Vol. 83: 1–13, 2019
https://doi.org/10.3354/ame01901

AQUATIC MICROBIAL ECOLOGY
Aquat Microb Ecol

Published online May 2

Symbiont frequency predicts microbiome composition in a model bryozoan-bacterial symbiosis

Nastassia V. Patin[1,*], Sara Locklear[1], Frank J. Stewart[1], Nicole B. Lopanik[1,2]

[1]School of Biological Sciences and [2]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

ABSTRACT: The bryozoan *Bugula neritina* hosts an uncultured gammaproteobacterial symbiont, '*Candidatus* Endobugula sertula' (*E. sertula*), which also produces the defensive compounds known as bryostatins. These compounds are unpalatable and protect the larvae from predation. Recent studies have shown that the symbiont can be present or absent in *B. neritina* colonies. The factors determining the symbiotic state of the host are unknown and may include abiotic or biotic variables, including other members of the *B. neritina* microbiome, which are poorly studied. Here, we explored the relationships between host genotype, environmental factors, *E. sertula* frequency, and microbiome composition in 141 *B. neritina* colonies collected from the east coast of North America. We found that latitude, temperature, and host genotype did not correlate with symbiont frequency, but they did impact the overall microbiome composition of *B. neritina*. Furthermore, we found that symbiont frequency predicts the composition of the non-symbiont microbiome. We hypothesize that chemical or ecological interactions mediated by *E. sertula* determine the presence or absence of other microbial taxa, which are additionally impacted by environmental factors and host biology.

KEY WORDS: Symbiosis · Microbiome · Bryozoan · Chemical ecology

1. INTRODUCTION

Sessile marine invertebrates often harbor complex communities of bacteria (Harder 2008, Hentschel et al. 2012), which may increase host fitness through nutrient cycling (Dubilier et al. 2008, Freeman & Thacker 2011) or by providing chemical defense from predation (Engel et al. 2002, Lopanik 2014). The rich chemical repertoire of invertebrates and their associated microbes has historically generated great biotechnological interest (Haygood et al. 1999). More recently, these associations have become important targets of study for understanding the ecology and evolution of host-microbiome relationships (Newton et al. 2013, McFall-Ngai 2014, Puglisi et al. 2014, Cantley & Clardy 2015). These studies often rely on culture-independent methods to identify microbial partnerships, as most host-associated microbes have not been cultivated. For example, metagenomic analyses have linked specific microbes and their biosynthetic genes to the production of defensive bioactive metabolites (Schmidt et al. 2005, Oh et al. 2009, Rath et al. 2011, Flórez et al. 2017). While certain invertebrate-microbe associations may involve just one or a few specific taxa unique to the animal (Schmitt et al. 2012, Dishaw et al. 2014, Reveillaud et al. 2014, Cahill et al. 2016), others contain both a host-specific component composed of microbes that are consistently enriched or occur only in a host type (a 'core' microbiome) (Kott et al. 1984, Moss et al. 2003, López-Legentil et al. 2011) and a flexible component composed of microbes present in diverse hosts and other environments. Indeed, many sessile invertebrates, such as corals, sponges, and bryozoans,

*Corresponding author: nastassia.patin@biosci.gatech.edu

actively filter microbes from the water for nutrition, making it difficult to distinguish transitory from resident (and potentially symbiotic) members and to identify the ecological contributions of different microbial partners.

Understanding the nature of the symbiont-host relationship in terms of stability, mechanism, and benefits to both partners is a primary goal of symbiosis research. Meeting this goal is often challenging because the symbiont-host relationship can change over time and may involve interactions with multiple microbial taxa (Hussa & Goodrich-Blair 2013). In some symbioses, microbe-microbe interactions may be essential for maintaining the benefit to the host. For example, oligochaete worms obtain nutrition (carbon) from chemosynthetic sulfur-oxidizing endosymbionts, which depend on sulfide produced by co-occurring sulfate-reducing endosymbionts (Dubilier et al. 2001, Blazejak et al. 2005). Further, a metagenomic analysis of the tunicate *Lissoclinum patellum*, which harbors the cyanobacterial symbiont *Prochloron didemni*, revealed a complex microbiome contributing to the production of diverse primary and secondary metabolites (Donia et al. 2011). These studies, along with evidence that microbial chemical interactions are pervasive in diverse natural systems (Hay et al. 2017), suggest that for most symbioses, even those classically described as a 2-partner interaction, it is likely that the primary symbiont both influences and is influenced by co-occurring microbes. Baseline knowledge of such interactions in invertebrate microbiomes is critical for understanding the entire holobiont and developing new strategies for cultivating symbionts.

The association between the marine bryozoan *Bugula neritina* and its toxin-producing endosymbiont, '*Candidatus* Endobugula sertula' (*E. sertula*) (Davidson & Haygood 1999), is a valuable model for exploring microbial defensive symbioses. The host bryozoan forms colonies in temperate coastal waters worldwide, often attached to docks or pilings (Keough & Chernoff 1987, Mackie et al. 2006). On the Atlantic coast of North America, 2 *B. neritina* sibling species, designated as 'Type S' and 'Type N,' exist without rigid biogeographical boundaries (Linneman et al. 2014). Within each genotype, some colonies, but not all, associate with the uncultured gammaproteobacterium *E. sertula*. *E. sertula* produces the bryostatins, a family of macrocyclic lactones with potent cytotoxicity, including anticancer activity (Davidson et al. 2001, Hildebrand et al. 2004, Linneman et al. 2014). Larval *B. neritina*, which lack the protective chitin layer found in adult animals, contain high levels of bryostatins, which provide a chemical defense against predators (Lopanik et al. 2004, 2006). *E. sertula* does not exhibit the genome reduction typical of obligate endosymbionts and therefore likely also exists in a free-living form, although the bacterium is rarely detected in the water column (Miller et al. 2016a, Li et al. 2019). While other diverse bacterial taxa have been found in *B. neritina* (Miller et al. 2016b, Li et al. 2019), the extent of microbiome diversity and the factors driving microbiome variation are largely unknown. Notably, it remains unclear why some colonies contain *E. sertula* and others do not, and how the presence or absence of the symbiont and its bioactive metabolites affect the rest of the microbiome.

Microbiome science, which typically involves analysis of taxonomic composition by 16S rRNA gene amplicon sequencing, has several characteristics that make rigorous statistical analysis and ecological interpretation challenging. Technical biases can arise through differences in sample processing steps, including DNA extraction, PCR amplification, and sequencing library preparation (Zhou et al. 2011, Pollock et al. 2018). The common practice of clustering similar 16S rRNA gene sequences into operational taxonomic units (OTUs) can combine genetically and ecologically distinct organisms into unresolved groups and obscure diversity estimates (Patin et al. 2013). Microbiome sequence data are typically highly multi-dimensional, with the number of taxa much higher than the number of samples. They are also sparse, as most taxa are observed in only a small number of samples, leading to highly zero-inflated data tables. Further, because feature counts represent relative abundances and must sum to 1 in each sample, the data are compositional and not unconstrained in Euclidean space. These features render traditional ecological statistical methods for data normalization and differential abundance analysis inappropriate (Debelius et al. 2016, Thorsen et al. 2016, Tsilimigras & Fodor 2016, Gloor et al. 2017, Weiss et al. 2017). These underlying statistical assumptions can lead to spurious results when used to detect co-occurring microbes (Connor et al. 2017, Mainali et al. 2017). However, recent advances in data science, including Bayesian approaches (Wadsworth et al. 2017, Lee & Sison-Mangus 2018) and machine learning tools (Johnson et al. 2016, Min et al. 2017, Belk et al. 2018, Oudah & Henschel 2018), can overcome some of these problems. The Random Forest algorithm is a supervised machine learning tool designed to generate models of complex datasets based on decision trees. It can be used for both classification

and regression models and is gaining traction in microbiome science as a way to test the relationship between set variables and community composition.

Here, we used these methods to analyze the taxonomic composition of *B. neritina* microbiomes from >140 samples collected along a latitudinal gradient along the east coast of North America. The samples represent both Type S and N host genotypes and a range of *E. sertula* proportional abundances and temperature conditions. The data were analyzed by treating each unique amplicon as a distinct taxon, rather than clustering similar amplicons into OTUs. Along with nonparametric statistical methods commonly used in ecology, we applied the Random Forest algorithm to investigate patterns of *E. sertula* relative abundance and its relationship to *B. neritina* microbiome diversity and composition. Our findings indicate that a complex interplay of host biology and environmental conditions affects the overall microbiome and that the presence of *E. sertula* shapes the non-symbiont community.

2. MATERIALS AND METHODS

2.1. Sample collection

Bugula neritina colonies were collected over a period of 2 yr (March 2015 to November 2017) from 16 locations along the Atlantic Coast of the United States, ranging from Florida to Delaware, from 29.75° to 38.61° N in latitude (see Fig. S1 in Supplement 1 at www.int-res.com/articles/suppl/a083p001_supp.pdf and Table S1 in Supplement 2 at www.int-res.com/articles/suppl/a083p001_supp.xls). Each sample consisted of approximately 0.1 to 1 g of biomass, which comprised either an entire small colony or part of a larger colony. All samples were collected from docks by hand and immediately stored in RNA preservation buffer (25 mM sodium citrate, 10 mM EDTA, 5.3 M ammonium sulfate, pH 5.2; 1 ml in a 1.6 ml centrifuge tube). At each site, surface water salinity and temperature were measured using a refractometer and hand-held thermometer. Following transport to the lab, samples were frozen at −20°C and kept frozen until processing.

In November 2017, 4 replicate seawater samples (1 l each) were collected from 3 locations for microbial community analysis. These locations were chosen based on differences in *B. neritina* populations (host/symbiont genotype) assessed in previous samplings. Each 1 l seawater sample was filtered through a sterile 0.2 µm filter (Supor 200, Pall Life Sciences) within 1 to 7 h of collection, and the filters were immediately extracted for microbial DNA using the Qiagen DNeasy PowerWater Kit according to the manufacturer's instructions. The DNA was frozen at −20°C until further processing as described below.

2.2. *B. neritina* DNA extraction, host molecular characterization, and symbiont screening

Each *B. neritina* sample was thawed on ice, and approximately 30 to 50 mg of tissue was removed and placed in phosphate-buffered saline. DNA was extracted using the ZR Fungal/Bacterial miniprep kit (Zymo Research) following the manufacturer's protocol. DNA was quantified and assessed for purity using a Thermo Fisher Scientific NanoDrop 1000 spectrophotometer and Qubit 2.0 fluorometer (Thermo Fisher Scientific).

All colonies were classified by genotype (Type S/Type N) and by symbiont presence or absence as described by Linneman et al. (2014). Briefly, genotyping was performed by amplification of the mitochondrial cytochrome oxidase I gene using the *B. neritina*-specific primers BnCOIf and BnCOIr. Restriction fragment length polymorphism (RFLP) analysis was performed on amplicons using the restriction endonucleases *Dde*I and *Hha*I, which cut only Type S and N amplicons, respectively. The presence of *Endobugula sertula* was determined by visual detection of amplicons following PCR using primers specific to the *bryB* and *bryR* genes, which are part of the bryostatin polyketide synthase gene cluster (Sudek et al. 2007). PCR amplicons and restriction-digested products were visualized under UV light after agarose gel electrophoresis.

2.3. Microbiome library preparation and sequencing

B. neritina samples were selected for sequencing to represent approximately equal numbers of *B. neritina* colonies and host genotypes (Type S/N) per site. Of those selected, all *B. neritina* and water samples with a DNA concentration ≥10 ng μl^{-1} and a 260/280 nm ratio of ~1.8 were used for PCR amplification of the V4 region of the 16S rRNA gene using Earth Microbiome Project primers 515FB and 806RB in a protocol adapted from Kozich et al. (2013). In total, 148 *B. neritina* and 12 water column samples were sequenced (Table S1). Briefly, amplicons were gener-

ated (1 reaction per sample, plus a negative control reaction with Milli-Q water instead of DNA template) using Platinum PCR SuperMix (Life Technologies) and EMP primers that were barcoded and appended with Illumina-specific adaptors. Thermal cycling consisted of denaturation at 94°C (3 min); 30 cycles of denaturation at 94°C (45 s), primer annealing at 55°C (45 s), primer extension at 72°C (90 s); and final extension at 72°C for 10 min. Amplicons were analyzed by gel electrophoresis to verify size of the amplification product and no amplification of the negative control. Reactions were purified using RapidTip2 PCR purification tips (Diffinity Genomic), quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific), and pooled in equimolar concentrations to generate 2 libraries for sequencing (Table S1). These libraries were sequenced on 2 runs of an Illumina MiSeq using a v2 500 cycle kit and a 2 × 250 bp paired-end protocol according to the manufacturer's instructions. Unprocessed reads were submitted to the NCBI Sequence Read Archive (BioProject accession number PRJNA496491).

2.4. Sequence processing

Raw sequences were quality-filtered using Trimmomatic v. 0.36 (Bolger et al. 2014) using the following parameters: removal of bases at the start of the read below quality 3 (LEADING:3), removal of bases at the end of a read below quality 3 (TRAILING:3), read scanning with a 4-base sliding window cutting when average quality per base drops below 25 (SLIDINGWINDOW:4:25), and minimum read length of 150 bp (MINLEN:150). Post-filtered sequences were analyzed using the QIIME2 pipeline (qiime2.org; Caporaso et al. 2010) in a workflow based on the Deblur algorithm (Amir et al. 2017), which retains individual sequence variants (SVs) rather than clustering amplicons into operational taxonomic units. Seven samples that contained <1000 reads were removed from all subsequent analyses (Table S1). The final sample set (samples containing >1000 sequences post-Deblur processing) included 12 water column and 141 *B. neritina* samples.

Taxonomy was assigned using a Bayes classifier trained on the 'Silva 132 99% OTUs from the 515FB/806RB region' database (https://docs.qiime2.org/2019.1/data-resources/), and sequences classified as chloroplasts were removed, along with sequences that were unassigned at the domain level. A phylogeny was generated with FastTree 2 and used for calculating phylogeny-based metrics, including weighted and unweighted Unifrac. Samples were rarefied to 5021 reads for statistical analyses based on the approximate plateauing of alpha rarefaction curves generated in QIIME2 (Fig. S2). This step excluded a further 26 *B. neritina* samples, bringing the total count to 115 *B. neritina* data sets and 12 water column data sets (Table S1). *B. neritina* samples were classified as 'symbiont-enriched (+)' if *E. sertula* sequences composed >1% of total amplicons; samples with <1% *E. sertula* sequences were classified as 'symbiont-reduced or absent (–)'. For certain analyses described below, *E. sertula* sequences were removed from the SV table and the data sets (without *E. sertula*) were rarefied to 5012 reads to assess patterns in the non-symbiont community.

2.5. Phylogeny of the *Crenothrix* sp. and *Endozoicomonas* sp. SVs

To assess the accuracy of the taxonomic assignments of the genera *Crenothrix* and *Endozoicomonas*, 2 of the most dominant taxa in the *B. neritina* microbiomes, we built a phylogeny of full-length amplicons (251 bp, including flanking regions that were trimmed during Deblur processing) and reference 16S rRNA gene sequences downloaded from NCBI. We aligned the most abundant *Crenothrix* and *Endozoicomonas* sequence variants from our dataset with 2 reference *Crenothrix polyspora* clone sequences, 2 *Endozoicomonas* sp. isolate sequences, 2 other gammaproteobacterial reference sequences, and an alphaproteobacterial sequence as an outgroup, using MUSCLE (Edgar 2004). We then built a maximum likelihood phylogeny with a General Time Reversible + gamma model and 999 bootstrap replicates in MEGA (Kumar et al. 2016).

2.6. Statistical and Random Forest analyses

We first tested for correlations between latitude, seawater temperature, and the frequency of *E. sertula* in rarefied amplicon datasets using the SciPy library in IPython to generate Spearman unranked correlations. Only *B. neritina* samples were used for this analysis (water samples were excluded). We also tested for correlations between frequencies of *E. sertula* and other prevalent taxa, including *Endozoicomonas* spp. and unclassified *Gammaproteobacteria* spp.

We next assessed how the taxonomic composition of the *B. neritina* microbiome varied based on host and environmental factors, including host genotype (Type S/Type N), temperature, latitude, and collection site, using unconstrained ordination analyses. The SV table consisted of 115 *B. neritina* samples after rarefaction. SV counts were transformed with a variance stabilizing transformation and normalized by sample size using the varianceStabilizingTransformation function in the R DESeq2 library (Love et al. 2014) to obtain a matrix of counts that were approximately homoscedastic. This matrix was used to perform a principal components analysis (PCA) based on pairwise Euclidean distances to identify corresponding variables without *a priori* assumptions. Points on the PCA plot were colored according to metadata variables (host genotype, latitude, salinity, surface water temperature, and *E. sertula* frequency) to assess separation in ordination space. To assess the effect of *E. sertula* frequency on the composition of the *B. neritina* microbiome, we repeated the PCA analyses as described above but using the data set from which *E. sertula* sequences were excluded (see Section 2.4). All ordination analyses were performed in Primer 7 (Quest Research).

Based on the PCA clustering results, we performed a 2-factor permutational multivariate analysis of variance (PERMANOVA) to test the effect of symbiont presence/absence and temperature on both the full and non-symbiont (*E. sertula* sequences excluded) *B. neritina* microbiome. *E. sertula* enrichment (>1% frequency) or absence (0–1% frequency) was assigned as a fixed factor and temperature as a random factor (Tables S2 & S3).

To circumvent the pitfalls of many traditional statistical methods when applied to microbiome data, we applied a machine learning tool, the Random Forest algorithm, to test the importance of variables in shaping *B. neritina* microbiome composition. We tested factors that appeared to have the biggest effect on composition based on PCA ordination (*E. sertula* enrichment/absence and frequency, and temperature) as well as host genotype to determine whether patterns undetected by PCA were apparent in Random Forest. We performed these analyses using rarefied and transformed SV tables (115 *B. neritina* samples). Models based on *E. sertula* presence and frequency were run both with and without *E. sertula* sequences. In QIIME2, the Random Forest algorithm uses 80% of the data with known classification (e.g. '+' or '−') to train the model, and then tests the model against the remaining 20% of the data to determine its accuracy. We used the 'sample-classifier classify-samples' and 'sample-classifier regress-samples' commands to develop models for categorical (presence/absence) and continuous (frequency) *E. sertula* data, respectively. Classifier and regression models were also developed using host genotype as a categorical variable and temperature as a continuous variable, respectively.

2.7. Network analysis

To search for linkages between *E. sertula* and other microbial taxa in *B. neritina* microbiomes, we used the rarefied and transformed SV table (115 *B. neritina* samples) without water samples in a network analysis based on the Maximal Information Coefficient (MIC) between each pair of SVs. We generated the MIC value for each pairwise interaction using Maximal Information-based Nonparametric Exploration (MINE) (Reshef et al. 2011) with the 'equitability' and 'allPairs' parameters. Finally, we applied a cutoff for MIC values corresponding to a p-value < 0.001.

2.8. Comparison with published *B. neritina* microbiome data

We compared our data with 2 recent studies that provided some of the first characterizations of the *B. neritina* microbiome to assess consistencies and discrepancies among geographically separated hosts and different sequencing approaches. The first study presented microbiomes of 15 *B. neritina* colonies collected from coastal China (Li et al. 2019), and we used these data for a direct qualitative comparison of our *B. neritina* microbiomes. The second study applied shotgun metagenomics to generate 7 metagenome-assembled genomes (MAGs) from diverse taxa in *B. neritina* colonies also collected from the East Coast of North America (Miller et al. 2016b), 5 of which had associated 16S rRNA gene sequences. Only 2 of those 5 ('AB1_phaeo' and 'AB1_endozoicomonas') contained the gene region that was amplified in our study. We aligned those 2 sequences with the full-length (251 bp) reads of the top 10 taxa driving each of the Random Forest models (*E. sertula* frequency in the full data set, *E. sertula* frequency in the filtered data set, temperature, and host genotype; Table S4) (Miller et al. 2016b) using MUSCLE (Edgar 2004). These 2 sequences were also included in the phylogenetic analysis described above.

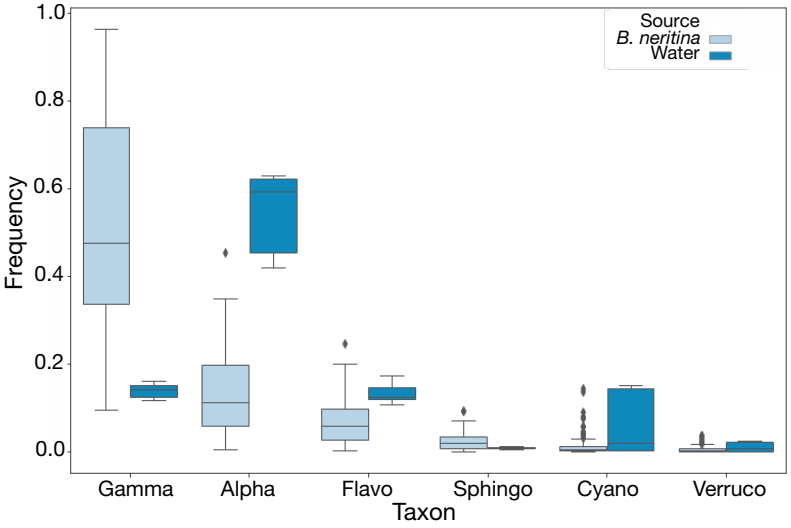

Fig. 1. The composition of host (*Bugula neritina*) microbiomes differed from those of the surrounding water column. The plot shows the frequency of sequences (as a % of total sequences) representing the 6 most abundant subphyla and classes in *B. neritina* and water column microbiomes. Gamma: *Gammaproteobacteria*; Alpha: *Alphaproteobacteria*; Flavo: *Flavobacteriia* (Phylum *Bacteroidetes*); Sphingo: *Sphingobacteriia* (Phylum *Bacteroidetes*); Cyano: *Cyanobacteria* (Phylum *Cyanobacteria*); Verruco: *Verrucomicrobiae* (Phylum *Verrucomicrobia*). The box shows the quartiles of the data set and the whiskers extend to show the rest of the distribution, except for outliers (values outside the inter-quartile range)

3. RESULTS

We collected *Bugula neritina* colonies from 16 sites along the Atlantic coast from Florida to Delaware. The colonies comprised 71 Type S and 70 Type N host genotypes. Because we intentionally selected representatives of each genotype at each site, our analysis does not allow us to test the relationship between host genotype and latitude. Further, as expected, latitude significantly correlated with surface seawater temperature ($p < 0.001$, $R = 0.46$), and temperature is related to season, which varied over the collection period; thus, latitude, temperature, and seasonality are confounded as explanatory variables.

Within this sampling framework, the taxonomic composition of *B. neritina* microbiomes varied considerably. Most colonies were dominated by the subphylum *Gammaproteobacteria*, although the frequency of this group varied widely, from 9.6 to 96% of sequences (Fig. 1). *Alphaproteobacteria* were also abundant in *B. neritina* microbiomes, representing <1% to 45% of sequences and >30% in 6 colonies (Fig. 1). After *Proteobacteria*, *Bacteroidetes* was the second most abundant phylum in *B. neritina* microbiomes, represented mainly by the classes *Flavobacteriia* and *Sphingobacteriia* (Fig. 1). The phyla *Cyanobacteria* and *Verrucomicrobia* were consistently present at low levels (Fig. 1). In 63 (45%) of the colonies, *Gammaproteobacteria* made up >50% of sequences and consistently included members of the orders *Methylococcales*, *Oceanospirillales*, and *Alteromonadales* (to which *Endobugula sertula* belongs). Sequences matching the order *Methylococcales* were dominated by those assigned to the genus *Crenothrix*, a type of filamentous methane-oxidizing bacteria (Stoecker et al. 2006, Oswald et al. 2017). However, a phylogenetic investigation of the full-length (251 bp) 'Crenothrix' amplicon sequence variant showed these sequences were more closely related to the genus *Endozoicomonas* than to *Crenothrix* (Fig. S4 in Supplement 1). When the incorrect classification was accounted for, the genus *Endozoicomonas* was present in all *B. neritina* microbiomes and comprised up to 89% of the community.

Although represented by some of the same broad taxonomic groups, seawater microbiomes were compositionally distinct from host-associated microbiomes (Fig. 1, Fig. S3). The major taxa whose frequencies differed between seawater and *B. neritina* samples included *Alpha-* and *Gammaproteobacteria*, *Flavobacteriia*, *Sphingobacteriia*, and *Cyanobacteria* (Fig. 1). In contrast to the host-associated microbiomes, which were dominated by *Gammaproteobacteria*, the seawater datasets were dominated by *Alphaproteobacteria*, including members of the ubiquitous SAR11 clade. Alpha diversity (Faith's phylogenetic diversity) was also significantly higher in the *B. neritina* microbiomes relative to the water column communities (Fig. S2).

The frequency of the symbiont *E. sertula* varied widely. The symbiont was detected in 89% of *B. neritina* amplicon datasets (126 of 141) at frequencies

Table 1. The frequency of '*Candidatus* Endobugula sertula' (*E. sertula*) sequences (as a percentage of total sequences) varied widely among *Bugula neritina* colony microbiomes

E. sertula frequency (%)	# *B. neritina* colonies
<1	61
1–10	29
11–20	23
21–30	15
31–53	13

from 0.003% to 53% (Table 1, Table S1 in Supplement 2) but was absent from all water column datasets. *E. sertula* frequency was significantly negatively correlated with both latitude ($p < 0.001$, $R = -0.42$) and temperature ($p < 0.02$, $R = -0.19$). Our threshold for classifying an amplicon dataset as 'symbiont-enriched' (*E. sertula* sequences > 1% of total) was consistent with results from the more traditional PCR-based screening for symbiont detection, and we therefore chose 1% as a cut-off value. PCR screening yielded a visible symbiont band in all host samples classified as symbiont-enriched based on amplicon data (Table S1). PCR detection of *E. sertula* in data sets in which symbiont frequency was between 0 and 1% was less reliable (37 out of 47 samples classified as 'symbiont-absent' by PCR contained *E. sertula* sequences). There was one false positive PCR result, in which a host data set with no *E. sertula* sequences was classified as symbiont-enriched (+).

The presence of *E. sertula* strongly influenced overall microbiome diversity. Alpha diversity differed significantly between *B. neritina* microbiomes with low (–, <1%) and higher (+, >1%) *E. sertula* titers (Fig. S2). PCA of microbiome composition also showed that datasets classified as symbiont-enriched (>1% *E. sertula*) clustered apart from those classified as symbiont-absent (<1%), although the clusters are not entirely discrete and the 2 axes comprised <20% of the observed variation (Fig. 2, Fig. S3). This effect was observed both in analyses that included *E. sertula* sequences (Fig. S3) and in those from which *E. sertula* was removed (Fig. 2). The clustering of *B. neritina* microbiomes based on symbiont presence corresponded with differences in seawater temperature at the collection sites (Fig. 3). Indeed, both variables interacted significantly to affect composition (evaluated both with and without symbiont sequences; $p = 0.02$ and $p = 0.04$, respectively; Tables S2 & S3). In contrast, host genotype (Type S/Type N) did not affect overall microbiome composition (Fig. S5).

We further explored the above patterns by generating Random Forest models of microbiome composition based on the following factors: *E. sertula* enrichment (+/–), *E. sertula* frequency, temperature, and *B. neritina* host genotype. When provided the full dataset (*E. sertula* sequences included), the model accurately predicted '+' and '–' samples 93 and 100% of the time, respectively (Table S4). When *E. sertula* sequences were removed from the dataset, the classifier accuracy dropped to 79% for '+' samples and was unable to accurately identify '–' samples (Table S4). However, the regression model displayed strong correlation between predicted and actual *E. sertula* frequencies when both the full dataset was applied ($R^2 = 0.93$, $p < 0.001$; Fig. 4A) and when *E. sertula* sequences were removed ($R^2 = 0.81$, $p < 0.001$; Fig. 4B). In both cases (with and without *E. sertula*), the microbial taxa driving the 2 models were

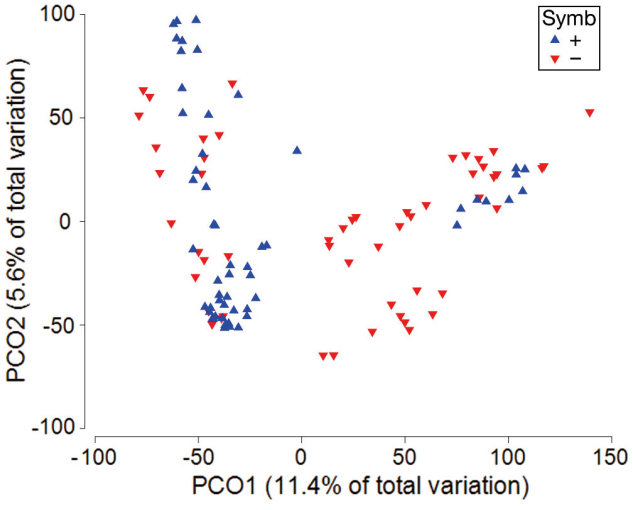

Fig. 2. The composition of host (*B. neritina*) microbiomes varied based on the frequency of *Endobugula sertula* (>1%, '+'; 0–1%, '–'). The plot shows results from a principal components analysis of *B. neritina* microbiomes, and symbols differ based on symbiont (Symb) enrichment/absence. The plot is based on the Euclidean distances among rarefied and transformed data with *E. sertula* sequences removed

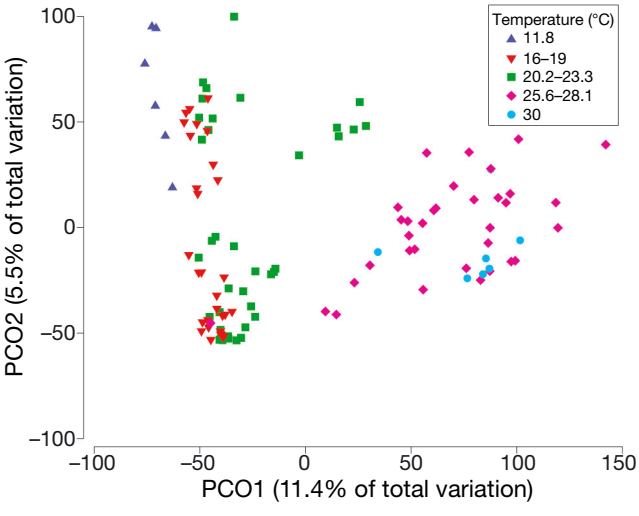

Fig. 3. The composition of host (*B. neritina*) microbiomes varied based on the surface seawater temperature at the time of collection. The plot shows results from a principal components analysis of full *B. neritina* microbiomes based on Euclidean distances among rarefied and transformed data. Points are colored based on temperature groupings as shown in the legend

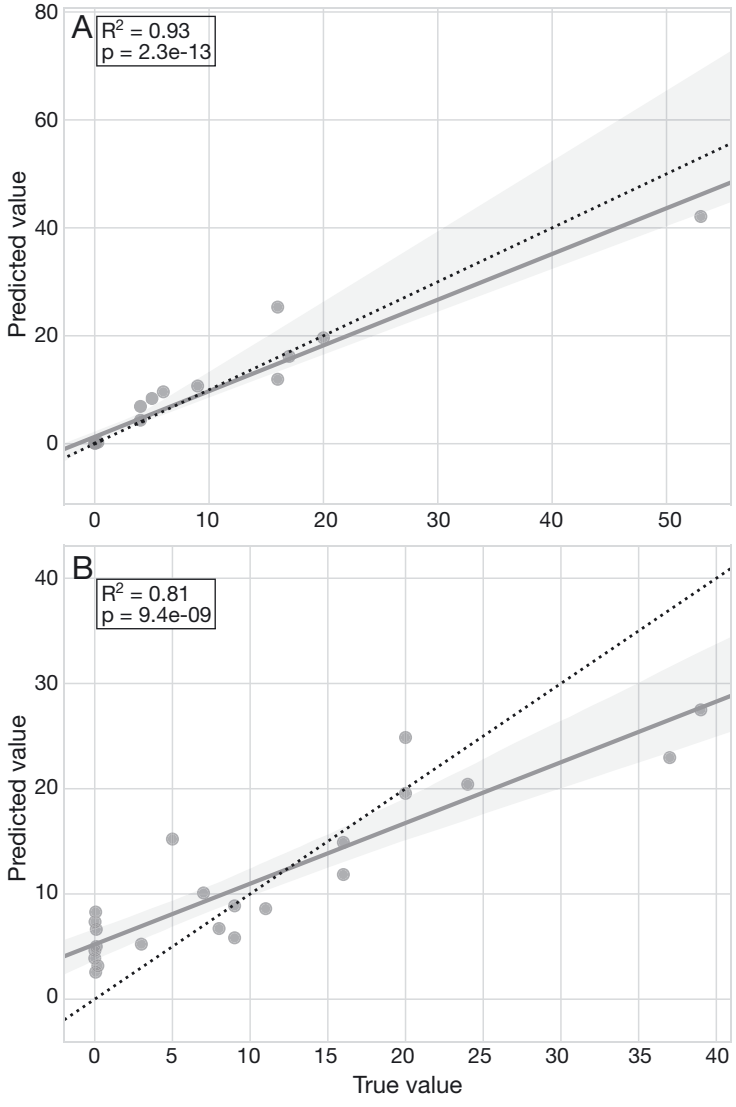

Fig. 4. *B. neritina* microbiome composition can predict the frequency of *E. sertula* in the community. Plots show the relationship between true and predicted *E. sertula* frequency in microbiome datasets based on a Random Forest-based regression analysis using (A) the full sequence dataset and (B) a dataset from which *E. sertula* sequences were excluded. The continuous line represents the linear regression line fitted to the data with 95% confidence intervals (gray shading). The dotted line represents the true 1:1 ratio between predicted and true values for comparison

similar but not identical, and in many cases could only be classified at the domain, class, or order level (Table S5).

The Random Forest analysis confirmed that both temperature and host genotype influence overall microbiome composition. The model trained to predict temperature was highly accurate (R^2 = 0.88, $p < 0.001$) (Fig. 5), and the taxa driving the model included a *Vibrio* sp. and 2 species of *Endozoicomonas*, one of which was also an SV driving the *E. sertula* frequency model using the full microbiome data set (Table S5). Similarly, the model trained to predict host genotype based on microbiome composition displayed high accuracy (92% for Type S; 91% for Type N).

Network analysis showed significant correlations between *E. sertula* and 36 other SVs in the *B. neritina* microbiomes (Table S6). The strongest connection was to an SV that could only be classified to the family *Oceanospirillaceae* (MIC = 0.521). Other SVs that correlated with *E. sertula* included those classified as *Synechococcus* spp., *Kordiimonas* spp., and members of the *Delta-* and *Epsilonproteobacteria*. *E. sertula* was not correlated with SVs of the abundant *Endozoicomonas* genus.

Of the 2 MAG 16S rRNA gene sequences from (Miller et al. 2016b) that contained the region amplified by our primers, we found that the MAG sequence classified as *Phaeobacter* sp. ('AB1_phaeo) appeared in most (87%) of our *B. neritina* samples, where it ranged from 1×10^{-5} to 25% in frequency, and was absent from all water column samples. We found no amplicons with 100% identity to the second available MAG sequence ('AB1_endozoicomonas').

4. DISCUSSION

Since the discovery of the symbiotic relationship between the bryozoan *Bugula neritina* and the gammaproteobacterium '*Ca*. Endobugula sertula' (hereafter *E. sertula*), studies over several decades have illuminated many features of the association, including symbiont localization (Sharp et al. 2007), the molecular basis of bryostatin production (Hildebrand et al. 2004, Sudek et al. 2007), and its potential ecological roles (Lopanik et al. 2004, 2006). However, it has also become clear that there are many caveats and nuances to the symbiosis. Until recently, the 2 species of *B. neritina* found in the western Atlantic, Type S and Type N, were thought to be separated geographically and differentiated by the presence of *E. sertula* (present in Type S, absent in Type N) (McGovern & Hellberg 2003, Lopanik et al. 2004). However, recent findings show that the 2 species can co-occur and that colonies of both Type S and Type N can be found with or without *E. sertula* (Linneman et al. 2014). In the eastern Pacific, the 2 genotypes (Type D and Type S) are separated by depth and show evidence of co-speciation with the

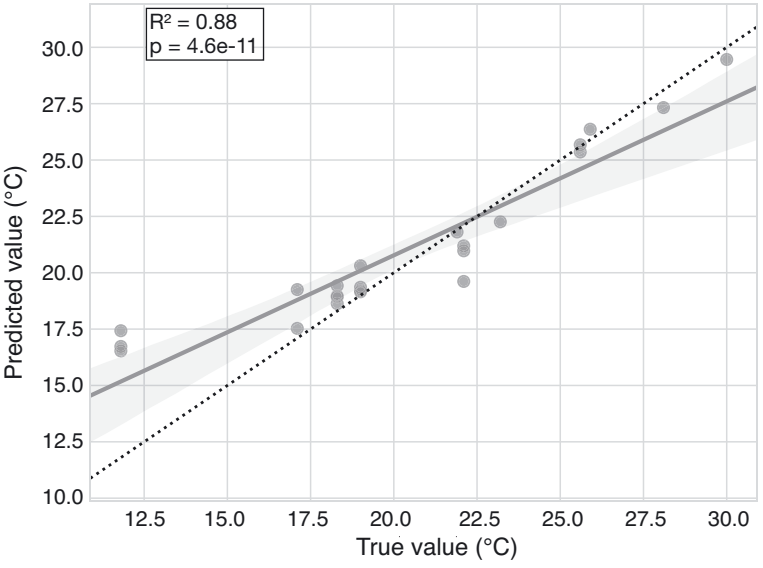

Fig. 5. *B. neritina* microbiome composition can predict seawater temperature. The plot shows the relationship between true and predicted seawater temperature based on a Random Forest-based regression analysis using the full microbiome sequence data. The continuous line represents the linear regression line fitted to the data with 95% confidence intervals (gray shading). The dotted line represents the true 1:1 ratio between predicted and true values for comparison

symbiont (Lim-Fong et al. 2008), while in the western Atlantic, there is little genetic variation between *E. sertula* sequences found in Type S and Type N hosts (Linneman et al. 2014). In this study, we expand on this growing body of knowledge by characterizing the diversity of the complex microbiome associated with *B. neritina*. Understanding this 'non-symbiont' microbiome and how it varies with environmental conditions, host genotype, and the abundance of *E. sertula* is a necessary first step to identifying the full set of interactions that affect the ecology and evolution of this symbiosis.

From our assessment of *E. sertula* frequencies in >140 *B. neritina* microbiomes, it is clear that host colonies should not be categorized as either symbiont-positive or symbiont-negative. Rather, symbiont frequencies vary along a wide spectrum among colonies. Only ~11% of the colony datasets lacked any *E. sertula* sequences, and these included colonies of both sibling species, Type S and N. An additional 46 colonies (nearly one-third of the samples) contained symbiont reads at levels below 1% (Fig. 1, Table S1). These results raise the question of whether there is a minimum frequency above which *E. sertula* confers a fitness advantage to adult *B. neritina* colonies. Alternatively, adult colonies may maintain the symbiont merely to transmit the bacteria to their larvae, and the proportion of *E. sertula* is irrelevant. This scenario raises the possibility that the symbiont-produced defensive bryostatins are not advantageous under all conditions. Indeed, hosting *E. sertula* could impose a physiological cost on the bryozoan colony. If so, it would be advantageous if the host could regulate the number of *E. sertula* cells, for example to increase their abundance during periods of heightened predation pressure on the host. While host regulation of symbiont numbers has been shown in other classic 2-partner symbioses, including plants with nitrogen-fixing rhizobial symbionts (Kiers et al. 2003) and the symbiosis between bioluminescent *Vibrio* bacteria and the Hawaiian bobtail squid (Wier et al. 2010), the potential that *B. neritina* actively regulates *E. sertula* frequency remains untested.

The observed range of *E. sertula* frequency agrees with a recent study on the microbiome of *B. neritina* colonies of unknown genotypes from the southeastern coast of China, in which 15 environmental samples featured *E. sertula* frequencies ranging from 0 to 54% (Li et al. 2019). As in our study, *Proteobacteria* dominated *B. neritina* microbiomes in coastal China, and *Gammaproteobacteria* were particularly prevalent. However, we observed much higher levels of *Endozoicomonas* sp. compared to (Li et al. 2019). Also, we did not detect sequences matching '*Candidatus* Endoacteinascidia sp.', an uncultured ascidian endosymbiont that composed up to 40% of the Chinese *B. neritina* microbiomes. These differences suggest substantial global variation in microbes associated with *B. neritina*, although the symbiosis with *E. sertula* appears cosmopolitan. Similarly, the *Phaeobacter* sp. sequence extracted from the metagenomes in (Miller et al. 2016b), which were also generated from *B. neritina* colonies collected in the western Atlantic, was present in our samples and ranged from 0 to 25% of the *B. neritina* microbiome (data not shown). The metagenomic analysis by Miller et al. (2016b) also generated draft microbial genomes whose 16S rRNA gene sequences contained mismatches with the 'universal' primers used in our study, suggesting that these sequences may not be detectable with standard amplicon protocols and showing the value of addressing host-microbiome questions using a variety of molecular and bioinformatics tools. Our strategy of assessing microbiomes at the individual sequence variant level, rather than clustering sequences into OTUs, allowed for detection of fine-scale microdiversity, particularly of the

genus *Endozoicomonas*, some of whose members were initially incorrectly classified.

The drivers of *B. neritina* microbiome composition are complex and potentially variable, but likely include both environmental and biological factors. Here, temperature/latitude and host genotype had varying effects on *E. sertula* frequency but stronger effects on overall microbiome composition. While temperature correlated with the frequency of *E. sertula* ($p < 0.02$), the dependence between variables was relatively modest ($R = -0.19$). In contrast, temperature correlated with, and could be accurately predicted by, overall *B. neritina* microbiome composition (Figs. 3 & 5, Random Forest analysis). In 3-way PERMANOVA analyses, temperature had the most significant influence on microbiome composition, although it interacted significantly with both symbiont presence and host genotype (Tables S2 & S3). This suggests that host-associated microbial taxa are differentially influenced by temperature. However, we cannot rule out that composition was being shaped by a diverse set of factors that co-vary with temperature. Indeed, both temperature and *E. sertula* frequency also correlated with latitude ($p < 0.001$, $R = 0.46$). This result is consistent with a recent study showing that most colonies at higher latitudes lack the symbiont (Linneman et al. 2014); however, we did not quantify the percentage of symbiont-enriched colonies per site in our study. We also did not observe a significant difference in symbiont frequency between Type S and Type N host genotypes, and genotype did not appear to correlate with microbiome composition in principal components ordination space (Fig. S5). These results contrast with those of the Random Forest analysis, which showed that microbiome composition could nonetheless predict host genotype with high accuracy. These findings highlight the utility of adding machine learning methods to help determine biologically relevant factors in microbiome composition. Here, a number of additional potential factors deserve attention in future studies. Our colonies came from environmental collections and likely varied by colony age, exposure to predation pressure, and other environmental parameters (e.g. nutrient/food availability), all of which were unmeasured. It is therefore likely that an interplay of environmental factors, host biology, and potentially as-yet-unknown ecological influences determines host microbiome composition and whether or not *E. sertula* will be associated with *B. neritina*, and to what degree.

Along with temperature and host genotype, symbiont frequency is an important factor affecting *B. neritina* microbiome composition (Figs. 2 & 4). Microbiomes with lower symbiont frequencies had significantly higher alpha diversity than symbiont-enriched colonies (Fig. S2), suggesting that *E. sertula* constrains the non-symbiont microbiome to a more phylogenetically narrow group of taxa. Further, the interaction effect of symbiont frequency and temperature on microbiome composition was significant (PERMANOVA $p < 0.05$ for both full and non-symbiont data sets; Tables S2 & S3), which supports a role for *E. sertula* in structuring composition. Finally, microbiome composition can be used to accurately predict symbiont frequency (Fig. 4). The accuracy of the Random Forest model generated with the dataset lacking *E. sertula* sequences implies that the non-symbiont community has certain features when *E. sertula* is present and that these features distinguish it from a community from which *E. sertula* is absent. It is possible that frequencies in the non-symbiont microbiome were artificially skewed when *E. sertula* sequences were removed, particularly in samples with high symbiont levels, thereby producing misleading model results. Nevertheless, our approach was appropriate for assessing community composition without *E. sertula* frequency as a confounding variable. As expected, predictions based on the full data set (*E. sertula* included) were driven largely (93%) by *E. sertula* frequency, with none of the remaining taxa contributing >0.3% of the model's predictive power (Table S5). In contrast, predictions of symbiont frequency based on the non-symbiont microbiome data set were driven by a combination of many taxa of low importance (Table S5). These results, combined with the network analysis, suggest there is no single taxon or group of taxa that are tightly linked to *E. sertula* presence in the microbiome. Rather, the symbiont's presence may exert weak and variable effects on diverse groups of bacteria, creating a range of compositional states depending on symbiont abundance.

Our results present a picture of a dynamic host-microbiome relationship that varies significantly with environmental conditions, as well as the abundance of a toxin-producing microbial symbiont. These findings set the stage for exploring the mechanism(s) by which *E. sertula* influences (or indirectly co-varies with) the holobiont community. Several hypotheses can be put forth. First, bryostatins synthesized by *E. sertula* actively antagonize certain co-occurring microbes, with *E. sertula* frequency affecting the level of antagonism. While it is unknown whether *E. sertula* constitutively produces bryostatins while in association with *B. neritina*, the compounds have

been identified in adult colonies (Lopanik et al. 2006, Sharp et al. 2007). Even if production only occurs during certain *B. neritina* life stages or in response to particular stimuli, bryostatins may still prevent the establishment of certain microbial taxa within the *B. neritina* microbiome. Second, *E. sertula* depends on co-occurring microbes for its own survival or growth and therefore may vary in frequency depending on the abundance or presence/absence of these taxa. Our network analysis detected significant co-variation between *E. sertula* and 36 other SVs (Table S6), although no single connection stood out in strength and none of the SVs corresponded to those driving the Random Forest models. This finding seems to preclude the possibility that *E. sertula* co-varies with just one or a few other microbial taxa in a biological partnership or consortium. Third, *E. sertula* directly affects host physiology, which indirectly structures the co-occurring microbiome. Fourth, the age of the *B. neritina* host influences both *E. sertula* frequency and microbiome composition, such that the latter 2 variables appear correlated but are in fact conflated with one another. Finally, environmental factors like temperature and host genotype are the primary determinants of microbiome composition, regardless of *E. sertula* presence or frequency. Of these variables, temperature most strongly predicts microbiome composition, suggesting a major role for abiotic factors in driving overall microbiome structure. These hypotheses are not necessarily mutually exclusive. Differentiating between these and potentially other mechanisms, quantifying their relative importance in shaping the host-bacteria relationship, and determining the chemical exchanges involved in the interactions remain important challenges in understanding this model of marine symbiosis. Meeting these challenges will add to a growing body of knowledge (Donia et al. 2011, Fan et al. 2012, Bourne et al. 2013, Douglas & Werren 2016) suggesting that symbiont-host interactions can only be fully understood in the context of the co-occurring microbiome.

Acknowledgements. This work was supported by the Simons Foundation (award 346253 to F.J.S.), the NSF Advances in Bioinformatics Program (award 1564559 to N.B.L. and F.J.S.), the NSF Biological Oceanography Program (award 1608709 to N.B.L.), and the Teasley Endowment to the Georgia Institute of Technology.

LITERATURE CITED

Amir A, McDonald D, Navas-Molina JA, Kopylova E and others (2017) Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2:e00191-16

Belk A, Xu ZZ, Carter DO, Lynne A, Bucheli S, Knight R, Metcalf JL (2018) Microbiome data accurately predicts the postmortem interval using random forest regression models. Genes (Basel) 9:104

Blazejak A, Erséus C, Amann R, Dubilier N (2005) Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (oligochaeta) from the Peru margin. Appl Environ Microbiol 71:1553–1561

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

Bourne DG, Dennis PG, Uthicke S, Soo RM, Tyson GW, Webster N (2013) Coral reef invertebrate microbiomes correlate with the presence of photosymbionts. ISME J 7: 1459

Cahill PL, Fidler AE, Hopkins GA, Wood SA (2016) Geographically conserved microbiomes of four temperate water tunicates. Environ Microbiol Rep 8:470–478

Cantley AM, Clardy J (2015) Animals in a bacterial world: opportunities for chemical ecology. Nat Prod Rep 32: 888–892

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K and others (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336

Connor N, Barberán A, Clauset A (2017) Using null models to infer microbial cooccurrence networks. PLOS ONE 12: e0176751

Davidson SK, Haygood MG (1999) Identification of sibling species of the bryozoan *Bugula neritina* that produce different anticancer bryostatins and harbor distinct strains of the bacterial symbiont '*Candidatus* Endobugula sertula'. Biol Bull 196:273–280

Davidson SK, Allen SW, Lim GE, Anderson CM, Haygood MG (2001) Evidence for the biosynthesis of bryostatins by the bacterial symbiont '*Candidatus* Endobugula sertula' of the bryozoan *Bugula neritina.* Appl Environ Microbiol 67:4531–4537

Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R (2016) Tiny microbes, enormous impacts: What matters in gut microbiome studies? Genome Biol 17:217

Dishaw LJ, Flores-Torres J, Lax S, Gemayel K and others (2014) The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. PLOS ONE 9:e93386

Donia MS, Fricke WF, Partensky F, Cox J and others (2011) Complex microbiome underlying secondary and primary metabolism in the tunicate-*Prochloron* symbiosis. Proc Natl Acad Sci USA 108:E1423–E1432

Douglas AE, Werren JH (2016) Holes in the hologenome: why host-microbe symbioses are not holobionts. MBio 7: e02099-15

Dubilier N, Mülders C, Ferdelman T, De Beer D and others (2001) Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. Nature 411: 298–302.

Dubilier N, Bergin C, Lott C (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol 6:725–740

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Engel S, Jensen PR, Fenical W (2002) Chemical ecology of marine microbial defense. J Chem Ecol 28:1971–1985

Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, Thomas T (2012) Functional equivalence and evolutionary convergence in complex communities of micro-

bial sponge symbionts. Proc Natl Acad Sci USA 109: E1878–E1887
Flórez LV, Scherlach K, Gaube P, Ross C and others (2017) Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. Nat Commun 8:15172
Freeman CJ, Thacker RW (2011) Complex interactions between marine sponges and their symbiotic microbial communities. Limnol Oceanogr 56:1577–1586
Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ (2017) Microbiome datasets are compositional: and this is not optional. Front Microbiol 8:2224
Harder T (2008) Marine epibiosis: concepts, ecological consequences and host defence. In: Flemming HC, Murthy PS, Venkatesan R, Cooksey K (eds) Marine and industrial biofouling. Springer Series on Biofilms, Vol 4. Springer, Berlin, Heidelberg, p 219–231
Hay M, Beatty D, Stewart F (2017) Chemical ecology: the language of microbiomes. In: National Academies of Sciences, Engineering, and Medicine. Division on Earth and Life Studies; Board on Chemical Sciences and Technology; Chemical Sciences Roundtable (ed) The chemistry of microbiomes: Proceedings of a seminar series, Chap 6. National Academies Press (US), Washington DC, available at www.ncbi.nlm.nih.gov/books/NBK447349/
Haygood MG, Schmidt EW, Davidson SK, Faulkner DJ (1999) Microbial symbionts of marine invertebrates: opportunities for microbial biotechnology. J Mol Microbiol Biotechnol 1:33–43
Hentschel U, Piel J, Degnan SM, Taylor MW (2012) Genomic insights into the marine sponge microbiome. Nat Rev Microbiol 10:641–654
Hildebrand M, Waggoner LE, Liu H, Sudek S and others (2004) bryA: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan *Bugula neritina.* Chem Biol 11:1543–1552
Hussa EA, Goodrich-Blair H (2013) It takes a village: ecological and fitness impacts of multipartite mutualism. Annu Rev Microbiol 67:161–178
Johnson HR, Trinidad DD, Guzman S, Khan Z and others (2016) A machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval. PLOS ONE 11:e0167370
Keough MJ, Chernoff H (1987) Dispersal and population variation in the bryozoan *Bugula neritina.* Ecology 68: 199–210
Kiers ET, Rousseau RA, West SA, Denison RF (2003) Host sanctions and the legume-rhizobium mutualism. Nature 425:78–81
Kott P, Parry DL, Cox GL (1984) Prokaryotic symbionts with a range of ascidian hosts. Bull Mar Sci 34:308–312
Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120
Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
Lee J, Sison-Mangus M (2018) A Bayesian semiparametric regression model for joint analysis of microbiome data. Front Microbiol 9:522
Li H, Mishra M, Ding S, Miyamoto MM (2019) Diversity and dynamics of '*Candidatus* Endobugula' and other symbiotic bacteria in Chinese populations of the bryozoan, *Bugula neritina.* Microb Ecol 77:243–256
Lim-Fong GE, Regali LA, Haygood MG (2008) Evolutionary relationships of '*Candidatus* Endobugula' bacterial symbionts and their *Bugula* bryozoan hosts. Appl Environ Microbiol 74:3605–3609
Linneman J, Paulus D, Lim-Fong G, Lopanik NB (2014) Latitudinal variation of a defensive symbiosis in the *Bugula neritina* (Bryozoa) sibling species complex. PLOS ONE 9: e108783
Lopanik NB (2014) Chemical defensive symbioses in the marine environment. Funct Ecol 28:328–340
Lopanik N, Lindquist N, Targett N (2004) Potent cytotoxins produced by a microbial symbiont protect host larvae from predation. Oecologia 139:131–139
Lopanik NB, Targett NM, Lindquist N (2006) Ontogeny of a symbiont-produced chemical defense in *Bugula neritina* (Bryozoa). Mar Ecol Prog Ser 327:183–191
López-Legentil S, Song B, Bosch M, Pawlik JR, Turon X (2011) Cyanobacterial diversity and a new *Acaryochloris*-like symbiont from Bahamian sea-squirts. PLOS ONE 6:e23938
Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550
Mackie JA, Keough MJ, Christidis L (2006) Invasion patterns inferred from cytochrome oxidase I sequences in three bryozoans, *Bugula neritina, Watersipora subtorquata*, and *Watersipora arcuata.* Mar Biol 149:285–295
Mainali KP, Bewick S, Thielen P, Mehoke T and others (2017) Statistical analysis of co-occurrence patterns in microbial presence-absence datasets. PLOS ONE 12: e0187132
McFall-Ngai MJ (2014) The importance of microbes in animal development: lessons from the squid-*Vibrio* symbiosis. Annu Rev Microbiol 68:177–194
McGovern TM, Hellberg ME (2003) Cryptic species, cryptic endosymbionts, and geographical variation in chemical defences in the bryozoan *Bugula neritina.* Mol Ecol 12: 1207–1215
Miller IJ, Vanee N, Fong SS, Lim-Fong GE, Kwan JC (2016) Lack of overt genome reduction in the bryostatin-producing bryozoan symbiont '*Candidatus* Endobugula sertula'. Appl Environ Microbiol 82:6573–6583
Miller IJ, Weyna TR, Fong SS, Lim-Fong GE, Kwan JC (2016b) Single sample resolution of rare microbial dark matter in a marine invertebrate metagenome. Sci Rep 6: 34362
Min S, Lee B, Yoon S (2017) Deep learning in bioinformatics. Brief Bioinform 18:851–869
Moss C, Green DH, Pérez B, Velasco A, Henríquez R, McKenzie JD (2003) Intracellular bacteria associated with the ascidian *Ecteinascidia turbinata*: phylogenetic and in situ hybridisation analysis. Mar Biol 143:99–110
Newton ILG, Sheehan KB, Lee FJ, Horton MA, Hicks RD (2013) Invertebrate systems for hypothesis-driven microbiome research. Microbiome Sci Med 1:1
Oh DC, Poulsen M, Currie CR, Clardy J (2009) Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. Nat Chem Biol 5:391–393
Oswald K, Graf JS, Littmann S, Tienken D and others (2017) *Crenothrix* are major methane consumers in stratified lakes. ISME J 11:2124–2140
Oudah M, Henschel A (2018) Taxonomy-aware feature engineering for microbiome classification. BMC Bioinformatics 19:227

Patin NV, Kunin V, Lidström U, Ashby MN (2013) Effects of OTU clustering and PCR artifacts on microbial diversity estimates. Microb Ecol 65:709–719

Pollock J, Glendinning L, Wisedchanwet T, Watson M (2018) The madness of microbiome: attempting to find consensus 'best practice' for 16S microbiome studies. Appl Environ Microbiol 84:e02627-17

Puglisi MP, Sneed JM, Sharp KH, Ritson-Williams R, Paul VJ (2014) Marine chemical ecology in benthic environments. Nat Prod Rep 31:1510–1553

Rath CM, Janto B, Earl J, Ahmed A and others (2011) Metaomic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. ACS Chem Biol 6:1244–1256

Reshef DN, Reshef YA, Finucane HK, Grossman SR and others (2011) Detecting novel associations in large data sets. Science 334:1518–1524

Reveillaud J, Maignien L, Eren MA, Huber JA, Apprill A, Sogin ML,Vanreusel A (2014) Host-specificity among abundant and rare taxa in the sponge microbiome. ISME J 8:1198–1209

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J (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

Schmitt S, Tsai P, Bell J, Fromont J and others (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. ISME J 6:564–576

Sharp KH, Davidson SK, Haygood MG (2007) Localization of *Candidatus* Endobugula sertula and the bryostatins throughout the life cycle of the bryozoan *Bugula neritina.* ISME J 1:693–702

Stoecker K, Bendinger B, Scoehning BR, Nielsen PH and others (2006) Cohn's Crenothrix is a filamentous methane oxidizer with an unusual methane monooxygenase. Proc Natl Acad Sci USA 103:2363–2367

Sudek S, Lopanik NB, Waggoner LE, Hildebrand M and others (2007) Identification of the putative bryostatin polyketide synthase gene cluster from '*Candidatus* Endobugula sertula', the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina.* J Nat Prod 70:67–74

Thorsen J, Brejnrod A, Mortensen M, Rasmussen MA and others (2016) Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. Microbiome 4:62

Tsilimigras MCB, Fodor AA (2016) Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann Epidemiol 26:330–335

Wadsworth WD, Argiento R, Guindani M, Galloway-Pena J, Shelburne SA, Vannucci M (2017) Erratum to: An integrative Bayesian Dirichlet-multinomial regression model for the analysis of taxonomic abundances in microbiome data. BMC Bioinformatics 18:185

Weiss S, Xu ZZ, Peddada S, Amir A and others (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5:27

Wier AM, Nyholm SV, Mandel MJ, Massengo-Tiasse RP and others (2010) Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. Proc Natl Acad Sci USA 107:2259–2264

Zhou J, Wu L, Deng Y, Zhi X and others (2011) Reproducibility and quantitation of amplicon sequencing-based detection. ISME J 5:1303–1313

Editorial responsibility: Mya Breitbart, St. Petersburg, Florida, USA

Submitted: October 29, 2018; Accepted: February 20, 2019
Proofs received from author(s): April 24, 2019