Title: Microbiome divergence of marine gastropod species separated by the Isthmus of Panama

Authors: Alexander T. Neu^{1,2*}, Mark E. Torchin², Eric E. Allen^{3,4} and Kaustuv Roy¹

1 - Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA

2 - Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama

3 - Section of Molecular Biology, Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA

4 - Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

*Corresponding author: Alexander T. Neu, Section of Ecology, Behavior and Evolution,
University of California San Diego, 9500 Gilman Drive, MC 0116, La Jolla, CA 92093
Email: aneu@ucsd.edu

Keywords

host-microbiome; Isthmus of Panama; geminate species; divergence; vicariance

Abstract

The rise of the Isthmus of Panama ~3.5 mya separated populations of many marine organisms, which then diverged into new geminate sister species currently living in the Eastern Pacific Ocean and Caribbean Sea. However, we know very little about how such evolutionary divergences of host species have shaped their microbiomes. Here, we compared the microbiomes

of whole-body and shell-surface samples of geminate species of marine gastropods in the genera *Cerithium* and *Cerithideopsis* to those of congeneric outgroups. Our results show that the effects of the Isthmus on microbiome composition varied among host genera and between sample types within the same hosts. In the whole-body samples, microbiome compositions of geminate species pairs in the focal genera tended to be similar, likely due to host filtering, although the strength of this relationship varied among the two groups and across similarity metrics. Shell-surface communities showed contrasting patterns, with co-divergence between the host taxa and a small number of microbial clades evident in *Cerithideopsis*, but not *Cerithium*. These results suggest that (i) the rise of the Isthmus of Panama affected microbiomes of geminate hosts in a complex and clade-specific manner and (ii) host-associated microbial taxa respond differently to vicariance events than the hosts themselves.

Introduction

The formation of the Isthmus of Panama (IP) ~3.5 mya [1,2] led to a mixing of terrestrial faunas between North and South America, commonly known as the Great American Biotic Interchange [3,4]. In the marine environment, however, the rise of the Isthmus separated populations of many species, which then evolved to form geminate species pairs [5], with one currently inhabiting the nutrient-rich upwelling zone of the Eastern Pacific Ocean and the other inhabiting the warmer, more oligotrophic Caribbean Sea [6]. Geminate species pairs provide a powerful set of natural evolutionary experiments that have been used to investigate how closely related species evolve under different environmental conditions [7–9] and to better constrain rates of molecular evolution [e.g., 1,10,11]. However, it remains unclear how the microbiomes of geminate hosts have diverged in response to this vicariance event. Although there is a framework

for examining how biogeographic isolation and environmental change may structure the microbiomes of geminate hosts [12], host-associated microbial divergence across the IP has so far only been studied using eggs of geminate species of echinoids [13].

Tests of concordance between divergence patterns of microbiomes and those of their hosts rely on correlations between the phylogenetic relatedness of host taxa and the similarity of their microbiomes, a pattern termed “phylosymbiosis” [14,15]. Under this model, geminate taxa are expected to harbor microbial communities that are significantly more similar to each other than to more distantly related hosts currently living in the same environment. This pattern may arise either as the result of co-phylogenetic divergence between hosts and their associated microbes, host filtering, or both [15]. In the first case, geminate hosts and their microbiomes (or a subset thereof) are expected to diverge in concert with one another after the closure of the IP, resulting in co-phylogenetic relationships between the geminate host clades and associated microbial taxa. In the second case, aspects of the host environment such as morphology, immune function and diet (e.g., nutrient and metabolite availability) may exclude certain microbial taxa while providing highly suitable conditions for others [16]. Since closely related hosts, such as geminates, often share many traits, they would be predicted to select for more compositionally similar microbial communities. While they are distinct processes, co-speciation and host filtering are not mutually exclusive pathways to phylosymbiosis, and have been shown to act in concert to generate concordance between host evolutionary history and microbiome composition in some host groups, such as corals [17].

Here we use geminate species from two genera of marine gastropods, *Cerithium* and *Cerithideopsis*, along with congeneric outgroups in each case, to test for phylosymbiosis resulting from vicariance due to the rise of the IP. Among our focal host species, *Cerithium*

lutosum and *Cerithium stercusmuscarum* form a well-supported geminate species pair, with *Cerithium atratum* as an outgroup [11; Fig. 1]. *Cerithideopsis mazatlanica* and *Cerithideopsis valida* diverged in the Eastern Pacific after the rise of the IP and are a geminate clade to *Cerithideopsis pliculosa* in the Caribbean, with the Eastern Pacific *Cerithideopsis montagnei* as a congeneric outgroup [11; Fig. 1]. Both outgroups were sampled in the same geographic locations as our geminate taxa (Fig. 1). The *Cerithium* species used here are typically found in near-shore environments with close proximity to mangroves. *Cerithium stercusmuscarum* largely occupies rocky shores in the Eastern Pacific, while its geminate *Cerithium lutosum* and the outgroup *Cerithium atratum* are found in coral rubble or sandy flats in the Caribbean. The *Cerithideopsis* species sampled here all occupy shallow, muddy environments and associated mangrove habitats [11]. Though all of these species forage in the mud, *Cerithideopsis montagnei* differs from its congeners by climbing mangrove roots during high tides.

For each host, we used soft body tissues (hereafter “whole-body”) as well as swabs of the shell-surfaces of multiple individuals (Table S1) to test the hypothesis that geminate species pairs would harbor microbiomes that are both compositionally and phylogenetically more similar to one another than to their sympatric congeners.

Methods

Sampling:

We sampled multiple individuals of each of seven different marine gastropod species from our focal genera at four sites across the IP (Figure 1, Table S1). We swabbed the shell-surface of each individual in the field using a sterile 152mm swab (Grenier Bio-One, Frickenhausen, Germany), avoiding any visible patches of mud. At each site, we also sampled

the top 3 cm of sediment in a sterile 50 mL conical tube and collected two replicate seawater samples by filtering 500mL of surface seawater through 0.22µm Sterivex filters (MilliporeSigma, Burlington, MA, USA). Samples were transported on ice to the Naos Marine Laboratory at the Smithsonian Tropical Research Institute (STRI), where they were kept at -20°C. Subsequently, the whole-body of each individual was removed from the shell, triple-rinsed with 95% EtOH and preserved at -20°C. Samples were later shipped to the University of California, San Diego for DNA extraction and sequencing.

Sequencing:

We used the Qiagen Blood and Tissue kit (Qiagen GmbH, Hilden, Germany) to extract DNA from our samples, following the manufacturer's instructions. We used the protocol of [18] to amplify and purify the V4 region of the 16S rRNA gene. The resulting amplicons were sequenced on the MiSeq platform (Illumina, 2x250bp output) at the DNA Technologies Core, University of California Davis. In order to verify host identities, we also amplified the cytochrome oxidase *c* subunit I (COI) gene from one individual of each host species using the PCR protocol of [11]. The PCR products were cleaned using an Exo-SAP approach and Sanger sequenced at Eton Bioscience, Inc. (San Diego, CA, USA).

Data Analyses:

We used 'DADA2' v1.18.0 [19] in R v4.0.3 [20] to process the 16S rRNA gene sequences. The resulting amplicon sequence variants (ASVs) were assigned taxonomy using the Silva database v138 [21] and all sequences not matching Bacteria or Archaea, as well as any chloroplasts and mitochondria, were removed. We aligned the remaining sequences using MAFFT v7.310 [22] and constructed microbial phylogenies for further analyses with FastTree v2.1.11 [23] using the GTR + CAT model and midpoint rooting.

For each host genus, we tested for differences in alpha diversity among the hosts using two sets of metrics. First, we used the entire, unrarefied microbial dataset to estimate Hill numbers with q of 0 and 1 for each sample using the package `hilldiv` v1.5.1 [24]. Using Hill numbers with $q = 0$ does not account for relative abundance and is equivalent to the number of observed ASVs, while $q = 1$ weights each ASV by its relative abundance and is equivalent to the exponential of Shannon's index (H') [25]. Second, we generated 100 ASV tables, each rarefied to 3,000 sequences per sample, using the package `metagMisc` v0.0.4 [26]. We then calculated the number of observed ASVs and Shannon's H' for each sample from each ASV table and used the average values to account for any differences in individual random draws. In both cases, we tested for differences between host species using Kruskal-Wallis tests and further investigated significant differences using Dunn's tests in the package `dunn.test` v1.3.5 [27].

All subsequent analyses used both the unrarefied dataset, normalized via total sum scaling (TSS) [30], as well as the 100 rarefied datasets (3,000 sequences/sample), unless otherwise specified. We tested for differences in microbiome composition between host species using permutational multivariate analyses of variance (PERMANOVAs) in the package `phyloseq` v1.34.0 [29] with Bray-Curtis Dissimilarity (BCD) and unweighted UniFrac Distance [30] as our beta diversity metrics. We use both of these metrics in our analyses as they emphasize different attributes; BCD includes relative abundance information, while unweighted UniFrac includes a measure of phylogenetic relatedness of the sampled microbial taxa. We also conducted PERMANOVAs on the TSS dataset to determine whether the host-associated microbiomes were significantly different from the microbial communities in the sediment and water samples. We tested for homogeneity of dispersions between species and sample types using the `betadisper` function in `vegan` v2.5-7 [31] for all iterations.

To investigate the role of host phylogenetic distance on microbiome composition more directly, we pooled together all microbial sequences from each individual host species. Since *Cerithium stercusmuscarum* was collected from two sites, we pooled all sequences from this species together but also pooled each site individually. We then clustered the samples using unweighted paired group with arithmetic means (UPGMA), using both BCD and unweighted UniFrac, and the TSS and rarefied datasets. To test the robustness of clustering, we calculated Dunn's Index for each node using the package cIValid v0.7 [32]. We used the normalized Robinson-Foulds metric (nRF) in the package phangorn v2.5.5 [33] to test for congruency between the UPGMA plots and the host phylogeny generated from COI sequences using FastTree v2.1.11 [23] with the GTR + γ model. For some samples, relationships varied across different rarefied datasets, so we generated boxplots showing the BCD and unweighted UniFrac distance between species across all 100 rarefied datasets.

To identify microbial clades contributing to similarities between geminate host species, we generated network plots using sparse partial least squares discriminant analysis (SPLS-DA) in the package mixOmics v6.14.0 [34] at taxonomic levels ranging from ASV to class. This analysis requires a log-ratio transformation and was only conducted on the unrarefied dataset using a centered log-ratio transformation. Microbial taxa that were positively correlated with the geminate species ($r > 0.6$) or negatively correlated with the congeneric outgroups ($r < -0.6$) were considered to contribute to phyllosymbiosis resulting from the rise of the IP [35]. To further investigate whether the taxa contributing to the differentiation between the geminate clades and outgroups also demonstrated potential co-phylogenetic relationships with the hosts, we used the parafit function in ape v5.4.1 [36].

Results

Alpha Diversity

Cerithium: The whole-body microbiomes of all three species of *Cerithium* showed similar Hill numbers at $q = 0$ (Fig. 2a) and observed ASVs after rarefying (Fig. S1a), and there were no significant differences in total ASV richness between the geminate and outgroup species (Tables S2, S3). However, at $q = 1$, the Hill numbers of *C. atratum* were significantly higher than those of the geminate pair *C. lutosum* and *C. stercusmuscarum* (Fig. 2c, Table S2), as confirmed by Shannon's H' values in the rarefied dataset (Fig. S2a, Table S3). In the *Cerithium* shell-surface microbiome, no significant differences were found between geminate and outgroup species using any of the alpha diversity metrics (Figs. 2b, 2d, S1b, S2b, Tables S2, S3).

Cerithideopsis: In this group, whole-body samples showed significant differences in Hill numbers at $q = 0$ (Fig. 2e; Table S2) and marginally significant differences in observed ASV richness (Fig. S1c, Table S3). This was driven by the low diversity of *C. montagnei*, which was significantly different from *C. mazatlanica* and *C. pliculosa* (Dunn's tests: $p < 0.05$), but not *C. valida*. Interestingly, the geminate and outgroup species did not show significant differentiation at $q = 1$ (Fig. 2f; Table S2) or using Shannon's index (Fig. S2c, Table S3), largely due to the presence of a single *C. montagnei* sample with high diversity (Fig. 2e, Supplement 2). *C. montagnei* showed significantly lower diversity than *C. mazatlanica* and *C. pliculosa* (Dunn's test: $p = 0.0084$ and $p = 0.005$, respectively), but not *C. valida*, when this high diversity individual was removed. In the *Cerithideopsis* shell-surface swabs, Hill numbers at $q = 0$ did not show significant differences between the host species (Fig. 2g; Table S2), again due to a single high diversity sample of *C. montagnei* (Fig. 2g, Supplement 2), though the rarefied observed ASV count showed marginally significant differentiation (Fig. S1d, Table S3). Removing the

high diversity sample from the Hill number analyses resulted in significant differences between *C. montagnei* and the higher diversity *C. pliculosa* and *C. mazatlanica* (Dunn's test: $p = 0.0129$ and $p = 0.0135$, respectively), but not *C. valida*. When $q = 1$, significant differentiation (Table S2) was driven by the difference between *C. pliculosa* and *C. montagnei* (Dunn's test: $p = 0.0244$), as *C. mazatlanica* and *C. valida* were not significantly different from either of these hosts (Fig. 2h). These results were confirmed by Shannon's H' from the rarefied dataset (Fig. S2d, Table S3).

Taxonomic Composition

Cerithium: Whole-body samples of individual host species were largely comprised of Actinobacteriota, Alphaproteobacteria, Cyanobacteria and Gammaproteobacteria (Fig. 3a). Interestingly, the most common group found in the geminate species *C. lutosum* and *C. stercusmuscarum* was "Unclassified Bacteria", as determined by the Silva database. A single ASV within this group (ASV10677) comprised, on average, greater than 20% of the relative abundance of the microbiome in these hosts. The closest named relative to this microbe is a member of the phylum Chlorobi, *Chloroherpeton thalassium* (81.89% sequence similarity using NCBI BLASTn). The two populations of *C. stercusmuscarum* whole-body samples were differentiated by the relative abundances of the phyla Crenarchaeota, which was higher at Bique, and Cyanobacteria, which was higher at Punta Culebra (Fig. 3a). *Cerithium* shell-surface microbiomes were comprised mainly of Alphaproteobacteria, Bacteroidota, Cyanobacteria, Gammaproteobacteria and Planctomycetota (Fig. 3b).

Cerithideopsis: Whole-body samples of these host species contained high relative abundances of Alphaproteobacteria, Bacteroidota and Gammaproteobacteria. However, *C.*

montagnei showed higher abundance of Actinobacteriota while the geminate clade including *C. mazatlanica*, *C. valida* and *C. pliculosa* had increased relative abundances of Planctomycetota and Verrucomicrobiota (Fig. 3c). *Cerithideopsis* shell-surface microbiomes showed high relative abundances of Alphaproteobacteria, Bacteroidota, Gammaproteobacteria and Planctomycetota (Fig. 3d). However, *C. montagnei* shell-surfaces showed higher levels of Actinobacteriota and a notable decrease in Cyanobacteria compared to the geminate species (Fig. 3d).

Microbial community compositions of sediment and water samples varied among the collection sites, but nearly all were dominated by members of the class Gammaproteobacteria and were compositionally distinct from the microbiomes of all host species (Fig. 3e).

Host evolutionary history and microbiome composition

Results of PERMANOVAs show that in both host genera, the microbiome compositions of individual host species are distinct, regardless of tissue type, normalization method or beta diversity metric used (Tables S4, S5), and significantly different from the environmental samples (Table S6).

Cerithium: UPGMA clustering of whole-body samples was consistent with the phylosymbiosis hypothesis when using BCD and either normalization method, with microbiomes of geminate species *C. lutosum* and *C. stercusmuscarum* being compositionally more similar to each other than to *C. atratum* (Figs. 4a, S3a). However, when *C. stercusmuscarum* individuals were separated by collection site, *C. stercusmuscarum* samples from Punta Culebra were more similar to *C. lutosum* than to the *C. stercusmuscarum* samples from Bique (Fig 5a). Analyses using unweighted UniFrac showed a different pattern, with *C. lutosum* microbiomes clustering closest to those from *C. atratum* with the *C. stercusmuscarum* populations as outgroups (Figs.

4a, 5a, S4a). UPGMA clustering of shell-surface samples from *Cerithium* did not match the host phylogeny, regardless of distance metric or normalization method (Figs. 4b, 5b, S3b, S4b). Instead, microbiome composition appeared to be determined by the local environment, as the Caribbean species *C. lutosum* and *C. atratum* were more similar to each other than to either population of the Eastern Pacific *C. stercusmuscarum*.

Cerithideopsis: When using BCD and the TSS dataset, UPGMA clustering of host microbiome compositions mirrored the host phylogeny (Fig. 4c). However, BCD of samples rarefied to 3,000 sequences produced mixed results depending on which of the 100 resampled ASV tables was used (Fig. 4c). In 71% of the datasets, results were consistent with the hypothesis of phyllosymbiosis, but in the other 29% of cases *C. valida* microbiomes were shown to be more similar to those of *C. montagnei* than to those of its sister species, *C. mazatlanica*, or the geminate *C. pliculosa*. This appears to be driven by the compositional dissimilarity between individuals of *C. valida*, as the BCD between this species and other *Cerithideopsis* encompass a large range of values (Fig. S3c). Clustering the TSS dataset using unweighted UniFrac distance does not support the phyllosymbiosis hypothesis. However, the rarefied OTU tables once again presented split patterns, with 86% of the results being consistent with predicted phyllosymbiosis (Fig. 4c, S4c). In contrast to the whole-body samples, shell-surface microbiomes of the *Cerithideopsis* geminate clade consistently clustered together, with *C. montagnei* as the outgroup, regardless of the distance metric or normalization method used (Figs. 4d, S3d, S4d).

Unique microbial associates of geminate hosts

Cerithium: Only one ASV (ASV10677) significantly differentiated whole-body samples of *C. lutosum* and *C. stercusmuscarum* from *C. atratum* (Fig. S5). As discussed above, this ASV

was the most dominant member of the *C. lutosum* and *C. stercusmuscarum* communities but was not highly abundant in the *C. atratum* microbiome (Fig. S5). In the shell-surface communities, no individual taxa could significantly differentiate the geminate pair from *C. atratum*.

Cerithideopsis: These samples contained several microbial taxa that significantly differentiated the geminate clade of *C. mazatlanica*, *C. valida* and *C. pliculosa* from the outgroup, *C. montagnei*. In the whole-body, two genera and four families were significant drivers of this differentiation (Fig. S6, Table S7), while shell-surface swabs contained four genera, five families, four orders and two classes of microbes which separated the geminates from *C. montagnei* (Fig. S7, Table S7).

Host-microbe co-phylogenetic relationships

Cerithium: We were unable to test for potential co-phylogenetic relationships between *Cerithium* samples and the microbial clades driving differentiation between geminates and the outgroup, as only a single ASV was responsible for this pattern in the whole-body samples, and none differed significantly between the geminate and outgroup shell-surface samples (Table S7).

Cerithideopsis: Whole-body samples of this genus did not show any co-phylogenetic relationships across multiple levels of microbial taxonomy (Table S7). In the shell-surface swabs, however, two families (A4b and Thermoanaerobaculaceae), two orders (SBR1031 and Thermoanaerobaculales) and the class Gammaproteobacteria all showed significant co-phylogenetic relationships with the host (Table S7).

Discussion

Whole-body and shell-surface samples of tropical marine gastropod hosts from the genera *Cerithium* and *Cerithideopsis* house diverse, complex and host-specific microbial communities. The richness of ASVs was surprisingly similar across host genera in both the whole-body and shell-surface samples, with the exception of relatively low diversity in *Cerithideopsis montagnei*. This may reflect the unique ecology of *C. montagnei*, which forages in the mud like the other *Cerithideopsis* species, but climbs mangroves during high tide, potentially reducing its exposure to taxa present in the seawater that surrounds its congeners, resulting in reduced diversity. Differences in microbial diversity between the geminate taxa and their congeneric outgroups are most evident using diversity metrics that account for relative abundances. The geminate pair *Cerithium lutosum* and *C. stercusmuscarum* had significantly lower Hill numbers at $q = 1$ and Shannon's H' compared to their congeneric outgroup, *C. atratum*. These metrics suggest that fewer taxa dominate the microbiomes of these geminate species as compared to *C. atratum*. Within *Cerithideopsis*, the trend is reversed, with *C. montagnei* harboring reduced diversity compared to its geminate congeners using these metrics.

Compositionally, the whole-body and shell-surface microbiomes of *Cerithium* and *Cerithideopsis* species were distinct from one another but were comprised of taxa commonly associated with marine invertebrate hosts. Many of the dominant microbial clades in the whole-body samples, including Alphaproteobacteria, Gammaproteobacteria and Planctomycetota, have been found previously in marine molluscan tissues [18,37]. The shell-surfaces of *Cerithium* and *Cerithideopsis* species also contained a number of clades previously found on shells of other marine invertebrates, including Alphaproteobacteria, Bacteroidota and Cyanobacteria [38,39]. In contrast, our environmental samples were largely dominated by Gammaproteobacteria, which are commonly found in seawater and intertidal sediment communities [40,41]. Thus the microbiome

compositions of all of our host species are distinct from that of the ambient environment, as is common in marine invertebrates [e.g., 36,43].

Similarities of microbiomes of geminate species varied considerably depending on the host genus, the type of sample and the metric used. *Cerithium* whole-body samples conformed to a phylosymbiosis pattern when using BCD, as *C. lutosum* was more similar to its geminate *C. stercusmuscarum* than to *C. atratum* samples from the same environment. However, when using unweighted UniFrac, *C. lutosum* clustered more closely with *C. atratum* than *C. stercusmuscarum*. These results suggest that the relative abundances of certain microbial taxa are more similar between the geminate species than the congeneric outgroup, while microbial taxa occurring on the same side of the Isthmus tend to be closer evolutionarily. A single ASV, ASV10677, was primarily responsible for the similarity between the geminates *Cerithium lutosum* and *C. stercusmuscarum* observed in the BCD dataset. Although this ASV was present in all three *Cerithium* hosts, its relative abundance was nearly an order of magnitude higher in the geminates when compared to *C. atratum*. ASV10677 was also found in all of our sediment and water samples, but at <1% relative abundance, suggesting selective enrichment in *C. lutosum* and *C. stercusmuscarum*. The reason for this is unclear, though the closest named relative to this taxon is *Chloroherpeton thalassium*, a green sulfur bacterium [43], suggesting a potential role in sulfur cycling. The *Cerithium* geminate species are found associated with mangrove sediments, where sulfur cycling occurs [44], while *C. atratum* is often found further out on the reef flat.

When the two sampled populations of *Cerithium stercusmuscarum* were analyzed separately, the BCD dataset showed a different pattern. Though the geminate species were separated from the congeneric outgroup, *C. atratum*, *C. stercusmuscarum* populations did not cluster together, potentially due to their positions on opposite sides of the Panama Canal (Fig. 1).

This shows that intraspecific, site-level differences in microbiome composition can be greater than the differences observed between species and should be accounted for in future studies.

Shell-surface microbiomes of *Cerithium* did not show any significant geminate signal, regardless of the metric used. Thus, the ambient environment may have a greater impact on shell-surface microbiome composition in this group than evolutionary history, especially as *C. atratum* and *C. lutosum* were collected from the same site.

In the whole-body samples of *Cerithideopsis*, compositional similarities of geminate hosts varied depending on the method used. In the TSS dataset, geminate species clustered together if BCD was used, but not unweighted UniFrac, a result similar to that in *Cerithium*. In our rarefied datasets, the outcomes were even more variable, with the geminate clade clustering together in 71% and 86% of random draws using the BCD and unweighted UniFrac metrics, respectively. This highlights two important points. First, although rarefaction is commonly used, and often recommended [28,45], for standardization of microbiome data, our results show that in some cases, hypothesis testing using rarefied datasets can produce different results than other normalization approaches. Second, using a single rarefied ASV table to test hypotheses, as is common in microbial ecology, can be problematic given the stochastic nature of individual rarefaction draws. Instead, we recommend using multiple iterations to test the robustness of results using rarefied data. Finally, our network analyses identified multiple taxa that helped differentiate the geminate clade from the outgroup, *C. montagnei*, but none of these showed significant co-phylogenetic relationships with the host group. Taken together, our results suggest that in the *Cerithideopsis* whole-body samples, host filtering and/or host ecology likely drive any observed similarity between the microbiomes of the geminate species. As stated above, the climbing behavior of *C. montagnei* may limit its exposure to seawater microbes thus

differentiating the microbiome composition of this species from its entirely mud-dwelling relatives.

In contrast to the whole-body samples, shell-surface microbiomes of *Cerithideopsis* species consistently mirrored the host phylogeny. Several microbial clades were responsible for this differentiation between the geminate species and *C. montagnei*, some of which showed co-phylogenetic relationships with these hosts. These include members of the microbial orders Thermoanaerobaculales and SBR1031, and the class Gammaproteobacteria.

Thermoanaerobaculales is an order within the phylum Acidobacteriota, recently described from freshwater hot springs [46]. Little is currently known about this clade, but members have been found in high abundance in intertidal sands [47] and are present in many of our seawater and sediment samples, suggesting that these microbes may be transferred from the environment to shell-surfaces. SBR1031 is an order of nitrifying bacteria within the phylum Chloroflexi previously found in association with marine invertebrates [48,49] and the mangrove rhizosphere [50]. Since shell-surfaces are sites of active nitrogen cycling and nitrification in the marine environment [38,39], and members of this order were present in our sediment and seawater samples, this association may also constitute an acquisition from the environment.

Gammaproteobacteria is a highly diverse clade common in the marine environment and many marine hosts [40,51], so in this case the processes driving a co-phylogenetic pattern with *Cerithideopsis* are unclear. In general, it is surprising that significant co-phylogenetic divergence between geminate hosts and their microbiomes was only evident in shell-surface biofilms of *Cerithideopsis*, as these communities are expected to be more influenced by the external environment than the biology of host [16,52].

While the formation of the IP led to predictable patterns of evolutionary divergence of marine animal species on either side, our results show that this geologic event affected host-associated microbiomes in a much more complex and context-dependent manner. Depending on the geminate clade and host tissue sampled, we found evidence for (i) phylosymbiosis driven by environmental filtering, (ii) co-phylogenetic divergence of hosts and their microbial associates and (iii) no discernable impact of host evolutionary history on microbiomes of geminate species. These results could reflect multiple causes. First, an expectation of co-divergence between hosts and their microbiomes due to the rise of the IP assumes that populations of individual microbial taxa were originally present in both oceans and subsequently split by the closure of the IP, leaving representatives on both sides. However, microbiome compositions of marine invertebrate hosts commonly differ substantially between populations of the same host species [e.g., 36,46] as shown by our samples of *Cerithium stercusmuscarum*. If populations of hosts did not share many microbes prior to being separated by the IP, it is perhaps unsurprising that observed phylosymbiosis patterns are driven by a small number of taxa, rather than being a community-wide trend. Second, for co-phylogenetic divergence of hosts and their microbes to be detectable millions of years after a vicariant event, hosts and microbial taxa must diverge at similar rates. However, due to their short generation times and differences in selective pressures between the Caribbean and Eastern Pacific after the closure of the IP, the divergence rates of microbial taxa in separate oceans were likely much higher than those of their hosts. Such differences in evolutionary rates can decrease the phylosymbiosis signal and create closer evolutionary relationships between microbial taxa currently present on the same side of the IP, as we found in multiple cases. Unfortunately, current data limitations make this difficult to test, as ancestral

distributions of microbial taxa remain unknown and calibrating microbial phylogenies also remains challenging [53].

Overall, our results suggest that ecological and evolutionary responses of host-associated microbial communities to major vicariant events such as the rise of the Isthmus of Panama, are complex and taxon-specific. Future comparative analyses of microbiomes of other geminate hosts would be useful for better understanding how the microbiome compositions of marine hosts have been shaped by major geological events.

Acknowledgements

We thank E. Caballero and G. Castellanos-Galindo for assistance with sampling and C. Schloeder, I. Ochoa and R. Collin, for sample storage, shipping and logistical support. ATN was funded by a STRI Short-Term Research Fellowship. All samples were collected in compliance with permits #SC/A-1-19 (collections) and #SEX/A-33-19 (export) from the Ministerio de Ambiente de Panamá.

Data Accessibility Statement

16S rRNA gene sequence data and associated metadata are available from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under Bioproject accession number PRJNA744215. The ASV tables used for analyses will be made available via Dryad.

References

1. O'Dea A *et al.* 2016 Formation of the Isthmus of Panama. *Sci. Adv.* **2**, e1600883.

2. Jaramillo CA. 2018 Evolution of the Isthmus of Panama: biological, palaeoceanographic and palaeoclimatological implications. *Mt. Clim. Biodivers.* , 323–338.
3. Simpson GG. 1980 Splendid isolation: the curious history of South American mammals.
4. Leigh EG, O’Dea A, Vermeij GJ. 2014 Historical biogeography of the isthmus of panama. *Biol. Rev.* **89**, 148–172. (doi:10.1111/brv.12048)
5. Jordan DS. 1908 The Law of Geminate Species. *Am. Nat.* **42**, 73–80.
6. Lessios HA. 2008 The Great American Schism: Divergence of Marine Organisms After the Rise of the Central American Isthmus. *Annu. Rev. Ecol. Evol. Syst.* **39**, 63–91. (doi:10.1146/annurev.ecolsys.38.091206.095815)
7. Lessios HA. 1990 Adaptation and phylogeny as determinants of egg size in echinoderms from the two sides of the Isthmus of Panama. *Am. Nat.* **135**, 1–13. (doi:10.1086/285028)
8. Marko PB, Jackson JBC. 2001 Patterns of morphological diversity among and within arcid bivalve species pairs separated by the Isthmus of Panama. *J. Paleontol.* **75**, 590–606. (doi:10.1017/s0022336000039688)
9. Thacker CE. 2017 Patterns of divergence in fish species separated by the Isthmus of Panama. *BMC Evol. Biol.* **17**, 111. (doi:10.1186/s12862-017-0957-4)
10. Knowlton N, Weigt LA. 1998 New dates and new rates for divergence across the Isthmus of Panama. *Proc. R. Soc. B Biol. Sci.* **265**, 2257–2263. (doi:10.1098/rspb.1998.0568)
11. Miura O, Torchin ME, Bermingham E. 2010 Molecular phylogenetics reveals differential divergence of coastal snails separated by the Isthmus of Panama. *Mol. Phylogenet. Evol.* **56**, 40–48. (doi:10.1016/j.ympev.2010.04.012)
12. Wilkins LGE *et al.* 2019 Host-associated microbiomes drive structure and function of marine ecosystems. *PLoS Biol.* **17**, e3000533. (doi:10.1371/journal.pbio.3000533)

13. Carrier TJ, Lessios HA, Reitzel AM. 2020 Eggs of echinoids separated by the Isthmus of Panama harbor divergent microbiota. *Mar. Ecol. Prog. Ser.* **648**, 169–177. (doi:10.3354/meps13424)
14. Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. 2016 Phyllosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLoS Biol.* **14**, e2000225. (doi:10.1371/journal.pbio.2000225)
15. Lim SJ, Bordenstein SR. 2020 An introduction to phyllosymbiosis. *Proc. R. Soc. B Biol. Sci.* **287**, 20192900. (doi:10.1098/rspb.2019.2900)
16. Mazel F, Davis KM, Loudon A, Kwong WK. 2018 Is Host Filtering the Main Driver of Phyllosymbiosis across the Tree of Life? *mSystems* **3**, e00097-18.
17. Pollock FJ, McMinds R, Smith S, Bourne DG, Willis BL, Medina M, Thurber RV, Zaneveld JR. 2018 Coral-associated bacteria demonstrate phyllosymbiosis and cophylogeny. *Nat. Commun.* **9**, 4921. (doi:10.1038/s41467-018-07275-x)
18. Neu AT, Hughes I V., Allen EE, Roy K. 2021 Decade-scale stability and change in a marine bivalve microbiome. *Mol. Ecol.* **30**, 1237–1250. (doi:10.1111/mec.15796)
19. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016 DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583. (doi:10.1038/nmeth.3869)
20. R Core Team. 2020 R: A language and environment for statistical computing.
21. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013 The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596. (doi:10.1093/nar/gks1219)
22. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7:

Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.

(doi:10.1093/molbev/mst010)

23. Price MN, Dehal PS, Arkin AP. 2010 FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490. (doi:10.1371/journal.pone.0009490)

24. Alberdi A, Gilbert MTP. 2019 Hilldiv: An R package for the integral analysis of diversity based on Hill numbers. *bioRxiv* (doi:10.1101/545665)

25. Alberdi A, Gilbert MTP. 2019 A guide to the application of Hill numbers to DNA-based diversity analyses. *Mol. Ecol. Resour.* **19**, 804–817. (doi:10.1111/1755-0998.13014)

26. Mikryukov V. 2018 metagMisc: Miscellaneous functions for metagenomic analysis. R package version 0.0.4.

27. Dinno A. 2017 dunn.test: Dunn’s test of Multiple Comparisons Using Rank Sums. R package version 1.3.5.

28. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. 2019 Methods for normalizing microbiome data: An ecological perspective. *Methods Ecol. Evol.* **10**, 389–400. (doi:10.1111/2041-210X.13115)

29. McMurdie PJ, Holmes S. 2013 Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217. (doi:10.1371/journal.pone.0061217)

30. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. 2011 UniFrac: An effective distance metric for microbial community comparison. *ISME J.* **5**, 169–172. (doi:10.1038/ismej.2010.133)

31. Oksanen J *et al.* 2020 vegan: Community Ecology Package. R package version 2.5-7.

32. Brock G, Pihur V, Datta S, S. D. 2008 clValid: An R package for cluster validation. *J.*

481 *Stat. Softw.* **25**.

482 33. Schliep KP. 2011 phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593.

483 (doi:10.1093/bioinformatics/btq706)

484 34. Rohart F, Gautier B, Singh A, Lê Cao KA. 2017 mixOmics: an R package for ‘omics

485 feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752.

486 (doi:10.1101/108597)

487 35. Demko AM. 2021 Chemical ecology of marine microbial communities: An assessment of

488 bacterial diversity and dynamics in tropical marine sediments.

489 36. Paradis E, Schliep K. 2019 ape 5.0: an environment for modern phylogenetics and

490 evolutionary analyses in R. *Bioinformatics* **35**, 526–528.

491 37. Neu AT, Allen EE, Roy K. 2019 Diversity and composition of intertidal gastropod

492 microbiomes across a major marine biogeographic boundary. *Environ. Microbiol. Rep.* **11**,

493 434–447. (doi:10.1111/1758-2229.12743)

494 38. Pfister CA, Meyer F, Antonopoulos DA. 2010 Metagenomic profiling of a microbial

495 assemblage associated with the california mussel: a node in networks of carbon and

496 nitrogen cycling. *PLoS One* **5**, e10518. (doi:10.1371/journal.pone.0010518)

497 39. Arfken A, Song B, Bowman JS, Piehler M. 2017 Denitrification potential of the eastern

498 oyster microbiome using a 16S rRNA gene based metabolic inference approach. *PLoS*

499 *One* **12**, e0185071. (doi:10.1371/journal.pone.0185071)

500 40. Sunagawa S *et al.* 2015 Structure and function of the global ocean microbiome. *Science*

501 **348**, 1261359. (doi:10.1126/science.1261359)

502 41. Wang Y, Sheng HF, He Y, Wu JY, Jiang YX, Tam NFY, Zhou HW. 2012 Comparison of

503 the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by

using millions of illumina tags. *Appl. Environ. Microbiol.* **78**, 8264–8271.

(doi:10.1128/AEM.01821-12)

42. Chu ND, Vollmer S V. 2016 Caribbean corals house shared and host-specific microbial symbionts over time and space. *Environ. Microbiol. Rep.* **8**, 493–500. (doi:10.1111/1758-2229.12412)

43. Gibson J, Pfennig N, Waterbury JB. 1984 Chloroherpeton thalassium gen. nov. et spec. nov., a non-filamentous, flexing and gliding green sulfur bacterium. *Arch. Microbiol.* **138**, 96–101. (doi:10.1007/BF00413007)

44. Varon-Lopez M, Dias ACF, Fasanella CC, Durrer A, Melo IS, Kuramae EE, Andreote FD. 2014 Sulphur-oxidizing and sulphate-reducing communities in Brazilian mangrove sediments. *Environ. Microbiol.* **16**, 845–855. (doi:10.1111/1462-2920.12237)

45. Weiss S *et al.* 2017 Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27. (doi:10.1186/s40168-017-0237-y)

46. Dedysh SN, Yilmaz P. 2018 Refining the taxonomic structure of the phylum Acidobacteria. *Int. J. Syst. Evol. Microbiol.* **68**, 3796–3806. (doi:10.1099/ijsem.0.003062)

47. Taylor HB, Kurtz, Jr. HD. 2020 Microbial community structure shows differing levels of temporal stability in intertidal beach sands of the grand strand region of South Carolina. *PLoS One* **15**, e0229387. (doi:10.1371/journal.pone.0229387)

48. Cleary DFR, Becking LE, de Voogd NJ, Pires ACC, Polónia ARM, Egas C, Gomes NCM. 2013 Habitat- and host-related variation in sponge bacterial symbiont communities in Indonesian waters. *FEMS Microbiol. Ecol.* **85**, 465–482. (doi:10.1111/1574-6941.12135)

49. Morrow KM, Fiore CL, Lesser MP. 2016 Environmental drivers of microbial community shifts in the giant barrel sponge, *Xestospongia muta*, over a shallow to mesophotic depth

gradient. *Environ. Microbiol.* **18**, 2025–2038. (doi:10.1111/1462-2920.13226)

50. Xie H *et al.* 2021 Chemotaxis-selective colonization of mangrove rhizosphere microbes on nine different microplastics. *Sci. Total Environ.* **752**, 142223.

(doi:10.1016/j.scitotenv.2020.142223)

51. Thompson LR *et al.* 2017 A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* **551**, 457–463. (doi:10.1038/nature24621)

52. Woodhams DC *et al.* 2020 Host-associated microbiomes are predicted by immune system complexity and climate. *Genome Biol.* **21**, 23. (doi:10.1186/s13059-020-01955-y)

53. Louca S, Shih PM, Pennell MW, Fischer WW, Parfrey LW, Doebeli M. 2018 Bacterial diversification through geological time. *Nat. Ecol. Evol.* **2**, 1458–1467.

(doi:10.1038/s41559-018-0625-0)

Figures and Figure Legends

Figure 1. Map of sampling area. The map of Panama (bottom inset) shows the location of the detailed map of the sampling area (left). Blue dots indicate sampling sites in the Eastern Pacific, while pink dots indicate sampling sites in the Caribbean. Phylogenies of the species sampled from each host genus (right) have colors to represent the side of the IP each species was sampled from.

Figure 2. Hill numbers of $q = 0$ (A,B,E,F) and $q = 1$ (C,D,G,H) from whole-body (A,C,E,G) and shell-surface (B,D,F,H) samples. Blue boxes represent samples collected from the Eastern Pacific and pink boxes represent samples collected from the Caribbean. Dashed boxes represent geminate species and solid boxes represent congeneric outgroups. Asterisks indicate samples which were found to be significantly different by Dunn's test at $p < 0.05$ (*) or $p < 0.005$ (**).

Figure 3. Barplots of the relative abundances of microbial phyla in the whole-body (A,C), shell-surface (B,D) and environmental (E) samples. The phylum Proteobacteria is separated into

classes for increased resolution. In the top-left corner of each plot, the shape refers to the genus (squares and circles for *Cerithium* and *Cerithideopsis*, respectively), the color refers to the ocean basin of sampling (blue and pink for Eastern Pacific and Caribbean, respectively) and geminate species are open while congeneric outgroups are represented by a diagonal line.

Figure 4. Comparing host phylogeny to microbiome composition in the whole-body (A,C) and shell-surface (B,D) communities of *Cerithium* (A,B) and *Cerithideopsis* (C,D) species. For each sample type, we show the host phylogeny (left) and UPGMA plots generated using BCD and TSS (top middle), BCD and rarefying (top right), unweighted UniFrac and TSS (bottom middle) and unweighted UniFrac and rarefying (bottom right). Each UPGMA plot includes the normalized Robinson-Foulds value (nRF) and Dunn's index (DI) in blue. Multiple DI values

provide the clustering strength when the UPGMA plot is divided into two and three clusters, respectively. Bolded numbers in (C) indicate the proportion of rarefied datasets which produce the pattern shown. Tip shape refers to the genus (squares and circles for *Cerithium* and *Cerithideopsis*, respectively) and color refers to the ocean basin of sampling (blue and pink for Eastern Pacific and Caribbean, respectively).

Figure 5. Comparing host phylogeny to microbiome composition in the whole-body (A) and shell-surface (B) communities of *Cerithium* species, separated by collection site. For each sample type, we show the host phylogeny (left) and UPGMA plots generated using BCD and TSS (top middle), BCD and rarefying (top right), unweighted UniFrac and TSS (bottom middle) and unweighted UniFrac and rarefying (bottom right). Tip color refers to the ocean basin of sampling (pink and blue for Eastern Pacific and Caribbean, respectively).