* PCC 73102, the genome of a cyanobacterial strain isolated from the coralloid roots of a *C. revoluta* plant, termed *Nostoc cycadae* WK-1 (Kanesaki et al. 2018), was recently released. WK-1 strain, isolated and kept under laboratory conditions for almost four decades before its sequencing, was not included in our analyses as it does not group with the symbiotic cyanobacteria reported herein (Gagunashvili and Andrésson 2018). This latter report deals with cyanolichens genomes, but includes some of our genomic sequences previously released. Despite these indications of a close relationship of our isolates with species of the genus *Nostoc* we refer to our strains as Cyanobacterium as they could also be members of the genera *Calothrix*, *Fortiea*, or *Tolypothrix* (fig. 3).

Long branches of 3335mG and RF15115, suggesting divergence from their closely related microorganisms, are also obvious (fig. 3). To further characterize this observation, trees based on gene sets from the core proteome, which correspond to 16 metabolic subsystems, were reconstructed, as different phylogenetic signatures between them may contribute to long branches. All of the subsystem trees showed divergent branches for both 3335mG and RF15115, albeit with

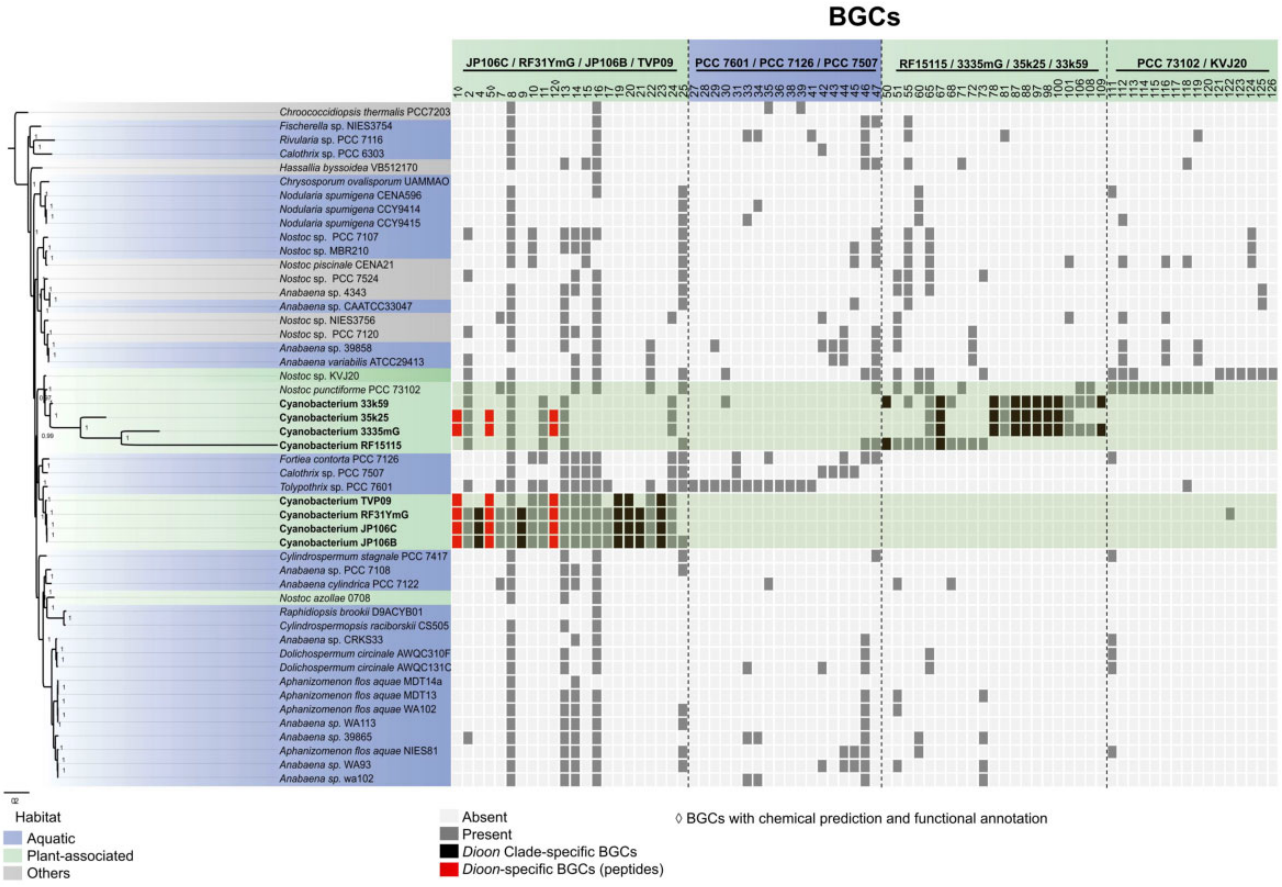

Fig. 3.—Phylogenomic tree of selected Nostocales species and their Niche-specific BGCs. Tree reconstructed with 198 conserved proteins to place the eight cyanobacterial genomic sequences, JP106B, JP106C, TVP09, RF31YmG plus RF15115, 33k59, 35k25, and 3335mG. Habitats for each species are indicated with colored bullets. The closely related microorganisms showing a sign of metabolic specialization, are highlighted with a green box. BGCs that are *Dioon* clade-specific (black boxes) and *Dioon*-specific (red boxes), with chemical structure prediction and functional annotation (◊) are shown.

similar topologies, with only two exceptions: nitrogen metabolism, consisting of only one regulatory protein, and virulence disease and defense (supplementary fig. s5, Supplementary Material online). Thus, even in the absence of a pending detailed taxonomic description of our cyanobacterial isolates, we can conclude that 1) these cycad endophytes belong to the nitrogen-fixing endophytic Nostocales clade; 2) they are closely related despite their different *Dioon* host species; and 3) within *Dioon*, there are two phylogenetic groups that include divergent microorganisms. To gain functional insights from these evolutionary patterns we did an in-depth metabolic annotation as described next.

Identification of BGCs in the Sub-community Metagenomes Suggests Metabolic Specialization of *Dioon* Cyanobacteria

Genome mining of specialized metabolites with an emphasis in all of the genomic sequences from endophythes of the coralloid roots (JP106C, TVP09, JP106B, RF15115, 35k25, 33k59, and 3335mG) plus five taxonomically closely related microorganisms identified in the previous section (KVJ20, PCC 73102, PCC 7126, PCC 7507, and PCC 7601), unveiled a total of 248 antiSMASH hits (supplementary table s8, Supplementary Material online). A total of 77 nonredundant and well-defined BGCs (supplementary table s9, Supplementary Material online) were used to analyze the distribution of these loci (fig. 3). This analysis shows that most of these BGCs are uniquely present within plant-associated cyanobacterial genomes, including 43 BGCs contributed solely by the *Dioon*-specific cyanobacteria. Along these lines, we found that only four of these 77 BGCs are fully conserved BGCs present in all cyanobacteria. These include BGC 13, the well-characterized heterocyst glycolipid (Soriente et al. 1992); a ladderane lipids BGC (Sinninghe Damsté et al. 2002), termed BGC 16; and BGC 8 and 14, putatively directing the synthesis of a terpene and a bacteriocin–lantipeptide, respectively. The remaining BGCs could be overall annotated as five systems coding for NRPSs, of which some may contain additional PKSs (32%), bacteriocins (32%), terpenes (8%), lantipeptides (4%), and indoles (3%). Moreover, sixteen BGCs (21%) were classified as Others.

It should be noted that from the BGCs analyzed, 23% are only present in the *Dioon* cyanobacteria isolates, including three *Dioon*-specific BGCs (1, 5, and 12) and 15 *Dioon*-clade-specific BGCs (4, 9, 19, 20, 21, 23, 50, 67, 78, 87, 88, 97, 98, 100, and 109), which are only present in cycads with coralloid roots and in the liverworth *Blasia* (Liaimer et al. 2016). Remarkably, microorganisms taxonomically closer to our strains, such as *N. punctiforme* PCC 73102 isolated from an Australian cycad of the genus *Macrozamia*, whose genome has been recently resequenced (Moraes et al. 2017), shows a reduced biosynthetic potential. Thus, we focus on the biosynthetic systems that are *Dioon*-specific, which have NRPS or PKS components, facilitating the prediction of their chemical structure. As mentioned, these include three BGCs exclusively found in our strains, that is, BGCs 1, 5, and 12 (supplementary fig. s6, Supplementary Material online). The overall sequence identity between these BGCs, which may provide examples of specialized metabolites with functional implications within the coralloid root, is of 85%, 89%, and 87%, respectively.

First, BGC 1, predicted to encode a bacteriocin, shares 53% of sequence identity with the biosynthetic genes of a bacteriocin isolated from the cyanobacteria *Pleurocapsa minor*. BGC 1 includes two key domains: 1) the YcaO-like domain, involved in the peptidic biosynthesis of azoline, and 2) the cyclodehydratase domain, which could be a cyclodehydratase involved in the synthesis of a thiazole/oxazole-modified microcins (Dunbar et al. 2012). The sequence similarities detected are relevant as *P. minor* forms an intertwined community with *Calothrix* spp. (Paerl 2012). Second, BGC 12 may code for an assembly line capable of synthesizing an N-terminal acylated hexapeptide with several modifications, such as the epimerization of four of its residues, the N-acylation of its second amidic bond, and the reduction of its C-terminal end to yield an aldehyde group. The N and C terminal modifications on this peptide would resemble small peptide aldehyde protease inhibitors, which have been previously reported on cyanobacteria (Fewer et al. 2013). Alternatively, the product of BGC 12 may be a siderophore, as iron-related genes were found next to the NRPS coding-genes, including a reductase domain implicated in the synthesis of iron chelators, for example, myxochelin biosynthesis (Li et al. 2008).

Third, BGC 5, which codes for a NRPS system putatively directing the synthesis of a tripeptide consisting of valine, isoleucine and tyrosine residues. This prediction also includes informative modifications, such as N-terminal acylation, N-methylation at an amide bond of the isoleucine residue, and modification of the tyrosine residue by a domain of unknown function. A search for peptides containing the abovementioned modifications, performed with the server PRISM with the feature for dereplication of known chemical structures (Skinnider et al. 2015), directed our attention to nostoginins, specialized metabolites originally isolated from *Nostoc* spp. (Ploutno and Carmeli 2002). Interestingly, the biosynthetic pathway of this natural product remains to be identified, but it is known that nostoginins A and B, as well as their congeners microginins FR1/SD755 and oscillaginins A/B (Sano and Kaya 1997), have protease inhibitory activity. In agreement with this observation, BGC 5 includes a peptidase-coding gene (supplementary fig. s6, Supplementary Material online).

It is clear from these results that the *Dioon* cyanobionts have distinctive metabolic capabilities, perhaps as a result of the co-evolution of these microorganisms within a diverse microbiome present in the coralloid root. In analogy with the pairwise interdependencies reported for *Pleurocapsa* and *Calothrix* (Paerl 2012), it is tempting to speculate about similar cases occurring within the *Dioon* coralloid roots microbiome. The predicted functions may be relevant for such microbial interactions, in addition to the interactions with its *Dioon* host species. As a first step towards the biochemical and functional characterization of our strains, with regards to the occurrence and role(s) of specialized metabolites, we focused in the cyanobacteria–*Caulobacter* putative interaction.

Identification of *Dioon*-specific *Caulobacter* BGCs

During the course of our investigations, we consistently found *Caulobacter* species associated with cyanobacteria throughout our various methods, even when efforts to obtain completely axenic cyanobacterial cultures were carried out. In contrast, *Caulobacter* strains, whose genomes were sequenced from monocultures, could be isolated relatively easy. We isolated and sequenced two *Caulobacter* strains, D5 and D4A, from the sub-community co-cultures of *D. edule* (*t1*). The genome sequences of these microorganisms show similarity with *Caulobacter* sequences extracted from the metegenomic sequence 35k35. The *Caulobacter* phylogenomic tree, reconstructed using 82 proteins from 22 *Caulobacter* genomes and *H. jannaschiana* VP2 as outgroup, places both of our strains together with *Caulobacter vibrioides* LNIY01 forming a monophyletic clade. Little is known about this latter microorganism other than the fact that it was isolated from a selenium mining area, and that it may be capable of producing the phytohormone 3-indoleacetic acid (Wang et al. 2016).

After genome mining for the prediction of BGCs, the genus *Caulobacter* shows reduced biosynthetic potential, with only 90 BGCs from all 22 genomes. All the BGCs could be overall annotated as six systems coding for NRPSs, of which some may contain additional PKSs (4%), bacteriocins (37%), terpenes (20%), lassopeptides (19%), H-serlactone (14%), and fatty acid-saccharides (6%). We detected a BGC, classified as Other, that is present in the monophyletic clade but absent from all other *Caulobacter* genomes (fig. 4*A*). This BGC, termed *Caulobacter* BGC 1, shows an overall 80%

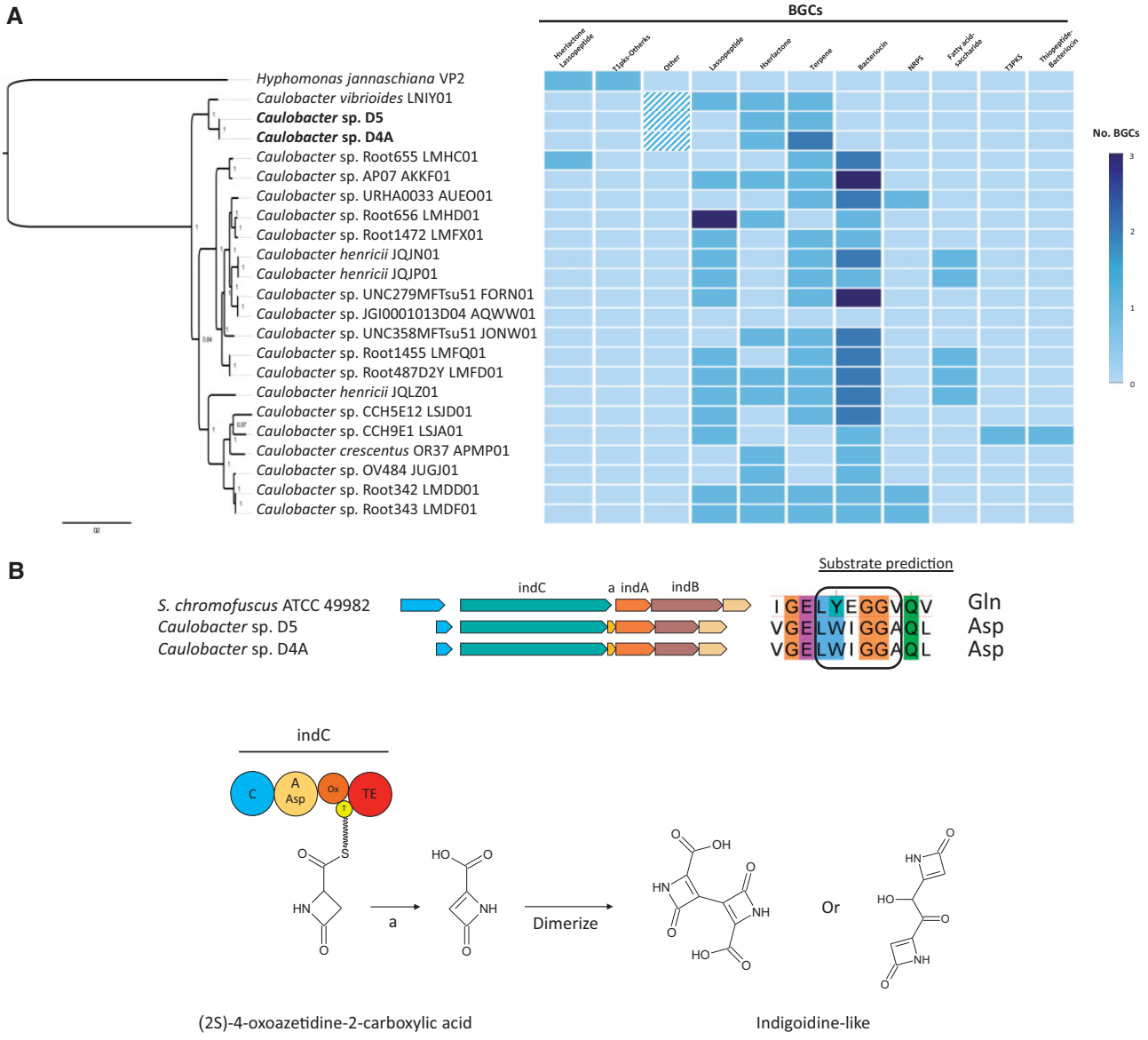

Fig. 4.—Phylogenomic tree of *Caulobacter* species and their BGCs distribution. (*A*) Tree reconstructed with 82 conserved proteins to place our two *Caulobacter* strains, D5 and D4A, showing a sign of metabolic specialization. (*B*) Chemical structure prediction and functional annotation of the aspartic acid-derived indigoidine-like metabolite.

sequence similarity with respect to the indigoidine BGC, encoding a blue pigment described previously in various actinobacterial species, including *Streptomyces chromofuscus* (Yu et al. 2013), *Streptomyces aureofaciens* (Novakova et al. 2010), *Streptomyces lavendulae* (Kurniawan et al. 2014); but also in the distantly related Gammaproteobacteria, and plant pathogen, *Erwinia chrysanthemi* (Reverchon et al. 2002) and in the Alphaproteobacteria, and marine bacterium, *Phaeobacter* sp. strain Y4I (Cude et al. 2012).

Comparative analysis between *Caulobacter* BGC 1 and the indigoidine BGC shows similar organization, including four biosynthetic genes, one of which is an NRPS, two are transporter genes and one is a regulatory gene. A key difference was found in the adenylation domain of the NRPS, for which we predicted it to recognize an aspartic acid to form (2S)-4-oxoazetidine-2-carboxylic acid, instead of a glutamine residue as in *bona fide* indigoidine (Reverchon et al. 2002). Additionally, *Caulobacter* BGC 1 has a gene coding for an enoyl reductase domain, which may form a double bond in the oxoazetidine ring. This compound could dimerize to form an indigoidine-like compound (fig. 4*B*). There are no reports of this predicted indigoidine-like metabolite in other bacteria,

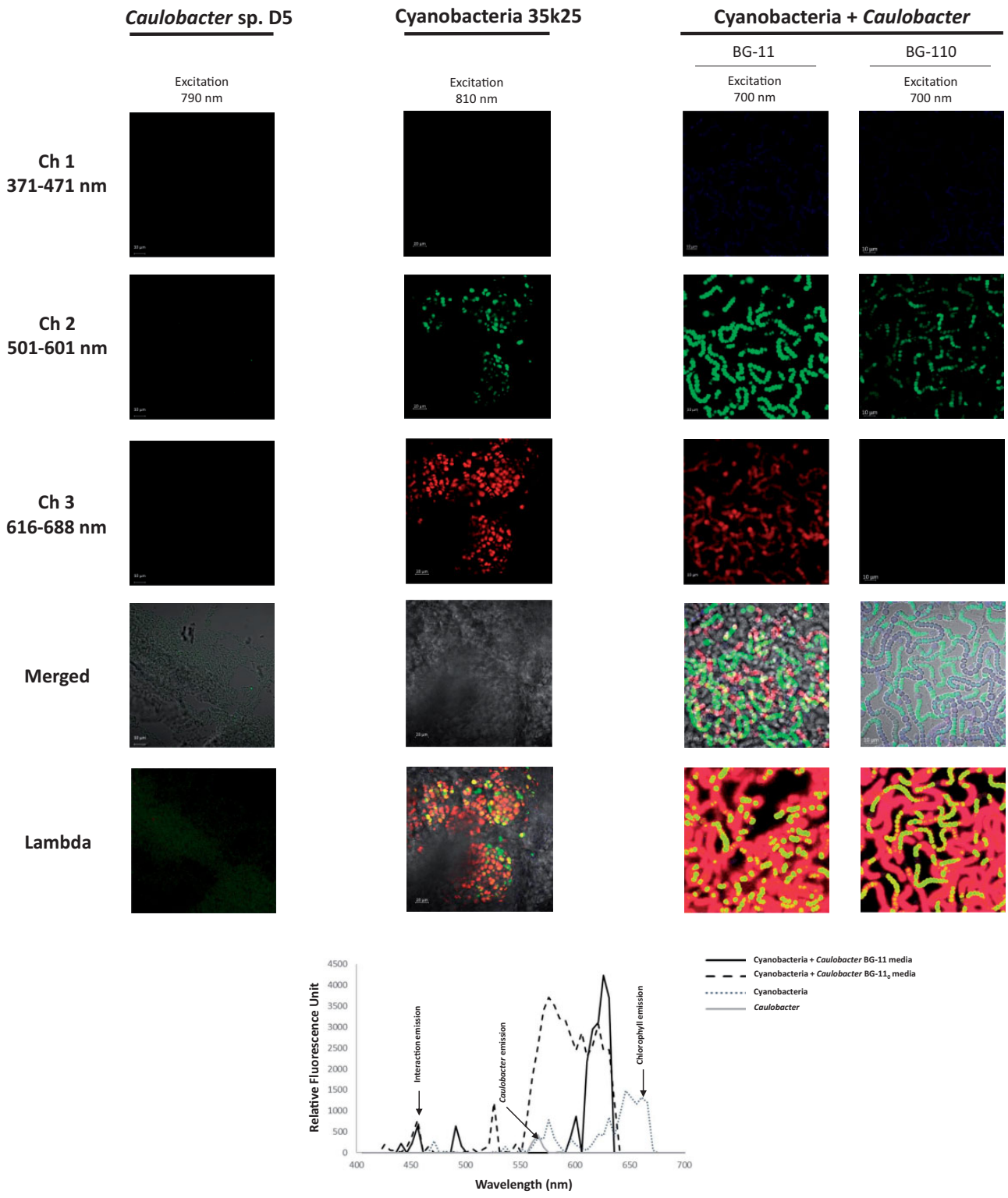

Fig. 5.—Pigment emissions in vitro of the Cyanobacteria–*Caulobacter* association. *Caulobacter* sp. D5 and Cyanobacterium sp. 33k25 microcommunity, grown independently on plates containing BG-11$_0$ medium, and together on plates containing BG-11$_0$ and BG-11 media. *Caulobacter* sp. D5, by its own, showed a unique and relevant pigment emission between 550 and 570 nm (green light). Cyanobacterium 33k25, showed three emissions corresponding to chlorophyll (650–720 nm, red light), the *Caulobacter* species (550–570 nm, green light) and a signal proposed to be related to the predicted indigoidine-like metabolite (430–480 nm, blue light). In both BG11 and BG-11$_0$ synthetic bacterial co-cultures, three emission signals were detected corresponding to chlorophyll (650–720 nm), the *Caulobacter* species (550–570 nm), and remarkably, an increase of the signal proposed to be related with the indigoidine-like metabolite (430–480 nm). Refer to the text for further details.

although synthesis of related congeners with antibacterial activity has been reported (O'Dowd et al. 2008). The distribution of the indigoidine BGC in divergent bacterial lineages, suggests horizontal gene transfer to fulfill biological functions yet-to-be determined.

Multiphotonic Microscopy Suggests That Cyanobacteria–*Caulobacter* Association May Be Mediated by Specialized Metabolites

To test if our *Caulobacter* strains co-exist with our cyanobacterial strains, pigment emissions based in multiphotonic microscopy were obtained (fig. 5). For these experiments, microbial cultures of *Caulobacter* sp. D5 and Cyanobacterium sp. 35k25, both isolated from *D. edule*, were used. Three different emissions were detected, Ch1 (371-471nm), Ch2 (501-601nm), and Ch3 (616–688 nm), after exciting samples at different excitation energies (700–850 nm). To dissect these emissions, we first observed the characteristic emission signals when each sample or strain was cultivated separately. The axenic culture of *Caulobacter* sp. D5 showed a unique and relevant metabolic emission between 550–570 nm (green light). In contrast, for the *Cyanobacterium* sp. 33k25 sample, in agreement with the fact that it is a microcommunity rather than a monoculture, three emissions were detected corresponding to chlorophyll (650–720 nm, red light), the above mentioned *Caulobacter* species (550–570 nm, green light), and a small signal located in the range of blue light (430–480 nm, blue light). We suggest that the latter emission is related to the presence of the indigoidine-like compound predicted after genome mining, but this remains to be experimentally confirmed.

To further investigate into the suggested association between *Caulobacter* and the cyanobacteria, potentially mediated by the predicted indigoidine-like metabolite, strains D5 and 33k25 were cultivated together in plates containing BG11 and BG-11_0 media. As expected from previous results, irrespective of the media, the synthetic bacterial co-culture emitted signals in the same wavelengths as chlorophyll (650–720 nm), the *Caulobacter* species (550–570 nm) and that suggested to have a relationship with the indigoidine-like metabolite (430–480 nm). Interestingly, the intensity recorded for each signal, in relative fluorescence units, was different for the two different media used. Specifically, the chlorophyll emission decreased in BG-11_0, which may be due to the large number of heterocyst formed under nitrogen limitation. However, the emission intensity of the blue signal remained the same in both conditions, but became notably greater in the synthetic bacterial co-culture. These results suggest that the interaction between the cyanobacteria contained in 33k25 and the *Caulobacter* spp. induces the production of specialized metabolites.

Discussion

Although our findings are congruent with recent cycad taxonomic diversity reports (Zheng et al. 2018; Suarez-Moo et al. 2018), our focus was to investigate the relationship between key taxa of the coralloid roots symbiotic communities and provide information on their possible functional roles. Our results support previous studies in which different species of cycads host multiple shared Nostocales cyanobacteria (Zimmerman and Rosen 1992; Zheng et al. 2002; Gehringer et al. 2010; Yamada et al. 2012), yet the monophyletic placement of our cyanobacterial samples clearly indicates a different evolutionary trajectory of *Nostoc* species associated with *Dioon* from their free-living counterparts. This phylogenetic separation is only clear when the whole genome is taken into account, as fewer genetic markers do not seem to have enough resolution to resolve the clade (Yamada et al. 2012). For instance, free-living *Nostoc* PCC 7120 grouped distantly to strains of symbiotic or plant-associated origin, whereas the also distant genome reported as *Nostoc azollae* is now taxonomically *Trichormus azollae* (Baker et al. 2003). Similarly, other noncycad plant-associated *Nostoc* are more related to *Anabaena* and *Aphanizomenon*, congruent with previous reports (Papaefthimiou et al. 2008).

The newly isolated cyanobacterial strains are part of a clade that includes symbiotic species able to colonize specialized nitrogen-fixing cavities through plant tissues or inside coralloid roots. This observation is congruent with previous phylogenies that include hormogonia-producing symbiotic species as clustered together (Gagunashvili and Andrésson 2018; Warshan et al. 2018). The ability of different strains isolated from one plant species of infecting phylogenetically distant hosts, as previously reported (Johansson and Bergman 1994; Papaefthimiou et al. 2008; Paerl 2012), remains to be tested for the *Dioon*-specific cyanobacterial strains. We hypothesize that species within this phylogenetic clade could cross-infect different plant hosts, although their metabolic capacities and contributions to each cycad genus is likely to vary given their unique BGCs. This idea is consistent with previous cross-infection reports with the cycad symbiont *N. punctiforme* PCC 73102 and the bryophyte *Anthoceros* (Enderlin and Meeks 1983), but goes deeper into possible mechanisms mediated by specialized metabolites.

Differences in genera identified with 16S rRNA and shotgun metagenomics of sub-communities are explained in part due to the biases in 16S rRNA markers, which can erroneously diagnose the taxonomic composition of endophytes in the cycad host (Louca et al. 2018). These differences can also be explained by other scenarios: 1) rare groups present in low abundance can only be recovered in sub-community co-cultures on which they increase in biomass; 2) some microorganisms are fast growers irrespective of media, and will dominate in OTUs, simply by chance; 3) some groups are more media-specific; and/or 4) groups in BG-11_0 (*t1*) are

recovered as a result of functional interactions in cyanobacteria-associated groups. Also, metagenomic data biases could be due to the type of DNA extraction protocol, which has been shown to confound diversity assessment of endophytic communities (Maropola et al. 2015).

The long-term one-year co-culture (*t2*) allowed us to explore at least some of these possibilities, as the initial amount of nitrogen available in these co-cultures became a limiting factor over time. Hence, the establishment of stable sub-communities after one year (with emerging and surviving taxa) suggests that autotrophic, nitrogen fixation and photosynthesis, may be a driving force in the assembly of the coralloid root community. On the basis of this observation it is possible to set out comparative and systematic analyses of the diazotrophic potential of these communities. Coupling the metagenomic analysis with traditional microbiology increased our ability to accurately identify interacting endophytes in the sub-communities, and may help us characterize microbial life strategies and their phylogeny, a challenge that has received increasing attention in complex systems (Ho et al. 2017).

Notably, genome mining of the cyanobacterial strains revealed three BGCs unique to the *Dioon*-specific cyanobionts, in agreement with the hypothesis of a specialized community driven by metabolic adaptation regardless of the taxonomic plasticity. Of the three BGCs found in our isolates, BGC 12 and BGC 5 putatively coding for small peptide protease inhibitors, or a siderophore, could have a functional bearing on the evolution and biology of the *Dioon*-bacteria symbiosis. Peptide production may imply that proteolysis is involved in the cyanobacteria–cycad interaction, linked to the reconfiguration of the root architecture or the filtering of the microbiome, as it has been observed in arbuscular mycorrhiza and legumes (Takeda et al. 2007). A starting point to explore this possibility would be to confirm the predicted gene-metabolite relationship proposed here, by sequencing the genomes of nostoginin-producing microorganisms (Ploutno and Carmeli 2002) together with their functional characterization. If a siderophore, BGC 12 could be mediating bacterial community interactions as previously reported (Cruz-Morales et al. 2017), and/or interactions with the host, adding to the growing notion that symbiotic relations occur under heavy influence of metabolic exchange.

The association between *Caulobacter*, a small Alphaproteobacterium adapted to nutrient-poor conditions (Landt et al. 2010), and the *Dioon* cyanobacterial endophyte strains, is also remarkable. Both our two *Caulobacter* isolates are able to produce a specific specialized metabolite, which could mediate the association between either microorganisms or their host. Hence, the *Caulobacter*-Nostocales consortium could be playing a previously unknown role inside the coralloid roots of cycads by producing few but specific natural products, making them specialized metabolites. The pigment emission detected using multiphotonic microscopy, a novel approach in many ways, opens the possibility of looking into these presumed interactions *in planta*, and warrants more detailed chemical analysis of the metabolic exchange occurring within the microbiome of the cycad coralloid root.

In sum, our study confirmed the broad taxonomic bacterial diversity of the coralloid root. Monophyletic Nostocales with at least three unique BGCs and an associated *Caulobacter* with a unique BGC, could be a sign of the constraints imposed by the transition from a free-living lifestyle, perhaps initially even as a bacterial consortium outside the plant, into a symbiotic lifestyle inside a plant's specialized organ. As data are gathered from more genomes of bacterial cycad symbionts isolated from different environments of the rhizosphere-coralloid root ecological niche, it will be possible to test for deeper co-evolutionary relationships. Genomes from other endophytic genera identified here, could inform the evolutionary and ecological nature of cycad–bacterial and bacterial–bacterial interactions, which in turn could direct detailed functional characterization of these ancient plant specialized communities.

Supplementary Material

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

Acknowledgments

This study was supported by CONACyT #169701 and FON.INST./265/2016 (SA-FC2015-2-901) to A.C.J. CONACyT #179290 and #177568 to F.B.G. We acknowledge Juan Palacios and Antonio Hernández for help during field collections, as well as Flor Zamudio for technical support.

Literature Cited

Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9(1):75.

Bachmann BO, Ravel J. 2009. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. In: Hopwood, DA, editor. Complex enzymes in microbial natural product biosynthesis, Part A: Overview articles and peptides, vol. 458, methods in enzymology, Elsevier: Academic Press. p. 181–217.

Baker JA, Entsch B, McKay DB. 2003. The cyanobiont in an *Azolla* fern is neither *Anabaena* nor *Nostoc*. FEMS Microbiol Lett. 229(1):43–47.

Baulina OI, Lobakova ES. 2003. Atypical cell forms over producing extracellular substances in populations of cycad cyanobionts. Microbiology 72(6):701–712.

Berg KA, et al. 2009. High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. ISME J. 3(3):314–325.

Bergersen FJ, Kennedy GS, Wittmann W. 1965. Nitrogen fixation in the coralloid roots of *Macrozamia communis* L. Johnson. Aust J Bot. 18(6):1135–1142.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30(15):2114–2120.

Bunt JS. 1961. Blue-Green Algae: growth. Nature 192(4809):1274.

Cruz-Morales, P, et al. 2017. Actinobacteria phylogenomics, selective isolation from an iron oligotrophic environment and siderophore

functional characterization, unveil new desferrioxamine traits. FEMS Microbiol Ecol. 93(9).
Cibrián-Jaramillo A, Barona-Gómez F. 2016. Increasing metagenomic resolution of microbiome interactions through functional phylogenomics and bacterial sub-communities. Front Genet. 7:4.
Costa JL, Paulsrud P, Lindblad P. 1999. Cyanobiont diversity within coralloid roots of selected cycad species. FEMS Microbiol Ecol. 28(1):85–91.
Costa JL, Romero EM, Lindblad P. 2004. Sequence based data supports a single *Nostoc* strain in individual coralloid roots of cycads. FEMS Microbiol Ecol. 49(3):481–487.
Cuddy WS, Neilan BA, Gehringer MM. 2012. Comparative analysis of cyanobacteria in the rhizosphere and as endosymbionts of cycads in drought-affected soils. FEMS Microbiol Ecol. 80(1):204–215.
Cude WN, et al. 2012. Production of the antimicrobial secondary metabolite indigoidine contributes to competitive surface colonization by the marine roseobacter *Phaeobacter* sp. strain Y4I. Appl Environ Microbiol. 78(14):4771–4780.
De Luca P, Sabato S, Torres MV. 1981. *Dioon merolae* (*zamiaceae*), a new species from Mexico. Brittonia 33(2):179–185.
Dunbar KL, Melby JO, Mitchell DA. 2012. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. Nat Chem Biol. 8(6):569–575.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.
Enderlin CS, Meeks JC. 1983. Pure culture and reconstitution of the *Anthoceros-Nostoc* symbiotic association. Planta 158(2):157–165.
Fewer DP, et al. 2013. New structural variants of aeruginosin produced by the toxic bloom forming cyanobacterium *Nodularia spumigena*. PLoS ONE 8(9):1–10.
Foster RA, Goebel NL, Zehr JP. 2010. Isolation of *Calothrix rhizosoleniae* (Cyanobacteria) strain Sc01 from *Chaetoceros* (Bacillariophyta) spp. diatoms of the subtropical North Pacific Ocean. J Phycol. 46(5):1028–1037.
Gagunashvili AN, Andrésson ÓS. 2018. Distinctive characters of *Nostoc* genomes in cyanolichens. BMC Genomics 19(1):434.
Gehringer MM, et al. 2010. Host selection of symbiotic cyanobacteria in 31 species of the Australian cycad genus: *macrozamia* (*Zamiaceae*). Mol Plant-Microbe Interact. 23(6):811–822.
Grilli Caiola M. 1980. On the phycobionts of the cycad coralloid roots. New Phytol. 85(4):537–544.
Grobbelaar N, Scott W, Hattingh W, Marshall J. 1987. The identification of the coralloid root endophytes of the Southern African cycads and the ability of the isolates to fix dinitrogen. S Afr J Bot. 53(2):111–118.
Grove TS, O'Connell AM, Malajczuk N. 1980. Effects of fire on the growth, nutrient content and rate of nitrogen fixation of the cycad *Macrozamia riedlei*. Aust J Bot. 28(3):271–281.
Gutiérrez-Ortega, JS, et al. 2017. The phylogeography of the cycad genus Dioon (Zamiaceae) clarifies its Cenozoic expansion and diversification in the Mexican transition zone. Ann Bot. 121(3), 535–548.
Halliday J, Pate JS. 1976. Symbiotic nitrogen fixation by coralloid roots of the cycad *Macrozamia riedlei*: physiological characteristics and ecological significance. Aust J Plant Physiol. 3:349–358.
Ho A, Di Lonardo DP, Bodelier PL. 2017. Revisiting life strategy concepts in environmental microbial ecology. FEMS Microbiol Ecol. 93(3), fix006.
Huang TC, Grobbelaar N. 1989. Isolation and characterization of endosymbiotic *Calothrix* (Cyanophyceae) in *Encephalardtos hildebrandtii* (Cycadales). Phycologia 28(4):464–468.
Huerta-Cepas J, Dopazo J, Gabaldón T. 2010. ETE: a python environment for tree exploration. BMC Bioinformatics 11:24.
Johansson C, Bergman B. 1994. Reconstitution of the symbiosis of *Gunnera manicata* Linden: cyanobacterial specificity. New Phytol. 126(4):643–652.
Kanesaki Y, et al. 2018. Draft genome sequence of the nitrogen-fixing and hormogonia-inducing cyanobacterium Nostoc cycadae strain WK-1, isolated from the coralloid roots of Cycas revoluta. Genome Announc. 6(7): e00021–18.
Kurniawan YN, Kitani S, Maeda A, Nihira T. 2014. Differential contributions of two SARP family regulatory genes to indigoidine biosynthesis in *Streptomyces lavendulae* FRI-5. Appl Microbiol Biotechnol. 98(23):9713–9721.
Landt SG, Lesley JA, Britos L, Shapiro L. 2010. CrfA, a small noncoding RNA regulator of adaptation to carbon starvation in *Caulobacter crescentus*. J Bacteriol. 192(18):4763–4775.
Lane DJ. 1991. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic Acid Techniques in Bacterial Systematics. John Wiley and Sons: West Sussex. p. 115–175.
Lázaro-Zermeño JM, González-Espinosa M, Mendoza A, Martínez-Ramos M. 2012. Historia natural de *Dioon merolae* (Zamiaceae) en Chiapas, México. Bot Sci. 90(1):73–87.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760.
Li Y, Weissman KJ, Müller R. 2008. Myxochelin biosynthesis: direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. J Am Chem Soc. 130(24):7554–7555.
Liaimer A, Jensen JB, Dittmann E. 2016. A genetic and chemical perspective on symbiotic recruitment of cyanobacteria of the genus Nostoc into the host plant *Blasia pusilla*. Front Microbiol. 7:1693.
Lobakova ES, Orazova MK, Dobrovolskaya TG. 2003. Microbial complexes occurring on the apogeotropic roots and in the rhizosphere of cycad plants. Microbiology 72(5):628–633.
Lobakova ES, Orazova MK, Dobrovolskaya TG. 2003. The structure of cyanobacterial communities formed during the degradation of apogeotropic roots of cycads. Microbiology 72(5):634–637.
Louca S, Doebeli M, Parfrey LW. 2018. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome 6(1):41.
Maropola MKA, Ramond JB, Trindade M. 2015. Impact of metagenomic DNA extraction procedures on the identifiable endophytic bacterial diversity in *Sorghum bicolor* (L. Moench). J Microbiol Methods. 112:104–117.
Massalha H, Korenblum E, Tholl D, Aharoni A. 2017. Small molecules below-ground: the role of specialized metabolites in the rhizosphere. Plant J. 90(4):788–807.
Meeks JC. 2009. Physiological adaptations in nitrogen-fixing nostoc–plant symbiotic associations. Berlin, Heidelberg (Germany): Springer Berlin Heidelberg, p. 181–205.
Moraes LE, et al. 2017. Resequencing and annotation of the *Nostoc punctiforme* ATTC 29133 genome: facilitating biofuel and high-value chemical production. AMB Express 7(1):42.
Narechania A, et al. 2012. Random addition concatenation analysis: a novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. Genome Biol Evol. 4(1):30–43.
Newton JW, Herman AI. 1979. Isolation of cyanobacteria from the aquatic fern, *Azolla*. Arch Microbiol. 120(2):161–165.
Norstog KJ, Nicholls TJ. 1997. The biology of the cycads. Ithaca, NY: Cornell University Press. p. 104–112.
Novakova R, Odnogova Z, Kutas P, Feckova L, Kormanec J. 2010. Identification and characterization of an indigoidine-like gene for a blue pigment biosynthesis in *Streptomyces aureofaciens* CCM 3239. Folia Microbiol. 55(2):119–125.
O'Dowd H, et al. 2008. Novel antibacterial azetidine lincosamides. Bioorgan Med Chem Lett. 18(8):2645–2648.
Ow MC, Gantar M, Elhai J. 1999. Reconstitution of a cycad–cyanobacterial association. Symbiosis 27(2):125–134.
Papaefthimiou D, Van Hove C, Lejeune A, Rasmussen U, Wilmotte A. 2008. Diversity and host specificity of *Azolla* cyanobionts. J Phycol. 44(1):60–70.

Paerl HW. 2012. Marine Plankton. In: Whitton BA, editor. Ecology of cyanobacteria II: their diversity in space and time. Dordrecht: Springer. p. 127–153.
Ploutno A, Carmeli S. 2002. Modified peptides from a water bloom of the cyanobacterium *Nostoc* sp. Tetrahedron 58(50):9949–9957.
Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14(9):817–818.
Quast C, et al. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41(D1):D590–D596.
Reverchon S, Rouanet C, Expert D, Nasser W. 2002. Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. J Bacteriol. 184(3):654–665.
Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542.
Sano T, Kaya K. 1997. A 3-amino-10-chloro-2-hydroxydecanoic acid-containing tetrapeptide from *Oscillatoria agardhii*. Phytochemistry 44(8):1503–1505.
Scherlach K, Hertweck C. 2018. Mediators of mutualistic microbe–microbe interactions. Nat Prod Rep. 35(4):303–308.
Damsté Sinninghe, et al. 2002. Linearly concatenated cyclobutane lipids form a dense bacterial membrane. Nature 419:708–712.
Skinnider MA, et al. 2015. Genomes to natural products prediction informatics for secondary metabolomes (PRISM). Nucleic Acids Res. 43(20):9645–9662.
Soriente A, Sodano G, Cambacorta A, Trincone A. 1992. Structure of the "heterocyst glycolipids" of the marine cyanobacterium *Nodularia harveyana*. Tetrahedron 48(25):5375–5384.
Storey WB. 1968. Somatic reduction in cycads. Science 159(3815):648–650.
Suarez-Moo PDJ, Vovides A, Griffith MP, Barona-Gómez F, Cibrián-Jaramillo A. 2018. Unlocking a high bacterial diversity in the coralloid root microbiome from the cycad genus Dioon. bioRxiv. 381798.
Takeda N, et al. 2007. Proteases in plant root symbiosis. Phytochemistry 68(1):111–121.
Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 56(4):564–577.
Tan LT. 2007. Bioactive natural products from marine cyanobacteria for drug discovery. Phytochemistry 68(7):954–978.
Thajuddin N, et al. 2010. Morphological and genetic diversity of symbiotic cyanobacteria from cycads. J Basic Microbiol. 50(3):254–265.
Wang Y, et al. 2016. Genome sequence of selenium-solubilizing bacterium *Caulobacter vibrioides* t5m6. Genome Announc. 4(1).
Warshan D, et al. 2018. Genomic changes associated with the evolutionary transitions of *Nostoc* to a plant symbiont. Mol Biol Evol. 35(5):1160–1175.
Weber T, et al. 2015. antiSMASH 3.0-A comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43(W1):W237–W243.
Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15(3):R46.
Yamada S, Ohkubo S, Miyashita H, Setoguchi H. 2012. Genetic diversity of symbiotic cyanobacteria in *Cycas revoluta* (Cycadaceae). FEMS Microbiol Ecol. 81(3):696–706.
Yu D, Xu F, Valiente J, Wang S, Zhan J. 2013. An indigoidine biosynthetic gene cluster from *Streptomyces chromofuscus* ATCC 49982 contains an unusual IndB homologue. J Ind Microbiol Biotechnol. 40(1):159–168.
Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18(5):821–829.
Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome Res. 16(9):1099–1108.
Zheng W, Song T, Bao X, Bergman B, Rasmussen U. 2002. High cyanobacterial diversity in coralloid roots of cycads revealed by PCR fingerprinting. FEMS Microbiol Ecol. 40(3):215–222.
Zheng Y, Chiang TY, Huang CL, Gong X. 2018. Highly diverse endophytes in roots of *Cycas bifida* (Cycadaceae), an ancient but endangered gymnosperm. J Microbiol. 56(5):337–345.
Zimmerman WJ, Rosen BH. 1992. Cyanobiont diversity within and among cycads of one field site. Can J Microbiol. 38(12):1324–1328.
Zvyagintsev DG, et al. 2010. Morphological and physiological modifications of cyanobacteria in experimental cyanobacterium-actinomycete associations. Microbiology 79(3):314–320.

Associate editor: Tal Dagan