











OPEN Emergent community architecture despite distinct diversity in the global whale shark (*Rhincodon typus*) epidermal microbiome

Michael P. Doane^{1,14}, Michael B. Reed^{2,14}, Jody McKerral¹, Laís Farias Oliveira Lima³, Megan Morris⁴, Asha Z. Goodman³, Shaili Johri⁵, Bhavya Papudeshi¹, Taylor Dillon³, Abigail C. Turnlund⁶, Meredith Peterson³, Maria Mora³, Rafael de la Parra Venegas⁷, Richard Pillans⁸, Christoph A. Rohner⁹, Simon J. Pierce⁹, Christine G. Legaspi¹³, Gonzalo Araujo^{10,11}, Deni Ramirez-Macias¹², Robert A. Edwards¹ & Elizabeth A. Dinsdale¹

Microbiomes confer beneficial physiological traits to their host, but microbial diversity is inherently variable, challenging the relationship between microbes and their contribution to host health. Here, we compare the diversity and architectural complexity of the epidermal microbiome from 74 individual whale sharks (*Rhincodon typus*) across five aggregations globally to determine if network properties may be more indicative of the microbiome-host relationship. On the premise that microbes are expected to exhibit biogeographic patterns globally and that distantly related microbial groups can perform similar functions, we hypothesized that microbiome co-occurrence patterns would occur independently of diversity trends and that keystone microbes would vary across locations. We found that whale shark aggregation was the most important factor in discriminating taxonomic diversity patterns. Further, microbiome network architecture was similar across all aggregations, with degree distributions matching Erdos–Renyi-type networks. The microbiome-derived networks, however, display modularity indicating a definitive microbiome structure on the epidermis of whale sharks. In addition, whale sharks hosted 35 high-quality metagenome assembled genomes (MAGs) of which 25 were present from all sample locations, termed the abundant ‘core’. Two main MAG groups formed, defined here as Ecogroup 1 and 2, based on the number of genes present in metabolic pathways, suggesting there are at least two important metabolic niches within the whale shark microbiome. Therefore, while variability in microbiome diversity is high, network structure and core taxa are inherent characteristics of the epidermal microbiome in whale sharks. We suggest the host-microbiome and microbe-microbe interactions that drive the self-assembly of the microbiome help support a functionally redundant abundant core and that network characteristics should be considered when linking microbiomes with host health.

¹Flinders University, Bedford Park, SA, Australia. ²North Carolina Agricultural and Technical State University, Greensboro, NC, USA. ³San Diego State University, San Diego, CA, USA. ⁴Lawrence Livermore National Laboratory, Livermore, CA, USA. ⁵Hopkins Marine Station, Department of Biology, Stanford University, Pacific Grove, CA, USA. ⁶Australian Centre for Ecogenomics, University of Queensland, St Lucia, QLD, Australia. ⁷Ch’ooj Ajauil AC, Cancún, Centro, Mexico. ⁸CSIRO, Brisbane, Australia. ⁹Marine Megafauna Foundation, West Palm Beach, FL, USA. ¹⁰Department of Biological and Environmental Sciences, Qatar University, Doha, Qatar. ¹¹Marine Research and Conservation Foundation, Lydeard St Lawrence, Somerset, UK. ¹²Tiburón Ballena Mexico de Conciencia Mexico, La Paz, Baja California Sur, Mexico. ¹³Independent Researcher, Quezon City, Philippines. ¹⁴These authors contributed equally: Michael P. Doane and Michael B. Reed. ✉email: michael.doane@flinders.edu.au; elizabeth.dinsdale@flinders.edu.au

For eukaryotic organisms, the diverse species that makeup their microbiomes are more than just passengers¹: they affect metabolic and immune processes² and confer physiological functions beyond the host's innate capabilities, promoting health^{3,4}. These services result from a multitude of interactions between microbes and host cells that become established over time⁵. Yet, microbiome species composition and diversity are highly variable in space and time and across individuals from similar species⁶, often with no apparent consequence to the host. These observations suggest that the function or the services rendered by the microbiome are not attributed simply to the presence or abundance of individual species⁷. The self-organization, or interaction network of microbiome communities, is one explanation for how consistent services that support healthy microbiomes are maintained⁸, however, whether host microbiomes exhibit emergent structure remains an outstanding question.

Self-organization of the microbiome refers to the collective behaviour of the microbial members, captured as population patterns of microbial groups relative to each other, described as co-occurrence: a pattern described as microbiome architecture. Architectural properties (i.e., number of nodes and associated edges or number of clusters formed by interacting microbes) of the microbiome arise in response to ecological and evolutionary processes⁹, which influence the host-microbe and microbe-microbe interactions⁷. Together, these eco-evolutionary processes drive population dynamics of these microbes which result in the microbiome architectural properties. Therefore, the rearrangement of microbial populations can subsequently alter emergent ecosystem functions¹⁰. As networks are inherently hierarchical¹¹, quantifying the different attributes of these levels, such as network complexity, modularity, and individual microbial group interactions, can reveal important ecological insight driving community structure¹². For instance, fluctuation in microbiome architectural complexity in response to different abiotic stress results in changes to emergent community properties⁸, while microbial organisms that form sub-networks suggest similar environmental preferences^{13,14}. Microbes with many connections to other microbial groups are predicted to be important in niche formation and considered keystone organisms¹⁵. For these reasons, determining the microbiome's network architecture is fundamental to understanding microbiome structure and function.

In this study, we examine epidermal microbiomes in whale sharks (*Rhincodon typus*) to test our hypothesis that the host-microbiome relationship relies on self-organization of microbiome members for required functions, rather than being dependent on specific taxonomic groups. Whale sharks are an ideal non-model host system to address this fundamental microbiome assembly hypothesis for the following reasons: (1) They are part of the ancient extant vertebrate lineage of Chondrichthyan fishes, which places them at a pivotal evolutionary point for understanding the vertebrate host-associated microbiome relationship; (2) they are one of a few animals species which have a global distribution, occupying tropical/subtropical waters between latitudes 30° N and 35° S¹⁶; and (3) these animals form predictable seasonal aggregations at specific sites taking advantage of plankton blooms. Using random shotgun metagenomics, we quantified the microbiome architecture from the epidermis of 74 individual whale sharks distributed across five aggregations from around the world, spanning every major ocean basin. Our results indicate that microbiome diversity and composition correspond with aggregation; however, microbiome architecture is a fundamental feature of the epidermal microbiome across the globe. In addition, abundant core microbiome members, identified as Metagenome Assembled Genomes (MAGs) discriminated into two main groups, termed Ecogroup 1 and 2 based on potential gene functions, revealing at least two distinct ecological niches within the epidermal microbiome. We predict these distinct functions support important metabolic processes that may help to stabilise the microbiome architecture of whale shark.

Results

Epidermal microbiomes were collected from the dorso-lateral skin surface, in line with the first dorsal fin of 74 whale sharks at five aggregations distributed globally (Fig. 1; SI Table 1). The mean library size ranged from 335,820 ($\pm 46,523$) reads at Cancun to 701,829 ($\pm 70,068$) at Ningaloo (SI Fig. 1a; SI Table 1). The proportion of reads with taxonomic assignments was lowest at Ningaloo (taxa: 16.5% \pm 1.5%) and highest in the Philippines (54.4 \pm 1.5%) (SI Fig. 1b).

Taxonomic diversity structure of whale shark epidermal microbiomes. The 20 most relative abundant microbial families accounted for 66% to 82% of total proportional abundance (Fig. 2a). Proportionally abundant families included Alteromonadaceae (max: Cancun mean of 18.7 \pm 2.9%—min: Philippines 7.8 \pm 1.7%), Flavobacteriaceae (max: Cancun 16.1 \pm 2.2%—min: Ningaloo 5.0 \pm 0.9%); Pseudoalteromonadaceae (max: Ningaloo 11.5 \pm 2.8%—min: Cancun 4.5 \pm 1.2%) and Pseudomonadaceae (max: Cancun 9.8 \pm 1.9%—min: Ningaloo 4.5 \pm 0.8%). A few families showed high relative abundance on sharks at one aggregation but low abundance elsewhere; for example, Sphingomonadaceae had a mean of 10.5 (\pm 1.8) % in the Philippines, and Pelagibacteraceae a mean of 14.6 (\pm 1.9) % at Ningaloo (Fig. 2a).

Family level effective diversity was significantly different across aggregations (Kruskal–Wallis $\chi^2_{df=4} = 30.02$; $p < 0.01$; Fig. 2b; SI Table 2) with Cancun having the least diverse microbiomes (Effective diversity; e^{H1} : 20.4 \pm 1.21) and Ningaloo the most diverse (47.9 \pm 4.25). Richness of microbial families was also significantly different across aggregations (Kruskal–Wallis $\chi^2_{df=4} = 17.1$; $p < 0.01$; min: La Paz 338 \pm 17.6 families – max: Tanzania 384 \pm 4.6 families, SI Fig. 2a; SI Table 3) as was evenness (Kruskal–Wallis $\chi^2_{df=4} = 30.3$; $p < 0.01$; Pielou's evenness; min: 0.51 \pm 0.01 at Cancun—max: 0.6 \pm 0.02 at Ningaloo SI Fig. 2b; SI Table 4). Therefore, both the number of microbial families and the relative abundance of each microbial family affected the diversity patterns of microbiomes across aggregations.

Compositional patterns of whale shark epidermal microbiomes also varied. We first tested the whale shark microbiomes to water samples collected from each location. Whale shark epidermal microbiomes did vary from the water column (PERMANOVA: Pseudo-F $_{df=1,79} = 2.43$, $p < 0.05$; $R^2 = 0.03$, SI Fig. 3). We then analyzed several factors to determine the best predictor of microbiome compositional difference across sharks. Aggregation was

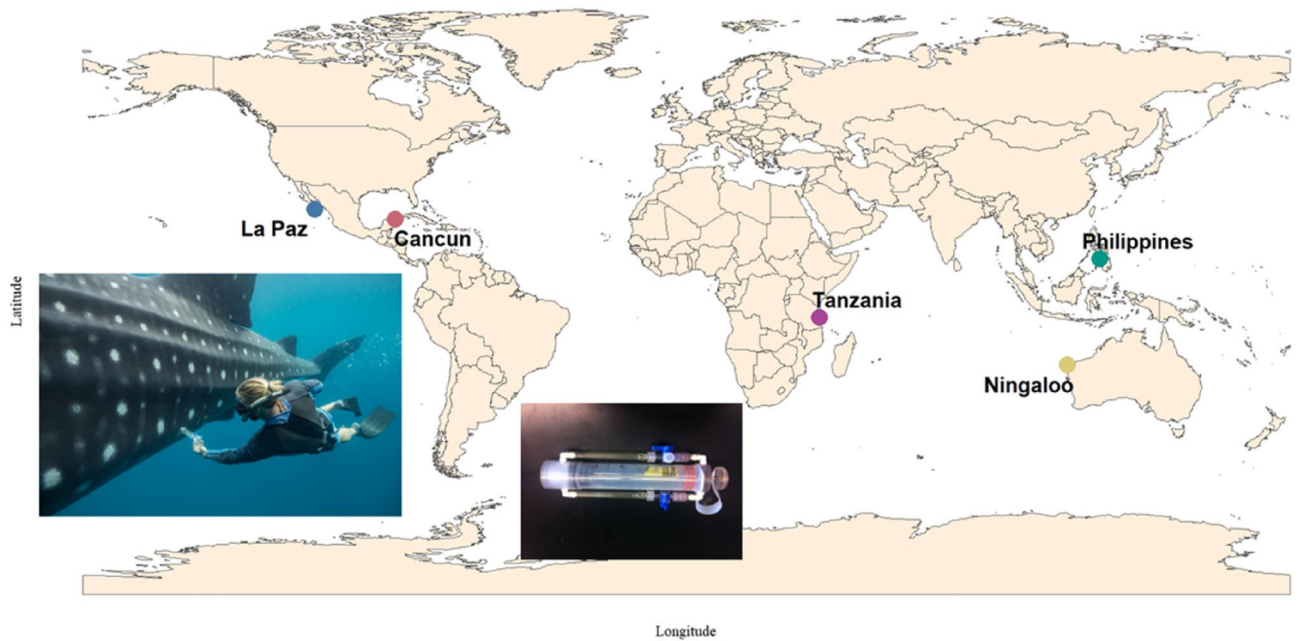


Figure 1. The location and the sampling methods used to analyse the skin microbiome of whale sharks. The sample locations are from aggregation globally distributed and denoted by colored dots. Inserts demonstrate how sampling was performed and the two-way syringe device for flushing sterile seawater over the skin surface of the whale sharks, isolating the epidermal microbiome from the surrounding water column.

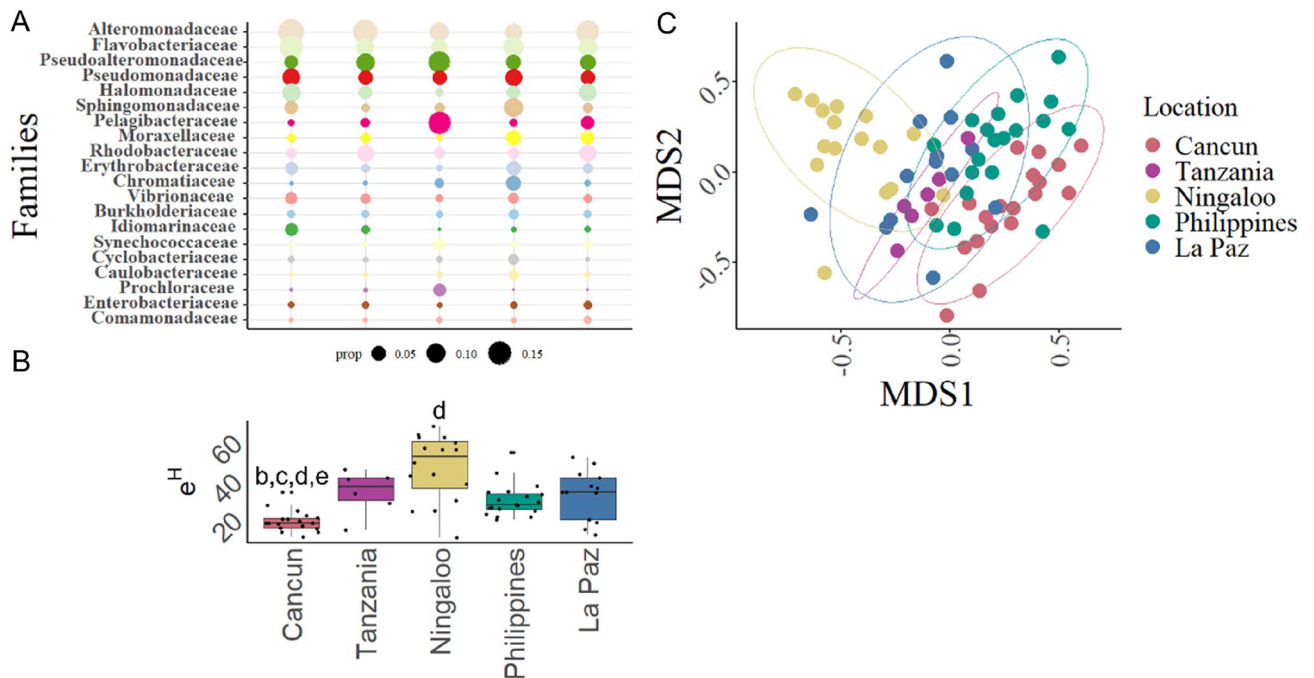


Figure 2. Microbiome taxonomic diversity and compositional patterns across whale sharks at five locations. (A) Mean distribution of the 20 most abundance taxa at the family level found across all locations. The size of the dot corresponds to the relative proportion each family. Dot size proportion is indicated below the figure. (B) Taxonomic alpha diversity measured as effective number of microbial families. Dots over boxes demonstrate the distribution of diversity values for each location. Letters correspond to the location which comparison is significantly different from at $p < 0.05$. Figure A and B share location labels. (C) Whale shark taxonomic microbiome compositional patterns based on Bray–Curtis dissimilarity. Each point represents a metagenomic sample, and the ellipse notes the 95% confidence for sample spread at each location in multivariate space.

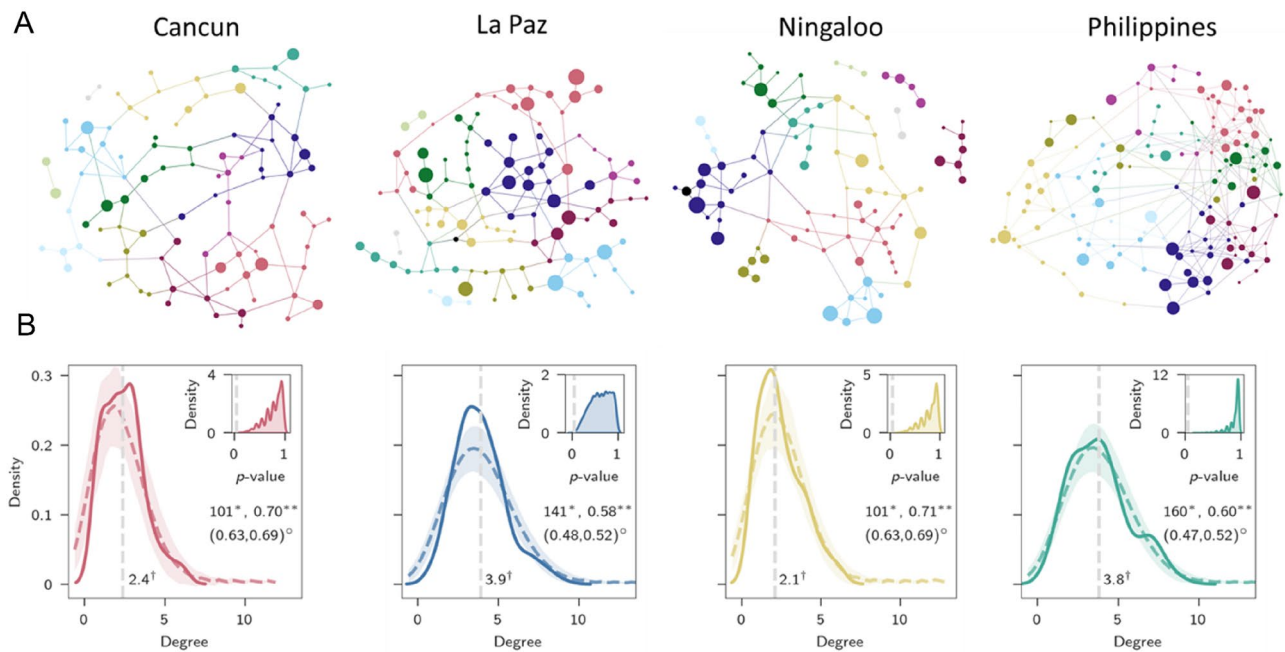


Figure 3. Network structure of microbial taxonomy from whale sharks across the four locations, including Cancun, La Paz, Ningaloo, and Philippines microbiomes. **(A)** Networks were calculated with the SpiecEasi algorithm. Network nodes have been oriented such that nodes with large degrees (more edges) are centralized while nodes with lower degree are peripheral. Color of nodes represents the assigned cluster. **(B)** Degree distribution for microbiome-derived networks compared against modelled degree distribution patterns. Degree distributions for the microbiome-derived networks are signified as a bold line in bi-plots, while the shaded region represents the estimated densities of 5000 generated Erdos–Renyi networks. The dotted vertical line (†) demonstrates the mean degree (number of edges) of all nodes in the microbiome-networks. Inserts within each bi-plot indicate p -value distributions of 5000 tests. The dotted line within the p -value distribution represents 0.05 cut-off. Symbols in figure include: † mean degree value of whale shark derived network; * the number of nodes in whale shark derived network; ** modularity for whale shark derived networks; ° lower and upper modularity values for modelled networks.

the strongest predictor of microbiome variation (PERMANOVA: Pseudo-F_{df=4,69} = 4.09, $p < 0.001$; $R^2 = 0.19$; Fig. 2c). Ocean basin from which the whale shark aggregation occurred was also significant, but with less explanatory power (PERMANOVA: Pseudo-F_{2,71} = 4.52, $p < 0.001$; $R^2 = 0.11$). We additionally tested population structure of whale sharks by grouping locations together on well established whale shark population estimations¹⁶ and found significant groupings, but with much less explained variation (PERMANOVA: Pseudo-F_{1,72} = 6.55, $p < 0.001$; $R^2 = 0.08$), relative to the aggregation factor. A pairwise PERMANOVA was performed on aggregation and indicated the taxonomic composition of microbiomes from Cancun whale sharks were different to all other aggregations ($p < 0.01$). La Paz whale shark microbiomes were different to the Philippines and Tanzania ($p = 0.04$), but interestingly, there was no difference between Tanzania, Ningaloo, and Philippine whale shark microbiomes, which are in the Indian ocean region.

Taxonomic network structure of whale shark epidermal microbiome communities. To determine emergent microbiome patterns, network architecture was compared across four of the five aggregations (Tanzania was excluded due to low sample size) (Fig. 3a, b). The number of taxonomic families (nodes) in each network ranged from 101 (Cancun and Ningaloo) and 160 (Philippines) with mean node degrees of 2.1, 2.4, 3.8, and 3.9 (mean number of co-occurrences with other microbial families; Ningaloo, Cancun, Philippines and La Paz, respectively; Fig. 3a). Degree distribution curves of the microbiomes at the four locations were consistent (Fig. 3b). As there is no agreed upon way for comparing networks distances¹⁷ and our objective was to compare the statistical properties of each network to one another, we made null model comparisons. Each microbiome-created network was compared to 5000 bootstrapped $G(m, n)$ random networks generated based on the number of nodes and edges for each location (Two-sample Kolmogorov Smirnov: $p > 0.05$ in 19,997 out of 20,000 total bootstraps; Fig. 3b). These microbiome networks from all aggregations shared similar global network characteristic (i.e., the number of co-varying microbial families), each consistent with that of an Erdos–Renyi network. In other words, each microbiome network was consistent with the null model comparison. Interestingly, sub-network structure of each microbiome-created network displayed modularity, an attribute that random networks do not possess. Modularity scores of the microbiome-created networks (Cancun: 0.70; La Paz: 0.58; Ningaloo: 0.71; Philippines: 0.60) were higher than the upper 95% confidence threshold for modularity scores of the randomised networks (upper CI; Cancun: 0.69; La Paz: 0.52; Ningaloo: 0.69; Philippines: 0.52), indicating a signifi-

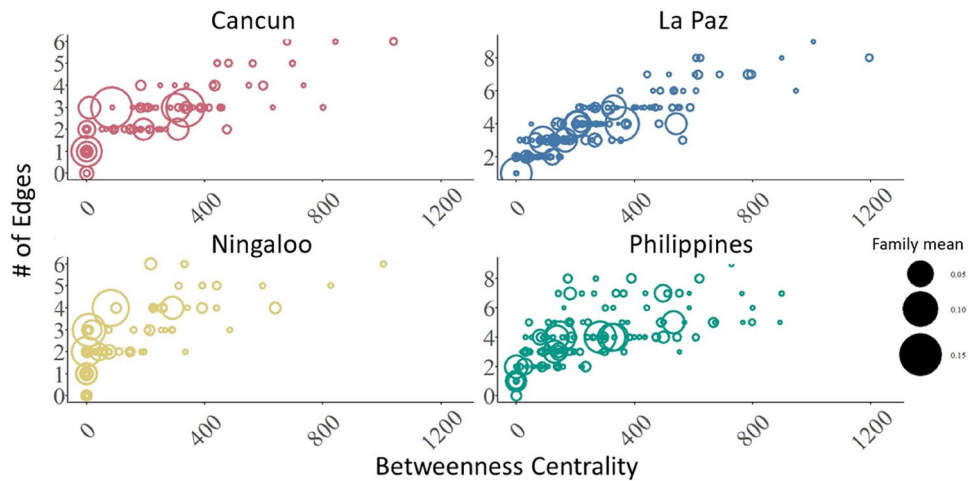


Figure 4. The distribution of betweenness centrality scores for each microbial family plotted against its number of edges for each location. The size of the symbols corresponds to the mean relative abundance of each microbial family.

cant difference, therefore demonstrating underlying community structure in the microbiome network topology. Modularity indicates some microbial families have stronger co-occurrence with a set of microbial families relative to the rest of the microbes, therefore forming clusters within the network. Each network had ~ 11 clusters of microbial families (Cancun—12; La Paz—11; Ningaloo—12; Philippines—10). However, family membership within a cluster was not consistent across locations. We also determined whether the importance of microbial families, measured as betweenness centrality (those with greater number of edges) corresponded with the relative abundance of the microbial family (Fig. 4). Interestingly, increases in the number of edges did not correspond with an increase in relative abundance of reads in each microbial family; a consistent pattern at each location. Therefore, microbes that are of low relative abundance in the microbiome play a disproportionately large role in microbiome architecture, suggesting these are keystone microbiome species.

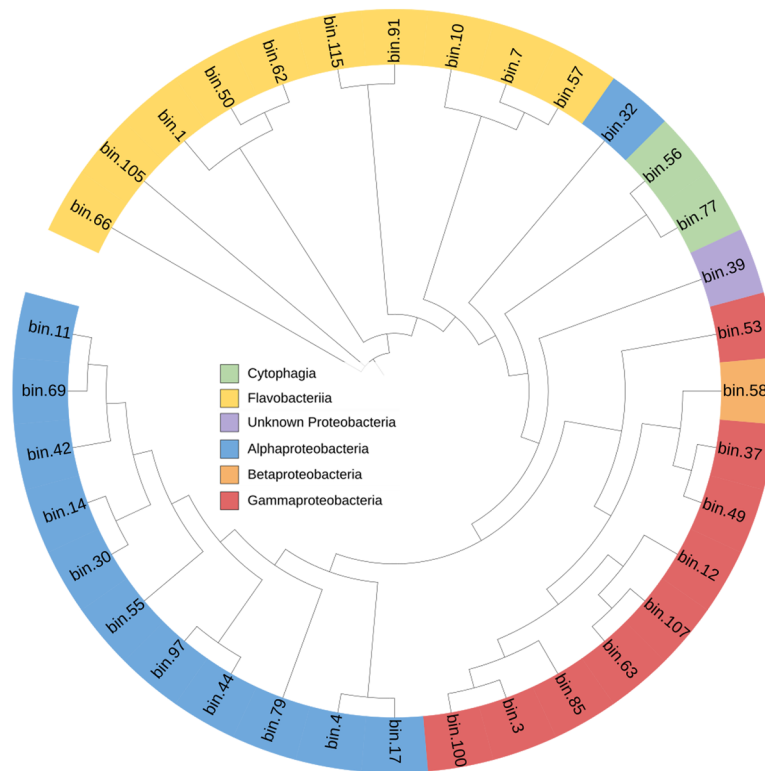
Metagenome assembled genomes (MAGs) indicate a core group of whale shark epidermal microbes that occur across the world.

To further assess the variation in important taxa across host microbiomes, we constructed Metagenome Assembled Genomes (MAGs). A total of 118 MAGs were generated, 35 of which were high quality with completeness of $\geq 70\%$ and contamination $\leq 5\%$ (Fig. 5a, b)¹⁸. There were 268 contigs $> 50,000$ bp and the longest contig was 430,630 bp. The contribution of reads to the MAGs was high across all aggregations (Philippines—64.2%; Cancun—54.4%; Tanzania—38.7%; La Paz—28.5%; Ningaloo—11.1%). MAGs were annotated to five microbial classes including Alpha-, Gamma-, and Beta-proteobacteria; Cytophagia; and Flavobacteria (Fig. 5a). MAG-bin 39 was only annotated to the Proteobacteria phyla while only a single MAG (Bin 107—*Pseudomonas stutzeri*) could be annotated to the species level, suggesting microbial members in the whale shark microbiome are quite novel. There were several MAGs in which all aggregations contributed reads equally (e.g., Bin 91- Flavobacteriaceae, Bin 107—Pseudomonadaceae, Bin 42—Erythrobacteraceae) and others where a single location over contributed (e.g., Bin 62; Cancun—Flavobacteriaceae, Bin 100; Cancun—Pseudoalteromonadaceae, Bin 3; Philippines—Chromatiaceae). Of the 35 MAGs, 25 were found across all aggregations, while eight were missing from Ningaloo and two each from Philippines and Cancun. All 35 MAGs were represented in La Paz and Tanzania whale shark microbiomes. The family identity of the MAGs corresponds with several of the most relative abundant families identified through short read annotations, including, Flavobacteriaceae, Pseudoalteromonadaceae, and Pseudomonadaceae. Alteromonadaceae, the most abundant short read annotated family, was not represented in the high-quality MAGs suggesting this family may have high diversity at the species level, with species in relatively low abundance.

Functional metabolic characterization of the MAGs.

Clustering the gene abundance within functional subsystems defined five primary groups of MAGs (Fig. 6). Ecogroup 1 was a Bacteroidetes cluster with MAGs identified to Flavobacteria and Cytophagia classes and mostly present on whale sharks at all locations (6 of 9 MAGs). Bin 7 was however not found at Ningaloo, and bins 1 and 57 were not found at Cancun. The overrepresented gene functions in the Bacteroidetes group included nitrogen, iron acquisition, amino acid and motility and chemotaxis functions, suggesting this group's role in trace nutrient metabolisms. Ecogroup 2 was more diverse with MAGs being identified as Alpha-, Beta-, Gammaproteobacteria, and Flavobacteriia. Again, the majority of MAGs in this cluster were distributed across the four locations (16 of 23 MAGs). Gene functions overrepresented in this group included metabolism of aromatic compounds; type I, II, and III secretion systems; and sphingolipids. Carbon metabolism, including mono-, di- and oligo-saccharides utilization were also overrepresented, suggesting this groups role in carbon metabolisms. There were three MAGs that did not cluster, indicating each had unique gene functions. MAGs of these groups were annotated as Pelagibacteraceae

A



B

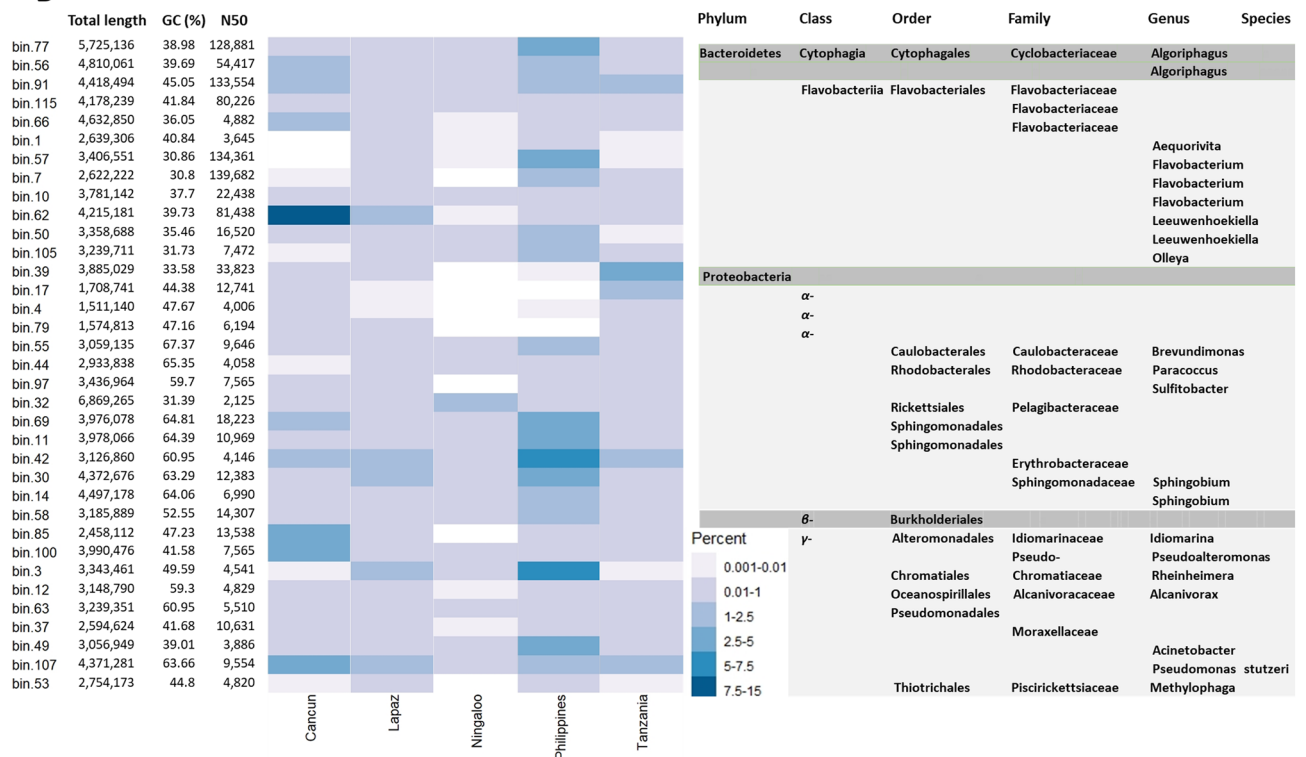


Figure 5. The 35 most complete metagenome assembled genomes (MAGs). **(A)** MAG phylogenetic placement relative to other MAGs with leaf colour indicating Class level taxonomic assignment. **(B)** Table and heatmap include MAG statistics, with the colors corresponding to the mean relative percent of reads from metagenomes at each location contributing to the MAGs and the hierarchical taxonomic identify of each MAG. Shading of the taxonomic identify of each MAG is based on microbial classes. Heatmap was generated using pheatmap v. 1.0.12 (<https://cran.r-project.org/web/packages/pheatmap/index.html>) in the R package.

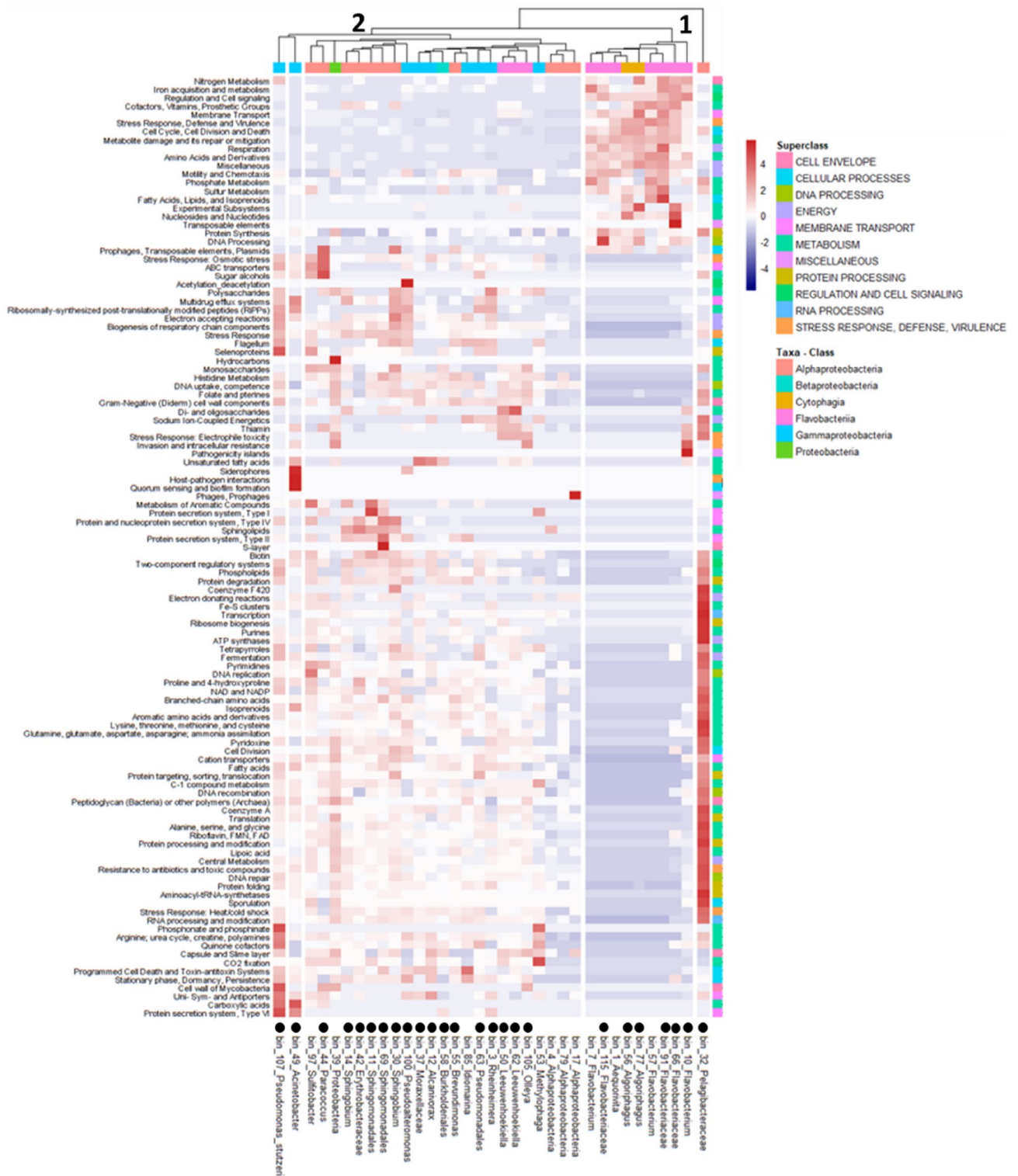


Figure 6. Heatmap of the 35 high quality MAGs and the relative abundance of genes associated with specific metabolic functions. Red indicates a higher abundance of genes whereas blue a lower representation within each functional Subsystem obtained from the annotation of the MAGs in the PATRIC database. Colors along the top x-axis corresponds to the taxonomic class each MAG was assigned and colors along the y-axis correspond to the Superclass each Subsystem belongs. Black dots along the bottom x-axis represent those MAGs found in all locations. Heatmap was generated using pheatmap v. 1.0.12 (<https://cran.r-project.org/web/packages/pheatmap/index.html>) in the R package.

(Bin 32), *Acinetobacter* genera (Bin 49) and *Pseudomonas stutzeri* (Bin 107); each bin was present at all locations. Pelagibacterceae had an overrepresentation of housekeeping genes such as a range of amino acid metabolisms, carbon transport, and NAD to NADP pathways and interestingly, lacked photorhodopsin genes. *Acinetobacter* had an overrepresentation of genes associated with host–pathogen interactions, quorum sensing, siderophores, multidrug resistance efflux pumps, carbonic acids, and protein secretion system type VI. *Pseudomonas stutzeri* had an overrepresentation of genes involved in polysaccharide metabolism, stress response, multidrug resistance efflux pumps, toxin systems, phosphonate, selenoproteins, transporter genes, and flagella for motility.

Discussion

The microbiome is increasingly viewed as more than the sum of its parts, and here we corroborate this view that microbiomes from discrete groups of whale sharks from around the world have consistent microbiome architectural characteristics. We confirm that the pattern of emergent architecture arises independently of microbiome diversity and composition. This suggests a fundamental relationship between the host and its microbiome that modulate microbial architecture rather than diversity alone. Disruption to these architectural properties thus, may provide an early indicator of dysbiosis or unhealthy microbiomes. Further, two microbial ecogroups were identified across all locations and defined as the abundant core despite some MAGs only being present at some locations. Therefore, potential gene functions from the two ecogroups may help stabilise or facilitate consistent network architecture by mineralizing highly abundant substrates in the whale shark microbiomes.

Similar network architectures within the microbiome across whale shark aggregations indicates self-organisation within these communities, and therefore similar ecological niches^{19, 20}. Self-organization is not dependent upon the taxonomic identities of important community members (those with high number of predicted interactions within the network), consistent with results found in bioreactor experiments that maintained carbon cycling activities¹². Therefore, the whale shark microbiomes are likely performing similar functions at each location, suggesting an association with the host. Further, network architecture is inherently hierarchical¹¹, therefore characterising attributes of network levels, including at the whole network, sub-network, and node level reveal ecological change in microbiome functioning. For instance, the network complexity of soil microbial communities decreased with increased cropping intensity²¹ and rhizosphere networks became more organized and complex as the plants matured²². Therefore, the consistency of whale shark microbiome networks, regardless of location, suggest that similar ecological processes are shaping the epidermal microbiomes of the world's largest shark; perhaps a phenomenon driven by a host feature such as the filtering effects of the mucus secreted from the epidermal surface and/or of the dermal denticles²³.

Complex microbiomes form cross-feeding networks which are expected to mediate sub-network structure, for instance modularity²⁴. Whale sharks of the different aggregations consistently had modularity, or subnetwork structure of the epidermal microbiome despite the distinct geographical locations. Modularity suggests niche differentiation that results from strong selection, such as host filtering²⁵, or microbe-microbe interactions in the form of cooperation and/or competition¹². In addition, networks showed that important microbial families, defined as those with many interactions, are in low abundance. In bioreactor experiments, when the microbial communities stabilised, the community became highly connected and keystone (highly connected) microbes were in low abundance⁸. While the role of the keystones was not explicitly examined, we do suggest this group has a disproportionately large impact on total community structure by performing specialized metabolic processes²⁶. Experimental evidence from a multiplicative community culture demonstrated that keystone microbes had specialized metabolisms, including cellulose and chitin degradation, each metabolically expensive to utilize as an energy source²⁷. Therefore, microbes occupying the keystone position are crucial to network structure and further research will reveal the role of these microbes in the whale shark epidermal microbiome.

The abundant core was identified by assembling the single reads into Metagenome Assembled Genomes (MAGs). The number of MAGs characterized in our study (35 high quality) was similar to two other studies focused on shark microbiomes, one which found 54 high/medium quality MAGs from the skin surface of leopard sharks (*Triakis semifasciata*)²⁸ and another that found 27 MAGs from fecal samples of two shark species²⁹. These abundant core microbes may be supporting microbiome architectural patterns by providing metabolic processes. For instance, abundant microbial groups cultured from natural rainwater-filled tree holes drove bulk processes including respiration, metabolic potential, and cell yield of the total community²⁷. The abundant core of the whale shark microbiome was further partitioned into two ecogroups based on the presence of potential gene functional subsystems. MAGs within the two ecogroups exhibit variation in their abundance across the sample locations, supporting results observed in the networks that important microbial groups need not be the same taxonomic groups and that functional redundancy is occurring. Ecogroup 1 was comprised of nine MAGs identified as Flavobacteriia and Cytophagia microbes. Interestingly, microbes of these taxonomic groups are commonly associated with fish mucus microbiomes³⁰ and may indicate a similar role in the epidermal microbiome. Flavobacteraceae MAGs were found from the epidermal microbiome of the leopard shark (*T. semifasciata*)²⁸, suggesting this microbial family may be a symbiont of marine fishes (bony and cartilaginous). Gene pathways in Ecogroup 1 support this groups role as skin microbiome symbionts that utilize mucus. Mucus, which is secreted onto the skin surface in fishes³¹ is composed of brush-like fibers formed from glycoprotein backbones which are covered in O-linked amino acids, which are rich in essential nutrients necessary for microbial activity, including sulfur, nitrogen and phosphorous³². Ecogroup 1 MAGs were discriminated from Ecogroup 2 by the presence of gene pathways to mineralize the abundant amino acids found in this environment. Experimentally amending seawater with fish mucus was shown to trigger rapid microbial mineralization, as evidenced by a rapid increase in ammonium³³, demonstrating the microbial response to the components of mucus. Ecogroup 1 also had an overrepresentation of genes linked with motility and chemotaxis, suggesting this group is important

in the physical establishment of the microbial community, as seen in plant-root systems³⁴ and coral mucus³⁵. Chemotaxis may also be used by microbes to overcome inhibitory effects of the physical patterns produced by elasmobranch dermal denticles. In modelled systems, the surface topography of shark epidermis reduced biofilm formation of medically relevant microbial groups³⁶. In addition, two MAGs identified as Algoriphagus (bin 77 and 56) produce lipids that inhibit development of choanoflagellates³⁷, signalling a potential role in inhibiting or competing with micro-eukaryotes and other microbial groups for space. We predict the Flavobacteriia and Cytophagia are crucial microbiome members as they were found across all locations.

The Ecogroup 2 microbes were more diverse having 25 MAGs identified as Alpha-, Beta-, and Gammaproteobacteria; and Flavobacteriia. Gene subsystem present in this ecogroup were also highly diverse, and indicative of microbes driving microbial-host and/or microbe-microbe interactions along with making available shorter sugar oligomers for cross-feeding. For instance, overrepresentation of gene classes associated with metabolism of aromatic compounds, type I, II, and III secretion systems³⁸ and sphingolipids³⁹ suggest microbes are interacting with the surrounding environment and utilizing host substrates. Interestingly, sphingolipids have been suggested to interact with receptor on the skin of rainbow trout (*Oncorhynchus mykiss*) to modulate mucosal homeostasis⁴⁰, possibly suggesting the Alphaproteobacteria from which the sphingolipids genes were identified may have a similar role in whale shark skin, helping to modulate skin mucus. Sphingolipids were also identified in the epidermal microbiomes of leopard sharks (*Triakis semifasciata*) across 3 years⁴¹, thus sphingolipids may be an important member of the epidermal microbiome of elasmobranchs. In addition, the presence of gene subclasses representing carbon utilization and copolymer compounds, including poly- di- and oligo-saccharide metabolisms, suggests mucolytic capacity for energy and release of other limiting nutrients within mucus. MAGs constructed from the epidermal surface of two stingray species (*Myliobatis californica* and *Urolophus halleri*) contained pathways for the breakdown of long chain carbons (i.e. polysaccharides)⁴², suggesting this is a common pathway among elasmobranch epidermal microbiome members.

The three other MAGs that did not cluster with Ecogroup 1 or 2, each possessed highly diverse gene subsystems. The most notable is the Pelagibacteraceae MAG (Bin 32), which was found at each location and is a common pelagic microbial groups⁴³, therefore possibly being contamination from the surrounding water column. However, Pelagibacteraceae is in the epidermal microbiome of blacktip reef sharks (*Carcharhinus melanopterus*)⁴⁴ and coral mucus microbiome⁴⁵, and therefore may be an inhabitant of the whale shark microbiomes.

Diversity and compositional structure provides ecological insight of whale sharks as the host epidermal microbiome was best predicted by the whale shark aggregation, possibility revealing an environmental signature. The epidermal surface of whale sharks remains in contact with the water column, which itself hosts a distinct microbial community²³ and serves as an environmental reservoir, or regional pool of potential microbial colonizers. The whale shark microbial communities were different than the water column microbial communities, similar to other shark microbiomes²³. However, given their movement behaviour, whale sharks regularly expose their skin surface to environmental extremes. For instance, in the Gulf of Mexico, whale sharks move as far as 52.3 km/day⁴⁶ and dive to great depths, exposing the microbes to temperatures ranging from 30 °C at the surface to 4 °C at depth^{47,48}. Microbiome diversity patterns in whale sharks may also be shaped by inherent aggregation effects including diet, as is seen in other fish species⁴⁹. Whale sharks in the Philippines⁵⁰ and Tanzania⁵¹ primarily eat sergestid shrimp, while La Paz whale sharks feed on copepod blooms⁵², and the aggregation near Cancun ingests fish eggs⁵³. Genetic structure was also a moderate predictor of microbiome structure. Whale sharks have two inferred populations based on genetics⁵⁴: an Atlantic and Indo-Pacific group. Drivers of the microbiome at the population level may be due to inherent host-metabolic difference, or factors that co-vary with whale shark populations, such as global regions utilized by the sharks, as discussed above. Studies in other marine species show microbiome compositions that correspond to host population structure such as in some sponges⁵⁵, and to variation in haplotype structure within populations, such as the phytoplankton *Thalassiosira rotula*⁵⁶. Whale sharks have diverse haplotype structures⁵⁷ which may be an important predictor of microbiome structure observed across aggregations however we did not test for this relationship here.

In summary, we have identified the fundamental architecture of the epidermal microbiome on whale sharks from across the world's oceans. Our results provide evidence for an inherent assembly process, that microbiome diversity alone has not uncovered. By characterizing fundamental microbiome patterns from the whale shark, a member of the oldest extant group of vertebrates, we reveal an emergent microbiome structure which may be ubiquitous across vertebrate organisms, including human host. Thus, microbiome architecture provides a benchmark from which to interrogate healthy microbiome structure, more likely revealing underlying microbiome-host relationships. Diversity structure of the microbiome on the other hand is more informative for identifying ecological aspects of adaptations in response to plastic attributes of the host, such as environment or diet. In addition, two ecogroups emerged that suggest two ecological niches occur in the abundant core microbiome, and these groups may provide energy to other microbial groups such as keystones that support the microbiome architecture. The MAGs were identified as microbial groups commonly found in the skin microbiome of teleost fishes, however only one could be identified to the species level, which suggests novel microbial species are occupying ecologically important abundant core niche within the whale shark microbiome.

Methods

Sample locations and collection methods. Microbiomes were surveyed from the epidermis of 74 individual whale sharks from several locations, globally (Fig. 1). Five aggregations were sampled across 2017–2018, that included 14 whale sharks from La Paz (24°18'31.5" N; 110°37'42.5" W) in February 2017; 19 from Cancun, Mexico (21°23'49.59" N; 86°37'24.16" W) in July 2017; six from Mafia Island, Tanzania (7°52'56.67" S 39°39'22.35" E) in November 2017; 16 from Ningaloo Reef, Australia (22° 04'32.78" S; 113°39'06.26" E) in June 2018; and 19 from Oslob, Philippines (9°30'32.3 N; 123°25'01.8) in July 2018. The low number of samples from

Tanzania was due to the sharks aggregating later than predicted. Microbiomes were taken from the epidermal surface along the dorso-lateral surface in line with the first dorsal fin (Fig. 1). Samples were collected using a two-way syringe device that circulates filtered seawater (0.02 µm filter) over the skin surface before being drawn into the backside of the syringe^{58, 59}. This process enables the sampling of microbes from submerged whale sharks, whilst minimizing seawater microbiome contamination. From each shark we took four syringes, resulting in approximately 180 ml of sample water that was then passed through a 0.22 µm Sterivex filter (Millipore, USA), trapping all microbial life on the filter. Sterivex filters were sealed with parafilm and stored on ice until long term storage at - 20 °C. Epidermal microbiomes were the focus of the research, because they are minimally invasive, and one specific area of the shark was examined enabling the sharks to continue feeding/swimming with minimal interruption. Animal handling and ethics were reviewed at San Diego State University through IACUC under permit APF #14-05-011D, APF #17-11-010D, APF # 18-05-007D and all methods were conducted in accordance under IACUC permits.

DNA extraction, metagenomic construction and bioinformatic processing. Genomic material was extracted directly from Sterivex filters using a modified spin column purification protocol from Nucleospin Tissue kit (Macherey–Nagel, Allentown, PA, USA). Modification to extraction procedure included first incubating sealed Sterivex filters with 720 µl of T1 buffer and 90 µl of Proteinase K (2.5 mg/mL) at 55 °C with rotation overnight. Subsequent extraction followed manufacturer protocol. DNA was prepared for sequencing using the Accel-NGS 2s Plus DNA kit (Swift Biosciences, Ann Arbor, MI, USA) for paired-end sequencing with the Illumina MiSeq v3 600 cycles (San Diego, CA, USA). Samples were barcoded and the whale shark microbiomes were mixed in with a range of microbiome samples (e.g., kelp, fish, rays, and seagrass microbiome samples) and run on several sequencing runs by the undergraduates in San Diego State University ecological metagenomics class.⁶⁰

The 74 raw fastq files were first quality controlled using PRINSEQ⁶¹ with parameters set to retain reads with a minimum length of 100 basepairs, have no ambiguous bases (N), a minimum quality score of 20 and no exact duplicates. After quality control, an average of $86.4 \pm 0.8\%$ of reads were retained. Only forward reads were annotated to generate taxonomic and gene function annotations. The taxonomic identity was assigned using Focus⁶², a tool designed for rapid annotation by matching k-mer profiles calculated from metagenomic reads to precalculated k-mer profiles of reference databased genomes (Date of reference genome database: 2018).

Network construction. Networks were constructed to examine the microbial family level architecture of the epidermal microbiome of whale sharks from different aggregations. Networks attempt to reconstruct co-occurrence patterns based on the abundance relationships of organisms in a community. The architecture of the networks is built on the distribution of the number of connected members (degree distribution) and the formation of these members into groups (modularity)⁶³. Metagenomic samples are compositional and suffer from low sample size relative to the number of co-occurrences (i.e., Species 1 to Species 2 covariation), thus to overcome these limitations, we constructed networks using the SpiecEasi algorithm (version 1.0.7)⁶⁴ which combines transformations specific to compositional values and an underlying graphical model that assumes sparse data. Due to the large number of rare microbial groups inherent to metagenomic community datasets, we removed taxa that had less than a mean of 100 total reads across all metagenomes. Network analysis is sensitive to sample size; therefore, we excluded the Tanzania samples ($n=6$). The parameter values for the SpiecEasi pipeline were 'mb', 150 and $5e-2$ for the neighborhood inference scheme, $n\lambda$ and $\lambda\text{.min. ratio}$, respectively. These values were selected to maximize network stability (0.05 per SpiecEasi recommendation) and obtained using the *getStability* function in the SpiecEasi package. Resulting adjacency matrices were converted to igraph objects with the *adj2igraph* function in the SpiecEasi package. Network visualisations were conducted within Gephi 0.9.2 and analysis was executed using the Python networkx package (2.6), where community detection was undertaken with Clauset–Newman–Moore modularity maximisation⁶⁵ prior to calculating the modularity score. To ascertain whether the network statistics (such as modularity) were driven by the degree distributions or smaller scale structures in the graph, for each empirical network (four locations) we bootstrapped 5000 degree-preserved randomisations and calculated the same statistics to find their means and confidence intervals within the random ensembles.

We determined whether the network degree distributions fit classical generative models of Erdos–Renyi (also referred to as a random network) or scale-free networks. A network is considered Erdos–Renyi if an edge is equally likely to connect any two nodes, or scale free if most nodes are not highly connected but there are 'hubs' where a single node connects to many others, creating a power law degree distribution⁶⁶. Therefore, to compare the empirical degree distributions to a null generative model, for each aggregation, we sampled 5000 $G(m, n)$ random graphs, with m nodes and n edges as per the empirical network; the 5000 networks are drawn uniformly at random from the set of all possible $G(m, n)$ graphs. Two-sample Kolmogorov Smirnov tests between the empirical and random ensemble network's degree distribution were used to generate a p -value distribution to accept or reject the null hypothesis that the empirical network's degree distribution is consistent with a $G(m, n)$ random graph. Note that the properties of $G(m, n)$ random graphs are, in this case, rather similar to $G(n, p)$ random graphs as we have $m \approx \binom{n}{2} p$, where $p \approx 0.025$. A Jupyter notebook implementing the network analysis is available at: <https://github.com/jcmckerral/whalesharknetworks>

Construction of metagenome assembled genomes (MAGs). To explore whether whale sharks have a core microbiome, despite the large geographic separation among aggregations, metagenome assembled genomes (MAGs) were constructed from the forward and reverse reads of the 74 metagenomes⁶⁷. In brief, R1 and R2 reads were concatenated for assembly into contigs and the quality of assembly checked. Contigs were

then binned and the quality of each MAG checked for completeness and contamination using CheckM^{68,69}. Reads from each sample were then mapped back to each MAG to identify how each sample contributed to MAG construction. Taxonomy of the MAGs and MAG gene functions were identified with PATRIC version 3.6.9 using the RAST tool kit (RASTtk)⁷⁰.

Statistical analysis. Metagenomes were compared using proportional abundance, a more robust approach to rarefaction⁷¹. Alpha diversity was compared using several metrics to evaluate how abundance and richness influenced diversity patterns. Richness (S) was calculated as the total number of families found in each sample, and evenness as the quotient of Shannon's index (H') and S, each calculated in Vegan version 2.5.7 with the *diversity* function. Effective diversity was calculated across samples as $e^{(H')}$ ⁷². Diversity distribution was compared across aggregations with non-parametric Kruskal test with Dunn pairwise tests using the *dunn.test* function with `kw = TRUE` in the package *dunn.test* version 1.3.5.

Dissimilarity matrices were generated with *vegdist* function from the Vegan package, with `method = 'bray'` for Bray–Curtis dissimilarity, to test differences in taxonomic family level composition across aggregations. MDS ordination was used to visualize beta diversity patterns and permutational anova (PERMANOVA) to evaluate compositional patterns across aggregations using *adonis2* function with default settings⁷³. Microbiomes were evaluated using three independent tests that included aggregation ($n = 5$), ocean ($n = 3$), and population ($n = 2$). Pairwise permutational anova was conducted using the package *pairwiseAdonis* (version 0.0.1) and the *pairwiseadonis* function. All visualizations were conducted using the GGplot package (version 3.3.3) or pheatmap package (version 1.0.12). All analyses were conducted using R version 3.6.0. Code is freely available at https://github.com/mpdoane2/whaleshark_analysis.

Ethical approval. Animal handling and ethics were reviewed at San Diego State University through IACUC under permit APF #14-05-011D, APF #17-11-010D, APF # 18-05-007D.

Data availability

The raw metagenomic datasets generated for this project are available under BioProject accession PRJNA808622.

Received: 18 October 2022; Accepted: 20 July 2023

Published online: 07 August 2023

References

- Apprill, A. Marine animal microbiomes: Toward understanding host–microbiome interactions in a changing ocean. *Front. Mar. Sci.* **4**, 1–9 (2017).
- Ottman, N., Smidt, H., de Vos, W. M. & Belzer, C. The function of our microbiota: Who is out there and what do they do?. *Front. Cell. Infect. Microbiol.* **2**, 104 (2012).
- McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *PNAS* **110**, 3229–3236 (2013).
- Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327–336 (2011).
- McKenney, E. A., Koelle, K., Dunn, R. R. & Yoder, A. D. The ecosystem services of animal microbiomes. *Mol. Ecol.* **27**, 2164–2172 (2018).
- Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
- Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: Networks, competition, and stability. *Science* **350**, 663–666 (2015).
- Guo, B. *et al.* Microbial co-occurrence network topological properties link with reactor parameters and reveal importance of low-abundance genera. *Nat. Biofilms Microbiomes* **3**, 1–13 (2022).
- Layeghifard, M., Hwang, D. M. & Guttman, D. S. Disentangling interactions in the microbiome: A network perspective. *Trends Microbiol.* **25**, 217–228 (2017).
- Lean, C. H. Can communities cause?. *Biol. Philos.* **34**, 59 (2019).
- Lau, M. K., Borrett, S. R., Baiser, B., Gotelli, N. J. & Ellison, A. M. Ecological network metrics: Opportunities for synthesis. *Ecosphere* **8**, e01900 (2017).
- de Jesús Astacio, L. M., Prabhakara, K. H., Li, Z., Mickalide, H. & Kuehn, S. Closed microbial communities self-organize to persistently cycle carbon. *PNAS* **118**, e2013564118 (2020).
- Faust, K., Lahti, L., Gonze, D., de Vos, W. M. & Raes, J. Metagenomics meets time series analysis: Unraveling microbial community dynamics. *Curr. Opin. Microbiol.* **25**, 56–66 (2015).
- Lurgi, M., Thomas, T., Wemheuer, B., Webster, N. S. & Montoya, J. M. Modularity and predicted functions of the global sponge-microbiome network. *Nat. Commun.* **10**, 992 (2019).
- Shi, Y., Delgado-baquerizo, M., Li, Y., Yang, Y. & Zhu, Y. Abundance of kinless hubs within soil microbial networks are associated with high functional potential in agricultural ecosystems. *Environ. Int.* **142**, 105869 (2020).
- Vignaud, T. M. *et al.* Genetic structure of populations of whale sharks among ocean basins and evidence for their historic rise and recent decline. *Mol. Ecol.* **23**, 2590–2601 (2014).
- Hartle, H. *et al.* Network comparison and the distance. *Proc. R. Soc. A* **476**, 20190744 (2020).
- Levin, D. *et al.* Diversity and functional landscapes in the microbiota of animals in the wild. *Science* **372**, eabb5352 (2021).
- Tecon, R. & Or, D. Cooperation in carbon source degradation shapes spatial self-organization of microbial consortia on hydrated surfaces. *Sci. Rep.* **7**, 1–11 (2017).
- Liu, Z. *et al.* The self-organization of marine microbial networks under evolutionary and ecological processes: Observations and modeling. *Biology (Basel)* **11**, 592 (2022).
- Karimi, B. *et al.* Biogeography of soil bacterial networks along a gradient of cropping intensity. *Sci. Rep.* **9**, 3812 (2019).
- Shi, S. *et al.* The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecol. Lett.* **19**, 926–936 (2016).
- Doane, M. P., Haggerty, J. M., Kacev, D., Papudeshi, B. & Dinsdale, E. A. The skin microbiome of the common thresher shark (*Alopias vulpinus*) has low taxonomic and gene function β -diversity. *Environ. Microbiol. Rep.* **9**, 357–373 (2017).
- Lücken, L., Lennartz, S. T., Froehlich, J. & Blasius, B. Emergent diversity and persistent turnover in evolving microbial cross-feeding networks. *Front. Netw. Physiol.* **2**, 1–17 (2022).

25. Ortiz-Álvarez, R. *et al.* Network properties of local fungal communities reveal the anthropogenic disturbance consequences of farming practices in vineyard soils. *mSystems* **6**, 00344–21 (2021).
26. Giovannoni, S. J. & Stingl, U. Molecular diversity and ecology of microbial plankton. *Nature* **437**, 343–348 (2005).
27. Rivett, D. W. & Bell, T. Abundance determines the functional role of bacterial phylotypes in complex communities. *Nat. Microbiol.* **3**, 767–772 (2018).
28. Goodman, A. Z. *et al.* Epidermal microbiomes of leopard sharks (*Triakis semifasciata*) are consistent across Captive and wild environments. *Microorganisms* **10**, 2081 (2022).
29. Pratte, Z. A. *et al.* Microbiome structure in large pelagic sharks with distinct feeding ecologies. *Anim. Microbiome* **4**, 1–16 (2022).
30. Smith, C. J., Danilowicz, B. S. & Meijer, W. G. Bacteria associated with the mucus layer of *Merlangius merlangus* (whiting) as biological tags to determine harvest location. *Can. J. Fish. Aquat. Sci.* **66**, 713–716 (2009).
31. Shephard, K. L. Functions for fish mucus. *Rev. Fish Biol. Fish.* **4**, 401–429 (1994).
32. Bansil, R. & Turner, B. S. The biology of mucus: Composition, synthesis and organization. *Adv. Drug Deliv. Rev.* **124**, 3–15 (2018).
33. Molina, V. & Fernández, C. Bacterioplankton response to nitrogen and dissolved organic matter produced from salmon mucus. *Microbiologyopen* **9**, 1–11 (2020).
34. Feng, H. *et al.* Chemotaxis of beneficial rhizobacteria to root exudates: The first step towards root-microbe rhizosphere interactions. *Int. J. Mol. Sci.* **22**, 6655 (2021).
35. Hernandez-Agreda, A., Leggat, W. & Ainsworth, T. D. A place for taxonomic profiling in the study of the coral prokaryotic microbiome. *FEMS Microbiol. Lett.* **366**, 1–9 (2019).
36. Chien, H. W., Chen, X. Y., Tsai, W. P. & Lee, M. Inhibition of biofilm formation by rough shark skin-patterned surfaces. *Colloids Surfaces B Biointerfaces* **186**, 110738 (2020).
37. Woznica, A. *et al.* Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *PNAS* **113**, 7894–7899 (2016).
38. Nazir, R., Mazurier, S., Yang, P., Lemanceau, P. & van Elsas, J. D. The ecological role of type three secretion systems in the interaction of bacteria with fungi in soil and related habitats is diverse and context-dependent. *Front. Microbiol.* **8**, 1–14 (2017).
39. Heaver, S. L., Johnson, E. L. & Ley, R. E. Sphingolipids in host-microbial interactions. *Curr. Opin. Microbiol.* **43**, 92–99 (2018).
40. Sepahi, A., Cordero, H., Goldfine, H., Esteban, M. Á. & Salinas, I. Symbiont-derived sphingolipids modulate mucosal homeostasis and B cells in teleost fish. *Sci. Rep.* **6**, 1–13 (2016).
41. Doane, M. P. *et al.* The epidermal microbiome within an aggregation of leopard sharks (*Triakis semifasciata*) has taxonomic flexibility with gene functional stability across three time-points. *Microb. Ecol.* **85**, 747–764 (2023).
42. Lieber, L. *et al.* Mucus: Aiding elasmobranch conservation through non-invasive genetic sampling. *Endanger. Species Res.* **21**, 215–222 (2013).
43. Giovannoni, S. J. SAR11 bacteria: The most abundant plankton in the oceans. *Ann. Rev. Mar. Sci.* **9**, 231–255 (2017).
44. Pogoreutz, C. *et al.* Similar bacterial communities on healthy and injured skin of black tip reef sharks. *Anim. Microbiome* **1**, 9 (2019).
45. Marchioro, G. M. *et al.* Microbiome dynamics in the tissue and mucus of acroporid corals differ in relation to host and environmental parameters. *PeerJ* **8**, 1–26 (2020).
46. Hoffmayer, E. R. *et al.* Seasonal occurrence, horizontal movements, and habitat use patterns of whale sharks (*Rhincodon typus*) in the Gulf of Mexico. *Front. Mar. Sci.* **7**, 598515 (2021).
47. Araujo, G., Labaja, J., Snow, S., Huvneers, C. & Ponzio, A. Changes in diving behaviour and habitat use of provisioned whale sharks: Implications for management. *Sci. Rep.* **10**, 1–12 (2020).
48. Tyminski, J. P., De La Parra-Venegas, R., Cano, J. G. & Hueter, R. E. Vertical movements and patterns in diving behavior of whale sharks as revealed by pop-up satellite tags in the Eastern Gulf of Mexico. *PLoS ONE* **10**, e0142156 (2015).
49. Landeira-Dabarca, A., Sieiro, C. & Álvarez, M. Change in food ingestion induces rapid shifts in the diversity of microbiota associated with cutaneous mucus of Atlantic salmon *Salmo salar*. *J. Fish Biol.* **82**, 893–906 (2013).
50. Araujo, G. *et al.* Population structure and residency patterns of whale sharks, *Rhincodon typus*, at a provisioning site in Cebu, Philippines. *PeerJ* **2014**, 1–20 (2014).
51. Rohner, C. A. *et al.* Whale sharks target dense prey patches of sergestid shrimp off Tanzania. *J. Plankton Res.* **37**, 352–362 (2015).
52. Clark, E. & Nelson, D. R. Young whale sharks, *Rhincodon typus*, feeding on a copepod bloom near La Paz, Mexico. *Environ. Biol. Fishes* **50**, 63–73 (1997).
53. de la Parra Venegas, R. *et al.* An unprecedented aggregation of whale sharks, *Rhincodon typus*, in Mexican coastal waters of the Caribbean Sea. *PLoS ONE* **6**, e18994 (2011).
54. Schmidt, J. V. Genetic population structure of whale sharks. In *Whale Sharks: Biology, Ecology, and Conservation* (eds Dove, A. D. M. & Pierce, S. J.) 83–103 (CRC Press, 2021). <https://doi.org/10.1201/b22502-5>.
55. Easson, C. G., Chaves-Fonnegra, A., Thacker, R. W. & Lopez, J. V. Host population genetics and biogeography structure the microbiome of the sponge *Cliona delitrix*. *Ecol. Evol.* **10**, 2007–2020 (2020).
56. Ahern, O. M., Whittaker, K. A., Williams, T. C., Hunt, D. E. & Ryneerson, T. A. Host genotype structures the microbiome of a globally dispersed marine phytoplankton. *PNAS* **118**, 1–8 (2021).
57. Ramirez Macas, D., Vázquez-Haikin, A. & Vázquez-Juárez, R. Whale shark (*Rhincodon typus*) populations along the west coast of the Gulf of California and implications for management. *Endanger. Species Res.* **18**, 115–128 (2012).
58. Doane, M. P. *et al.* The skin microbiome of elasmobranchs follows phylosymbiosis, but in teleost fishes, the microbiomes converge. *Microbiome* **8**, 93 (2020).
59. Lima, L. F. O. *et al.* Modeling of the coral microbiome: The influence of temperature and microbial network. *MBio* **11**, 1–17 (2020).
60. Edwards, R. A. *et al.* Microbes, metagenomes and marine mammals: enabling the next generation of scientist to enter the genomic era. *BMC Genomics* **14**, 600 (2013).
61. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
62. Silva, G. G. Z., Cuevas, D. A., Dutilh, B. E. & Edwards, R. A. FOCUS: An alignment-free model to identify organisms in metagenomes using non-negative least squares. *PeerJ* **2**, e425 (2014).
63. Thébault, E. & Fontaine, C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329**, 853–856 (2010).
64. Kurtz, Z. D. *et al.* Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **11**, e1004226 (2015).
65. Clauset, A., Newman, M. E. J. & Moore, C. Finding community structure in very large networks. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **70**, 6 (2004).
66. Barabási, A. L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
67. Papudeshi, B. *et al.* Optimizing and evaluating the reconstruction of metagenome-assembled microbial genomes. *BMC Genomics* **18**, 1–13 (2017).
68. Davis, J. J. *et al.* The PATRIC bioinformatics resource center: Expanding data and analysis capabilities. *Nucleic Acids Res.* **48**, D606–D612 (2020).
69. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

70. Brettin, T. *et al.* RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 8365 (2015).
71. McMurdie, P. J. & Holmes, S. Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
72. Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427–2439 (2007).
73. Anderson, M. J. Permutational multivariate analysis of variance (PERMANOVA). *Wiley StatsRef Stat. Ref. Online* <https://doi.org/10.1002/9781118445112.stat07841> (2017).

Acknowledgements

We would like to thank Stephanie Lo and Ben Billings Global Shark Research and Conservation Fund. Field work in Tanzania was supported by Aqua-Firma, WWF Tanzania, the Shark Foundation, Waterlust and two private trusts, and we thank Mathias Igulu, Paul Kugopya, Liberatus Mokoki and Carlos Omari for their help with field logistics. In the Philippines we would further like to acknowledge the Tan-awan Oslob Sea Warden and Fishermen Association (TOSWFA) for assisting in this project and facilitating our work. The project was conducted under permission from the Municipality of Oslob, duly represented by Hon. Mayor Jon Tumalak, through a Prior Informed Consent. Export of water samples from the Philippines were under export permit BFAR-7 n from the Bureau of Fisheries and Aquatic Resources—Region 7—Department of Agriculture, Cebu City, Philippines. Sampling in Cancun and La Paz were under permit 03362 from the Secretaria de Medio Ambiente Y Recurso Naturales.

Author contributions

M.D. and E.D. conceived of project design. M.R. conducted the network analysis. M.D. and M.R. drafted manuscript. R.E. and B.P. provided bioinformatic analysis. M.D., E.D., M.R., L.F.O.L., J.M. performed data analysis. M.D., E.D., S.P., C.R., D.R., R.P., C.L., G.A., R.P.V., R.E., M.P., A.T., T.D. helped with sample collection and facilitated on site activities. All authors help with interpretation of results, provided feedback, and approved final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-39184-5>.

Correspondence and requests for materials should be addressed to M.P.D. or E.A.D.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023