ARTICLES
https://doi.org/10.1038/s41564-022-01125-9
nature microbiology

Microbiomes of microscopic marine invertebrates do not reveal signatures of phylosymbiosis

V. Boscaro[1,2,9]✉, C. C. Holt[1,2,9], N. W. L. Van Steenkiste[1,2,3], M. Herranz[1,2,3,4], N. A. T. Irwin[1], P. Àlvarez-Campos[5], K. Grzelak[6], O. Holovachov[7], A. Kerbl[8], V. Mathur[1], N. Okamoto[1,2], R. S. Piercey[2], K. Worsaae[4], B. S. Leander[1,3] and P. J. Keeling[1]✉

Animals and microorganisms often establish close ecological relationships. However, much of our knowledge about animal microbiomes comes from two deeply studied groups: vertebrates and arthropods. To understand interactions on a broader scale of diversity, we characterized the bacterial microbiomes of close to 1,000 microscopic marine invertebrates from 21 phyla, spanning most of the remaining tree of metazoans. Samples were collected from five temperate and tropical locations covering three marine habitats (sediment, water column and intertidal macroalgae) and bacterial microbiomes were characterized using 16S ribosomal RNA gene sequencing. Our data show that, despite their size, these animals harbour bacterial communities that differ from those in the surrounding environment. Distantly related but coexisting invertebrates tend to share many of the same bacteria, suggesting that guilds of microorganisms preferentially associated with animals, but not tied to any specific host lineage, are the main drivers of the ecological relationship. Host identity is a minor factor shaping these microbiomes, which do not show the same correlation with host phylogeny, or 'phylosymbiosis', observed in many large animals. Hence, the current debate on the varying strength of phylosymbiosis within selected lineages should be reframed to account for the possibility that such a pattern might be the exception rather than the rule.

All macroscopic organisms interact closely with microbes[1]. While pathogens receive most of the attention, host-associated prokaryotes are often harmless, beneficial or some complex combination of the three[2,3]. Discrete communities of microbes interacting with each other and with their shared host form distinctive microbiomes[4], which have become the subject of intense scrutiny over recent decades.

Data on metazoan-associated microbiomes are extensive but highly skewed towards certain host taxa, limiting our ability to answer basic questions about broader evolutionary trends. The two largest subgroups of arthropods, namely insects[5,6] and crustaceans[7], receive a lot of attention, as do corals[8] and sponges[9,10]. But most of the information comes from the gut microbial communities of mammals[11,12] and to a lesser extent of other vertebrates[13,14]. Mammalian surveys have often shown a strong correlation between host phylogeny and microbiome similarity[11,15], a pattern also observed in other animal groups and dubbed 'phylosymbiosis'[16], the prevalence of which is strongly debated[13,17–21]. Other factors often shown to play a role in shaping host-associated microbiomes include diet[14,22,23], physiology[20,24], social structure[25], surrounding environment[19,26] and geography[18,27].

How well conclusions drawn from model systems can be generalized to other metazoans is currently unclear because most of the 33 known extant phyla receive little attention[12,28]. Some recent studies have investigated microbiomes in lesser-known taxa[29,30] but generally focussed on a single phylum, and commonly only on a few species within that phylum, leaving most of animal diversity unexplored. Here we take a different approach and characterize the microbiomes of hundreds of animals belonging to the most diverse and least-studied category: microscopic marine invertebrates.

Most animal phyla include, and several entirely consist of[31,32], marine representatives smaller than 1–2 mm, which are numerically dominant and play crucial roles in the ecology of nearly all marine ecosystems[31]. Whether such tiny organisms even have the capacity to host complex, differentiated microbiomes is unclear, but a comprehensive study of their microbial associations would represent much more metazoan host diversity and evolutionary history than has been examined to date, and the nature of their microbial communities would provide critical points of comparison for more familiar model systems[33].

Results

Collection of 1,037 animals from 21 phyla. We isolated, imaged and preserved marine invertebrates measuring ~100–2,000 µm from five temperate (British Columbia, Canada) and tropical (Curaçao, Dutch Caribbean) locations (Fig. 1a). We collected 46 samples from three main habitats (sediment, water column and intertidal macroalgae), characterizing for each the background environmental microbial community as well as the bacteria associated with about 15–30 animals (Supplementary Tables 1 and 2). This totalled 1,037 individual specimens representing 21 phyla, documented by >11,000 high-quality pictures (Fig. 1b and Extended Data Fig. 1). Missing marine phyla are primarily symbiotic (for example, Cycliophora), require very specific sampling techniques (for example, Loricifera) or have generally larger representatives (for example, Brachiopoda). Detailed taxonomic identifications were

[1]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada. [2]Hakai Institute, Heriot Bay, British Columbia, Canada. [3]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada. [4]Department of Biology, University of Copenhagen, Copenhagen, Denmark. [5]Department of Biology, Universidad Autónoma de Madrid, Madrid, Spain. [6]Institute of Oceanology, Polish Academy of Sciences, Sopot, Poland. [7]Department of Zoology, Swedish Museum of Natural History, Stockholm, Sweden. [8]Centre of Natural History, University of Hamburg, Hamburg, Germany. [9]These authors contributed equally: V. Boscaro, C. C. Holt. ✉e-mail: vittorio.boscaro@gmail.com; pkeeling@mail.ubc.ca

performed on groups that were highly abundant in our survey, of particular interest or especially under-represented in the literature (Supplementary Table 2). Morphological identifications were confirmed by 18S ribosomal RNA gene barcoding on a subset of 232 animals, finding only a minority (9.5%) of below-phylum discrepancies (potentially also occurring in the other 805 non-barcoded specimens), whose removal or inclusion did not impact any of the analyses. Since the number of isolated organisms per species roughly reflects its abundance in each sample, taxa vary widely in frequency (Supplementary Table 2), altogether giving the dataset a great deal of taxonomic breadth, as well as considerable sampling depth for certain lineages.

Microscopic invertebrates harbour distinct microbiomes. Microbiomes associated with hosts were clearly distinct from environmental microbial communities. In a principal coordinates analysis (PCoA) of the entire dataset (Fig. 2a), the two groups clustered separately, their difference being the largest factor influencing the ordination. In support of this, a number of differentially abundant bacterial genera were detected in both animal and environmental bacterial communities (Supplementary Table 3). Animal-associated microbiomes (Fig. 2b) were only weakly and/or not significantly affected by environmental features such as habitat (analysis of similarities (ANOSIM) $R \approx 0$, $P = 0.476$) or location (ANOSIM $R = 0.172$, $P < 0.001$), which, unsurprisingly, more strongly impacted environmental communities (ANOSIM: habitat, $R = 0.662$, $P < 0.001$; location, $R = 0.366$, $P < 0.001$) (Fig. 2c).

Animal-associated microbiomes showed substantially lower Shannon-diversity and richness of amplicon sequence variants (ASVs) than their environmental counterparts (Fig. 2d–g and Extended Data Fig. 2). The Shannon index was clearly impacted by host phylum and location (two-way analysis of variance (ANOVA): host phylum, $P < 0.001$; location, $P < 0.001$); values within each phylum (Fig. 2d), as well as within smaller taxonomic units like orders in Annelida and Nematoda (Fig. 2e), varied extensively but were generally much lower than environmental microbial communities from the same location (Fig. 2f) or habitat (Fig. 2g). The same conclusions apply to the number of ASVs observed in each microbial community (two-way ANOVA: host phylum, $P < 0.001$; location, $P < 0.001$) (Extended Data Fig. 2). Neither Shannon-diversity nor ASV richness of invertebrate-associated microbiomes correlate with those of corresponding environmental communities (Extended Data Fig. 3), altogether highlighting differential and independent dynamics acting on the two types of microbial assemblages.

Guilds of non-specific host-associated bacteria. To identify the causes of the strong separation between animal-associated and environmental microbial communities, we measured the proportions of bacterial ASVs shared between individual invertebrates and their environment. These turned out to be relatively low (Fig. 3a–c and Extended Data Fig. 4), implying that a limited taxonomic overlap contributes to the separation of host-associated and environmental communities. Location and habitat were significant determinants of the extent of this overlap but there were no statistically significant differences among host phyla within comparable samples,

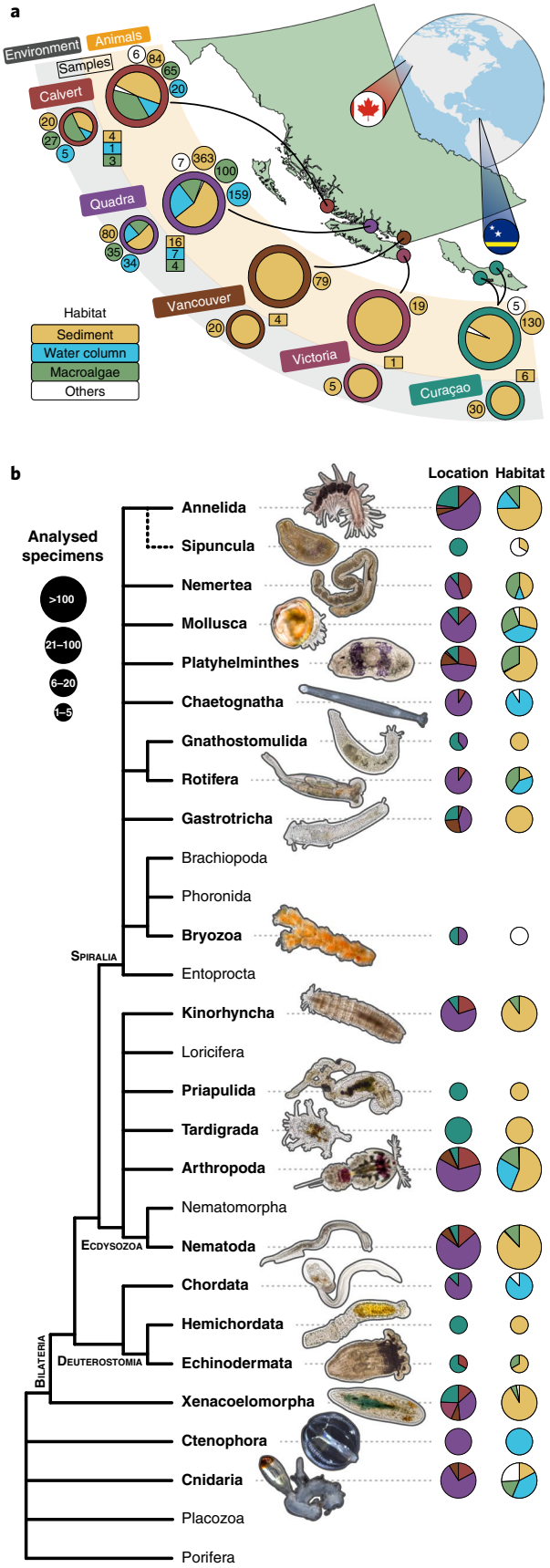


Fig. 1 | Collection of over 1,000 specimens representing most animal phyla. a, Surveyed locations in British Columbia and Curaçao. For each of the three marine habitats, numbers in rectangles represent collected samples; large pie charts in the inner ring represent numbers of specimens (exceptions not tied to main habitats are in white); small pie charts in the outer ring represent numbers of environmental aliquots from the same location. **b**, Specimen diversity. The cladogram only includes phyla with marine representatives that are free-living in at least one life stage. Those in bold and with images (see Supplementary Note 1 for descriptions) were covered by our survey. Pie charts depict the distribution of invertebrates from each phylum across locations and habitats, and chart size correlates with the number of specimens.

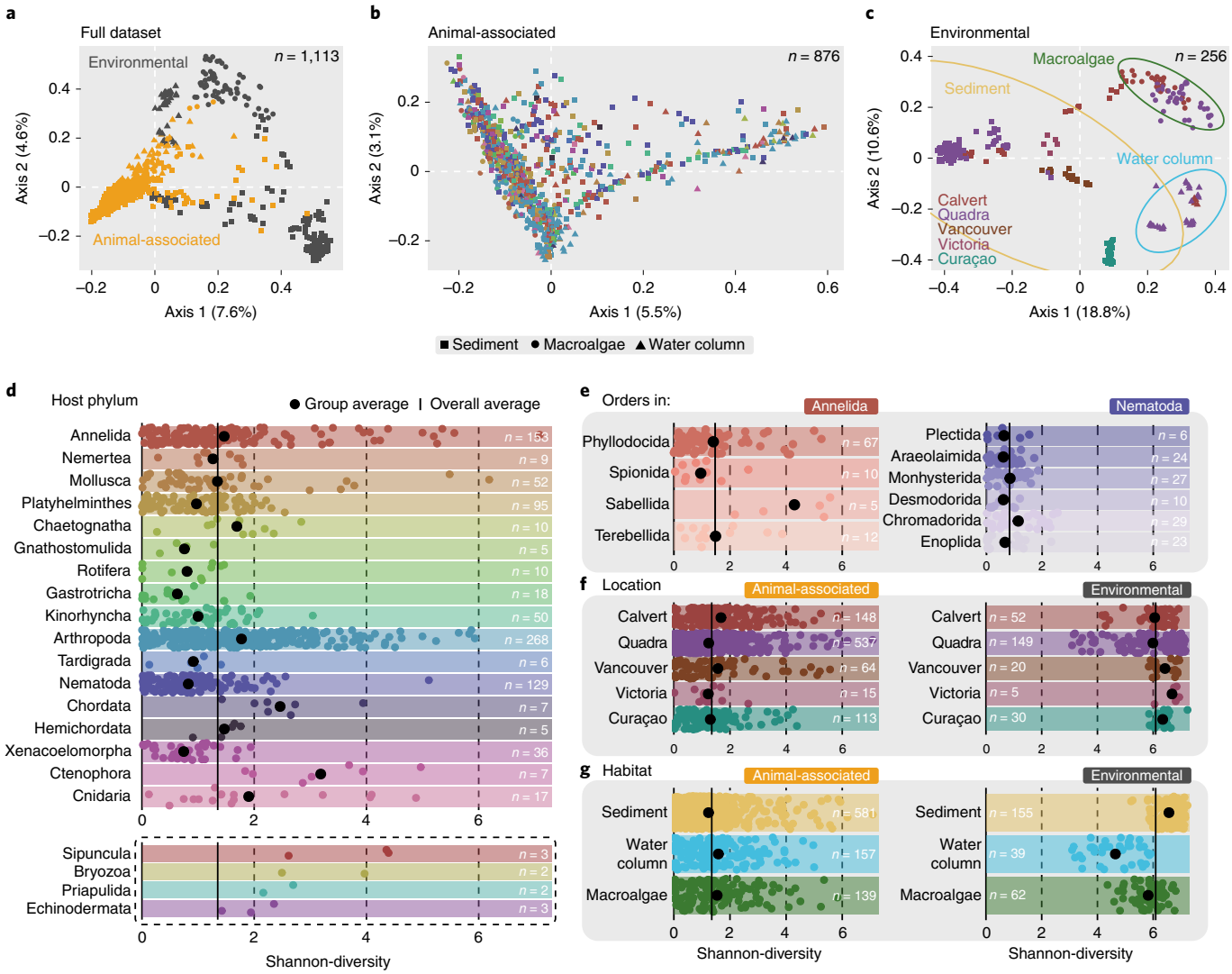


Fig. 2 | Impact of host taxonomy and environmental variables on microbiome composition. a–c, PCoA using Bray–Curtis dissimilarity; shapes correspond to habitats (squares for sediment, circles for macroalgae and triangles for the water column). **a**, Full dataset, with animal-associated and environmental microbial communities distinguished by colours. **b**, Animal-associated microbiomes, colour-coded according to host phyla as in **d**. **c**, Environmental communities, colour-coded according to location. Ellipses surround habitat groupings. **d–g**, Shannon-diversity of the bacterial communities, grouped according to: host phylum (**d**); host order within the phyla Annelida and Nematoda, respectively (**e**); location (for both animal-associated and environmental data) (**f**); and habitat (for both animal-associated and environmental data) (**g**). The dashed box in **d** separates phyla with less than five sampled specimens each. Larger black circles show group average. Solid lines show overall average of the dataset.

with the exception of communities from Calvert Island macroalgae (Supplementary Note 2). This suggests that, although the amount of bacterial ASV overlap between microscopic invertebrates and their surroundings widely varies in magnitude, it is not predominantly determined by the animals' taxonomic identity.

To assess the ecological importance of the relatively few bacteria shared by invertebrates and their wider environments, we first constructed environmental ecological networks to determine the keystone bacterial ASVs in each combination of habitat and location (Fig. 3d and Extended Data Fig. 5). These were identified and ranked according to their eigen-centrality in the network—that is, the number of nodes (ASVs) to which each node is linked plus additional links made by those to subsequent nodes. We then assessed the prevalence and relative abundance of the key environmental ASVs with the highest eigen-centrality in the microbiomes of animals from the same habitat and location. Both prevalence and abundance were found to be generally low in animal-associated communities and several keystone ASVs were altogether absent (Fig. 3e and Extended Data Fig. 5).

The microbiomes of microscopic invertebrates therefore share little overlap with the environment and probably differ from it in the composition of the most ecologically important bacteria. In fact, individual invertebrates shared a significantly higher proportion of bacterial ASVs with co-occurring animals, regardless of their phylum, than with environmental microbial communities from the same sample (Fig. 3f–h, Extended Data Fig. 4 and Supplementary Note 3). The proportions of ASVs shared among co-occurring specimens and with the environment weakly correlate (Pearson's $R = 0.265$, $P < 0.001$) but the increments are not driven by the same ASVs (Extended Data Fig. 6), which suggests that the trend is not simply attributable to ubiquitous generalists. This phenomenon was not only consistent across the spectrum of habitats but held true even after agglomerating ASVs into bacterial genera (Supplementary Figs. 1 and 2 and Supplementary Note 3) although with a predictable

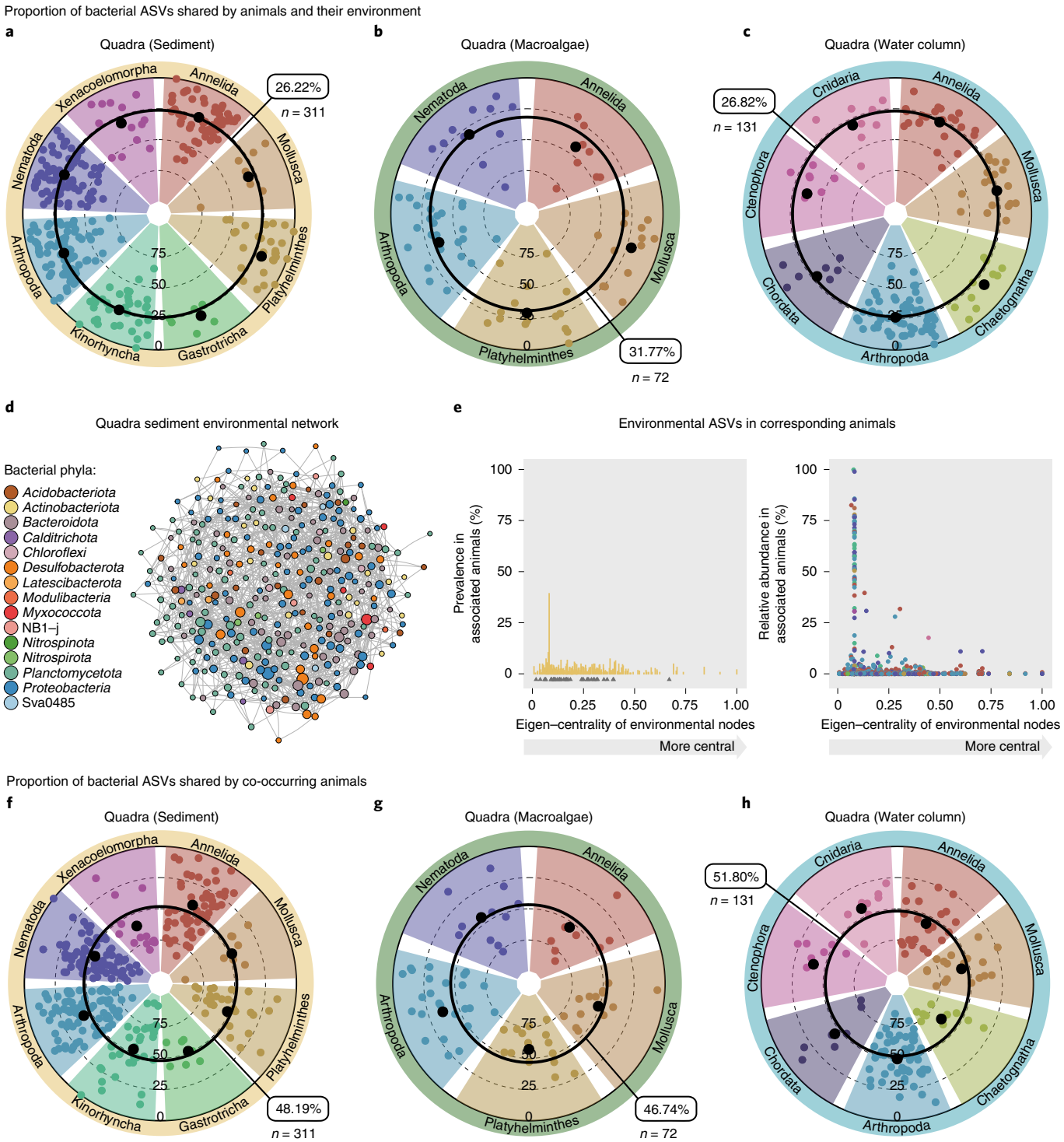


Fig. 3 | Microbiome overlap between animals and their environment. a–c, Proportion of bacterial ASVs shared between individual invertebrates from Quadra Island and their environment, separated by habitat: sediment (**a**); macroalgae (**b**); and water column (**c**). **d**, Co-occurrence network of 351 key environmental ASVs found in Quadra Island sediment samples. Each node represents a single ASV, whose taxonomic affiliation is shown on the left. Lines (edges) connecting two nodes indicate an association between the two ASVs. Node size is scaled to eigen-centrality, which considers the number of connecting nodes as well as their subsequent connections. **e**, Prevalence and abundance (both as percentage) of the environmental ASVs from **d** in animals from the same habitat and location. Individual ASVs (on the *x* axis) are ordered according to their eigen-centrality in the environmental network and may be represented by multiple datapoints in the abundance plot (on the right) to reflect their varying abundance in multiple host phyla, which are colour-coded as in Fig. 2d. Grey arrowheads in the prevalence plot (on the left) indicate environmental ASVs that are absent in host-associated microbiomes. **f–h**, Proportion of bacterial ASVs shared between individual specimens and all other animals in the same sample, separated by habitat: sediment (**f**); macroalgae (**g**); and water column (**h**). Solid, black lines in circular plots indicate overall average. Dashed lines indicate 25%, 50% and 75% thresholds, for scale. Black circles plot phylum average.

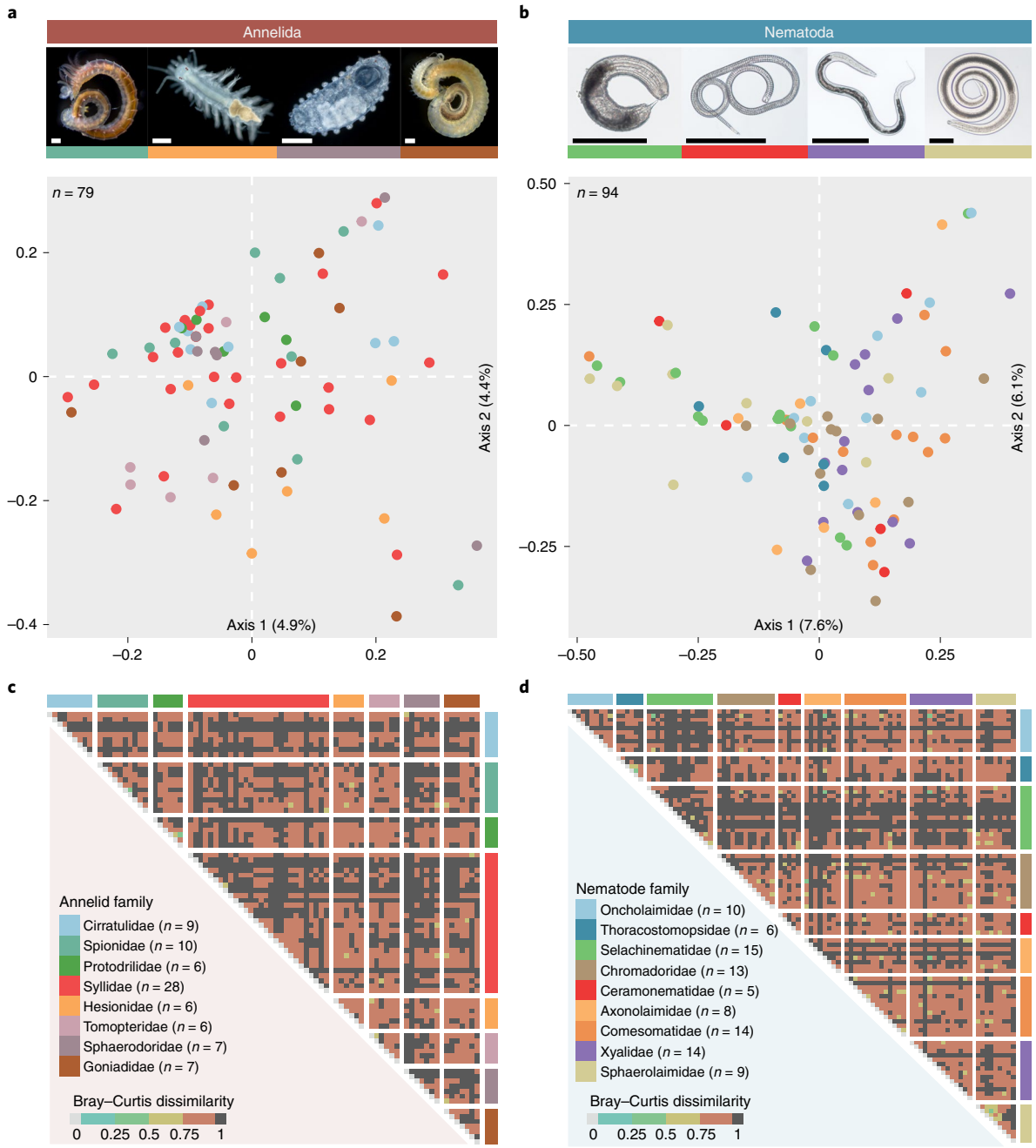


Fig. 4 | Microbiome composition differences do not reflect host family identity. a,**b**, PCoA of microbiomes according to annelid (**a**) and nematode (**b**) families with a minimum of five collected specimens (representative specimens for each family are shown on top; bars, 250 µm). **c**,**d**, Heatmaps showing pairwise comparisons of Bray–Curtis dissimilarity scores of the same animals. Tiles are coloured according to dissimilarity score; light grey boxes on the diagonals indicate self-self comparisons, with dissimilarity scores of 0; dark grey boxes correspond to a score of 1 and indicate absolute dissimilarity. *n*, number of specimens.

increase in overlap values. Overall, host-associated bacteria do not make up a substantial proportion of the environmental microbial community but tend to be shared among taxonomically unrelated animals, altogether suggesting the existence of guilds of bacteria that are preferentially host-associated but are not constrained by host taxonomy.

No evidence for phylosymbiosis. PCoA ordinations show no clustering of microbiomes from invertebrates of the same phylum (ANOSIM $R=0.012$, $P=0.173$) (Fig. 2b). Examining the relationship from the opposite perspective supports the same conclusion: microbial community compositions are extremely poor predictors of higher-level host taxonomy according to random forest models (Extended Data Fig. 7). These models can reliably distinguish animal-associated microbiomes from environmental microbial communities and fare reasonably well in discriminating animal-associated microbiomes from different locations and habitats (although these variables are predicted far better by environmental communities). In contrast, however, out-of-bag error rates are very high when attempting to classify microbiomes according to host phyla, classes or orders, overall providing no support for phylosymbiosis at these levels. Finally, when mapping our datapoints

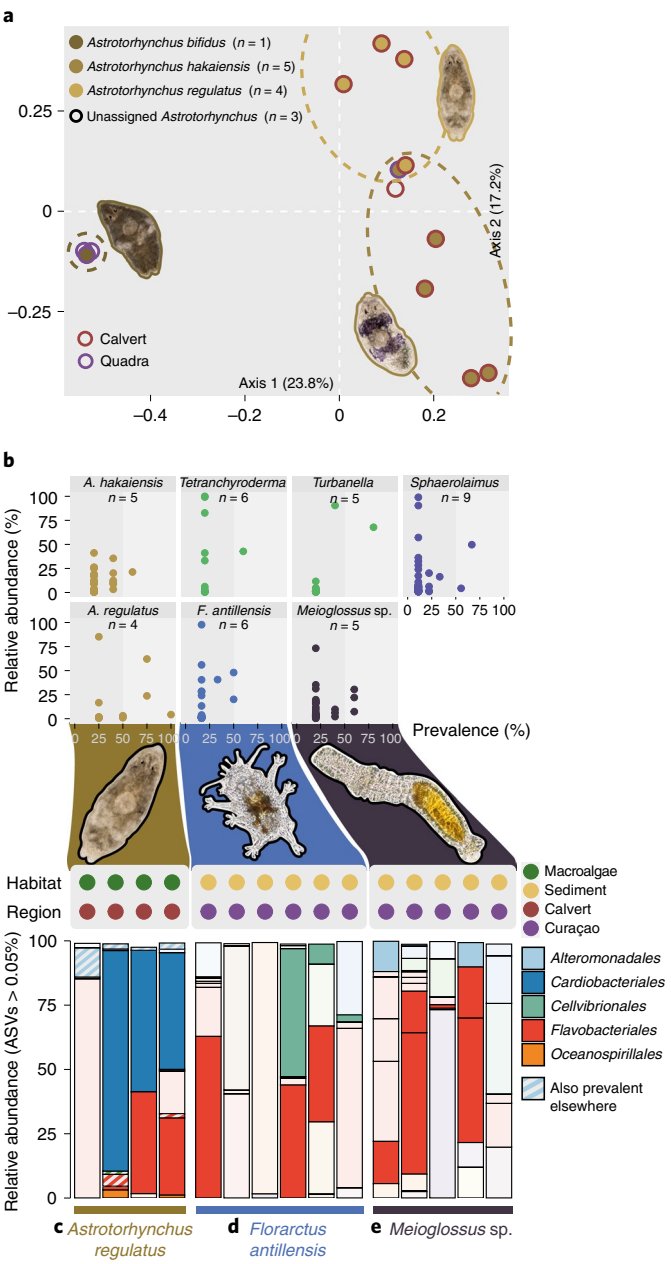


Fig. 5 | Weak signals of host genus- and species-specific bacteria. **a**, Microbiomes from *Astrotorhynchus* clustered according to host species in PCoA ordination. **b**, Average relative abundance of bacterial sequence variants plotted against their overall prevalence (as percentage of specimens) within host genera/species. **c**–**e**, Taxonomic composition of bacterial microbiomes from: *Astrotorhynchus regulatus* (Platyhelminthes) (**c**); *Florarctus antillensis* (Tardigrada) (**d**); and *Meioglossus* sp. (Hemichordata) (**e**). Solid colours indicate bacterial sequence variants with a prevalence of at least 50% in their respective host genus/species, which are potentially host-specific. Pale colours indicate non-prevalent variants. Striped bar segments identify variants that are also common in other hosts and/or environmental communities and, therefore, are not taxa-specific, regardless of prevalence in these hosts.

on published, reliable phylogenetic trees, the correlation observed between phylum phylogeny and microbiome dissimilarities is negligible (Mantel $R=0.141$, $P=0.001$) (Extended Data Fig. 8).

To investigate whether microbiome correlations became apparent with less inclusive host taxa, and hence at a shorter evolutionary scale, we focussed on two well-sampled and well-annotated lineages, Annelida and Nematoda (Fig. 4). Within these phyla, microbiomes did not cluster according to host family (ANOSIM R values were close to 0 and/or not significant: Annelida, $R=-0.087$, $P=0.948$; Nematoda, $R=0.093$, $P=0.002$) (Fig. 4a,b). Moreover, pairwise dissimilarity values were particularly high, with three-quarters of comparisons scoring above 0.96 and 0.92 in the two phyla, respectively (Fig. 4c,d). As seen with phyla, phylogenetic trees of the investigated families do not covary with dendrograms of microbiome dissimilarities (Mantel: Annelida, $R=0.131$, $P=0.002$; Nematoda, $R=0.036$, $P=0.094$) (Extended Data Fig. 8). Hence, typical signatures of phylosymbiosis are lacking even among families.

Since our goal was to cover the largest possible taxonomic diversity, we did not always have multiple specimens within low-level taxa identified, such as genera and species. When we did, some partial, weak phylosymbiosis signals were detected. For example, microbiomes from the flatworm *Astrotorhynchus* (Fig. 5a) as well as the kinorhynch *Echinoderes* (Extended Data Fig. 9) seemingly clustered according to host species. It is, however, worth noting that different species in these genera were sampled from different locations and/or habitats, making it difficult to distinguish whether the observed effect is due to host taxonomy or correlated environmental factors. Phylogeny-based phylosymbiosis test results were indeed inconclusive but similar to those obtained from higher taxonomic levels (Mantel: *Astrotorhynchus*, $R=0.120$, $P=0.074$) (Extended Data Fig. 8).

Potential symbionts and host taxa-specific associations. Phylosymbiosis involves trends that apply to the whole microbiome composition and its absence does not preclude the possibility that specific bacteria preferentially associate or co-differentiate with certain host taxa, as is the case in many symbioses. We attempted to predict 'core' bacterial taxa for each animal phylum present in the majority of specimens in that phylum but even examining broad targets, such as bacterial families (Extended Data Fig. 10), we could only find a small number of candidates in nine host phyla. For example, *Rhodobacteraceae* and/or *Flavobacteriaceae*, which are often found in marine invertebrates and environments, were commonly detected. Individual ASVs were almost never present in the majority of specimens belonging to the same phylum (Supplementary Fig. 3). Only in Ctenophora, Hemichordata and Tardigrada, each represented by relatively few specimens, were prevalent ASVs also phylum-specific.

To examine symbioses in lower-level host taxa, we plotted the relative prevalence of ASVs in well-sampled genera or species, considering as potential host-specific symbionts those that were present in at least half the specimens. This filter produced seven candidate host genera/species of interest, shown in Fig. 5b, which were then reduced to three after disregarding ASVs that were also present in other invertebrates (and therefore are not taxa-specific). Three ASVs were identified as belonging to candidate symbionts in *Astrotorhynchus regulatus* (Platyhelminthes) (Fig. 5c), two in *Florarctus antillensis* (Tardigrada) (Fig. 5d) and one in an undescribed *Meioglossus* species (Hemichordata) (Fig. 5e), all of which are similar to environmental and/or marine animal-related reference sequences but could not be identified below the order or family ranks. The exception was *Winogradskyella*, previously described in association with various marine invertebrates[34], found here in the microbiome of *Florarctus* (Supplementary Table 4). None of these ASVs were present in all specimens of the putative host, however. ASVs prevalent in, but not exclusive to, specific taxa, that could have been interpreted as symbionts in a less diverse survey, tend to be linked to environmental factors, providing an explanation for their distribution independent from host identity.

Our database also revealed the presence of a number of known bacterial symbionts inhabiting a wide range of host organisms. Members of the specialized intracellular order *Rickettsiales*, for

example, were present in 15 different host phyla and include genera known from invertebrate and protistan hosts[35,36] such as *Neorickettsia*, '*Candidatus* Aquarickettsia' and '*Candidatus* Megaira' (Supplementary Table 5). Other symbiotic lineages represented in several host phyla were *Chlamydiae*, *Holosporales* and the common animal symbiont *Endozoicomonas*[37] (Supplementary Table 5).

Discussion

Our data show that even the smallest marine invertebrates, some barely larger than protists, do harbour associated microbial communities just like more familiar models[7–10,38,39]. These microbiomes are distinct from environmental communities by both composition and keystone bacterial taxa. However, the animals' surroundings play a much larger role than the identity of the host in shaping the microbiomes of microscopic invertebrates. Co-occurring animals share a considerably higher proportion of bacterial species with each other than with their environment, suggesting that many bacterial taxa are preferentially host-associated but do not show a strong affinity for any specific host lineage. Those that do, have a patchy distribution and might be quantitatively sparse to the point where their signal is drowned by other components of the microbial community. So, while microbial communities are predictably impacted by the differences between invertebrates and their surrounding environment, they are surprisingly less affected by equally enormous differences in anatomy, diet, physiology and life-cycle stages of the animals that harbour them, at least in the investigated lineages. Hence, in the ecological relationships between bacteria and microscopic marine invertebrates, the main targets of selection are more likely to be the bacteria, rather than their hosts.

Phylosymbiosis, defined as a correlation between microbiome similarity and host phylogeny, undoubtedly exists, as it has been convincingly demonstrated in many systems through both descriptive and experimental approaches[8,15,16]. But our understanding of its frequency and underlying mechanisms is based on a small fraction of host diversity, almost exclusively represented by macroscopic metazoans with large and strongly insulated body compartments[12], the digestive system being the one most often investigated. Our survey does not distinguish microbes associated with different body parts of microscopic invertebrates, which have much higher surface area:volume ratios and a reduced capability to regulate their internal state. At the same time, our data encompass a greater phylogenetic range of animals than has ever been examined, including phyla where phylosymbiosis has never been tested[12]. What we found is that microbiome composition is predominantly influenced by the host's surroundings, microbiomes are not significantly more similar within than among host taxa (except possibly at the lowest end of the taxonomic scale) and there is no correspondence between host phylogeny and microbiome similarities. Such features are predicted proxies for the absence of phylosymbiosis[16], at least over long evolutionary times. The enormous variability we observe instead may be due to many factors, including stochastic processes driving microbial assemblages in aquatic environments[40], microbiome plasticity during invertebrate life-cycles[7], lack of an adaptive immune system[41] and the impact of host health and dysbiosis[2], the combination of which might obfuscate most of the potential host influence.

It has also been argued that constraints related to ecological factors mirror (and hence confuse) the phylogenetic signal even in systems where phylosymbiosis is observed, which would cast doubt on phylogeny as an independent factor and coevolution as a mechanism explaining the observed patterns[17]. In light of its overall absence across this wide taxonomic range of animal hosts and considering the growing number of studies with similar conclusions in specific lineages[9,12,18,19,26,29], we suggest that phylosymbiosis, however common within some well-studied groups, should not be the default assumption for all metazoans.

The scope of our analysis was intentionally very broad to: (1) complement other studies that mostly focus on a particular system, (2) address the general lack of information from most invertebrates and (3) provide a baseline and framework for future studies. Delving more deeply into any one lineage will clarify the frequency of shorter-term phylosymbiosis, allow for more precise molecular identifications and phylogenies of the hosts (possibly reconciling low-rank molecular and morphological ID discrepancies) and shed additional light on specific trends tied to different features of related animals, such as size, feeding habits or life-cycle stages. It is also important to stress the difference between overall host-associated microbiomes versus specific microbial symbionts. Examples of well-defined symbiotic associations between bacteria and small hosts, including marine invertebrates[42,43] and protists[35], abound and might display different evolutionary paths. Such relationships could be very common and yet still only account for a fraction of the total interactions between animals and bacteria they harbour, most of which we conclude are non-specific host-associated microbes.

Methods

Collection of samples and environmental aliquots. Samples were collected over several field trips from July 2017 to January 2019 in coastal sites around four locations in British Columbia, Canada (Calvert Island, Quadra Island, Vancouver and Victoria) and in Curaçao in the Dutch Caribbean. Sampled habitats included: sediments collected with a meiobenthic dredge (subtidal) or handpicked with a shovel (intertidal and subtidal); pelagic water collected via horizontal and vertical plankton tows using a 64 µm mesh; and macroalgae harvested from tidal pools and stored in sealable plastic bags filled with surrounding water. Samples were immediately transported to the laboratory, that is the Hakai Institute's Ecological Observatories on Calvert Island and Quadra Island, the Beaty Biodiversity Centre at the University of British Columbia (Vancouver) or the Caribbean Research and Management of Biodiversity Institute (CARMABI) Research Station on Curaçao, stored at 4 °C and inspected within 24 h under dissecting microscopes (see below). In total, 60 samples were collected but only the 46 with the highest abundance and diversity of microscopic invertebrates were processed. The number of samples processed at each location are shown in Fig. 1a; geographic details about each are listed in Supplementary Table 1. Five to nine environmental aliquots were collected for each processed sample to assess background microbial communities, specifically: five aliquots of ~1 ml of wet sediment for sediment samples; five bottles of 500 ml filled with seawater collected at equal time intervals along plankton tows, then filtered with 0.22 µm Sterivex filters (EMD Millipore) on a peristaltic pump; three filtered seawater aliquots and six surface swabs (Sterile Cotton Tipped Applicators, Puritan), three of which collected along the thallus of the algae and three over the anchoring rocks, for macroalgal samples. All environmental aliquots were stored in 1.5 ml plastic tubes and frozen at −20 to −80 °C until DNA extraction. Tools and containers used during sampling were sterilized with 10% bleach and 70% ethanol and rinsed with distilled water before each use.

Animal collection and identification. Portions of each sample were transferred to Petri dishes after concentrating the animals using standard methods: a gentle $MgCl_2$ treatment[44] for macroalgal and most sediment samples and the 'bubble and blot' protocol[45] for the remaining (muddier) sediment samples (Supplementary Table 1). Plankton tows did not need to be further concentrated. Fifteen to 29 (average 22) invertebrates in the target size range of ~100–2,000 µm were isolated from each sample using the same methodology, regardless of their life stage and in taxonomic proportions approximately representative of the invertebrate community observed in the sample. The reported size range refers to the length of the animal; it should not be intended as a good proxy for animal volumes across different phyla, which are more difficult to estimate. In Nematoda, a representative and well-sampled group where such measurements are relatively reliable, they varied from 1.5×10^{-4} to 6.5×10^{-2} mm^3. A few taxa, such as Ctenophora and Chaetognatha, slightly exceeded our maximum size threshold but were included for completeness. With few exceptions among the Annelida traditionally classified as 'Sedentaria' (Supplementary Table 2), none of the collected animals is infaunal (that is, lives mostly in a burrow or tube).

Live specimens were individually collected under the dissecting microscope (Zeiss Stemi 508) using Irwin Loops[46], transferred to droplets of sterilized marine water on glass slides and imaged on Zeiss Axioscope A1 or Leica DMIL microscopes (for British Columbia and Curaçao field trips, respectively) with Axiocam 503 color or Sony a6000 cameras. A minority of invertebrates were imaged directly under the dissecting microscope or under both dissecting and upright (Axioscope A1) microscopes. At this step, specimens were assigned to broad taxonomic groups and labelled with unique alphanumeric codes. After imaging, live specimens were collected again with the Irwin Loop, washed via successive transfers in at least three separate sterile water droplets and finally

dropped in 10 µl of sterile water in 1.5 ml plastic tubes, immediately frozen and stored at −20 °C until DNA extraction. All non-disposable tools were sterilized with 10% bleach and 70% ethanol before use.

A small fraction (1.9%) of animals were collected extemporaneously and are not tied to specific samples or environmental communities (Fig. 1), mostly because they belonged to lineages not easily collectable with the aforementioned protocols.

More detailed taxonomic identifications were later performed on target phyla, using the large photo database built during collection. Specimens were assigned by experts to the narrowest taxon that could unambiguously be inferred from the photos, avoiding uncertain over-specific assignments that would require staining or other types of morphological techniques for confirmation, especially considering that the surveyed regions are underexplored and that many collected specimens might belong to undescribed species or genera. The taxonomic classification in this paper follows that on the World Register of Marine Species (WoRMS, https://www.marinespecies.org/index.php) at the time of writing.

Molecular procedures. DNA extractions were performed on batches of stored, frozen animals and environmental aliquots soon after each sampling trip (Supplementary Table 2), using the DNeasy PowerSoil Kit (QIAGEN Gmbh) according to the manufacturer's instructions. For the final elution, 70 µl of UltraPure Distilled Water (Invitrogen) were used instead of the provided Elution Buffer. DNA concentrations were measured with a Qubit Fluorometer and the dsDNA HS Assay Kit (Life Technologies) for environmental aliquot extractions but were usually below the minimum detecting concentration threshold for extractions performed on stored animals. To confirm the success of the extraction, PCRs with universal eukaryotic primers[47] (forward A, 5′-AACCTGGTTGATCCTGCCAGT-3′; reverse B, 5′-TGATCCTTCYGCAGGTTCACCTAC-3′) for the 18S rRNA gene were performed on an initial batch of 30 invertebrates and their long amplicons Sanger sequenced by GENEWIZ to confirm the presence and identity of the animal host. This was later corroborated by short-read high-throughput sequencing of the V4 region of the 18S gene of >200 specimens, as reported in Supplementary Table 2 (amplification forward primer[48] 18S-EUK581-F, 5′-GTGCCAGCAGCCGCG-3′; amplification reverse primer[48] 18S-EUK1134-R, 5′-TTTAAGTTTCAGCCTTGCG-3′; PCR reagents and conditions as reported below for 16S PCR, except for the annealing temperature, which was 51 °C; sequencing as reported below for 16S amplicons but with primers[49] forward E572F, 5′-CYGCGGTAATTCCAGCTC-3′ and reverse E1009R, 5′-AYGGTATCTRATCRTCTTYG-3′). High-throughput reads were processed as described below for the main 16S data and sequences identified as Metazoa were extracted. The identifications of these 18S gene sequences were then manually refined by comparisons with the 100 best BLASTN hits against the GenBank database (February 2022). Species, genera and families were assigned when there was a consensus among all reference sequences sharing or exceeding 100%, 99% and 97% similarity (respectively) with the query. If no reference sequence was sufficiently similar, or if there was discordance among hits, ranks were putatively assigned on the basis of distance trees of the query and its 100 best hits. Overall, 18S gene sequences were not very useful in refining animal identifications due to the inadequacy of reference databases for this combination of taxa and marker, as well as the low amount of information carried by the 18S gene in animals in general (when both molecular and expertly curated morphological IDs were available, the latter showed the same or better resolution almost 80% of the time). The few cases (9.5%) displaying a significant discrepancy between morphological and molecular IDs (at or above family, always below phylum) were discarded out of caution from all analyses involving the relevant taxonomic information (Supplementary Table 2). None of the analysis conclusions were affected by this removal.

The V4–V5 region of the 16S rRNA gene was selected as the target for microbiome metabarcoding, with a standard method used for bacteria (and not necessarily suitable for the detection of potential archaea[50]). PCRs were performed using the Phusion High-Fidelity PCR Master Mix (New England BioLabs) and universal bacterial primers[51] (forward 27F, 5′-AGAGTTTGATCMTGGCTCAG-3′; reverse 1492R, 5′-CGGTTACCTTGTTACGACTT-3′) on DNA extracted from animals and environmental aliquots (in the latter case, after diluting the DNA solution down to the 0.1–5 ng µl^{-1} range). Reactions were performed in a total volume of 20 µl, using 2–4 µl of extracted DNA as input; a BioRad T100 thermocycler was used, with the following thermal profile: initial melting, 98 °C (30 s); 35 cycles, 98 °C (10 s), 55 °C (30 s), 72 °C (1 min); final extension, 72 °C (10 min). One negative control was added every 11 reaction tubes and amplicons were discarded if any corresponding controls during the same PCR showed a band in agarose gel electrophoretic runs. PCR dates are reported in Supplementary Table 2. Amplicons were purified with the QIAquick PCR Purification Kit (QIAGEN) and yields subsequently measured with a Qubit. Purified amplicons, including a subset of negative controls without visible bands in electrophoretic runs, were sent to CGEB—Integrated Microbiome Resource to be sequenced on the Illumina MiSeq platform (2×300 paired-end sequencing) after a nested amplification/library preparation with primers 515FB (5′-GTGYCAGCMGCCGCGGTAA-3′) and 926R (5′-CCGYCAATTYMTTTRAGTTT-3′)[52]. Sequencing batches are reported in Supplementary Table 2 and did not have a sizable effect on data clustering, considering that they were not fully randomized (ANOSIM $R=0.277$, $P=0.001$).

Characterization of ASVs. Primer sequences were removed from paired-end reads using Cutadapt (v.3.4)[53]. Reads were processed in R using the DADA2 package (v.1.14.1)[54,55]. First, reads were truncated according to their quality profiles and filtered using the parameters maxN = 0 and maxEE = c(2,2). Error rates were characterized using the first 100 million bases (the default) and libraries were inferred with 'pseudo' pooling to allow singletons, providing they are present in more than one library. Paired-end reads were then merged. Sequencing runs were analysed independently to obtain run-specific error models and combined after read merging. Next, chimaeras were detected and taxonomic classification was based on the SILVA database (v.138), using the DADA2 function assignTaxonomy which implements the RDP Naive Bayesian Classifier algorithm using *k*-mer sizes of 8 and 100 bootstraps[55,56]. The resulting ASV table and taxonomic assignment were combined with library metadata using the phyloseq (v.1.36.0) package[57].

Eukaryotic, chloroplast and mitochondria sequences were removed from the dataset, as were ASVs with a read count of 0 after library filtering. Libraries with <1,000 reads were removed from further analysis. ASVs were discarded if they were present in both PCR negative controls and animal-associated libraries but absent in environmental libraries, to account for a greater likelihood of contamination associated with low biomass (methods such as decontam[58] are discouraged for low-biomass or non-homogeneous libraries).

Seed value for all functions or plots involving random objects was set to 2209. Observed numbers of ASVs and Shannon-diversity were estimated with the phyloseq package and plotted with ggplot2 (v.3.3.5) (as were other plots in the paper, unless stated)[59]. PCoA using Bray–Curtis dissimilarity were performed on data transformed with the DESeq2 (v.1.34.0) package and the 'poscounts' estimator[60]. Before running the varianceStabilizingTransformation function (which also normalizes with respect to library size) for each plot, ASVs with a total relative abundance <0.001% in all libraries were removed, as were libraries that were subsequently left with <1,000 reads. This resulted in a small number of libraries included in the environmental-only ordination (Fig. 2c) being excluded from the same analysis of the entire dataset (Fig. 2a).

To calculate the proportion of ASVs shared between individual animals and their environment or other animals, the number of shared ASVs was calculated as a proportion of the total number of unique ASVs in each individual invertebrate. Herein a list of ASVs from each animal is compared to a combined list of unique ASVs from multiple environmental aliquots or animals from the same sample. Despite the number of animals per sample being greater than the number of environmental aliquots, most specimens contain a fraction of the number of ASVs found in any single environmental community, therefore a greater number of animals did not equate to a greater pool of potential shared ASVs. This method was chosen over averaged values to mitigate the dampening of large variances. The SPIEC-EASI (v.1.1.1) package, designed to deal with the sparsity associated with microbiome data, was used to calculate co-occurrence networks in environmental libraries from the same habitat and location, using a neighbourhood selection model (mb), lambda.min.ratio of 1×10^{-2} and nlambda = 20, according to recommended usage[61]. Models were constructed using 999 replications. Before network computations, ASVs were filtered on the basis of their total relative abundance across all libraries within the subset (as above) and a minimum prevalence of 30%. The igraph (v.1.2.7) package was used to plot networks and determine network statistics[62].

Random forest models were used to classify microbial communities using the variance-stabilized ASV counts (with the minimum relative abundance filter) as predictors. The randomForest function of the randomForest (v.4.6.14) package was implemented using the default mtry value (the square root of the total number of predictors) and 1,000 trees. The resulting model was assessed according to its out-of-bag error[63].

Pairwise comparisons of Bray–Curtis dissimilarity (on variance-stabilized counts) were plotted with the pheatmap function[64]. Relative prevalence was plotted against averaged relative abundance to identify host taxa-specific ASVs. ASVs with a relative abundance of <0.05% were discarded for plotting.

The microbiome (v.1.16.0) package was used to calculate 'core' bacterial families by aggregating ASVs to the family level and analysing prevalence at increasing relative abundance thresholds. ANCOMBC (v.1.2.2) was used to test for the differential abundance of bacterial genera in specimens and environments, using default parameters (with exception to struc_zero, neg_lb and conserve = TRUE)[65].

Statistical analysis. All statistical tests were performed in R. Test assumptions were tested where noted and individual datapoints were always shown. Permutation-based tests were performed with 999 replications. Statistical significance for distance-based analyses was determined with the (non-parametric) ANOSIM function in vegan[66] (v.2.5.7). ANOSIM tests whether the similarity between groups of datapoints is greater than the similarity within those groups. An ANOSIM statistic (*R* value) ranges from 0 to 1 (or −1 if data are patchy). If *R* is closer to 1, there is more similarity within groups and the grouping factor significantly (depending on *P* value) impacts microbiome composition. If $R=0$, there is no difference between groups. It should be noted, however, that the analyses of multivariate homogeneity of group dispersions were significant. The aov function was used to fit a series of linear models (ANOVA). Alternatively, the

t.test function was used to compare just two categories. Data were transformed before modelling (square root for Shannon index values and $\log_2$ for ASV counts) if the distribution of residuals was not normal.

The Mantel test (vegan) was used to compare the cophenetic distance of the host phylogeny (in this case, based on reliable published trees for animal phyla[32], annelids[67], nematodes[68] and *Astrotorhynchus*[69]) with the hierarchical clustering of a corresponding microbiome dissimilarity matrix (Bray–Curtis) (Extended Data Fig. 8). Mantel *R* values range from −1 (negative correlation) to 1 (positive correlation), with 0 representing no relationship between the two matrices. Higher positive values are indicative of phylosymbiosis.

Given the survey nature of the project, whose main goal was to process ~1,000 animals from as many phyla as possible, no statistical methods were used to predetermine sample size and animals/samples were not randomly assigned to experimental conditions. Data collection and analysis were not performed blind to the parameters of the survey.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All sequence data are deposited in the NCBI Short Read Archive under the BioProject accession number PRJNA746569. Specimen photographs are deposited at Dryad (https://doi.org/10.5061/dryad.ngf1vhhv6).

Code availability

No custom code has been used during this work. All analyses were conducted with publicly accessible packages in R and have been cited in the Methods.

Received: 16 July 2021; Accepted: 12 April 2022;
Published online: 26 May 2022

References

1. Gilbert, S. F., Sapp, J. & Tauber, A. I. A symbiotic view of life: we have never been individuals. *Q. Rev. Biol.* **87**, 325–341 (2012).
2. Bass, D., Stentiford, G. D., Wang, H.-C., Koskella, B. & Tyler, C. R. The pathobiome in animal and plant diseases. *Trends Ecol. Evol.* **34**, 996–1008 (2019).
3. Husnik, F. & Keeling, P. J. The fate of obligate endosymbionts: reduction, integration, or extinction. *Curr. Opin. Genet. Dev.* **58-59**, 1–8 (2019).
4. Berg, G. et al. Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**, 103 (2020).
5. Kwong, W. K. & Moran, N. A. Gut microbial communities of social bees. *Nat. Rev. Microbiol.* **14**, 374–384 (2016).
6. Hammer, T. J., Janzen, D. H., Hallwachs, W., Jaffe, S. P. & Fierer, N. Caterpillars lack a resident gut microbiome. *Proc. Nat Acad. Sci. USA* **114**, 9641–9646 (2017).
7. Holt, C. C., van der Giezen, M., Daniels, C. L., Stentiford, G. D. & Bass, D. Spatial and temporal axes impact ecology of the gut microbiome in juvenile European lobster (*Homarus gammarus*). *ISME J.* **14**, 531–543 (2020).
8. Pollock, F. J. et al. Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat. Commun.* **9**, 4921 (2018).
9. Thomas, T. et al. Diversity, structure and convergent evolution of the global sponge microbiome. *Nat. Commun.* **7**, 11870 (2016).
10. Engelberts, J. P. et al. Characterization of a sponge microbiome using an integrative genome-centric approach. *ISME J.* **14**, 1100–1110 (2020).
11. Ley, R. E. et al. Evolution of mammals and their gut microbes. *Science* **320**, 1647–1651 (2008).
12. Mallot, E. K. & Amato, K. R. Host specificity of the gut microbiome. *Nat. Rev. Microbiol.* **19**, 639–653 (2021).
13. Colston, T. J. & Jackson, C. R. Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown. *Mol. Ecol.* **25**, 3776–3800 (2016).
14. Levin, D. et al. Diversity and functional landscapes in the microbiota of animals in the wild. *Science* **372**, eabb5352 (2021).
15. Nishida, A. H. & Ochman, H. Rates of gut microbiome divergence in mammals. *Mol. Ecol.* **27**, 1884–1897 (2013).
16. Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J. & Bordenstein, S. R. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biol.* **14**, e2000225 (2016).
17. Mazel, F. et al. Is host filtering the main driver of phylosymbiosis across the tree of life? *mSystems* **3**, https://doi.org/10.1128/mSystems.00097-18 (2018).
18. Lutz, H. L. et al. Ecology and host identity outweigh evolutionary history in shaping the bat microbiome. *mBio* **4**, 6 (2019).
19. Grond, K. et al. No evidence for phylosymbiosis in Western chipmunk species. *FEMS Microbiol. Ecol.* **96**, fiz182 (2020).
20. Song, S. J. et al. Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. *mBio* **11**, 1 (2020).
21. Trevelline, B. K., Sosa, J., Hartup, B. K. & Kohl, K. D. A bird's-eye view of phylosymbiosis: weak signatures of phylosymbiosis among all 15 species of cranes. *Proc. R. Soc. B* **287**, 20192988 (2020).
22. Muegge, B. D. et al. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).
23. Youngblut, N. D. et al. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat. Commun.* **10**, 2200 (2019).
24. Amato, K. R. et al. Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *ISME J.* **13**, 576–587 (2019).
25. Moeller, A. H. et al. Social behavior shapes the chimpanzee pan-microbiome. *Sci. Adv.* **2**, e1500997 (2016).
26. Eckert, E. M., Anicic, N. & Fontaneto, D. Freshwater zooplankton microbiome composition is highly flexible and strongly influenced by the environment. *Mol. Ecol.* **30**, 1545–1558 (2021).
27. Yatsunenko, T. et al. Human gut microbiome viewed across age and geography. *Nature* **486**, 222–228 (2012).
28. Bik, H. M. Microbial metazoa are microbes too. *mSystems* **4**, e00109–e00119 (2019).
29. Schuelke, T., Pereira, T. J., Hardy, S. M. & Bik, H. M. Nematode-associated microbial taxa do not correlate with host phylogeny, geographic region or feeding morphology in marine sediment habitats. *Mol. Ecol.* **27**, 1930–1951 (2018).
30. Guidetti, R. et al. Further insights in the Tardigrada microbiome: phylogenetic position and prevalence of infection of four new Alphaproteobacteria putative endosymbionts. *Zool. J. Linn. Soc.* **188**, 925–937 (2020).
31. Giere, O. *Meiobenthology* (Springer-Verlag, 2009).
32. Laumer, C. E. et al. Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc. R. Soc. B* **286**, 20190831 (2019).
33. Hammer, T. J., Sanders, J. G. & Fierer, N. Not all animals need a microbiome. *FEMS Microbiol. Lett.* **366**, fnz117 (2019).
34. Alejandre-Colomo, C. et al. Cultivable *Winogradskyella* species are genomically distinct from the sympatric abundant candidate species. *ISME Commun.* **1**, 51 (2021).
35. Husnik, F. et al. Bacterial and archaeal symbioses with protists. *Curr. Biol.* **31**, R862–R877 (2021).
36. Salje, J. Cells within cells: Rickettsiales and the obligate intracellular bacterial lifestyle. *Nat. Rev. Microbiol.* **19**, 375–390 (2021).
37. Neave, M. J., Apprill, A., Ferrier-Pagès, C. & Voolstra, C. R. Diversity and function of prevalent symbiotic marine bacteria in the genus *Endozoicomonas*. *Appl. Microbiol. Biotechnol.* **100**, 8315–8324 (2016).
38. Weiland-Bräuer, N. et al. Composition of bacterial communities associated with *Aurelia aurita* changes with compartment, life stage, and population. *Appl. Environ. Microbiol.* **81**, 6038–6052 (2015).
39. Bik, E. M. et al. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. *Nat. Commun.* **7**, 10516 (2016).
40. Burns, A. R. et al. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J.* **10**, 655–664 (2016).
41. McFall-Ngai, M. Adaptive immunity: care for the community. *Nature* **445**, 153 (2007).
42. Ruehland, C. & Dubilier, N. Gamma- and epsilonproteobacterial ectosymbionts of a shallow-water marine worm are related to deep-sea hydrothermal vent ectosymbionts. *Environ. Microbiol.* **12**, 2312–2326 (2010).
43. Gruber-Vodicka, H. R. et al. Two intracellular and cell-type specific bacterial symbionts in the placozoan *Trichoplax* H2. *Nat. Microbiol.* **4**, 1465–1474 (2019).
44. Schockaert, E. R. in *Methods for the Examination of Organismal Diversity in Soils and Sediments* (ed. Hall, G. S.) 211–225 (CABI, 1996).
45. Higgins, R. P. in *Introduction to the Study of Meiofauna* (eds. Higgins, R. P. and Thiel, H.) 328–331 (SIP, 1988).
46. Schram, M. D. & Davison, P. G. Irwin Loops—a history and method of constructing homemade loops. *Trans. Kans. Acad. Sci.* **115**, 35–40 (1903).
47. Medlin, L., Elwood, H. J., Stickel, S. & Sogin, M. L. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**, 491–499 (1988).
48. Bower, S. M. et al. Preferential PCR amplification of parasitic protistan small subunit rDNA from metazoan tissues. *J. Eukaryot. Microbiol.* **51**, 325–332 (2004).
49. Comeau, A. M., Li, W. K. W., Tremblay, J.-E., Carmack, E. C. & Lovejoy, C. Arctic ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE* **6**, e27492 (2011).
50. Zhang, R.-Y. et al. Design of targeted primers based on 16S rRNA sequences in meta-transcriptomic datasets and identification of a novel taxonomic group in the Asgard archaea. *BMC Microbiol.* **20**, 25 (2020).
51. Lane, D. J. in *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M) 115–175 (Wiley, 1991).

52. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
53. Marcel, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).
54. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
55. Callahan, B. J. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
56. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
57. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
58. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
59. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
60. Love, M. I., Huber, W. & Anders, S. Moderate estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
61. Kurtz, Z. D. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **11**, e1004226 (2015).
62. Csardi, G. & Nepusz, T. *The igraph Software Package for Complex Network Research* (InterJournal, 2006).
63. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
64. Kolde, R. pheatmap: pretty heatmaps. R package version 1.0.12 https://CRAN.R-project.org/package=pheatmap (2015).
65. Lin, H. & Das Peddada, S. Analysis of composition of microbiomes with bias correction. *Nat. Commun.* **11**, 3514 (2020).
66. Oksanen, J. vegan: Community Ecology Package. R package version 2.5.7 https://CRAN.R-project.org/package=vegan (2020).
67. Rouse, G., Pleijel, F. & Tilic, E. *Annelida* (Oxford Univ. Press, 2022).
68. Ahmed, M. & Holovachov, O. Twenty years after De Ley and Blaxter—How far did we progress in understanding the phylogeny of the phylum Nematoda? *Animals* **11**, 3479 (2021).
69. Van Steenkiste, N. W. L., Herbert, E. R. & Leander, B. S. Species diversity in the marine microturbellarian *Astrotorhynchus bifidus* sensu lato (Platyhelminthes: Rhabdocoela) from the Northeast Pacific Ocean. *Mol. Phylogenet. Evol.* **120**, 259–273 (2018).

Acknowledgements

We wish to thank F. Bergmeier, B. Brenzinger, S. Fujimoto, U. Jondelius, M. Kolicka, C. Nielsen, A. Schmidt-Rhaesa, M. Sørensen and W. Sterrer for additional taxonomic expertise, the Hakai Institute and CARMABI and their helpful staff (in particular, N. Acharya-Patel, C. Janusson and C. Prentice from Hakai for assistance with DNA extractions) and C. Wall and G. Buckholtz for laboratory procedures at the University of British Columbia. This project was funded by the Tula Foundation's Hakai Institute (P.J.K. and B.S.L.), the Natural Sciences and Engineering Research Council (NSERC 2019-03896 to B.S.L. and NSERC 2019-03994 to P.J.K.), the Gordon and Betty Moore Foundation (https://doi.org/10.37807/GBMF9201, P.J.K.) and the Canadian Graduate Scholarship programme (N.A.T.I.).

Author contributions

The project was conceived and planned by V.B., N.W.L.V.S., M.H., B.S.L. and P.J.K. B.S.L. and P.J.K. are the senior authors and provided supervision and funding. V.B., N.W.L.V.S., M.H. and N.A.T.I. performed most of the field and laboratory work. C.C.H. conducted the analyses. V.M., N.O. and R.S.P. contributed to field and laboratory work. N.W.L.V.S., M.H., P.A.C., K.G., O.H., A.K. and K.W. identified the animals. V.B., C.C.H. and P.J.K. wrote the original draft. All authors edited iterative versions of the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-022-01125-9.

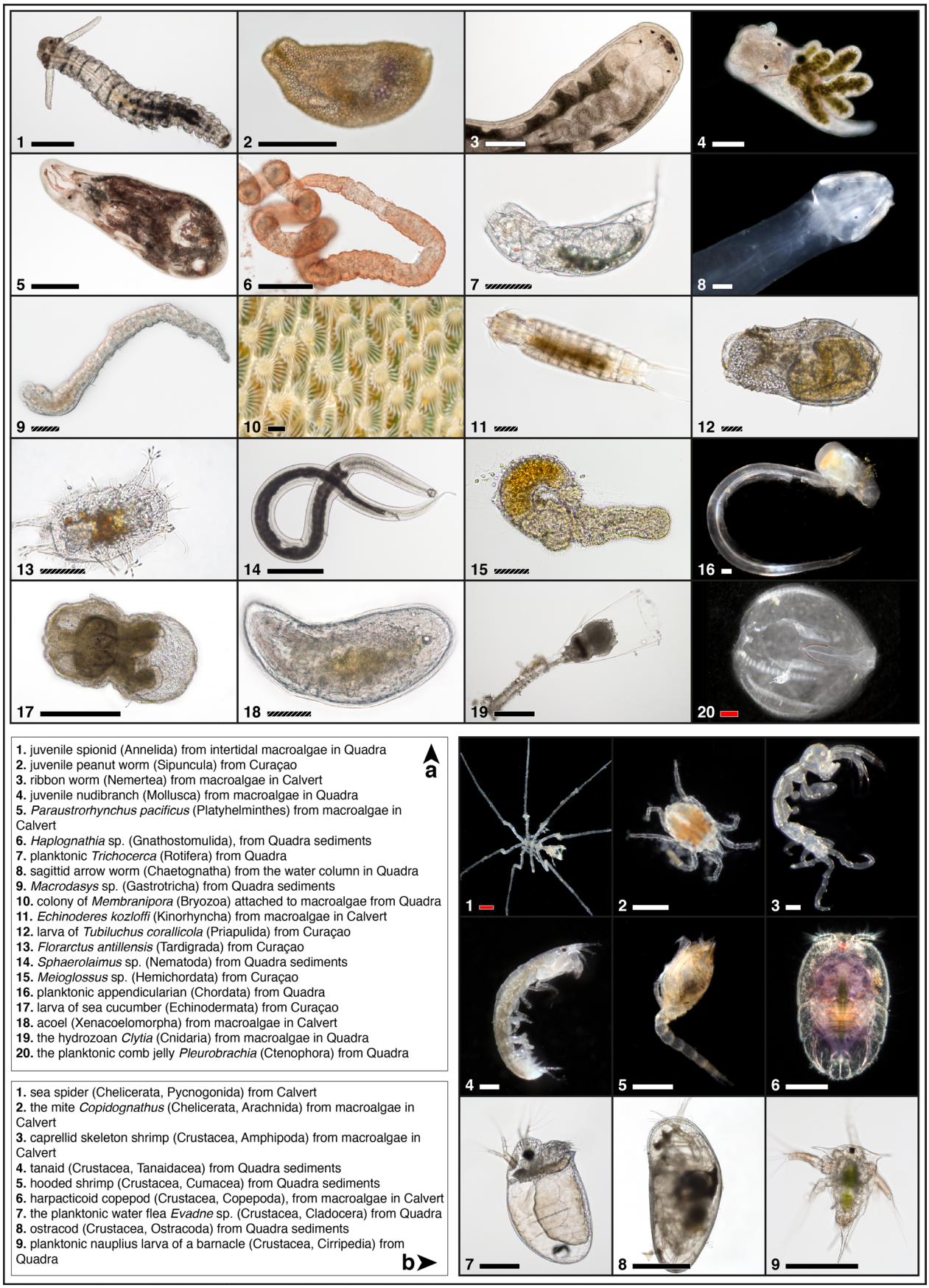
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-022-01125-9.

Correspondence and requests for materials should be addressed to V. Boscaro or P. J. Keeling.

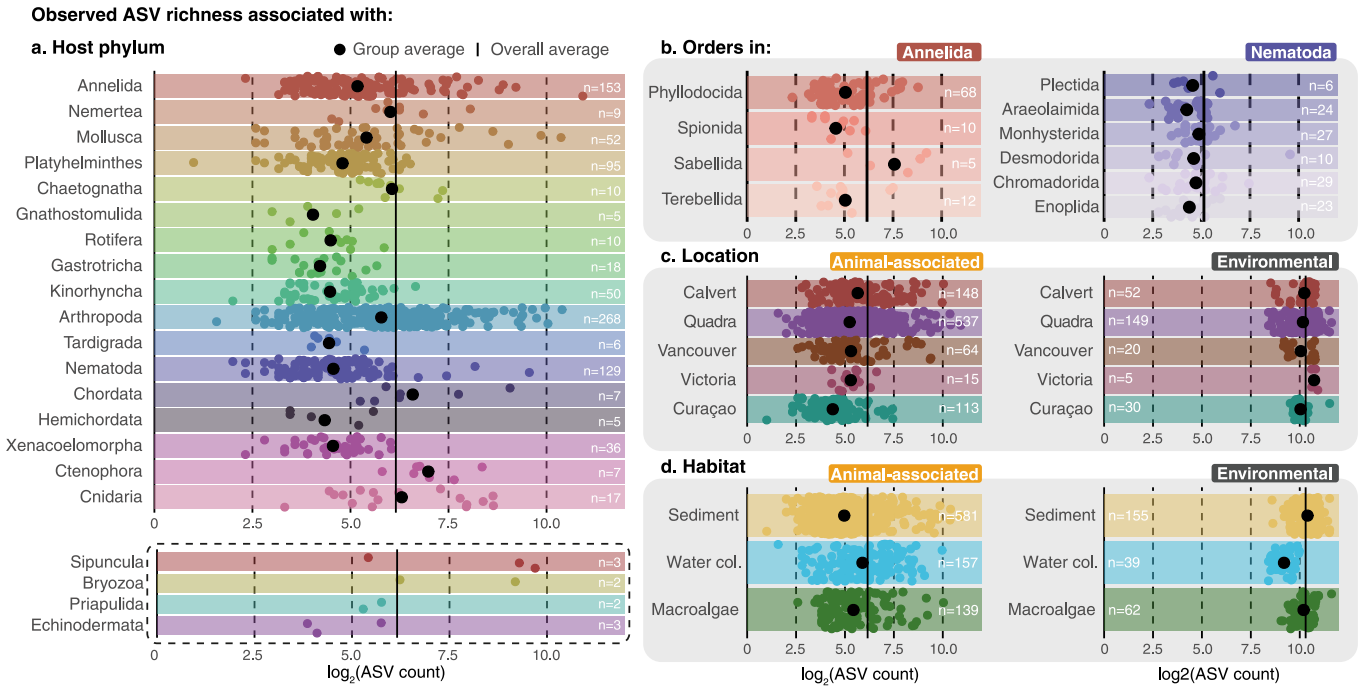
Peer review information *Nature Microbiology* thanks Diego Fontaneto, Holly Lutz, Owen Wangensteen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

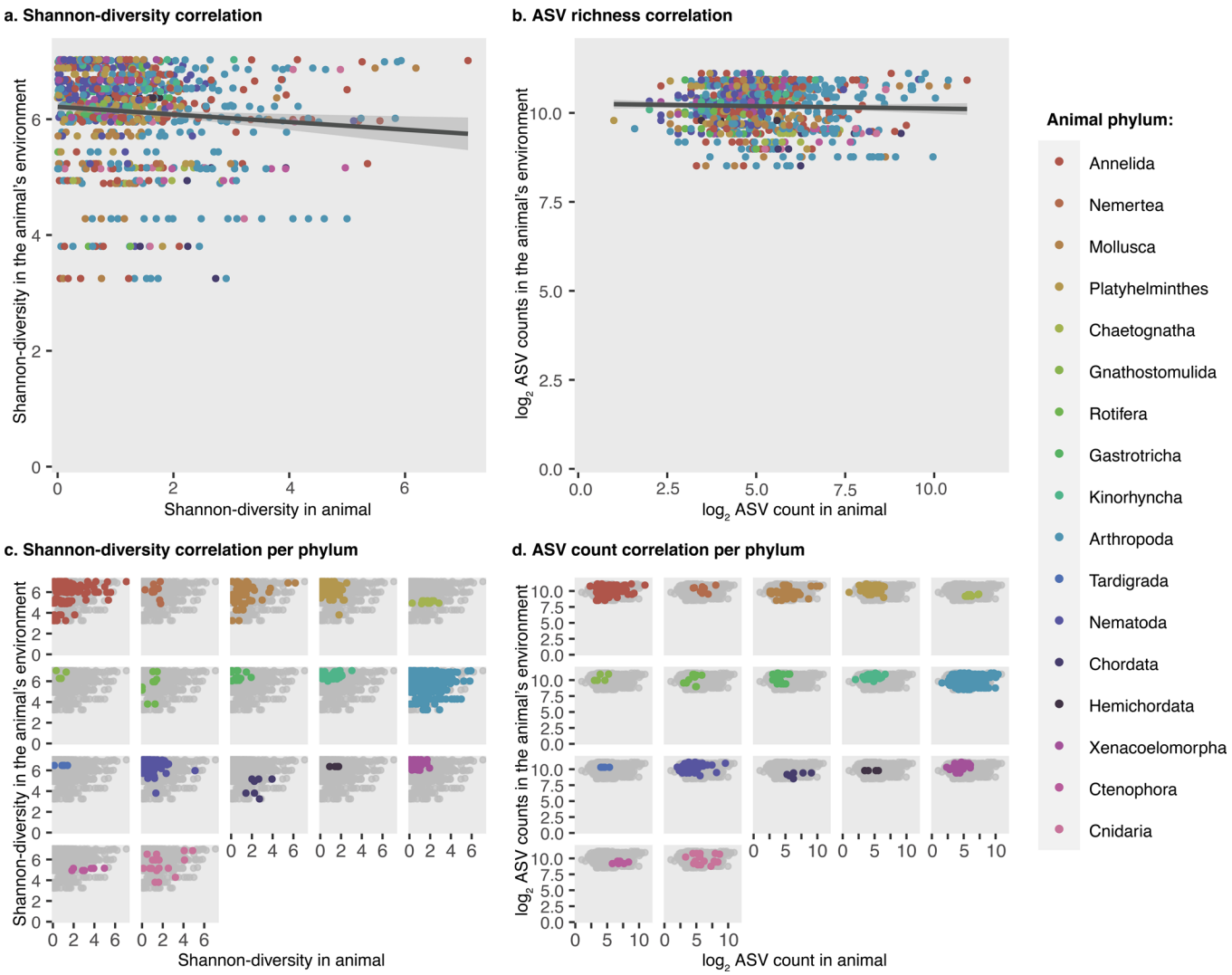
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



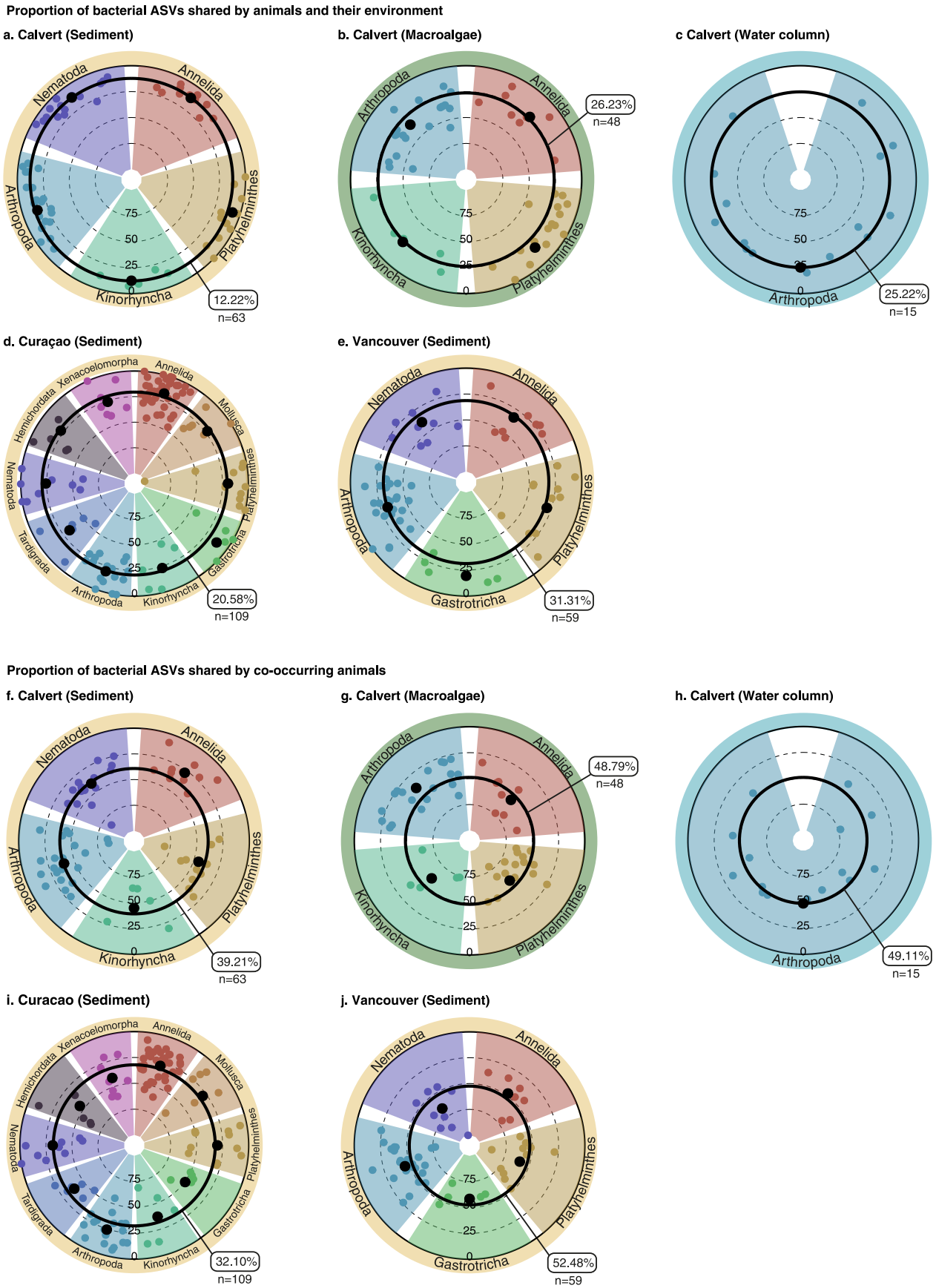
Extended Data Fig. 1 | Examples of animals collected during the survey. a, Examples of collected animals from non-arthropod phyla. **b**, Examples of collected animals from arthropod lineages. Bars stand for 250 μm, except for a7, a9, a11, a12, a13, a15, and a18, where they stand for 50 μm (hatched bars), and a20 and b1, where they stand for 1,000 μm (red bars).



Extended Data Fig. 2 | Observed number of ASVs in host-associated and environmental microbial communities. Bacterial richness, expressed as ASV counts, of the microbial communities, grouped according to: **a**, host phylum; **b**, host order within the phylum Annelida and Nematoda; **c**, location (for both animal-associated and environmental data); and **d**, habitat (for both animal-associated and environmental data). The dashed box in **a** separates phyla with less than five sampled specimens each. Microbial community Shannon-diversity (Fig. 2) and ASV numbers are considerably higher in environmental communities than in invertebrate hosts.

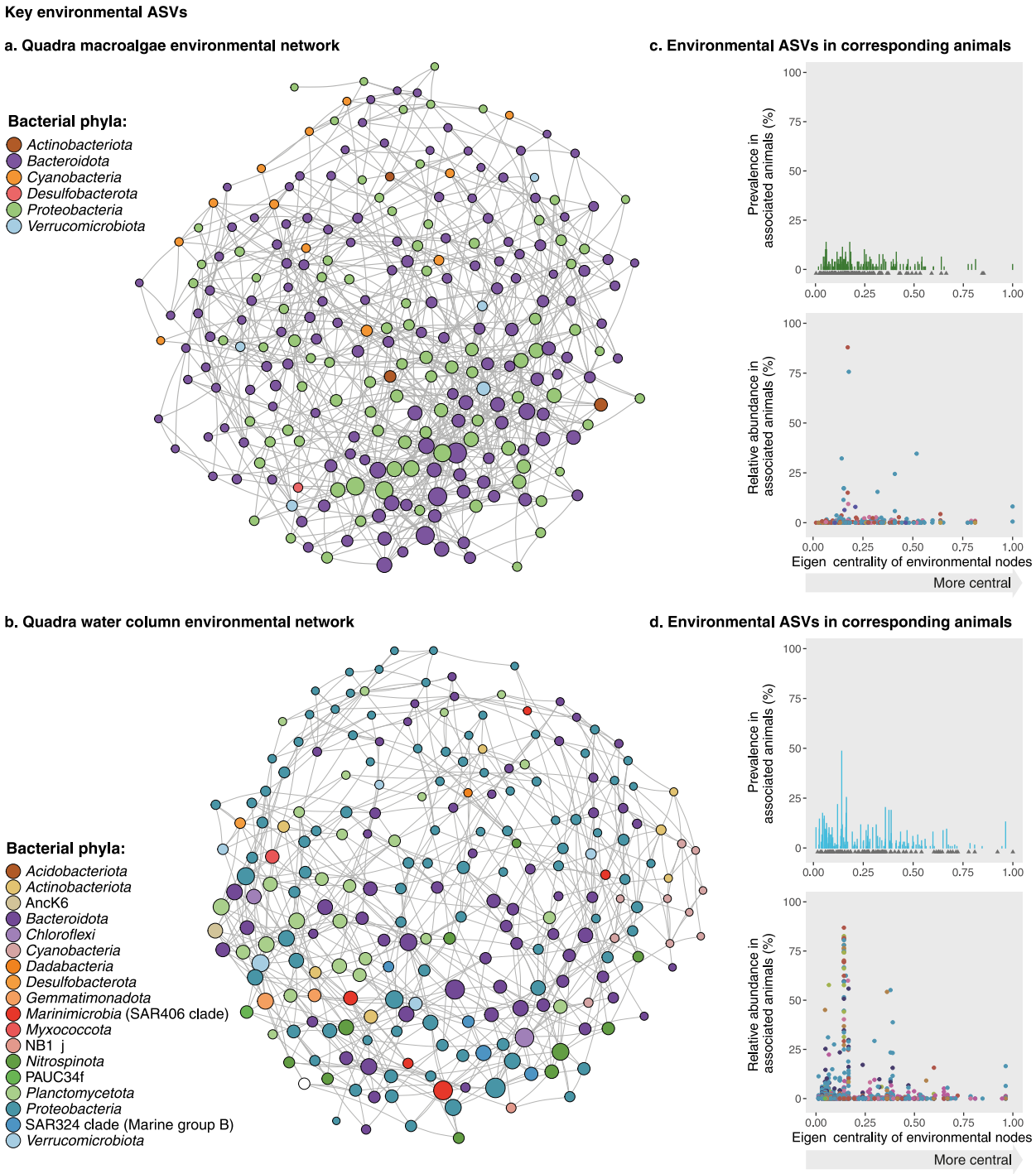


Extended Data Fig. 3 | Diversity and richness of host-associated and environmental microbial communities do not correlate. a, Individual animal microbiome Shannon-diversity index and **b**, ASV counts plotted against the averaged values from corresponding environmental communities. **c,d**, same plots separated according to host phylum. Differences in Shannon-diversity and richness of animal-associated microbiomes are not tied to Shannon-diversity and richness of background environmental communities. The grey area shows the 95% confidence interval (default geom_smooth se parameter). n = 877 specimens. n of specimens per phylum as in Fig. 2d.

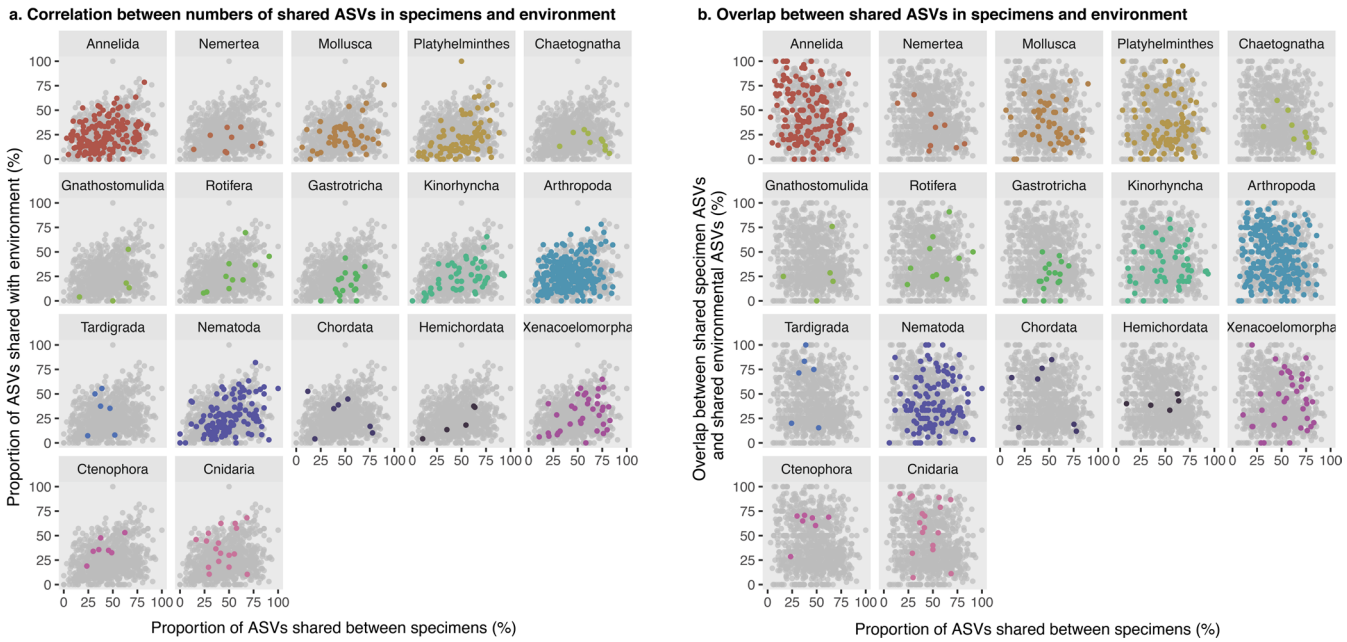


Extended Data Fig. 4 | See next page for caption.

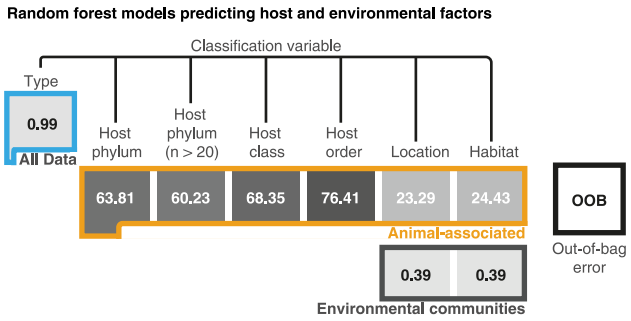
Extended Data Fig. 4 | Microbiome overlap between animals and their environment in locations other than Quadra Island. a-e, Proportion of bacterial Amplicon Sequence Variants shared between individual invertebrates and their environment, separated by habitat for locations with multiple sampled habitats. **f-j**, Proportion of bacterial ASVs shared between individual specimens and all other animals in the same sample, separated by habitat for locations with multiple sampled habitats. Solid, black lines in circular plots indicate overall average. Dashed lines indicate 25%, 50%, and 75% thresholds, for scale. Black circles plot phylum average.



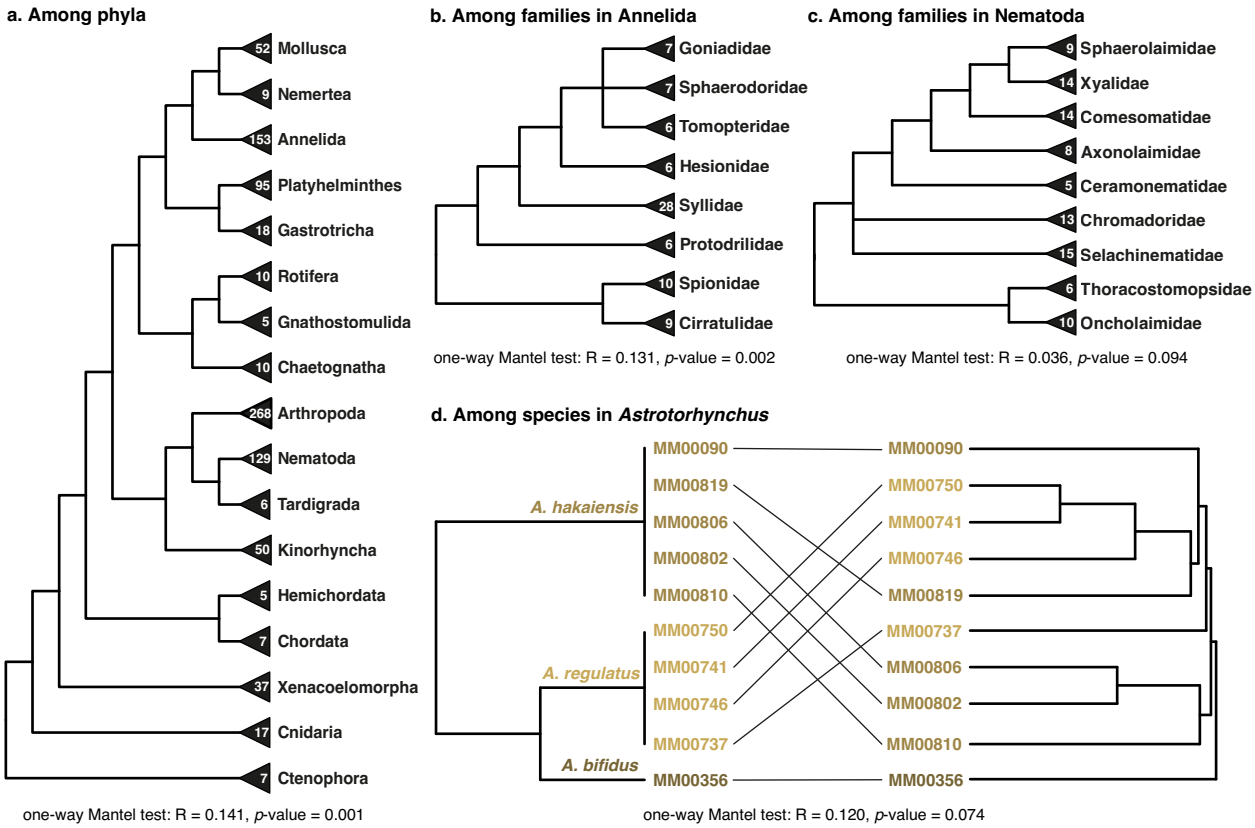
Extended Data Fig. 5 | Influence of keystone environmental bacteria in animal-associated microbiomes isolated from macroalgae and water. SPIEC-EASI co-occurrence network of key environmental ASVs found in Quadra Island **a**, macroalgae (n = 253 ASVs) and **b**, water column (n = 228 ASVs) samples. Each node represents a single ASV. Lines connecting two nodes (edges) indicate an association between the two ASVs. Node size is scaled to eigen-centrality, which considers the number of connecting nodes as well as their subsequent connections. **c,d**, prevalence and abundance (both as %) of the same environmental ASVs (respective of each habitat) in animals from the same habitat and location. Individual ASVs (on the x axis) are ordered according to their eigen-centrality in the environmental network, and may be represented by multiple datapoints in the abundance plot (on the right) to reflect their varying abundance in multiple host phyla. Grey arrowheads in prevalence plots indicate environmental ASVs that are absent in host-associated microbiomes. Point colour indicates host phyla. As is the case in sediments from the same location (see text), keystone environmental bacteria are not particularly abundant nor prevalent in animal-associated microbiomes.



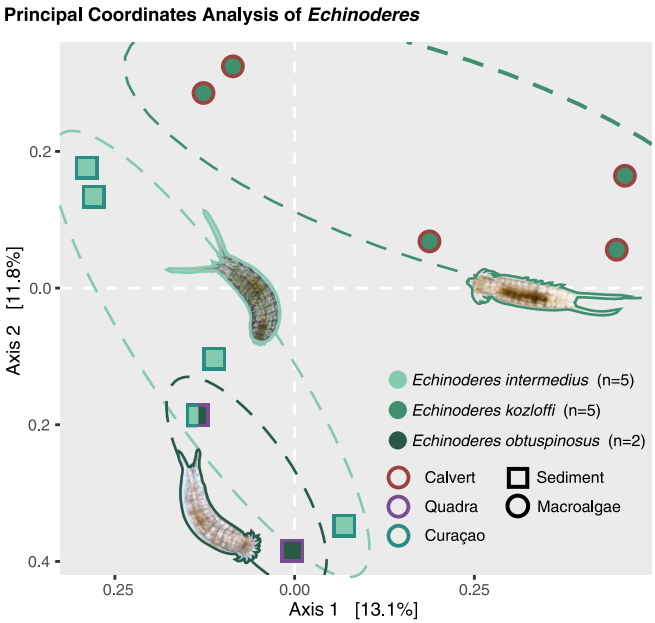
Extended Data Fig. 6 | Correlation of ASVs shared between animals and those shared between animals and the environment. The proportion of bacterial ASVs shared between individual invertebrates collected in Quadra Island and all other co-occurring animals in the same sample plotted against: **a**, the proportion of bacterial ASVs shared between animals and their environment; **b**, the proportion of shared ASVs between animals that are also shared with the environment. Both coloured and separated according to host phylum. While there is a tendency for co-occurring animals to share more ASVs in samples where more ASVs are also shared with the environment, the ASVs responsible for both overlaps do not increase in number accordingly, and hence are not necessarily the same. n of specimens per phylum as in Fig. 2d.



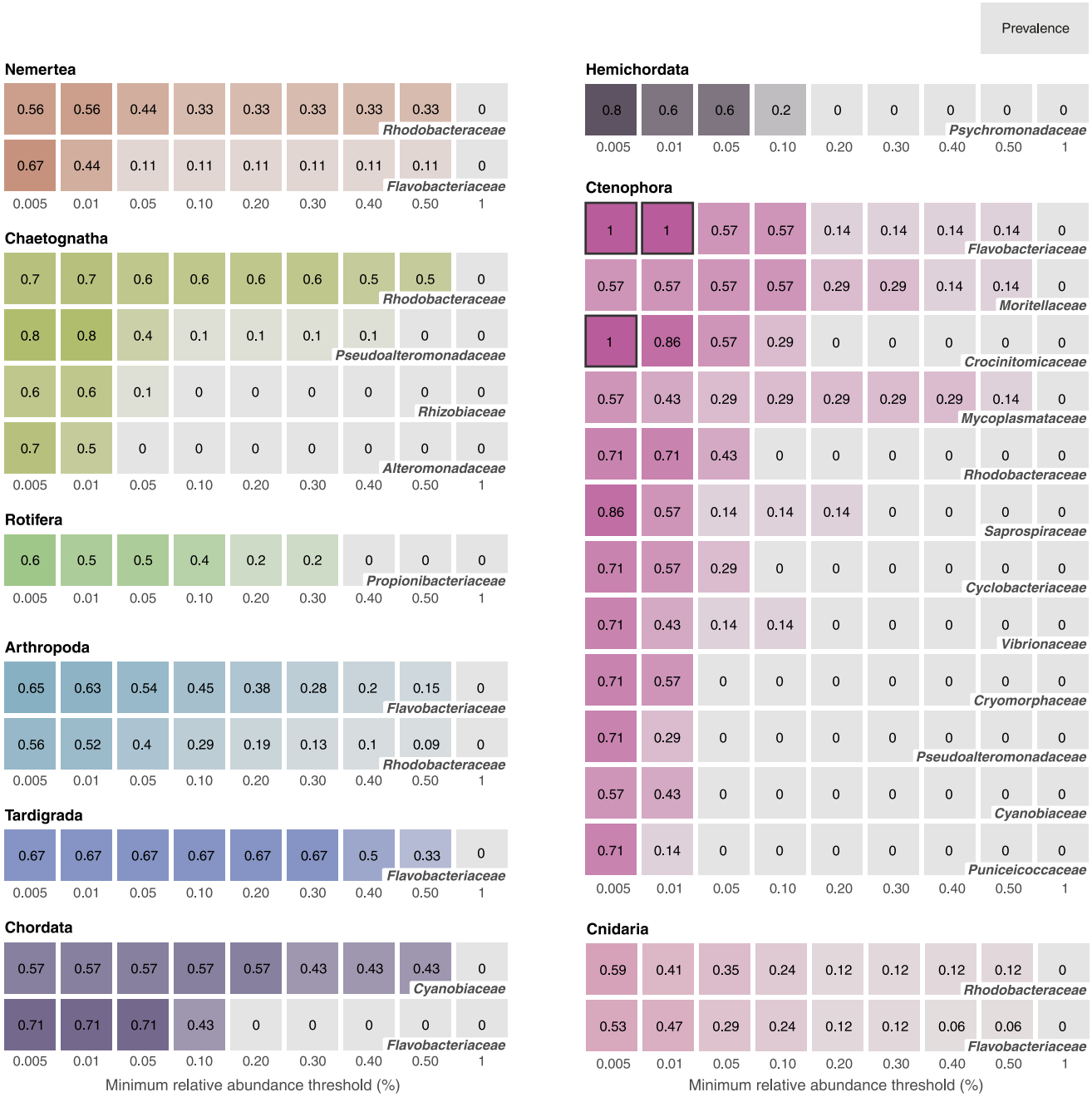
Extended Data Fig. 7 | Predicting host taxonomy and environmental factors with random forest models. Out-of-bag error rates of random forest models using microbial community ASVs to predict potential groupings. From left to right: all data, predicting community type (host-associated vs. environmental); animal-associated data, predicting host phylum, host phylum restricted to phyla that only include more than 20 specimens, host class, and host order; both animal-associated and environmental data, predicting location and habitat. The models can confidently discriminate microbial community type as well as environmental parameters from environmental communities. They fare poorly when discriminating any parameter from host-associated microbiomes, especially those related to host taxonomy.



Extended Data Fig. 8 | Phylogenetic relationships used in phylogeny-based phylosymbiosis tests. Phylogenetic trees (topology only) among **a**, phyla investigated in this work[32]. **b**, families in Annelida;[67] **c**, families in Nematoda;[68] **d**, species in *Astrotorhynchus*[69]. Mantel tests, results of which are shown under each panel, compare the phylogenetic topologies shown with similarity dendrograms of microbiomes from corresponding specimens, as shown in **d**. Microbiome data from specimens belonging to the same taxon (phylum, family, or species) were mapped in 0-branch lengths polytomies in each tree, as shown for *Astrotorhynchus* (numbers of specimens per polytomy are reported within dark triangles in **a–c**). All the performed analyses showed a very low degree of covariation (R value) between host phylogeny and microbiome similarity.



Extended Data Fig. 9 | Potential phylosymbiosis signal in *Echinoderes* (Kinorhyncha). Principal Coordinates Analysis using Bray–Curtis dissimilarity of microbiomes from *Echinoderes* specimens largely clustered according to host species. Ellipses group specimens of the same species.



Extended Data Fig. 10 | Core bacterial families found in invertebrate phyla. Prevalence of bacterial families at increasing relative abundance thresholds within each invertebrate phylum. Only families present in the majority of specimens (>50%) at or above 0.005% relative abundance are included. Actual prevalence values are included at each threshold with colour denoting degree of prevalence. Families occurring in all specimens at a given abundance threshold (prevalence value = 1) are indicated by a dark grey outline.

nature portfolio

Corresponding author(s): Vittorio Boscaro

Last updated by author(s): Apr 4, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect data in this study.
Data analysis	The following softwares were used for data analysis: Cutadapt v.3.4; the software package R, including packages DADA2 v.1.14.1; phyloseq v.1.36.; ggplot v.3.3.5; DESeq2 v.1.34.0; SPIEC-EASI v.1.1.1; igraph v.1.2.7; randomForest v.4.6.14; microbiome v.1.16.0; ANCOMBC v.1.2.2; vegan v.2.5.7

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All sequence data is deposited in the NCBI Short Read Archive under the BioProject accession number PRJNA746569. Specimen photographs are deposited at Dryad (doi: 10.5061/dryad.ngf1vhhv6).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We collected and characterized the microbiomes of about 1,000 microscopic marine invertebrates, sampled in five locations (two main areas: Canada and Curacao) and belonging to 21 different phyla
Research sample	We collected individual microscopic invertebrates (mostly in the 100-2,000 µm size range) from natural populations, targeting understudied groups in order to fill the gaps in our knowledge of animal-associated microbiomes (the vast majority of which was characterized only in vertebrates, selected insects, coral, and sponges). Specimens from 21 different phyla were picked from environmental samples, imaged, washed in sterile water and stored for DNA extraction and molecular procedures. Environmental controls of sediment, water, and rock/macroalgae swabs were also collected to characterize background microbial communities. Details on each individual specimen and control are provided in Supplementary Table 2
Sampling strategy	The goal of approximately 1,000 specimens was set in advance, in order to match and exceed most previous similar studies. The distribution of specimens across taxa is roughly representative of the taxonomic distribution in the original samples. A minimum of 15 (average: 22) specimens per sample were collected, compatibly with the time required to process the samples while the animals are still viable and with their original microbiome intact. Sediment samples were mostly collected with a dredge or manually (if intertidal). Macroalgal samples were collected manually. Water column samples were collected with plankton tows using a 64 µm mesh.
Data collection	Some of the authors (VB, NVS, MH, NATI) were present during sample and animal collection and collected all metadata in situ, updating electronic spreadsheets twice a day during collection trips.
Timing and spatial scale	Animals were collected from July 2017 to January 2019 during collection trips at equipped research stations, whenever conditions were favorable, from the start of the project until slightly more than 1,000 specimens had been collected. Sediment samples amounted usually to a few kilograms of material, macroalgal samples fit in 1-L bags and plankton tows were collected in 500 mL bottles.
Data exclusions	Obtained molecular libraries with less than 1,000 reads were excluded from further analysis after the characterization of Amplicon Sequence Variants. Specimens with discordant molecular and morphological ID at or above the family rank were removed from analysis where the changed ID was relevant.
Reproducibility	There was a considerable degree of variability in the animal communities observed in some of the sites collected multiple times. Otherwise, no systematic pattern of failure/success in different locations, areas, and field trips was observed. Since this was a survey where the main variable was the identity of the animal themselves, there were no experimental conditions to replicate or repeat.
Randomization	The goal of this survey was to collect as many and as varied understudied invertebrate taxa as possible, so this entry is not relevant.
Blinding	Data collection and analysis were not performed blind to the parameters of the survey.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Most sediment samples were collected either from a boat or on the intertidal area of beaches, under good weather conditions. Macroalgal samples were all collected in intertidal pools during low tides. Water samples were mostly collected by boat.
Location	Locations are detailed inn Supplementary Table 1. Briefly, we collected samples from four locations in British Columbia, Canada (Calvert Island, Quadra Island, Victoria, and Vancouver) and the Caribbean island of Curacao.
Access & import/export	All work was performed with the help and support of local research stations (the Hakai Institute in British Columbia and the Caribbean Research and Management of Biodiversity Institute (CARMABI) in Curacao, following all rules already in place for field work and sample collections in the investigated areas.
Disturbance	The study produced no measurable environmental disturbance, given the tiny amount of biological material needed and the protocols prepared in collaboration with local research stations. Most of the more involved operations, like sampling from boats, were coordinated with other experimental surveys conducted at the same time by the research stations, so to minimize the overall impact of research in the sites.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	No laboratory animals were used in this study.
Wild animals	More than 1,000 microscopic marine animals (below 2 mm long) were collected over the course of this study. Their details are reported in Supplementary Table S2. Each animal was isolated under the microscope while screening from the environmental sample, imaged, briefly washed, and frozen in sterile water for whole-body DNA extraction. This procedure kills the animal.
Field-collected samples	All animals came from field-collected samples around the Hakai Institute research stations, the Biodiversity Research Centre in British Columbia, and the CARMABI research station in Curacao. Housing and lab tools were provided by those institutes (with the exception of the Biodiversity Research Centre at the University of British Columbia, which is the workplace of all authors involved in sample collection).
Ethics oversight	No ethical approval is required for this kind of work on microscopic invertebrates.

Note that full information on the approval of the study protocol must also be provided in the manuscript.