Complex microbiome underlying secondary and primary metabolism in the tunicate-*Prochloron* symbiosis

Mohamed S. Donia[a,1], W. Florian Fricke[b], Frédéric Partensky[c,d], James Cox[e], Sherif I. Elshahawi[f], James R. White[b], Adam M. Phillippy[b,2], Michael C. Schatz[b,3], Joern Piel[g], Margo G. Haygood[f], Jacques Ravel[b], and Eric W. Schmidt[a,h,4]

[a]Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112; [b]Institute for Genome Sciences, Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201; [c]Station Biologique de Roscoff, Université Pierre et Marie Curie–Université Paris 6, 29680 Roscoff, France; [d]Oceanic Plankton Group, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7144, 29680 Roscoff, France; [e]Metabolomics Core Research Facility, University of Utah, Salt Lake City, UT 84112; [f]Division of Environmental and Biomolecular Systems, Institute of Environmental Health, Oregon Health and Science University, Beaverton, OR 97006; [g]Kekulé Institute of Organic Chemistry and Biochemistry, University of Bonn, D-53121 Bonn, Germany; and [h]Department of Biology, University of Utah, Salt Lake City, UT 84112

The relationship between tunicates and the uncultivated cyanobacterium *Prochloron didemni* has long provided a model symbiosis. *P. didemni* is required for survival of animals such as *Lissoclinum patella* and also makes secondary metabolites of pharmaceutical interest. Here, we present the metagenomes, chemistry, and microbiomes of four related *L. patella* tunicate samples from a wide geographical range of the tropical Pacific. The remarkably similar *P. didemni* genomes are the most complex so far assembled from uncultivated organisms. Although *P. didemni* has not been stably cultivated and comprises a single strain in each sample, a complete set of metabolic genes indicates that the bacteria are likely capable of reproducing outside the host. The sequences reveal notable peculiarities of the photosynthetic apparatus and explain the basis of nutrient exchange underlying the symbiosis. *P. didemni* likely profoundly influences the lipid composition of the animals by synthesizing sterols and an unusual lipid with biofuel potential. In addition, *L. patella* also harbors a great variety of other bacterial groups that contribute nutritional and secondary metabolic products to the symbiosis. These bacteria possess an enormous genetic potential to synthesize new secondary metabolites. For example, an antitumor candidate molecule, patellazole, is not encoded in the genome of *Prochloron* and was linked to other bacteria from the microbiome. This study unveils the complex *L. patella* microbiome and its impact on primary and secondary metabolism, revealing a remarkable versatility in creating and exchanging small molecules.

ascidian | metagenomics | photosynthesis | natural products

Ascidians, a class of tunicates (phylum: Chordata), are found in marine ecosystems worldwide, and some (e.g., *Ciona intestinalis*) are model organisms as "invertebrate" chordates. Within this large group, ascidians of the family Didemnidae are abundant components of shallow marine environments (Fig. 1) (1, 2). Didemnid ascidians comprise colonies of as many as thousands of individual zooids that filter seawater, consuming bacteria and other small particles (3). Zooids are embedded in a common tunic, which is composed largely of cellulose, and despite maintaining all the necessary organs for individual survival, they coordinate many of their functions (4). For example, zooids expel the filtered seawater through shared cavities. Colonies can fuse and divide in the course of days, and they exhibit coordinated movement of the tunic (5–7). In environments such as seagrass beds and coral reefs, didemnid ascidians are often major components of the ecosystem, and they have been shown to play roles in the environment such as smothering scleractinian corals (8, 9), potentially impacting reef structure. Didemnids have been extensively studied for their unique biology and also because they contain a rich array of natural products that have found clinical and preclinical applications (10, 11).

The survival of many didemnids directly depends upon the presence of symbiotic bacteria that fulfill primary and secondary metabolic functions. However, very few ascidians have been analyzed for their microbiological content. A particularly well studied example is a subset of didemnid ascidians obligately associated with *Prochloron didemni* and related cyanobacteria (Fig. 1) (12, 13). Although in these symbiotic associations the ascidians still filter-feed, *P. didemni* provides a large percentage (as much as 100% in some cases) of required carbon via photosynthesis (14–17). In addition, experimental evidence indicates that *P. didemni* participates in nitrogen recycling, and possibly even in obtaining nitrogen from nitrate or dinitrogen (12, 17, 18). An intriguing feature of *P. didemni* is the presence of chlorophyll (chl) *b*, like in plants, a quite unusual trait for a cyanobacterium (19, 20). The presence of both chl*a* and chl*b* extends the *P. didemni* light-harvesting capacity into the blue and orange/red regions of the visible light spectrum, which is useful in the shallow-water habitat of the host ascidians (12). *P. didemni* thus allows didemnid ascidians to thrive in nutrient-poor, light-rich environments of the tropics. Didemnids have also been found to contain other types of microbes, but unlike *P. didemni*, without extensive in vivo functional studies (21–29).

Aside from the use of ascidians as model organisms for studying eukaryotic biology and symbiosis, ascidians have also been extensively surveyed because they commonly contain secondary metabolites with pharmaceutical potential. Approximately 1,000 structurally diverse, bioactive natural products have been isolated from ascidians, including many from didemnid ascidians (10). Many of these compounds are thought to be produced by symbiotic bacteria, based in a few cases on molecular evidence. For example, it has been firmly established that

Fig. 1. *Prochloron*-bearing ascidians. (*A*) *L. patella* is shown at center, flanked by two other *Prochloron*-bearing species. (*B*) *D. molle*. (*C*) Light micrographs showing approximately 10-μm-diameter *P. didemni* cells examined in this study.

cyanobactin natural products, which are the major small molecules found in many didemnid ascidians, are actually produced by symbiotic *P. didemni* (30–32), and recently the producer of ecteinascidin-743 has been identified (33). The cyanobacterial symbiont also has the genetic capacity to produce a large number of other secondary metabolites, leading us to speculate that *Prochloron* was probably responsible for making most or all the potently bioactive natural products found in *Prochloron*-containing ascidians (34). However, the origin of the vast majority of "ascidian" natural products has yet to be studied. Clearly, *P. didemni* cannot synthesize most of them, as the genus *Prochloron* only occurs in a small subset of Ascidiacea.

To examine these problems in detail, we have been engaged in an effort to sequence *P. didemni* and the *L. patella* metagenome. Initially, we analyzed sample L1 collected in Palau in 2002 and have reported progress toward sequencing the *P. didemni* genome therein (which we named P1) (30, 34, 35). Since that time, we have also studied the *P. didemni* fractions P2 and P3 from the two metagenomic samples L2 and L3 collected in Fiji and the Solomon Islands in 2006. Preliminary analysis focusing only on secondary metabolism was previously published and showed that *P. didemni* populations within each ascidian host seemed to share the same genotype and that secondary metabolite gene clusters were among the main variable features between otherwise remarkably syntenic genomes from different ascidian hosts (34, 36).

Here, we report the final assembly, functional annotation, and detailed analysis and comparison of P1 and of the P2 and P3 draft genome sequences from the L2 and L3 metagenomic assemblies, with the aim to better understand the host/symbiont relationship. In addition, we present the newly sequenced P4 genome from *L. patella* collected in Papua New Guinea in 2007. We aimed to determine which genomic features are responsible for the unusual primary metabolic integration of host and bacteria, as well as to better define the largely unknown biosynthetic origin of marine natural products. We therefore used a comprehensive ascidian microbiome analysis that, besides deep metagenome sequencing, included amplicon sequencing, microscopy, cultivation, chemical analysis, and other methods. We report that, in addition to *P. didemni*, a significant diversity of other bacterial strains contributes to the observed chemical diversity in ascidian metagenomes. Taken together, these results explain the genomic basis of metabolic integration in the *L. patella* symbiotic model system and led to the discovery of many biosynthetic pathways to small molecules of pharmaceutical and biotechnological interest (e.g., for biofuel production).

Results

***P. didemni* Genomes.** We report the final assembly of the *P. didemni* P1 genome, consisting of 100 contigs on 32 scaffolds with a total length of 6,197,726 bp (accession no. AGRF00000000), as well as the three additional annotated draft genomes P2, P3, and P4 from metagenomic assemblies of three *L. patella* samples (L2–L4; accession nos. AFSJ00000000, AFSK00000000, and AGGA00000000). The improved *P. didemni* P1 genome assembly reported here was annotated and compared with the P2, P3, and

P4 *P. didemni* metagenome components (Fig. 2). The P1 assembly is noteworthy taking into consideration the large size of the genome, the complex bacterial diversity and host DNA present in the sample (as detailed later), and the abundance of repetitive elements as well as duplicated genes within the *P. didemni* genomes. Because most other well assembled genomes from human or insect gut metagenomes are approximately half the size of the *P. didemni* genome (37), to the best of our knowledge, this represents the largest assembled genome from an uncultivated organism.

Although *P. didemni* sequences have not been observed outside of certain symbiotic environments and *P. didemni* has resisted stable cultivation, the cyanobacterium exhibits several major differences from other described obligate symbionts (38). This includes the apparent absence of genome reduction, the presence of a full set of primary metabolic genes, and a G+C content relatively standard for free-living cyanobacteria (i.e., 41–42%). This is consistent with previous studies showing that vertical transmission is not the only way that hosts can acquire *P. didemni* (39–43). Although some host ascidians appear to require *P. didemni* for survival, our genome-sequencing data indicate that this relationship may be metabolically facultative for the bacteria. For example, despite reports that *P. didemni* may be auxotrophic for tryptophan (44), a full set of tryptophan anabolic genes was found in the *P. didemni* genomes.

Ascidian Microbiome. Beyond *P. didemni*, all sequenced samples contained DNA from other bacteria living with the ascidian and from the ascidian itself (35), which accounted for a significant portion of the total assembled metagenome in L2 (16.3 of 23.8 Mbp) and L3 (1.2 of 7.0 Mbp). Table 1 includes a summary of methods and sequencing data. The sequenced samples were obtained both by squeezing the animals to release bacteria (all samples) and by extracting DNA from whole pieces of tunic (L2). The bacteria are described in detail in ensuing sections, but, briefly, they are dominated by Proteobacteria, with other phyla contributing to varying degrees depending upon the sample.
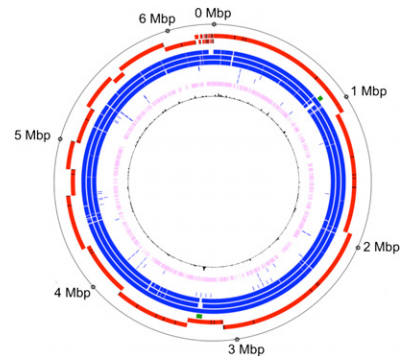
Fig. 2. The *P. didemni* genome. The P1 genome, shown as the outer red circle, is compared with genomes from different cyanobacteria (blue and pink circles). The staggering red circle fragments represent the 32 scaffolds of the P1 genome; contigs are separated by black lines. Continuous blue circles indicate regions of greater than 85% DNA sequence identity between P1 and the P4, P2, and P3 genomes (from outside to inside, respectively). Green bars show the location of insecticidal toxin at bottom and secondary metabolite gene clusters at upper right that are present in P1 and P4, but not P2 or P3. Scattered blue bars indicate regions with greater than 85% DNA sequence identity between *P. didemni* and the cyanobacteria *Cyanothece* sp. ATCC 51142 (GenBank accession no. CP000806), *Nostoc punctiforme* PCC 73102 (GenBank accession no. CP001037), and *Anabaena variabilis* ATCC 29413 (GenBank accession no. CP000117). The pink bars indicate regions of greater than 60% translated protein sequence identity between P1 and *Cyanothece* sp. ATCC51142, showing that these two cyanobacterial species share a large set of core proteins despite a lack of nucleotide sequence homology. The innermost black line shows trinucleotide composition (χ^2) of the P1 draft genome, indicating regions of possible lateral gene transfer. This analysis is biased as a result of the presence of sequence gaps.

Table 1. Summary of analytical methods used in this study

Method of analysis	Sample ID								
	L1/P1	L2/P2	L3/P3	P4	L5 05–019	06–014	06–028	06–029	07–021
16S PCR sequencing	Yes	Yes	Yes	No	Yes	No	No	Yes	No
454 sequencing reads	4942	4336	4140	—	—	—	—	—	—
Cyanobacteria	75%	85%	98%	—	—	—	—	—	—
Proteobacteria	19%	6%	1%	—	—	—	—	—	—
Actinobacteria	ND	1%	ND	—	—	—	—	—	—
Sanger sequencing reads	No	173	51	No	39	No	No	75	No
Whole metagenome sequencing	Yes	Yes	Yes	Yes	No	No	No	No	No
Sanger sequencing reads, thousands	108	No	No	No	—	—	—	—	—
454 read number, millions	1.29	2.84	0.39	1.30	—	—	—	—	—
454 sequencing nucleotides, Mbp	133	1,072	150	416	—	—	—	—	—
Total assembly, thousands of contigs*	NM	191	16	1.8	—	—	—	—	—
Total assembly, Mbp*	NM	69	10	6.1	—	—	—	—	—
Total assembly submitted (contigs > 500 bp), Mbp*	6.2	24	7.0	5.8	—	—	—	—	—
Fosmid end-sequencing read number	480	670	No	No	—	—	—	—	—
Prochloron in fosmids, %	82%	90%	NA	NA	—	—	—	—	—
Prochloron genome	Yes	Yes	Yes	Yes	No	No	No	No	No
Manual assembly	Yes	No	No	No	—	—	—	—	—
Manual annotation	Yes	Part	Part	No	—	—	—	—	—
No. of contigs	100	869	1,093	1,753	—	—	—	—	—
Total assembly, Mbp	6.20	6.00	5.27	6.07	—	—	—	—	—
Microbiome analysis	Yes	Yes	Yes	No	No	No	No	No	No
Functional binning	No	Yes	Yes	No	—	—	—	—	—
Phylogenetic binning	No	Yes	Yes	No	—	—	—	—	—
Biosynthetic gene analysis	Yes	Yes	Yes	Yes	Yes	No	No	No	No
P. didemni analyzed	Yes	Yes	Yes	Yes	—	—	—	—	—
Biosynthetic genes found	>50	276	>120	NM	—	—	—	—	—
KS number of amplicons (unique)	NM	44 (32)	25 (24)	NM	8 (2)	—	—	—	—
Microscopy	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes
Filamentous cyanobacteria	No	Yes	Yes	—	—	No	—	No	No
Eukaryotic microbes	No	Yes	No	—	—	No	—	No	No
Cultivated isolates	No	Yes	Yes	No	No	No	Yes	No	No
Salinispora	—	19	3	—	—	—	1	—	—
Verrucosispora	—	1	ND	—	—	—	ND	—	—
Streptomyces	—	ND	1	—	—	—	1	—	—
Chemical analysis	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
LC-MS analysis	Yes	Yes	Yes	—	Yes	Yes	Yes	Yes	—
Cyanobactins	Yes	Yes	Yes	—	Yes	No	Yes	No	—
Patellazoles	No	Yes	Yes	—	No	No	No	No	—
^{1}H NMR analysis	Yes	Yes	Yes	—	No	No	No	No	—

This table lists data acquired and should not be used comparatively, as samples were not treated identically. ND, not detected; NM, not measured.
*For P1, we had a much higher quality, manual P1 assembly that was submitted to GenBank. Therefore, this assembly focused only on *P. didemni*, whereas for P2 and P3, the assembly files submitted to GenBank included DNA for other bacteria and the ascidian as well.

Some of these bacteria are involved in persistent associations important to the chemistry of the holobionts. The Whole Genome Shotgun projects (representing fragments >500 bp in length) have been deposited in the DNA Data Base in Japan, European Molecular Biology Laboratory, and GenBank under the accession numbers AFSJ00000000 (L2) and AFSK00000000 (L3).

Impact of Symbiotic Bacteria on Primary Metabolism. Systematic screening of photosynthetic genes shows that *P. didemni* seemingly possesses a complete photosynthetic apparatus and carbon fixation machinery (Fig. 3). However, compared with model cyanobacteria, it exhibits a number of peculiarities, which are directly related to its specific environment. These features include an expanded family of integral membrane, chl *a/b*-binding light-harvesting complex proteins (usually called Pcb for "prochlorophyte chl-binding proteins") (45). Former studies had revealed the presence of two distantly related *pcb* genes organized in tandem (*pcbA* and *pcbC*) (46), as well as one *isiA*-like gene (45). Inspection of the *P. didemni* genome shows that the latter gene is itself organized in a (possibly iron stress-induced) cluster with a unique member of the *pcb* family (ORF00655 in P1) that we propose to call *pcbD*.

No phycobiliproteins, the components of the antenna complexes found in typical cyanobacteria (phycobilisomes), had hitherto been reported in *Prochloron* (reviewed in refs. 12, 47, 48). However, the *P. didemni* genome contains homologues of genes encoding a form of C-phycocyanin with an α-subunit (encoded by *cpcA*) showing several variations with respect to known sequences. This includes a unique 7-aa deletion between positions 65 and 71 and a modified amino acid environment around the phycocyanobilin chromophore-binding cysteinyl residue (Cys-77). Similarly, all sequenced *Prochlorococcus* strains possess genes encoding an atypical phycoerythrin as the sole phycobiliprotein, which was hypothesized to be a photoreceptor (49, 50). In *Prochlorococcus* cells, these genes were shown to be expressed at low levels, and the phycobiliprotein gene product was found in small but significant amounts (49, 50). Further analyses are therefore needed to check whether *P. didemni cpcBA* operon is also expressed at low levels.

Another peculiarity of the photosynthetic apparatus of *P. didemni* is the huge number (e.g., n = 31 in P1) and organization of *hli* genes encoding high light-induced proteins (51), also called small cab-like proteins or Scps (52). These small proteins, possessing one (and sometimes two) transmembrane helix, belong to
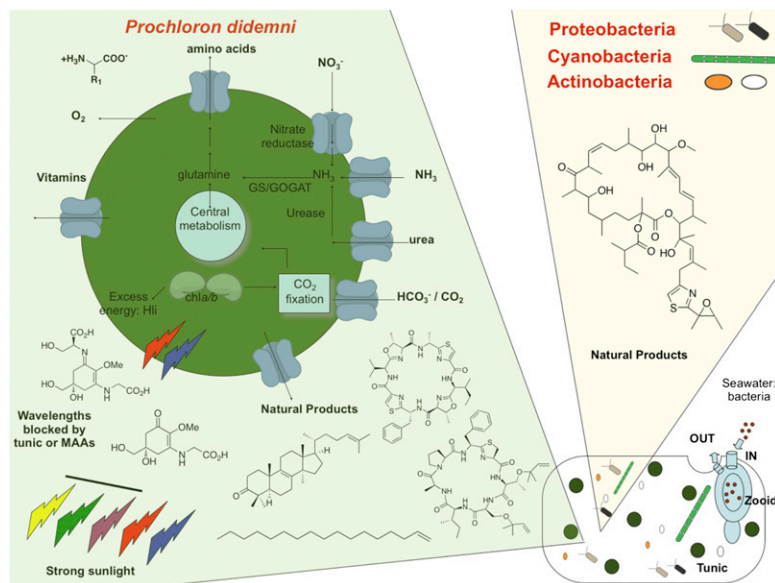
Fig. 3. Metabolic interactions in the ascidian metagenome. *L. patella* (*Lower Right*, schematic of tunic containing zooid and other bacteria) obtains nutrients and secondary metabolites from *P. didemni* and from other bacteria.

the plant Lhc protein family. They are thought to be involved in the absorption of excess excitation energy, allowing the cells to cope more efficiently with the production of reactive oxygen species (51), and/or in transient chl carriage (53). Despite their large number, there are only three or four *hli* gene types in each strain. One of them (e.g., ORF02209 in P1) is isolated in the genome and is most similar to the *hliD* gene of *Synechocystis* sp. PCC6803, whereas all others are homologues of *Synechocystis hliC* and are generally gathered into gene clusters. Their very close similarity at nucleotide level as well as organization in clusters strongly suggest that they result from recent events of gene duplication. We interpret this multiplication of *hli* genes in *P. didemni*, which is reminiscent of that previously reported for high light-adapted *Prochlorococcus* (54) as a photoprotective mechanism. As *Prochlorococcus* and *Prochloron* are taxonomically unrelated but share some photosynthetic features (39), this probably reflects functional convergence.

Like most cyanobacteria, *P. didemni* contains a form 1B 1,5-bisphosphate carboxylase/oxygenase that is protected from oxygen by β-carboxysomes (55). Consistent with its symbiotic association with an animal and the resulting constant availability of CO_2, *P. didemni* possesses only a minimal set of genes involved in CO_2 transport, consisting of a specialized NADH dehydrogenase complex (encoded by homologues of *Synechocystis ndhD4* and *ndhF4* genes) that takes up CO_2 with low affinity and converts it to HCO_3^- within the cell. It lacks many transporters found in other cyanobacteria, including those involved in high-affinity CO_2 uptake (NdhD3-F3) and the low-CO_2 inducible bicarbonate transporters SbtA and CmpA-D (55).

Early reports indicated that *P. didemni* may exchange its photosynthetically derived carbon with the host largely in the form of glycolate (56). However, a full contingent of glycolate oxidase and other glycolate metabolic genes are present in the genome. Based on estimated nitrogen budgets, *P. didemni*–ascidian associations rely on recycled nitrogen to supplement their filter-feeding diet (17). As predicted by experiment (57, 58), *P. didemni* contains genes required to take up ammonia and convert it into glutamine via the GS–GOGAT pathway. There is also an operon encoding urease and a complete urea transport system, indicating that *P. didemni* likely can take up urea, a major waste product of ascidians (59). In addition to using host sources of nitrogen, some strains of *P. didemni* have also been reported to incorporate nitrate (12) and to fix atmospheric nitrogen (18, 60). The genome-sequenced strains harbor complete nitrate reduction pathways, as would be required for conversion of nitrate to ammonium, but no nitrogenases or related genes

were found in the four sequenced metagenomes. *P. didemni* and hosts are therefore tightly coupled, with fixed carbon provided by the bacteria and recycled nitrogen originating with the host's filter-feeding diet, and not from fixation of atmospheric nitrogen.

Because *nifD* and *nifH* genes that are essential to nitrogen fixation were not found in *P. didemni*, we decided to extend the search to the remaining microbiomes. By using multiple published *nifD* and *nifH* sequences that explored the diverse *nif* subtypes, we performed tBLASTn searches on approximately 90 Mbp of total assembled sequence data in L2 and L3. No nitrogenase homologues were found in either sample, but the BLAST searches were robust, as revealed by numerous hits to more distantly related proteins such as protochlorophyllide reductase. In addition, BLAST searching of unassembled reads from the metagenomes did not reveal any putative nitrogenases in 5.9 million individual reads (average length of ∼400 bp). Based on the metagenome and the assembled *P. didemni* genomes, nitrogen recycling in *P. didemni* is likely to be the most important source of fixed nitrogen for this system, and nitrogen fixation is probably not possible except by a previously undiscovered class of nitrogenase enzyme or in a minor component of the microbiome. It remains possible that dinitrogen fixation is a variable property, and that bacteria in other *L. patella* samples fix nitrogen.

We used the automated CloVR-Metagenomics pipeline (61) to examine functional gene content compositions in the unassembled non-*Prochloron* metagenomic datasets based on assignment of sequence reads to the Clusters of Orthologous Genes (COG) database (62) by BLASTx (Fig. S1). Interestingly, the second most highly represented COG category was transporters, after proteins affecting nucleic acids. This observation may be a unique feature of this system, at least in comparison with other microbiome reports, reflecting the complex chemical exchange underlying this symbiosis.

Impact of Bacteria on Secondary Metabolism. *P. didemni* is responsible for many, but not all, of the secondary metabolites isolated from didemnid ascidians (34), and some of these are potent toxins (63). Here, we show that *P. didemni* also contains the genetic capacity to make several other compounds that greatly contribute to the overall chemical constitution of the holobiont, and that the remaining microbiome also contributes to the potent chemical arsenal of *L. patella* (Fig. 2).

Analysis of the *P. didemni* genome indicates that many compounds isolated from the whole animals are manufactured by the cyanobacterial symbionts. Pigments, lipids, and secondary metabolites significantly affect composition of the overall symbioses. For

example, *P. didemni* is protected from excess UV irradiation by the presence of mycosporine-like amino acids (MAAs), which are specifically localized to tunic bladder cells of the ascidian (64, 65). MAA biosynthetic genes are predominantly found in cyanobacteria, but also exist in the genomes of eukaryotes (66, 67). The *P. didemni* genome clearly encodes MAA biosynthesis. Other cyanobacterial pigments were encoded in the *P. didemni* genomes.

P. didemni also synthesizes the most abundant lipids found in the samples. We previously predicted that a *P. didemni* polyketide synthase (PKS) gene, *gaz*, was responsible for synthesizing terminal olefin lipids (34). Based on this information, recently Pfleger and coworkers reported that other cyanobacteria use a related gene to make nonadec-1-ene and derivatives (68). Following this analysis, here we used GC-MS and comparison with authentic standards to show that some of the *L. patella* samples contain abundant heptadec-1-ene (Fig. 4). This compound was found in several slices of tunic, which includes bacteria but is also rich in ascidian cells. This potential biofuel has not previously been detected in animals to the best of our knowledge. Further work is required to definitively connect this gene to the lipid product, but the accurate prediction and subsequent discovery of this compound is highly suggestive. GC-MS analysis of ascidian samples L1–L3 revealed that the animals also contain abundant alkanes, which were previously known from these animals and from *P. didemni* samples (69). Recent work shows that these alkanes are, in fact, common cyanobacterial products (70). The *P. didemni* genome thus defines a second major route to gasoline-like compounds in cyanobacteria and indicates that *P. didemni* may profoundly impact the membrane chemistry of the host animals. Although the compounds were not localized in this study, their relative abundance in whole tissue suggests that they might not be strictly localized to cyanobacteria.

As another example, we found a gene that was predicted to make lanosterol and oxidized lanosterol derivatives, with some similarities to pathways found in a small subset of other bacteria (71). Indeed, GC-MS analysis of *L. patella* tunic sections reveals that the holobiont contains abundant lanosterol and several oxidized lanosterol metabolites as the major sterols, with a rela-

tively smaller abundance of cholesterol and derivatives (Fig. 4). This is a very unusual finding, in that previous reports of ascidians describe largely cholesterol as the major sterol (69). To the best of our knowledge, lanosterol has not been previously described in ascidians, and its prediction in *P. didemni* on the basis of genome sequencing is highly suggestive that *P. didemni* synthesizes the major sterols isolated from the animals.

Although the biosynthesis of most compounds that could be identified in ascidian extracts could be assigned to *P. didemni*, there was one compound class that we were unable to assign to these cyanobacteria despite extensive effort: patellazoles (72). These potently cytotoxic compounds were found in L2 and L3, but not L1, and no biosynthetic genes explaining patellazoles could be identified in the *P. didemni* genome sequences. We thus hypothesized that patellazoles would be made by other symbiotic bacteria. As patellazoles have been isolated (albeit rarely) from *L. patella* samples since the early 1980s, such an association would represent a second dimension of stable bacterial symbiosis within these ascidians. Patellazoles also belong to a large family of biologically active ascidian metabolites, meaning that solving this problem would provide insight into important ascidian compounds from diverse species.

On the basis of structure, patellazoles are almost certainly hybrid PKS-nonribosomal peptide synthetase products. Furthermore, we predicted that patellazoles were likely made by the subclass of PKS enzymes in which the acyltransferase (AT) domain acts in trans (73). Therefore, we constructed a small database of known biosynthetic genes from polyketide, ribosomal, and nonribosomal peptide pathways and used it for tBLASTn searches against the L1, L2, and L3 metagenomes, consisting of raw reads before assembly. After filtering out *Prochloron* hits and genes that, on further analysis, were not relevant to natural products, we obtained 276 single reads and short contigs with predicted biosynthetic coding potential from sample L2 after extensive analysis (Fig. S2). In addition, at least 50 and 120 biosynthetic genes were identified by similar methods from L1 and L3, respectively, and as more sequence data were acquired, the number of new genes did not appear to reach a plateau. The resulting sequences were analyzed in detail as follows. Domains were classified based on Pfam (74) hits and on the taxonomy of the bacteria harboring the closest related sequence (Fig. S2). It should be emphasized that this phylogenetic call is weak for secondary metabolism genes, which often undergo extensive horizontal gene transfer.

Because of the large amount of total DNA sequence obtained by metagenome sequencing (1,072 Mbp, unassembled, in case of L2), we expected to achieve a plateau, indicating that more sequencing would not lead to greatly increased numbers of identified biosynthetic genes. However, more sequence provided ever more biosynthetic gene hits. Therefore, a nested degenerate PCR strategy was developed to amplify ketosynthase (KS) domains from PKS genes but not the predominant *P. didemni* PKS, *gaz*. PCR products obtained were cloned, and inserts were sequenced from 69 individual clones from L2 and L3, leading to 56 unique KS genes. Most of these genes were closely related to cyanobacterial and proteobacterial KSs and represented both *cis*- and *trans*-AT PKSs (Fig. 5). Interestingly, KS sequences were obtained with high protein sequence identity (>90%) to KS amplicons from other marine sources such as dinoflagellates and sponges. None of these amplicons overlapped with the metagenomic sequencing reads.

It was previously shown that, in some sponges, there is a relatively small number of *trans*-AT PKSs in comparison with the *cis*-AT PKSs, and this small group is often involved in synthesizing compounds that were previously isolated from the whole sponges (75). We used a recently developed phylogenetic method to classify short KS reads of the *trans*-AT group according to the chemistry of their products (Fig. 5) (75, 76). Some of these genes were only found in patellazole-containing ascidians, and not in other didemnid ascidians, including other samples of *L. patella* (Fig. 6). For example, by PCR, we showed that only L2 and L3
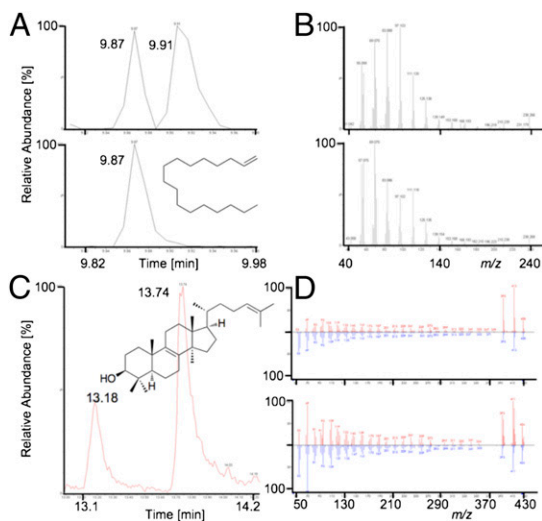
Fig. 4. Hydrocarbons and sterols in *P. didemni*. (*A*) GC-MS of an ascidian extract from a tissue slice containing the animal and bacteria (*Top*) and a standard of heptadec-1-ene (*Bottom*). (*B*) Fragmentation pattern of peak at 9.87 min in extract (*Top*) and control (*Bottom*), showing that these compounds are identical. The additional peak at 9.91 is an internal C_{17} olefin almost certainly resulting from the decarboxylative pathway to fatty acids. (*C*) GC-MS showing lanosterol-based lipids found in whole ascidian extracts. (*D*) Fragmentation pattern of *C*, showing observed spectra (above the lines in red) against expected spectra (below the lines in blue). These spectra indicate that lanosterol and an oxidized derivative are present in the ascidian sample.

contained the *trans*-AT gene PKS_11, whereas three other *L. patella* samples were both patellazole- and PKS_11-negative. The PKS_11 amplicon from L3 was 97% identical on the DNA level to that from L2. Similarly, several other gene products were discovered in the metagenome that could contribute to patellazole biosynthesis, such as PKS_100. Whether or not PKS_11 or these other genes are responsible for patellazole biosynthesis, this analysis showed that bacteria other than *Prochloron* form persistent associations with *L. patella* that contribute to diverse natural products from ascidians. Further analysis of the putative patellazoles biosynthetic gene cluster is in progress.

Origin of Chemical and Biological Diversity in Ascidians. The three most closely examined *P. didemni* genomes (P1, P2, and P3) were very similar with respect to gene synteny and total coding potential (Fig. 2). In particular, no significant differences were found in primary metabolism, indicating that all three strains have the genomic potential to contribute similarly to their hosts' nutrition. Differences between these genomes were dominated by hypothetical proteins and by families of overrepresented proteins that were present in many paralogous copies. These families included highly duplicated and repetitive sequences such as YD repeats, ankyrins, transposases, and lectins, as well as several apparently surface-associated proteins. These surface proteins are commonly variable and have been shown in other contexts to change rapidly in response to interactions with viruses or hosts (77). Comparative analysis and genome assembly was complicated by an unusually large number of repetitive sequences between and within genes that reduced the apparent similarity. By using our approach of 454 Life Sciences deep-sequencing, a large amount of sequence data were obtained for each metagenomic dataset. Like any other sequencing technology, 454 has an intrinsic error rate, so errors accumulate as more sequence data are collected. These errors should be removed during the assembly process, when multiple sequences are aligned to suppress errors. However, in highly sequence-repetitive regions, difficulties arise because the assembler has problems distinguishing between true sequence variations and those that arise from sequencing errors. Briefly, in repetitive regions, it is difficult to tell whether an observed difference be-

tween two highly related sequences reflects an actual difference in the genome, or whether it is a sequencing artifact, leading to problems in assembling highly repetitive regions. However, in the P1 *Prochloron* draft genome, most of these regions could be assembled correctly, as we also had extensive Sanger sequencing and manual assembly data, and only a small minority of repetitive regions were problematic (leading to 100 contigs in P1 instead of a closed genome). These methods enabled us to accurately identify many different repetitive sequences in the genomes, including long repeats, duplicated genes, and short repeats.

As one of many examples of short repeats, the *gaz* PKS leads to synthesis of terminal olefins described earlier and is found in all four samples. *gaz* contained the repeat sequence, CAAAA-CAA, previously observed as a phase-shifting repeat in *Neisseria gonorrheae* pilin biosynthesis (78). In genomes P1, P2, and P3, different copy numbers of this sequence element were observed within the KS and AT regions of *gaz*. The major metabolic gene differences among the *P. didemni* strains examined lie in secondary metabolism, as previously described (34), and also in the presence of other types of toxin genes such as a gene cluster encoding insecticidal toxin-like proteins, present in P1 and P4 but absent in P2 and P3.

Beyond *P. didemni*, ascidian hosts were found to possess a variable array of other bacteria, including those potentially responsible for patellazole biosynthesis. Microbial content of L1–L3 was examined by five methods: (*i*) PCR amplification of 16S genes from each sample and both pyrosequencing more than 4,000 amplicons and Sanger sequencing of 338 clones; (*ii*) shotgun metagenome sequencing of more than 3 Gbp of DNA isolated from the whole samples; (*iii*) observation by confocal microscopy; (*iv*) cultivation of Actinobacteria; and (*v*) end sequencing from fosmid libraries. All five methods indicated a highly diverse bacterial population, comprising approximately 10% to 20% of the bacterial content in L1–L3. The other 80% to 90% of bacteria were made up solely of clonal populations of *P. didemni*. Results from all five methods provided very similar and complementary data on the microbial diversity within ascidians.

16S PCR amplicons were sequenced by 454 and classified by using the automated CloVR-16S pipeline (61). Totals of 75%, 89%, and 98% of the sequences from L1, L2, and L3, re-
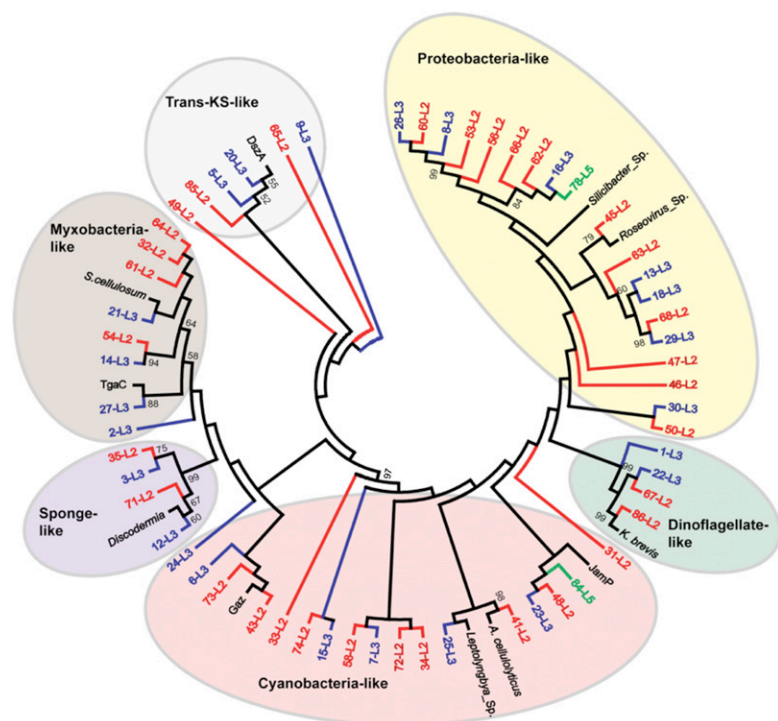
Fig. 5. Phylogenetic analysis of PCR-amplified KS domains. All unique (<95% sequence identical) KS clones obtained as described in *Materials and Methods* were used to generate a phylogenetic tree. Red indicates sequences derived from L2; blue indicates sequences derived from L3; green indicates sequences derived from L5; and black indicates previously described sequence relatives. Maximum parsimony analysis (Mega 5) was used to generate this phylogenetic tree with a bootstrap test of 1,000 replicates. Bootstrap values greater than 50% are shown. Clade names are provided based on the origin of previously identified sequences. For example, the sponge-like clade contains sequences that are most closely related to amplicons previously identified from sponges such as *Discodermia* (84, 96) (However, these are not the sponge-specific *sup* genes from *Poribacteria*). *trans*-AT PKS genes form their own ancient clade regardless of sequence origin (76).
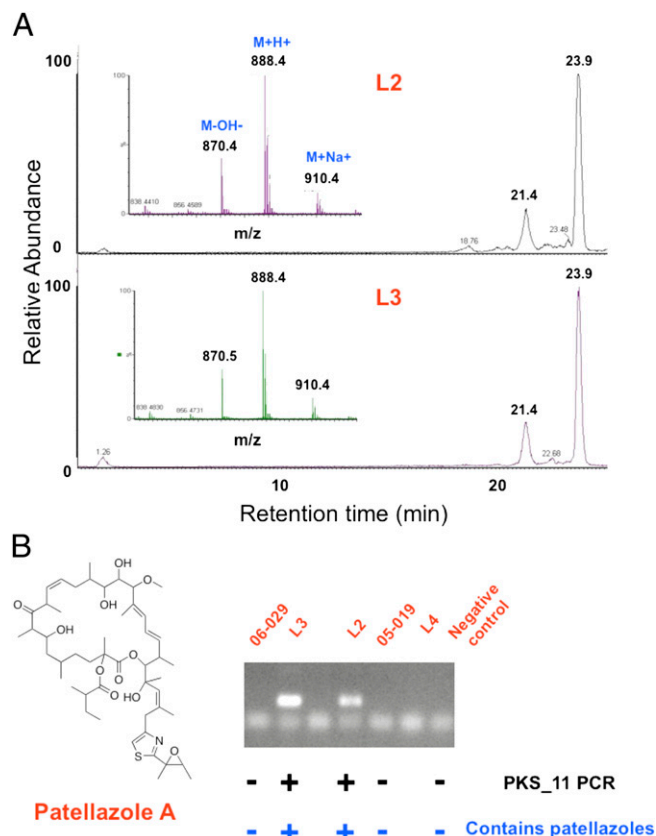
Fig. 6. Correlation of PKS genes with patellazoles. (*A*) Patellazoles were found only in L2 and L3 among our extensive ascidian collection examined between 2002 and 2007, which includes many *L. patella* samples and many other species of *Prochloron*-bearing didemnids. (*B*) As an example of the genetic approach, PKS_11 was found in ascidians containing patellazoles but not in ascidians lacking patellazoles, including adjacent colonies on the reef. Further work is required to definitively connect these genes to patellazoles biosynthesis.

spectively, clearly originated from the phylum Cyanobacteria, which was dominated by *P. didemni* sequences (Fig. 7). Proteobacteria was the second largest phylum in the three ascidians, representing 75%, 55%, and 50% of the noncyanobacterial sequences in L1, L2, and L3, respectively. Within this large group, the three ascidians contained mostly Alphaproteobac-

teria, followed by Gammaproteobacteria and Betaproteobacteria. 16S PCR products were also analyzed by Sanger sequencing for L2 and L3, which provided genus-level resolution (as detailed later).

Microbial community compositions of L2 and L3 were also explored based on total metagenomic sequences, by using the automated CloVR-Metagenomics pipeline (60), which makes phylogenetic assignments of individual sequence reads based on BLASTn searches against the National Center for Biotechnology Information Reference Sequence database of bacterial and archaeal genomes. For this classification, *P. didemni* reads were removed by *in silico* subtractive hybridization (BLASTn against the P1 genome, e-value of 1.0×10^{-50}). The CloVR-Metagenomics results correlated very closely to those from the 16S analysis, in that most of the sequences could be assigned to Cyanobacteria and Proteobacteria and smaller fractions to Firmicutes, Actinobacteria, and Spirochaetes (Fig. 8).

To scaffold the *Prochloron* genome, we constructed fosmid libraries from the same DNA samples, obtained by squeezing the ascidian sample, and end-sequenced 480 clones from L1 and 670 clones from L2. In L1, 82% of the clones were assigned to *Prochloron* and 18% to the host and other bacteria. In L2, approximately 90% of the clones were predicted to originate from *Prochloron* and approximately 10% were assigned to the host and other bacteria. These proportions resembled closely the community compositions determined by 16S rRNA and metagenome sequencing.

Sequencing data revealed that a large portion of the microbiome was cyanobacterial. Initially, it was thought that *P. didemni* was the only major cyanobacterium. However, analysis of biosynthetic reads and KS amplicons from L2 and L3 revealed genes that are clearly cyanobacterial in origin but are not part of the *Prochloron* genome. Because cyanobacterial cells and especially filaments are easily detectable by microscopy, we undertook a comparative approach by using fixed samples from L2, L3, and three other *Prochloron*-containing didemnid ascidians. Long filamentous cyanobacteria were observed in fixed specimens of L2 and L3 in an abundance of approximately 1 per 100 *Prochloron* cells (Fig. 9 and Fig. S3). These filaments were not observed in other related ascidians examined, including L1. In L2, fewer than 10% of cyanobacterial 454 16S reads originated from cyanobacteria other than *Prochloron*. These sequences appeared to be dominated by the Oscillatoriales, especially *Leptolyngbya* spp., and also by unclassified (uncultivated) cyanobacteria. *Calothrix* was also identified by Sanger sequencing of 16S PCR clone libraries. Filamentous cyanobacteria have previously been observed in *Prochloron*-bearing didemnids on several occasions, but not from *L. patella* (23–29). A few eukaryotic microbes, such as diatoms and dinoflagellates, were observed in L2 but not in L1 (Fig. 9 and Fig. S3).
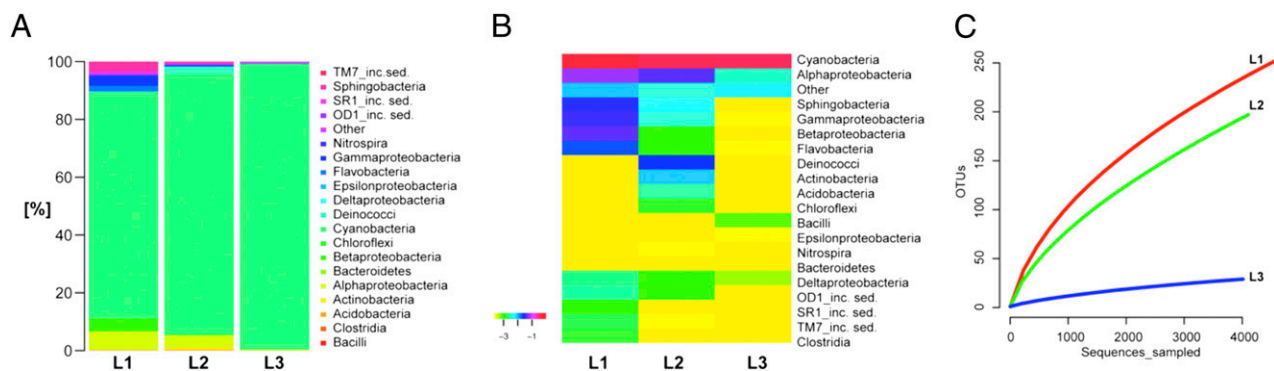

Fig. 7. 16S rRNA gene analysis. Visualization of the 16S analysis results by CloVR-16S: (*A*) Percentages of 16S sequences assigned to major taxonomic classes, with individual classes represented by different colors. All samples are dominated by Cyanobacteria, which includes *P. didemni*. (*B*) Complete-linkage (furthest neighbor) clustering of taxonomic classes based on log-normalized counts in each sample, as indicated by color in the scale bar. (*C*) Rarefaction curves showing that samples L1 and L2 contain more diverse and incompletely sampled microbial communities compared with L3.
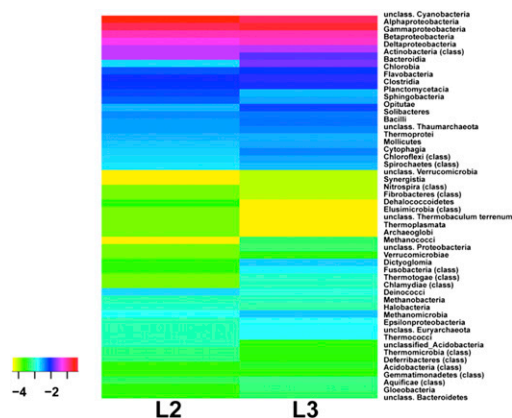
Fig. 8. Taxonomical classification of the non-*Prochloron* microbiome. *P. didemni* reads were subtracted from the total metagenome by BLASTn, and the remaining sequences were taxonomically assigned by using CloVR-Metagenomics. The results of complete-linkage (furthest neighbor) clustering of taxonomic classes based on relative counts are shown. The colors indicate the relative ratios of the reads in each sample on a logarithmic scale, as shown in the scale bar.

16S rDNA pyrosequencing revealed that more than 1% of L2 sequences originated in Actinobacteria, which are important producers of natural products. We therefore investigated the actinobacterial community using a cultivation-based approach with frozen tissue (Fig. 9 and Fig. S4). By 16S rRNA analysis, the major cultivable group in L2 (19 colonies) was assigned to the species *Salinispora arenicola* [99% identity to 16S rRNA genes from the sequenced strain CNS-205 (79)], and another colony belonged to *Verrucosispora*. Both genera are considered to be relatively rare components of marine sediments that are important natural product makers. The presence of *S. arenicola* in the L2 sample was further supported by comparative analysis of selected secondary metabolic gene clusters from our isolate, *S. arenicola* PZ-M17, with the previously reported genome of *S. arenicola* CNS 205. Amplicons were sequenced showing that *S. arenicola* PZ-M17 encodes several PKS and nonribosomal peptide synthetase genes that are unique to our strain and several that are found in the sequenced strain. These results suggest that the ascidian isolates resemble their sediment relatives but also contain additional unique biosynthetic gene clusters.

To confirm that these isolates resulted from a persistent bacteria–ascidian association, we next examined cultivated isolates from L3 and another Solomon Islands ascidian sample (*Didemnum molle* 06–028, collected within several meters of L3). From L3, we recovered strains with 99% 16S rRNA sequence identity to *S. arenicola*. From 06-028, we isolated a strain that is 99% identical (in its 16S rDNA sequence) to *Salinispora pacifica*. No *Salinispora* sequences were identified in the whole metagenome sequencing reads, either because of their low abundance or po-
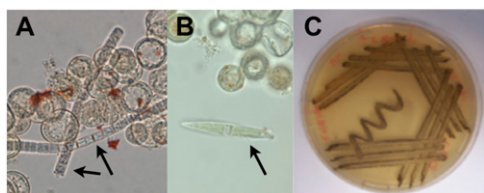
Fig. 9. Microbiological approaches. By light microscopy, other bacteria beyond *P. didemni* such as filamentous cyanobacteria (*A*) and diatoms (*B*) were found in L2 and L3, but not L1 or other ascidians. Cultivation analysis led to isolation of many natural product synthesizing strains, such as *Salinispora* (*C*) and *Verrucosispora*. Figs. S3 and S4 show more details.

tentially because of the relative difficulty of DNA extraction for these spore-forming Actinobacteria.

Discussion

L. patella harbors a complex microbiome consisting of multiple players that contribute to the primary and secondary metabolism of this ubiquitous tropical animal. Numerous methods were required to provide a comprehensive picture of the interactions between the ascidian host and its symbionts (Table 1). Here we present draft *P. didemni* genome sequences, obtained from complex metagenomic mixtures. The *P. didemni* genomes help to explain the photosynthetic and metabolic processes underlying the symbiosis and which have been shown to be critical for survival of the host animal. The whole *Prochloron* genomes obtained from four remote locations were remarkably syntenic and therefore displayed similar gene content, in particular factors that mediate primary and secondary metabolism. Despite some unusual features and the lack of stable cultivation, *P. didemni* contains intact and standard cyanobacterial primary metabolic pathways, and no genome reduction was observed. Apparently, nitrogen recycling from the host filter-feeding diet and photosynthesis are sufficient to provide these needed nutrients to the host. *P. didemni* also probably synthesizes some of the major lipids that are found in the holobionts, including potential biofuel molecules that were otherwise unanticipated and not discovered or sought in previous analyses of *L. patella* during the past 35 y, and that were only found through metagenome analysis. Within the complex *L. patella* environment, the non-*Prochloron* bacterial symbiosis appears to be largely mediated by previously underappreciated chemical interactions. For example, transporters were among the major functional groups identified in the non-*Prochloron* bacterial gene pool, an observation not previously recorded, to the best of our knowledge.

Ascidians differ in their metabolism, and a great variability is found in secondary metabolites. Previously, we showed that *P. didemni* exhibits sporadic variation of secondary metabolic genes, whereby whole pathways are present or absent in an otherwise conserved genomic background across the ocean (34). In this phenomenon, pathways are constantly exchanged between *P. didemni* cells across the ocean, providing the holobionts with a variable arsenal of metabolites. However, these results only partially explain the secondary metabolic variability found in ascidians. Here, we propose that symbiotic bacteria from different phyla synthesize abundant secondary metabolites isolated from single animals in great abundance. Candidate patellazole biosynthetic genes were discovered in non-*Prochloron* bacteria and were absent in *P. didemni* or host sequencing reads. Because patellazoles have been found in *L. patella* in different locations for the past 30 y, this indicates that non-*Prochloron* bacteria probably form persistent relationships with *L. patella*, expanding the known symbionts beyond merely *P. didemni*.

The genes associated with patellazoles biosynthesis belong to the *trans*-AT family of PKS genes. Interestingly, Riesenfeld et al. found an abundance of proteobacterial 16S sequences and *trans*-AT PKS genes in an Antarctic didemnid ascidian (80). This ascidian also contained palmerolide, a polyketide with some biosynthetic similarity to patellazoles and a diverse array of other ascidian polyketides. As only tropical ascidians contain *Prochloron*, it is clear that palmerolide, patellazoles, and related polyketides are likely synthesized by a distinct group of bacteria, which may possibly be Proteobacteria. One caveat is that the available sequences are not very similar between *L. patella* and the palmerolide sample. The PKS sequences are quite different at the DNA sequence level. From 16S sequences, only one of 34 palmerolide 16S clones in GenBank showed more than 95% DNA sequence identity with any of the more than 13,000 454 sequences from L1–L3. Of these, none were from L1, which exhibited a maximum of 92% identity with palmerolide clones. The single palmerolide sample clone with more than 95% identity to *L. patella* clones was similar to *Mesorhizobium* and other uncultivated symbionts from sponges. One sequence from

L2 and 19 sequences from L3 were more than 95% sequence identical to this clone.

In our samples, non-*Prochloron* bacteria fall into large taxonomic groups that are strongly correlated with natural product synthesis in these specific samples, as shown by a broad array of techniques. By using microscopy and Sanger sequencing, filamentous cyanobacteria were found associated with L2 and L3; Proteobacteria, especially Alphaproteobacteria, but also others, were the second most abundant group in L1–L3, as shown by all techniques used. Alphaproteobacteria are the most abundant group in open ocean (81), but are also extremely common, specific symbionts and pathogens (82, 83). Interesting sequences and isolates include those for myxobacteria and actinomycetes, which are renowned for natural product synthesis but have not been previously reported in ascidians. By applying numerous techniques, the biases inherent to individual methods were overcome, providing a comprehensive picture of bacterial symbiosis in ascidians.

The abundance of biosynthetic genes, including *trans*-AT genes, found in this study is noteworthy compared with those resulting from related studies in other environments. For example, by using the same degenerate KS primers we used here, Hochmuth et al. obtained 30,473 sequence reads using 454 analysis from two specimens of the marine sponge *Cacospongia mycofijiensis* (84). Only 118 unique genes were found, and 113 belonged to one group of sponge-specific poribacterial polyketide synthases; only five potentially synthesized more elaborate polyketides. Similar results were also obtained in a metagenome analysis of the sponge *Discodermia dissoluta*, in that, of 256 clones sequenced, 69 were different, as defined by less than 95% sequence identity (85). In our study with ascidians, we obtained 56 unique and highly sequence-diverse amplicons from sequencing only 69 clones. As a terrestrial example, a recent study described deep metagenomic sequencing of a leaf cutter ant colony (86). Although these ants are rich in antibiotic-producing Actinobacteria, from the provided information it appears that fewer than 100 secondary metabolism gene fragments were identified in the same categories used in our analysis. These results show a huge biosynthetic potential of marine ascidians in comparison with other animals.

In conclusion, *L. patella* harbors a remarkable array of microbial symbionts with the capacity to produce secondary metabolites and to contribute to host survival and primary metabolism. As didemnid ascidians themselves occupy many different habitats, and didemnids represent a large amount of genetic variation between and within species, these bacteria provide an important additional dimension to understanding biodiversity in the ocean.

Materials and Methods

Sample Collection, Processing, and Sequencing. Ascidians were collected and processed as previously described (34) from the following locations and of the following species: L1, North 7° 15′ East 134° 15′ (*L. patella*); L2, South 17° 55′ East 177° 16′ (*L. patella*); L3, South 8° 57.35′ East 159° 59.12′ (*L. patella*); L4, South 4° 45.601′ East 151° 25.320′ (*L. patella*); 05–019, South 10° 15.856′ East 150° 15.856′ (*L. patella*); 06–014, South 9° 0.13′ East 159° 15.69′ (*Prochloron*-bearing didemnid); 06–028, South 8° 57.35′ East 159° 59.12′ (*D. molle*); 06–029, South 8° 57.35′ East 159° 59.12′ (*Lissoclinum* sp.); and 07–021, North 7° 21′ 0.40′′ East 134° 34′ 29.08′′ (*Lissoclinum* sp.; Table 1). Sequencing methods are summarized in Table 1. All sequences, except for fosmids, were assembled using Celera (87), and the resulting contigs taxonomically sorted by using PhymmBL (88). Initial assembly and statistics of *P. didemni* genome sequencing efforts on the P1, P2, and P3 samples are provided in the authors' previous publication (34). Reduction to 32 scaffolds as reported here was accomplished through manual closure based on sequencing PCR product linking contigs.

Genome Sequence Annotation. Metabolic genes and many other gene subsets of the P1 genome were manually annotated by using Manatee. The annotation was then transferred to the assembled genomes of P2, P3, and P4. Secondary metabolic gene clusters were previously annotated (34).

Total Metagenomic DNA Sequence Analysis. Total metagenomic whole-genome shotgun DNA isolates of L2 and L3 were sequenced on the Roche/454 GS FLX Titanium platform as described before (34). Raw sequence reads were searched against the P1 draft genome sequence with BLASTN, using an e-value threshold of 1.0×10^{-50}. This relatively stringent threshold was applied to remove those reads that could be assigned to *P. didemni* with high confidence. Reads without matches to the P1 draft genome were processed with the CloVR-Metagenomics pipeline from the CloVR package for automated and cloud computing-enabled microbial sequence analysis (89), which provides taxonomic and functional composition analyses, using BLAST searches against the COG and National Center for Biotechnology Information Reference Sequence collection. Using custom R and Perl scripts that are part of the CloVR-Metagenomics pipeline, the results were displayed as distance histograms.

Other Experiments. Details of other experiments are provided in *SI Materials and Methods*.

1. Kott P (1980) Algal-bearing didemnid ascidians in the Indo-west-Pacific. *Mem Queensl Mus* 20:1–47.
2. Kott P (2001) The Australian Ascidiacea. Part 4, Aplousobranchia (3), Didemnidae. *Mem Queensl Mus* 47:1–407.
3. Brusca RC, Brusca GJ (2003) *Invertebrates* (Sinauer, Sunderland, MA), 2nd Ed.
4. Hirose E (2009) Ascidian tunic cells: Morphology and functional diversity of free cells outside the epidermis. *Invertebr Biol* 128:83–96.
5. Ryland JS, Wigley RA, Muirhead A (1984) Ecology and colonial dynamics of some Pacific reef flat Didemnidae (Asciadiacea). *Zool J Linn Soc* 80:261–282.
6. Cowan ME (1981) Field observations of colony movement and division of the ascidian *Didemnum molle*. *Mar Ecol Prog Ser* 6:335–337.
7. Birkeland C, Cheng L, Lewin RA (1981) Motility of didemnid ascidian colonies. *Bull Mar Sci* 31:170–173.
8. Bak RPM, Sybesma J, van Duyl FC (1981) The ecology of the tropical compound ascidian *Trididemnum solidum*. 2. Abundance, growth and survival. *Mar Ecol Prog Ser* 6:43–52.
9. Littler MM, Littler DS (1995) A colonial tunicate smothers corals and coralline algae on the Great Astrolabe Reef, Fiji. *Coral Reefs* 14:148–149.
10. Schmidt EW, Donia MS (2010) Life in cellulose houses: Symbiotic bacterial biosynthesis of ascidian drugs and drug leads. *Curr Opin Biotechnol* 21:827–833.
11. Davidson BS (1993) Ascidians: Producers of amino acid-derived metabolites. *Chem Rev* 93:1771–1791.
12. Lewin RA, Cheng L (1989) *Prochloron, a Microbial Enigma* (Chapman and Hall, London), 1st Ed.
13. Hirose E, Neilan BA, Schmidt EW, Murakami A (2009) Enigmatic life and evolution of *Prochloron* and related cyanobacteria inhabiting colonial ascidians. *Handbook on Cyanobacteria*, eds Gault PM, Marler HJ (Nova, New York), pp 161–189.

14. Koike I, Suzuki T (1996) Nutritional diversity of symbiotic ascidians in a Fijian seagrass meadow. *Ecol Res* 11:381–386.
15. Kremer BP, Pardy R, Lewin RA (1982) Carbon fixation and photosynthates of *Prochloron*, a green alga symbiotic with an ascidian, *Lissoclinum patella*. *Phycologia* 21:258–263.
16. Lewin RA, Pardy RL (1981) Photosynthetic activity of *Prochloron*, and translocation of metabolites to its ascidian host. *Phycologia* 20:109–109.
17. Koike I, Yamamuro M, Pollard PC (1993) Carbon and nitrogen budgets of two ascidians and their symbiont, *Prochloron*, in a tropical seagrass meadow. *Aust J Mar Freshwater Res* 44:173–182.
18. Odintsov VS (1991) Nitrogen fixation in *Prochloron* (Prochlorophyta)-ascidian associations - Is *Prochloron* responsible. *Endocytobiosis Cell Res* 7:253–258.
19. Lewin RA (1976) Prochlorophyta as a proposed new division of algae. *Nature* 261:697–698.
20. Lewin RA, Withers NW (1975) Extraordinary pigment composition of a prokaryotic alga. *Nature* 256:735–737.
21. Shimada A, Yano N, Kanai S, Lewin RA, Maruyama T (2003) Molecular phylogenetic relationship between two symbiotic photo-oxygenic prokaryotes, *Prochloron* sp. and *Synechocystis trididemni*. *Phycologia* 42:193–197.
22. Swingley WD, et al. (2008) Niche adaptation and genome expansion in the chlorophyll *d*-producing cyanobacterium *Acaryochloris marina*. *Proc Natl Acad Sci USA* 105:2005–2010.
23. Lewin RA, Cheng L (1975) Associations of microscopic algae with didemnid ascidians. *Phycologia* 14:149–152.
24. Miyashita H, et al. (1996) Chlorophyll *d* as a major pigment. *Nature* 383:402.
25. Larkum AWD, et al. (1987) A filamentous cyanophyte containing phycourobilin and symbiotic in sponges and an ascidian. *Mar Biol* 95:1–13.
26. Parry DL, Kott P (1988) Co-symbiosis in the Ascidiacea. *Bull Mar Sci* 42:149–153.

27. Hirose E, Uchida H, Murakami A (2009) Ultrastructural and microspectrophotometric characterization of multiple species of cyanobacterial photosymbionts coexisting in the colonial ascidian *Trididemnum clinides* (Tunicata, Ascidiacea, Didemnidae). *Eur J Phycol* 44:365–375.

28. Groepler W, Schuett C (2003) Bacterial community in the tunic matrix of a colonial ascidian *Diplosoma migrans*. *Helgol Mar Res* 57:139–143.

29. Schuett C, Doepke H, Groepler W, Wichels A (2005) Diversity of intratunical bacteria in the tunic matrix of the colonial ascidian *Diplosoma migrans*. *Helgol Mar Res* 59:136–140.

30. Schmidt EW, et al. (2005) Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proc Natl Acad Sci USA* 102:7315–7320.

31. Donia MS, Ravel J, Schmidt EW (2008) A global assembly line for cyanobactins. *Nat Chem Biol* 4:341–343.

32. Donia MS, et al. (2006) Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol* 2:729–735.

33. Rath CM, et al. (2011) Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. *ACS Chem Biol*, 10.1021/cb200244t.

34. Donia MS, Fricke WF, Ravel J, Schmidt EW (2011) Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLoS ONE* 6:e17897.

35. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.

36. Donia MS, Ruffner DE, Cao S, Schmidt EW (2011) Accessing the hidden majority of marine natural products through metagenomics. *ChemBioChem* 12:1230–1236.

37. Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190.

38. Nakabachi A, et al. (2006) The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science* 314:267.

39. Münchhoff J, et al. (2007) Host specificity and phylogeography of the prochlorophyte *Prochloron* sp., an obligate symbiont in didemnid ascidians. *Environ Microbiol* 9:890–899.

40. Bright M, Bulgheresi S (2010) A complex journey: Transmission of microbial symbionts. *Nat Rev Microbiol* 8:218–230.

41. Hirose E, Nakabayashi S (2008) Algal symbionts in the larval tunic lamellae of the colonial ascidian *Lissoclinum timorense* (Ascidiacea, Didemnidae). *Zoolog Sci* 25:1205–1211.

42. Hirose E, Fukuda T (2006) Vertical transmission of photosymbionts in the colonial ascidian *Didemnum molle*: The larval tunic prevents symbionts from attaching to the anterior part of larvae. *Zoolog Sci* 23:669–674.

43. Hirose E, Hirose M, Neilan BA (2006) Localization of symbiotic cyanobacteria in the colonial ascidian *Trididemnum miniatum* (Didemnidae, Ascidiacea). *Zoolog Sci* 23:435–442.

44. Patterson GM, Withers NW (1982) Laboratory cultivation of *prochloron*, a tryptophan auxotroph. *Science* 217:1034–1035.

45. La Roche J, et al. (1996) Independent evolution of the prochlorophyte and green plant chlorophyll *a/b* light-harvesting proteins. *Proc Natl Acad Sci USA* 93:15244–15248.

46. Chen M, Hiller RG, Howe CJ, Larkum AW (2005) Unique origin and lateral transfer of prokaryotic chlorophyll-*b* and chlorophyll-*d* light-harvesting systems. *Mol Biol Evol* 22:21–28.

47. Matthijs HCP, van der Staay GWM, Mur LR (1994) Prochlorophytes: the "other" cyanobacteria? *The Molecular Biology of Cyanobacteria*, ed Bryant DA (Kluwer, Dordrecht), pp 49–64.

48. Partensky F, Garczarek L (2003) The photosynthetic apparatus of chlorophyll b- and d-containing Oxychlorobacteria. *Photosynthesis in Algae, Advances in Photosynthesis Series*, eds Larkum AWD, Douglas SE, Raven JA (Kluwer, Dordrecht, The Netherlands), Vol 14, pp 29–62.

49. Steglich C, Frankenberg-Dinkel N, Penno S, Hess WR (2005) A green light-absorbing phycoerythrin is present in the high-light-adapted marine cyanobacterium *Prochlorococcus* sp. MED4. *Environ Microbiol* 7:1611–1618.

50. Hess WR, et al. (1996) Coexistence of phycoerythrin and a chlorophyll *a/b* antenna in a marine prokaryote. *Proc Natl Acad Sci USA* 93:11126–11130.

51. He QF, Dolganov N, Bjorkman O, Grossman AR (2001) The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *J Biol Chem* 276:306–314.

52. Kufryk G, et al. (2008) Association of small CAB-like proteins (SCPs) of *Synechocystis* sp. PCC 6803 with Photosystem II. *Photosynth Res* 95:135–145.

53. Storm P, Hernandez-Prieto MA, Eggink LL, Hoober JK, Funk C (2008) The small CAB-like proteins of *Synechocystis* sp. PCC 6803 bind chlorophyll. In vitro pigment reconstitution studies on one-helix light-harvesting-like proteins. *Photosynth Res* 98:479–488.

54. Kettler GC, et al. (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 3:e231.

55. Badger MR, Hanson D, Price GD (2002) Evolution and diversity of CO$_2$ concentrating mechanisms in cyanobacteria. *Funct Plant Biol* 29:161–173.

56. Fisher CR, Trench RK (1980) In vitro carbon fixation by *Prochloron* sp. isolated from *Diplosoma virens*. *Biol Bull* 159:636–648.

57. Goodbody I (1974) The physiology of ascidians. *Adv Mar Biol* 12:1–149.

58. Parry DL (1985) Nitrogen assimilation in the symbiotic marine algae *Prochloron* spp. *Mar Biol* 87:219–222.

59. Markus JA, Lambert CC (1983) Urea and ammonia excretion by solitary ascidians. *J Exp Mar Biol Ecol* 66:1–10.

60. Kline TC, Lewin RA (1999) Natural ^{15}N/^{14}N abundance as evidence for N$_2$ fixation by *Prochloron* (Prochlorophyta) endosymbiotic with didemnid ascidians. *Symbiosis* 26:193–198.

61. White JR, et al. (2011) CloVR-Metagenomics: Functional and taxonomic microbial community characterization from metagenomic whole-genome shotgun (WGS) sequences – standard operating procedure, version 1.0. *Nat Precedings*, 10.1038/npre.2011.5886.1.

62. Tatusov RL, et al. (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4:41.

63. Donia MS, Schmidt EW (2010) Cyanobactins – Ubiquitous Cyanobacterial ribosomal peptide metabolites. *Comprehensive Natural Products II Chemistry and Biology* (Elsevier, Oxford), Vol 2, pp 539–558.

64. Dionisio-Sese ML, Ishikura M, Maruyama T, Miyachi S (1997) UV-absorbing substances in the tunic of a colonial ascidian protect its symbiont, *Prochloron* sp., from damage by UV-B radiation. *Mar Biol* 128:455–461.

65. Maruyama T, Hirose E, Ishikura M (2003) Ultraviolet-light-absorbing tunic cells in didemnid ascidians hosting a symbiotic photo-oxygenic prokaryote, *Prochloron*. *Biol Bull* 204:109–113.

66. Balskus EP, Walsh CT (2010) The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science* 329:1653–1656.

67. Singh SP, Klisch M, Sinha RP, Häder DP (2010) Genome mining of mycosporine-like amino acid (MAA) synthesizing and non-synthesizing cyanobacteria: A bioinformatics study. *Genomics* 95:120–128.

68. Mendez-Perez D, Begemann MB, Pfleger BF (2011) A gene encoding a modular synthase is involved in alpha-olefin biosynthesis in *Synechococcus* sp. PCC7002. *Appl Environ Microbiol* 77:4264–4267.

69. Perry GJ, Gillan FT, Johns FB (1978) Lipid composition of a prochlorophyte. *J Phycol* 14:369–371.

70. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* 329:559–562.

71. Bode HB, et al. (2003) Steroid biosynthesis in prokaryotes: Identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Mol Microbiol* 47:471–481.

72. Zabriskie TM, Mayne C, Ireland CM (1988) Patellazole C: A novel cytotoxic macrolide from *Lissoclinum patella*. *J Am Chem Soc* 110:7919–7920.

73. Piel J (2010) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* 27:996–1047.

74. Finn RD, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38 (Database issue):D211–D222.

75. Fisch KM, et al. (2009) Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat Chem Biol* 5:494–501.

76. Nguyen T, et al. (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol* 26:225–233.

77. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011) Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* 474:604–608.

78. Power PM, et al. (2003) Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. *Mol Microbiol* 49:833–847.

79. Udwary DW, et al. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica. *Proc Natl Acad Sci USA* 104:10376–10381.

80. Riesenfeld CS, Murray AE, Baker BJ (2008) Characterization of the microbial community and polyketide biosynthetic potential in the palmerolide-producing tunicate Synoicum adareanum. *J Nat Prod* 71:1812–1818.

81. Yooseph S, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLoS Biol* 5:e16.

82. Batut J, Andersson SG, O'Callaghan D (2004) The evolution of chronic infection strategies in the alpha-proteobacteria. *Nat Rev Microbiol* 2:933–945.

83. Gruber-Vodicka HR, et al. (2011) *Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms. *Proc Natl Acad Sci USA* 108:12078–12083.

84. Hochmuth T, et al. (2010) Linking chemical and microbial diversity in marine sponges: Possible role for poribacteria as producers of methyl-branched fatty acids. *ChemBioChem* 11:2572–2578.

85. Schirmer A, et al. (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Appl Environ Microbiol* 71:4840–4849.

86. Suen G, et al. (2010) An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet* 6:pii:e1001129.

87. Miller JR, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24:2818–2824.

88. Brady A, Salzberg SL (2009) Phymm and PhymmBL: Metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* 6:673–676.

89. Angiuoli SV, et al. (2011) CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics* 12:356.