Diversity, structure and convergent evolution of the global sponge microbiome

Torsten Thomas, Lucas Moitinho-Silva, Miguel Lurgi, Johannes R. Björk, Cole Easson, Carmen Astudillo-García, Julie B. Olson, Patrick M. Erwin, Susanna López-Legentil, Heidi Luter, Andia Chaves-Fonnegra, Rodrigo Costa, Peter J. Schupp, Laura Steindler, Dirk Erpenbeck, Jack Gilbert, Rob Knight, Gail Ackermann, Jose Victor Lopez, Michael W. Taylor, Robert W. Thacker, Jose M. Montoya, Ute Hentschel & Nicole S. Webster

Sponges (phylum Porifera) are early-diverging metazoa renowned for establishing complex microbial symbioses. Here we present a global Porifera microbiome survey, set out to establish the ecological and evolutionary drivers of these host–microbe interactions. We show that sponges are a reservoir of exceptional microbial diversity and major contributors to the total microbial diversity of the world’s oceans. Little commonality in species composition or structure is evident across the phylum, although symbiont communities are characterized by specialists and generalists rather than opportunists. Core sponge microbiomes are stable and characterized by generalist symbionts exhibiting amensal and/or commensal interactions. Symbionts that are phylogenetically unique to sponges do not disproportionally contribute to the core microbiome, and host phylogeny impacts complexity rather than composition of the symbiont community. Our findings support a model of independent assembly and evolution in symbiont communities across the entire host phylum, with convergent forces resulting in analogous community organization and interactions.
Microbial symbionts are essential for the function and survival of multicellular eukaryotes, ranging from humans to invertebrates to plants. Most symbioses involve complex communities of microorganisms, often comprising a large phylogenetic breadth of microbial diversity associated with a single host organism. Many factors, including host-derived nutrients, chemico-physical characteristics (for example, pH) and host properties (for example, immune response), determine the composition and structure of symbiont communities over time and space. However, the evolutionary and ecological drivers of symbiont composition in animals and plants remain largely unknown.

Sponges are among the most ancient living Metazoa and generally form symbiotic relationships with complex communities of microorganisms. Sponges can maintain highly diverse, yet specific symbiont communities, despite the constant influx of seawater microorganisms resulting from their filter-feeding activities. These symbioses are known to be at least partially underpinned by metabolic exchange between symbiont and host, including nitrogen cycling, CO₂ fixation, secondary metabolite production, and uptake and conversion of dissolved organic matter. In this respect, sponge symbionts perform analogous functions to the symbionts found in mammalian guts and plants. Therefore, sponge-microbe symbioses represent an ecologically relevant example of host–microbe interactions in an early-diverging metazoan clade.

While the diversity of sponge symbionts has been extensively addressed using molecular tools, comparative work has been hindered due to methodological differences in sampling, sample processing and data analyses. Large-scale efforts, such as the Human Microbiome Project and the Earth Microbiome Project, have standardized these technical aspects to reliably and consistently describe patterns of microbial diversity and composition. These efforts have generated a large knowledge base for host-associated microorganisms of vertebrates, and especially humans, but equivalent data sets for invertebrates are missing. To gain critical insights into the evolution and complexity of symbiotic interactions, we require a greater understanding of the properties and origins of microbial symbioses in early-divergent Metazoa. Furthermore, microbiome research has primarily focused on within-species comparisons, in particular humans, or the comparative analysis of microbiomes of very disparate host organisms (for example, plants versus mammalian guts). However, to define important aspects for the evolution of microbial symbiosis, a deeper understanding of symbiont communities in closely related host species within defined phylogenetic clades (for example, a single phylum) is required.

Here we provide a comprehensive analysis of microbial symbiont communities associated with 81 species from the phylum Porifera. Throug}
**Species (environments)**

<table>
<thead>
<tr>
<th>Species</th>
<th>Environment</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xestospongia muta</td>
<td>(Water)</td>
<td>5</td>
</tr>
<tr>
<td>Carteriospongia foliacea</td>
<td>(Water)</td>
<td>7</td>
</tr>
<tr>
<td>S. m.</td>
<td>(Water)</td>
<td>5</td>
</tr>
<tr>
<td>X. 1055</td>
<td>(Water)</td>
<td>4</td>
</tr>
<tr>
<td>Halichondria panicea</td>
<td>(Water)</td>
<td>6</td>
</tr>
<tr>
<td>Ectyoplasia ferox</td>
<td>(Water)</td>
<td>4</td>
</tr>
<tr>
<td>Phyllospongia clathria</td>
<td>(Water)</td>
<td>3</td>
</tr>
<tr>
<td>C. muta</td>
<td>(Water)</td>
<td>3</td>
</tr>
<tr>
<td>C. baik</td>
<td>(Water)</td>
<td>4</td>
</tr>
<tr>
<td>C. bah</td>
<td>(Water)</td>
<td>3</td>
</tr>
<tr>
<td>C. aur</td>
<td>(Water)</td>
<td>5</td>
</tr>
</tbody>
</table>

**Figure 2 | Variability of symbiont communities.** Intraspecific community dissimilarity measured as distance of samples to group centroids for 16S rRNA gene composition of different sponge species (orange) and habitats (blue: seawater; brown: sediment). Vertical bar represent the median, the box represent the first to third quartiles and whiskers show the lowest or highest datum within 1.5 times the interquartile range of the lowest and upper quartile, respectively. Names in brackets represent the abbreviations used in all subsequent figures. The number behind the brackets refers to the number of individual samples analysed per sponge taxon, seawater or sediment.
dominant in most sponges species analysed here, with Chloroflexi, Cyanobacteria and Crenarchaeota occasionally reaching high relative abundances (~10%). Nevertheless, sponges host a high diversity of phyla (albeit at low relative abundances), with over 32 phyla and candidate phyla regularly reported to associate with sponges\textsuperscript{20} and a further 6 phyla and 14 candidate phyla recently reported as part of the rare community using a deep Illumina sequencing approach\textsuperscript{21}. In the current study, we detected 41 phyla (including candidate phyla) with all sponge species with more than 20 replicate samples as well as seawater microbial communities in sponges and seawater. (Fig. 3).

Sponges harbour an exceptional microbial diversity. High sample replication (n > 20) employed in this study facilitated estimation of total microbial richness for specific sponge species and seawater. Analysis of the 133 surface seawater samples (collected here from disparate geographic areas, including Spain, Florida, Puerto Rico, Sweden, Mexico, Bahamas and Australia) showed that the combined planktonic richness in these regions approaches 15,000 OTUs (at 97% sequence identity) (Fig. 4). This estimate lies between the ~20,000 and ~9,000 predicted OTUs (at 97% sequence identity) found in surface waters of the coastal and open ocean, respectively, as part of the International Census of Marine Microbes (ICoMM)\textsuperscript{22}, which was based on pyrosequencing analysis of the V6 region of the 16S rRNA gene. However, the estimated planktonic richness in this and our study is lower than the 29,457 OTUs (at 97% sequence identity) recently reported using Illumina amplicon sequencing of seawater\textsuperscript{21} or the 37,470 OTUs estimated from metagenomic sequencing of the global Tara Oceans samples\textsuperscript{23}, with the higher richness in the latter two studies likely explained by the inclusion of deep-water samples. Remarkably though, richness estimates show that a single sponge species can harbour as many different
Symbiont communities consist of generalists and specialists (that is, found in only one or a few sponge species), while only a few are truly cosmopolitan (that is, found across many sponge species). Importantly, the degree distribution for the subset of OTUs belonging to previously defined sponge-specific sequence clusters follows the same distribution as the whole (see below).

OTUs as might be expected from the surrounding seawater. For example, Carteriospongia foliacea and Ircinia variabilis (n ≥ 50 individuals across their sampled biogeographic distribution) contain more than 12,000 OTUs (Fig. 4). Similar richness projections were observed for the species Cliona delitrix, Ircinia strobilina, Ircinia oros, Mycale laxissima, Plakortis halichondrioides, Sarcotragus fasciculatus, Xestospongia sp. and Xestospongia muta, which were each sampled between 20 and 50 times (Fig. 4).

Limited overlap in microbiome structure was observed between different sponge species or between sponges and the seawater and sediment samples (Fig. 5). Thus, considering all OTUs discovered across the 804 sponge samples that included 81 different species, richness estimates approach a value of 40,000 OTUs (Fig. 4b). The 81 sponge species analysed here represent only a tiny fraction of the 8,553 described sponge species (and likely a higher number when considering undescribed species) suggesting that sponge-associated (and likely other host-associated) communities are a significant global source of unique microbial diversity.

Figure 6 | Cumulative degree distributions for OTUs (black dots, bottom and left axes) and sponges (red dots, top and right axes). Black dots correspond to the number of different host species (k) that contain a given OTU, represented as the cumulative probability of finding an OTU in the network with k or less-associated hosts (Pc(k)). Red dots correspond to the number of different OTUs (k) found in a given host species, represented as the cumulative probability of finding a sponge host with k or less-asssociated OTUs (Pc(k)). The OTU degree distribution followed a truncated power-law P(k) = k 0.32 e k/7.44, while the sponge degree distribution followed an exponential given by P(k) = e k/1.849. Blue and orange dots correspond to random degree distributions for OTUs and sponges, respectively, where the number of nodes and links from the empirical distribution is kept constant.

Figure 5 | Community similarity for microbial communities in sponges, seawater and sediments. Clustering was performed using multi-dimensional scaling of Bray–Curtis distances. Sponge species abbreviations are outlined in Fig. 2.
The cumulative probability of finding a sponge host with k or less-associated OTUs also follows a skewed degree distribution with exponential decay. A large fraction (> 50%) of species harbour a symbiont diversity between ~60 and ~1,800 distinct taxa, while a small fraction of sponge species can harbour up to ~7,000 OTUs (see also above). Skewed degree distributions have been identified in several types of ecological networks, and are linked to important properties of ecological communities, such as their robustness to species loss and their stability over time27. Our results suggest that ecological communities formed between microbial symbionts and their sponge hosts display similar patterns, which may be linked to their ability to maintain important functions at both the host and ecosystem levels12.

To further investigate the specialization of OTUs in our interaction network, we analysed how consistently they are found across individual replicates of any given host species. Both highly specialized (defined here as those found in less than five different host species) and generalist OTUs (defined here as those found in more than 50 different host species) are present in a large fraction of the biological replicates of their respective host species (Fig. 7; Supplementary Fig. 2). In contrast, a large proportion of OTUs with an intermediate degree of host association (between 5 and 50 host species) can be considered as opportunistic taxa, associated with only a few biological replicates of multiple host species. Thus, symbiont communities within the phylum Porifera are characterized by a combination of highly generalist and truly specialist community members. Our analysis showed that generalists are cosmopolitan not only qualitatively (that is, present in a large number of species), but also quantitatively (that is, consistently present in a large fraction of individuals of those host species). To our knowledge, such patterns have not previously been observed for ecological networks, as it has traditionally been difficult to undertake repeated measures of many individuals across multiple host species.

**Generalist symbionts comprise the core sponge microbiome.** Considering the existence of generalist (that is, cosmopolitan) sponge OTUs, we queried their relative contribution to the core microbiome of any individual species. Here we define a core membership as any OTU that is present in at least 85% of the replicates for any single host species. To effectively model population dynamics of these OTUs, we identified host species with a sufficiently large number of replicates (here ≥47) across the entire data set. We identified five host species (*Carteriospongia foliascens*, *Cliona delitrix*, *Ircinia oros*, *Ircinia variabilis* and *Sarcotragus fasciculatus*) that fit this requirement and observed cores ranging in size from 7 to 20 OTUs. The proportion of OTUs with a certain degree (number of connections to different sponge species) or higher, and the frequency distribution of degrees were compared for all OTUs present in the global bipartite network and the aggregated subset of OTUs present in all five core microbiomes (Fig. 8). The core OTUs aggregated from all five sponge species showed an uneven distribution of degree frequencies. Core OTUs are primarily generalist and cosmopolitan (high degree) OTUs, while specialist (low degree) and intermediate degree OTUs are underrepresented. This shows that highly connected OTUs in the global bipartite network also tend to comprise a larger fraction of the core for each of the host species investigated here.

**Strong density dependence and weak, unidirectional interactions.** Of particular interest is whether these core OTUs and their local interactions are important for the overall dynamics of the symbiont populations within each host species. For instance, density dependence (that is, the growth rate of a population is controlled by its density) has a strong effect on community dynamics, with stabilizing effects on population fluctuations26. To disentangle the complex nature of microbe–microbe interactions within our five sponge hosts described above, we applied a statistical framework27 that models population dynamics of the Lotka–Volterra type and allows us to decouple the variation in relative abundance of populations into contributions of (i) inter-specific interactions, (ii) density dependence and (iii) environmental stochasticity. Population dynamics are sensitive to both non-modelled environmental stochasticity and modelled fluctuating environmental conditions27. However, in this study, the environment is considered as fixed due to replicates being sampled from similar environments during the same time period (see Supplementary Data 2 and also (ref. 28) for temporal stability of symbionts in *Ircinia oros*, *Ircinia variabilis* and *Sarcotragus fasciculatus*), hence population dynamics are considered to be influenced solely by environmental stochastic processes and species interactions.

Density-dependent processes were found to explain the majority of variation in the relative abundance of core OTUs across biological replicates, followed by stochastic mechanisms (Supplementary Table 1). Only a small proportion of variance (3–8% across hosts) is explained by inter-specific interactions (Supplementary Table 1). It should, however, be noted that the contribution of inter-specific interactions may be larger because we are missing those interactions excluded from the cores (that is, interactions with more opportunistic OTUs).

Although inter-specific interactions contribute little to the dynamics of the core microbiomes, it is still important to investigate the nature and strength of these interactions as, for example, antagonism (that is, competition) and mutualism are known to differ in how they affect population and community...
Both empirical and theoretical studies in community ecology demonstrate that distributions skewed towards many weak and a few strong interactions enhance both population and community stability, and may arise during the assembly of persistent communities. Similarly, mutualism or skewed interactions only affecting one interacting partner (that is, amensalism and commensalism) have been shown to promote diversity and lead to community stability.

A number of indices were calculated for each core microbiome (Supplementary Table 2). Despite some variability in OTU number and linkages across different hosts, connectance (defined as the fraction of realized links among all possible links) was consistently low, ranging between 4.5 and 7.5%. We find that all cores are characterized by very few strong and many weak interactions (Supplementary Fig. 3). Moreover, cores are distinguished by a mixture of positive and negative interactions with amensalism and commensalism as a signature rather than competition and/or mutualism. Figure 9 illustrates this using the example of *Ircinia oros*; see Supplementary Figs 4–7 for further details and other sponge species. Across hosts, we observe that the most probable links are generally negative, although as the core size increases, the fraction of positive inter-specific interactions increases. Interestingly, we find that some OTUs, which are highly connected in the global bacteria-sponge (bipartite) network, are also highly connected within the core network. This suggests that OTUs that are present in a large number of different host species tend to be important for population dynamics within each particular host.

The low connectance, weak, and amensal and/or commensal interactions, together with strong density dependence found in most sponge species, suggest that symbiont communities in the phylum Porifera have stable cores. However, whether these stable cores play a role in the dynamics of remaining OTUs within individual microbiomes, and more importantly, whether this stability guarantees the homeostasis of host functionality requires further investigation.

**Sponge microbiomes are enriched in specific sequence clusters.** Many of the microbes inhabiting sponges have previously been found to fall into monophyletic clusters of ‘sponge-specific’ or ‘sponge- and coral-specific’ 16S rRNA gene sequences, with these clusters spanning 14 bacterial and archaeal phyla. The ecological and evolutionary significance of these monophyletic clusters remains unclear, yet it is noteworthy that this phenomenon has not been reported outside the phylum Porifera. Over 43% of all sponge-derived sequences from this global sponge analysis were assigned to previously defined monophyletic sponge-specific clusters. However, using deep sequencing and our extensive sampling, 2.7% of seawater sequences and 8.7% of sediment sequences were also assigned to these clusters, demonstrating some clusters are not strictly ‘sponge-specific’, but better described as ‘sponge-enriched’ (Supplementary Fig. 8). Importantly, these clusters contain generalists, specialists and opportunists (Fig. 7) indicating that...
the sponge-specific/enriched microbial sequence clusters have evolved multiple times, either early (that is, core) or late (that is, specialists and opportunists) in the assembly of symbiont communities.

Environmental and host factors are known to influence the composition of host-associated communities12,36,37; however, the impact of host evolutionary history on the structure and composition of symbiont communities has only recently been explored37. Considering the phylogenetic breadth of sponge species sampled here, we were able to evaluate the relationship between host phylogeny and microbial diversity. Diversity was assessed using the inverse Simpson’s index ($D$), while Blomberg’s $K$ was calculated using the phylosignal function in the R package picante38 (Fig. 10; see Methods for details). K values of 1

Host phylogeny and identity structure symbiont communities.

In this multi-gene phylogeny of host sponge species, 100% Bayesian posterior probabilities (PP) are indicated by black circles at internal nodes, while grey circles indicate 95–99% PP. Nodes with < 95% PP are not labelled. Black circles at the tips of the phylogeny are sized in proportion to the mean value of $D$ calculated for the symbiotic microbial community associated with each host species. Multiple clades of sponges contain either high (for example, Aplysina) or low (for example, Mycale) values of $D$. 

Figure 10 | Phylogenetic signal of the inverse Simpson’s index ($D$).
correlated with Bray–Curtis distance (Mantel test showed that host phylogeny was significantly among species, with the factor 'host species' accounting for chance.

findings indicate a significant signal of convergent evolution in expected if there was no phylogenetic signal. Combined, these at 0, AICc was calculated to further compare the similarity of covariances among species with the covariances expected, given a random process. The lambda value of 0.216 (AICc = 623.3; with \( z \) fixed at 0, AICc = 627.0) was significantly larger than what would be expected if there was no phylogenetic signal. Combined, these findings indicate a significant signal of convergent evolution in community structure, whereby sponges hosting more diverse communities are more phylogenetically related than expected by chance.

Beta-diversity analysis of symbiont communities (using Bray–Curtis distance) also indicated significant differences among species, with the factor 'host species' accounting for \( \approx 64\% \) of the observed variation among specimens. A partial Mantel test showed that host phylogeny was significantly correlated with Bray–Curtis distance (\( r = 0.442, R^2 = 0.195, P = 0.001 \)), as was host identity (\( r = 0.706, R^2 = 0.498, P = 0.001 \)). Testing for the effect of host phylogeny given host identity greatly reduced the explanatory power of host phylogeny (\( r = 0.223, R^2 = 0.050, P = 0.001 \)), although host phylogeny still had a significant effect.

Overall, the evolutionary history of the host plays a significant role in structuring the diversity of symbiont communities, but only a minor role in structuring community composition (that is, identity of microbial symbionts), where host identity (reflective of species-level forces) is the more important determinant. Thus, the evolutionary history of the host exerts a significant influence on microbial diversity despite strong selective forces for divergent microbiome composition, which might be critical for niche differentiation among closely related hosts.40

Methods

Sampling and sample processing. Samples were taken and processed according to the standard operating procedures to ensure maximum comparability. In brief, at least three specimens of each sponge species were subjected to standard protocols to remove planktonic or loosely associated microorganisms and cut into small pieces from which a random sub-sample of pieces was used for subsequent DNA extraction. Sediment samples were collected under water in close proximity to sponges. Sediments were scooped into sterile containers using sterile spatulas to wash, three times with sterile seawater to remove planktonic or loosely associated microorganisms and cut into small pieces from which a random sub-sample of pieces was used for subsequent DNA extraction. Sediment samples were processed according to the standard protocols (DNA extraction-protocol). Microbial communities in seawater were collected by passing 21 of seawater through 0.2-\( \mu \)m Sterivex filters and DNA was extracted from the filters as previously described.42 Samples were extracted at laboratories at the Australian Institute of Marine Sciences (Townsville, Australia), the University of California, Santa Barbara (USA) and the Technical University of Denmark (Kgs. Lyngby, Denmark) or the Nova Southeastern University (Dania Beach, FL, USA) to minimize shipment of frozen specimens. Aliquots of the specimens and DNA were kept at the three locations (and are available on request) and an aliquot of the extracted DNA was shipped to the University of Colorado, Boulder, CO, USA for sequencing of the 16S rRNA gene using standard procedures of the Earth Microbiome Project (http://www.earthmicrobiome.org/T, referred to as EarthMicrobiome/16s/).

Briefly, the V4 region of the 16S rRNA gene was amplified using the primer 515F–806Rb and sequenced using the HiSeq2500 platform (Illumina)44.

Analysis of sequencing data

We processed Illumina reads in mothur v.1.31.2 (ref. 45). First, quality-filtered, demultiplexed fastq sequences were trimmed according to quality (using the trim.seqs command: parameters qwindow/wavelength = 30, qwindow/window = 5, maxambig = 0, maxhomop = 8, mlength = 100). Files were reduced to non-identical sequence (uniq.seqs and count.seqs) to minimize computational effort. Non-redundant sequences were aligned (align.seqs: flip = 1) to a trimmed reference SILVA 102 (ref. 46) bacteria database (parg.seqs: start = 11894, end = 23519, keepdps = F), which was provided by mothur43. Only sequences that were aligned to the expected position were kept (classify.seqs: filter = 1968, end = 4, trim.seqs: threshold = 95). Aligned reads were reduced to non-redundant sequences (uniq.seqs). Chimeric sequences were detected using Uchime (chimera.uchime: dereplicate = t49, and filtered out (remove.seqs)). Pairwise distances between aligned sequences were calculated (dist.seqs: cutoff = 0.05) and used for clustering. Prior to clustering, aligned sequences were phylogenetically classified based on the relative neuronal distance. These sequences were clustered (cluster.split: split = f, count = c, taxonomy = t, splithit = classify, taxlevel = 4, cutoff = 0.03, hard = t, method = furthest) and converted to shared file format (make.sharedlist = c, label = 0.03). Finally, OTU representative sequences were retrieved based on the distance among the cluster sequences (get.oturep: list = c, label = 0.03, fasta = c, count = c) and were further classified based on SILVA. Greengenes (version gg_13_5_99 from May 2013), and RDPP taxonomies (classify.seqs: fasta = c, template = t, taxonomy = t, cutoff = 0.6,60,61). Furthermore, FastQ sequences from additional samples (n = 340) that were generated at a later time point (using the same sequencing procedure as described above) were pre-standradized using the same pipeline. These sequences were integrated into the shared file using QIIME 1.8.52, based on their similarity to the OTU representative sequences (parallel_pick_otus_uclust_ref.fy: --similarity 0.985 --optimal_uclust). Sequences that were similar to the OTU representative sequences were separately clustered with mothur and integrated into the previous files (shared and taximony files). The integrated OTU table (shared file) was filtered to remove low-abundance sequences (sequences < 0.001% across the whole data set) and chloroplasts (according to SILVA or Greengenes). In addition, counts from seawater-like OTUs (> 0.01% across all seawater samples) were removed from sponge samples. File manipulation and processing was carried out with python scripts (http://www.python.org). OTU tables are available in Supplementary Data 4.

Calculation of community metrics. Rarefaction curves were generated using the R package vegan 2.2-1 (ref. 53). Inter-sample rarefaction curves were generated by mothur (rarefaction.shared). Distances of the samples in a group (spoon species,
seawater or sediment) and their respective group centroids were calculated based on Bray–Curtis distances by the function betadisper from the vegan package in R (ref. 32). Richness indicators (Chao, Ace, Sobs) were also calculated with vegan. Nonmetric multidimensional scaling was calculated with vegan package based on Hellinger transformed OTU counts (45). Taxonomic profiles were obtained based on Greengenes, which provided more phylum-level assignments than the SILVA or RDP databases. Briefly, percentage OTU counts were averaged by species-environment with the package analogue 0.16-0 (ref. 35). Phylogenetic percentages were calculated by summing averaged OTU percentages. Bray–Curtis dissimilarities were calculated and heatmap was obtained using the package pheatmap 1.0.7 (ref. 56).

**Sponge-bacteria bipartite network analysis.** A bipartite interaction network was constructed using the presence of specific OTUs within each of the sponge species in the host. OTUs were considered part of the network only if they were found in at least 25 distinct samples from the whole data set. In this bipartite host–microbe interaction network, nodes represent sponge species (on one side) and OTUs (on the other); and links among them represent the presence of an OTU in the microbial community of the sponges to which it is linked. The network was constructed using a software script developed in R using the package igraph 0.7.1 (ref. 57) and interrogated using statistical tools to describe its properties. The degree distribution of sponges and OTUs was analysed to assess the heterogeneity of the network in terms of node connectivity. Degree distributions depict the statistical probability distribution of finding nodes with a certain degree (number of other nodes it is connected to). A variant of the degree distribution was employed: the cumulative degree distribution, which has the same probability distribution, but shows the probability of finding nodes with that degree or less. These probability distributions (one for the OTUs and another for the sponges) were obtained using the cumsum function in R (ref. 58). In addition, we fitted truncated power-law distribution to the degree distributions of the OTUs and sponges cumulatively degree distributions, respectively. This was achieved using the non-linear least squares (nls) function provided by R. This analysis reveals the pattern of connectivity between sponges and OTUs and facilitates determination of the balance between generalist and specialist species. The thresholds for specialization and generalism were chosen arbitrarily, but following basic requirements for this type of network analysis. First, neither of the groups contains the parameter that provides the characteristic scale at which the exponential cutoff occurs in the truncated power-law distribution. In our case, this value is 7.44 occurring in the truncated power-law distribution. In our case, this value is 7.44 in the parameter that provides the characteristic scale at which the exponential cutoff occurs in the truncated power-law distribution. In our case, this value is 7.44 (ref. 53). Richness indicators (Chao, Ace, Sobs) were also calculated with vegan.

The relationship between the fraction of samples within which a given OTU is found and a particular sponge species versus the total number of sponges in which that OTU was found (degree of the OTU in the network) was also assessed. This was achieved by obtaining the fraction of sponge samples in which a given OTU was found out of all the samples available for the sponge hosts to which that OTU is connected in the host–microbe network. This information was plotted against the degree of the OTU (interaction strength, and), and versa. Using all information in , we constructed a representative network for each host species as a mean of visualizing the most ‘credible’ network structure. This was done by mapping the posterior average number of links onto , and in doing so, extracting the links with the highest probability of non-zero interactions. This was done by custom-written R scripts. As a way of validating the structure of each representative network, we compared the connectance of each representative network with the posterior average connectance for each host species. The representative networks was plotted using the igraph package97 in R.

**Identification of sponge-specific and sponge/coral-specific clusters.** A representative sequence from each OTU was taxonomically assigned using a BLAST search against a combined ARB-SILVA database containing 178 previously identified sponge-specific clusters (SC) and 32 sponge/coral-specific clusters (SCC) (ref. 14). For each BLAST search, the 10 best hits were aligned to determine sequence similarities. The most similar OTU sequence to the reference sequence within the database was then assigned (or otherwise) to an SC or SCC based on the sequence identity and 75% similarity threshold (that is, a sequence read was only assigned to a cluster if it was more similar to the members of that cluster than sequences outside the cluster and its similarity to the most similar sequence within that cluster was above 75%). In cases where the assignment of the most similar sequences was inconsistent, a majority rule was applied, and the OTU within that cluster was above 75%). In cases where the assignment of the most similar sequences was inconsistent, a majority rule was applied, and the OTU was assigned to the cluster with the highest similarity to the most similar sequence within that cluster.

**Phylogenetic analysis of host species and correlation with symbiont communities.** Our phylogenetic analysis considered 61 sponge species for which at least one of three gene sequences (small subunit of nuclear ribosomal RNA [18S], large subunit of nuclear ribosomal RNA [28S] or mitochondrial cytochrome oxidase subunit 1 [cox1]) could be obtained from GenBank. For 39 of the 61 species (64%), sponge gene sequences were also obtained from the databases described before. The phylogeny was constructed with MrBayes version 3.2.1 (ref. 64), using the computational resources provided by CIPRES58. Within MrBayes, five partitions were used markers for sponge molecular systematics.
References


**Acknowledgements**

T.T. and N.S.W. were funded through Australian Research Council Future Fellowships FT140100197 and FT120100480, respectively. S.L.L. and P.M.E. were funded by the Spanish Government project MARSYMBIOMICS CTM2013-43287-P. R.C. was funded by the Portuguese Foundation for Science and Technology through the Investigator Grant IF/01076/2014 and the project UID/Multi/04326/2013. C.G.E. and R.W.T. were supported by grants from the US National Science Foundation (DEB-0829986 and DEB-1208340). A.C.-F. was supported by the UNESCO L’Oreal Fellowship for Young Women in Science. U.H. and L.M.S. received funding from the EU- FP7 Program (KBBE2012.3.2-01; grant no. 311932; SeaBioTech). J.M.M. was supported by the French Laboratory of Excellence Project ‘TULIP’ (ANR-10-LABX-41; ANR-11-IDEX-002-02) and by a Region Midi-Pyrénées Project (CNRS 121090). We thank Professor Russell Hill and members of his laboratory for providing samples and hosting a community data analysis workshop as part of the 2nd International Symposium for Sponge Microbiology. We also thank Emmanuelle Botte and Christine Gernert for DNA extractions and the Earth Microbiome Project for sequencing and sample preparation.

**Author contributions**


**Additional information**

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/


This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material.

To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/