Title: Comparative metagenomics evidence distinct trends of genome evolution between sponge-dwelling bacteria and their pelagic counterparts

Authors: Joseph B. Kelly[1,2*], David Carlson[1], Jun Siong Low[3,4], Robert W. Thacker[1,5].

Affiliations

[1]Stony Brook University, Department of Ecology and Evolution, Stony Brook NY, USA.

[2]Limnological Institute University Konstanz, Aquatic Ecology and Evolution, Konstanz, Germany.

[3]Fondazione per l'istituto di ricerca in biomedicina, Bellinzona, Switzerland.

[4]Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA.

[5]Smithsonian Tropical Research Institute, Box 0843-03092, Balboa, Republic of Panama.

*Correspondence to: joseph.b.kelly@stonybrook.edu.

Abstract. Prokaryotic associations with sponges are among the oldest host-microbiome relationships on Earth. In this study, we investigated how bacteria from several phyla have independently adapted to the sponge interior by comparing metagenome-assembled genomes of sponge-dwelling and pelagic bacteria sourced from broad phylogenetic and geographic samplings. We discovered that sponge-dwelling bacteria have more energetically expensive genomes and share patterns of depletion and enrichment for functional categories of genes that evidence evolution towards lower pathogenicity. We also identified a new defining genomic characteristic of sponge-dwelling bacteria that is virtually absent from pelagic bacteria, the presence of cassettes that contain eukaryotic steroid biosynthesis genes. Collectively, these results illuminate the trends in genome evolution that are associated with a sponge-dwelling life

history strategy and have implications for furthering our understanding of how sponge-microbial symbioses have persisted through deep evolutionary time.

Importance

Much attention has recently been devoted to investigating the evolution of microbes that live in symbiosis with sponge hosts using microbial metagenomic data. However, several biological questions regarding this symbiosis remain unanswered. Two questions that we address here are: 1) what are the long-term consequences of the symbiosis on the evolution of microbial symbiont genome size, protein content, and nucleotide content, and 2) how is the evolution of virulence in sponge-dwelling microbial symbionts, which generally undergo a mixed transmission modes (e.g. horizontal and vertical), related to long-term stability of the symbiosis? By employing the largest comparative metagenomic analysis to date in terms of host sponge species and geographic representation, we address these questions and provide further resolution into the evolutionary processes that are involved in mediating the crosstalk between sponge hosts and their microbial symbionts.

Introduction

The universal association of microbes with multicellular hosts is now recognized as an established principle of biology (1). In the past few decades, the combination of plummeting costs of high throughput sequencing and key statistical and computational advancements have driven a renaissance in microbiology focused on investigating the genetic underpinnings of microbial physiologies. Consequently, science has gained unprecedented insights into prokaryotic defense systems, molecular machinery related to virulence (a term that we use here

to refer to both pathogenicity, the process of causing harm to the host, and the promotion of the microbe's residence in the host), and the metabolisms of uncultured microbes. Understanding the ways in which microbes evolve in association with their hosts is central to understanding how microbes impact ourselves and our environment. In the current study, we seek to elucidate what the impacts of host-microbial associations have been on prokaryotic genome evolution in what is perhaps the oldest-standing relationship between microbes and multicellular life.

Prokaryotes are unusually successful in their colonization of sponges (phylum Porifera), a symbiosis established early in the evolutionary history of the phylum around 540 million years ago (2). Sponge microbiomes can account for up to 40% of the sponge's volume (3) and routinely comprise thousands of microbial taxa from dozens of eubacterial and archaeal phyla (4). The taxonomic richness of sponge-dwelling prokaryotes (SDPs) is met by their physiological diversity as numerous autotrophic and heterotrophic nutritional strategies have proliferated in these microbes in addition to their capacity to synthesize a multitude of bioactive secondary metabolites (5–8). Sponge microbiomes are assembled horizontally through transmission from the environmental microbial pool and vertically through parental inheritance in a phenomenon termed mixed-mode transmission (9,10). The longevity of the SDP-sponge symbiosis is at odds with the prevalence of horizontal transmission in the system, since competition among horizontally transmitted SDPs and their relaxed dependence on survival of the host would incentivize the evolution of higher virulence, resulting in greater damage to the host that would ultimately endanger the symbiosis (11–14). The mystery stands: among the distantly related SDPs that possess such metabolic diversity, and whose transmission mode presents an evolutionary paradox, what features do they have in common that account for their shared success within the environments of sponge bodies and the stability of the symbiosis?

The current study seeks to illuminate what the consequences of sponge inhabitation have been on SDP genome evolution by performing the largest comparative metagenomic analysis to date of SDPs and free-living prokaryotes. Since the host sponge species in our analysis were sourced worldwide from multiple biomes (Table S1), we decided to use the MAGs assembled from the Tara Oceans pelagic water samples (15) as our free-living dataset to standardize the comparison. We focus on two general genomic facets; the first is the evolution of general genomic characteristics and encompasses genome size and, particularly, how it scales with protein-coding and GC contents. In the open ocean, these features are intrinsically linked to nutrient availability, which drives genome streamlining of pelagic prokaryotic genomes and A-T biases in nucleotide composition (16–19). The genomes of at least one bacterial genus, *Synechococcus* (phylum Cyanobacteria), have been found to be of comparable size between sponge-dwelling and pelagic strains (20,21). Yet, we lack a model describing how genome size scales with protein and GC content across SDPs from distantly related prokaryotic clades. Second, we investigate the shared enrichment patterns of individual genes in SDPs with the goal of elucidating the sources of selection that are experienced across SDP taxa in multiple host species. Previous work has uncovered patterns of enrichment for functional categories of genes that are characteristic of SDPs and are likely adaptations to a symbiotic lifestyle including ATP-binding cassette (ABC) transporters, prokaryotic defense mechanisms (e.g. CRISPR-CAS and restriction-modification (RM) systems), and eukaryotic-like proteins (ELPs) (22–26). Other functional categories of genes that concern virulence and are thus relevant to how the SDPs interact with their hosts, such as cellular motility and chemotaxis, have been reported as being depleted in three sulfur-oxidizing gammaproteobacterial SDPs (27) and in SDPs belonging to *Rhodospirillaceae* (Alphaproteobacteria) (28). However, theoretical work has posited that

horizontally transmitted symbionts would benefit from increased motility and chemotaxis as these traits would result in increased recruitment to the host (29). Since mixed-mode transmission is a major driver of sponge microbiome assembly (9,10), we need to develop a more complete understanding of how virulence has evolved in SDPs to better model how environmental recruitment shapes the microbiomes of sponges.

Materials and Methods

DNA extractions and sequencing

New metagenome assembled genomes (MAGs) were produced for this study from two specimens of each of three nominal *Ircinia* species and seven *Ircinia* growth forms resulting in a sample size of 20. In Bocas del Toro, Panama, specimens of the 'Massive A pink' growth form were collected from prop roots of the mangrove hammock at Inner Solarte (latitude, longitude: 9.3058, -82.1732), individuals of the growth forms 'Massive A green' and 'Massive B' from seagrass beds near the Smithsonian Tropical Research Institute's Bocas del Toro Research Station (9.3517, -82.2590), and individuals of the 'Encrusting' growth form from patch reefs at Punta Caracol (9.3771, -82.3023). In Florida, specimens of *I. campana* and the 'Ramose' growth form were collected from a seagrass bed near the MOTE Marine Laboratory and Aquarium's Elizabeth Moore International Center for Coral Reef Research & Restoration (24.6609, -81.4563). In Belize, specimens of the 'BZ1' growth form were collected from the prop roots of mangrove hammocks near the Blue Ground site (16.8083, -88.1496), specimens of *I. strobilina* and the 'BZ2' were collected from the coral patch reefs of Blue Ground (16.8010, -88.1461), and specimens of *I. felix* were collected from the forereef of Carrie Bow Cay (16.8042, -88.0796). Thumb-size tissue fragments were cut from each sponge and stored in 90% EtOH that was

replaced at the 24 and 48 timepoints to ensure thorough inundation. Photos and descriptions of the growth forms and maps of the sampling locations can be found in Kelly et al. (2020). Note: the specimens used in the current study are present in the set of specimens used in Kelly et al. (2020).

DNA was extracted from each specimen using the Molzyme Molysis DNA isolation kit, which depletes eukaryotic host DNA in a sample resulting in an extraction that is enriched for microbial DNA. The absence of host DNA was evaluated by amplifying the Molzyme DNA isolations with the *Ircinia*-specific *cytochrome oxidase c subunit 1* primers and PCR conditions reported in Kelly & Thacker (2020). To provide a positive control, PCR amplifications were repeated for the same samples using bulk-tissue DNA isolations produced using the Wizard Genomic DNA Purification kit. Whole-metagenome shotgun sequencing was performed for the Molzyme DNA extractions after confirming the absence of an amplification band at the Yale Center for Genome Analysis on a NovaSeq6000 with a target of 30 million 2x150 reads per specimen.

Metagenomic analysis

The reads were filtered and trimmed using FastP (32) with default parameters. Since the DNA libraries were prepared in a facility that routinely handles vertebrate model organisms, bbsplit.sh was used to remove contaminating reads mapping to the *Mus musculus* and *Homo sapiens* genomes as a precautionary step in addition to removing the control PhiX reads. Metagenomic contigs were assembled independently for each specimen using Megahit v1.2.9 (33) and subsequently binned using MetaBAT 2 v2:2.15 (34) with default parameters. Contaminating contigs were identified on the basis of being outliers with regard to genomic

content (i.e., %GC and tetranucleotide content) and removed from the bins using refinem v0.0.25 (35). The qualities of the refined bins were assessed using checkm v1.1.2 (36); only bins with qualities equal to or higher than 40, calculated as genome completeness – 5x contamination (35), were retained for downstream analysis. The remaining bins were screened for uniqueness within host specimens using checkm.

To investigate the microbial mechanisms that promote their residence in sponge hosts, we compared a dataset of SDP MAGs with a dataset of previously assembled MAGs from the Tara Oceans expedition (15). The Tara Oceans dataset was produced from DNA isolated from filtered oceanic seawater, and thus we assume that most of these microbes are pelagic; these are referred to as the pelagic dataset throughout the manuscript (15). Prior to quality control and dereplication, the pelagic dataset consisted of 2307 MAGs (NCBI bioproject PRJNA391943, Table S2). The SDP MAG dataset was produced by combining our Caribbean *Ircinia* MAG dataset with 689 publicly available draft genomes of SDPs (Table S2). Prior to dereplication and further QC, the SDP dataset consisted of 1579 MAGs. The SDP and Tara Oceans datasets were independently dereplicated using drep v2.5.2 (37) with a cutoff of 96.5% identity, an estimate of maximum prokaryotic intraspecific genomic divergence backed by empirical evidence (38). We retained only MAGs with a quality equal to or greater than 40 and a completeness greater than 85% for downstream analyses. Taxonomy was again assigned and a whole-genome phylogenetic tree was produced for the combined datasets using GTDB-Tk v1.0.2 (39).

Protein predictions and subsequent KO annotations (http://www.kegg.jp/kegg/) were performed on the MAGs using Enrichm v0.5.0 (40). Domain PFAM annotations were performed on the MAG proteins using Interproscan v5.39-77 (41). For proteins that contained duplicated

identical domain annotations, only one domain was retained so as to avoid double counting proteins.

Statistical analysis

To identify features that were significantly enriched in one life history category relative to the other, we performed pairwise permutational t-tests, each run from 10,000 iterations, comparing the average abundance (e.g. copy number) of a feature in the pelagic vs. the SDP dataset. To factor for differences in genome size, we normalized feature counts by dividing them by the total number of predicted proteins prior to the t-tests. P-values were corrected using the FDR adjustment of Benjamini-Hochberg (42) using the total number of features within an annotation category (either KO or PFAM). Given the high number of multiple comparisons, we further decided to assign a conservative significance cutoff of $p<0.01$. Additionally, we omitted features that were present in less than 10% of both the SPD and pelagic MAGs.

Linear models (linear regressions) were inferred to describe the relationship between genome size vs. GC content and coding density. Statistical significance in differences between the linear models of SDP and pelagic microbes was inferred using ANCOVA. We tested for significance in genome size, GC content, coding density, and the total number of proteins between SDPs and pelagic MAGs within a phylogenetic framework using an ultrametric bacterial tree estimated with the software PATHd8 (43). The GTDB-Tk bacterial phylogeny was used as the input and node dating was calibrated using the divergence times estimates of Battistuzzi et al. (2004) for the cyanobacteria/actinobacteria, cyanobacteria/proteobacteria, and alphaproteobacteria/gammaproteobacteria splits. We then performed phylogenetic ANOVAs for

each of the aforementioned genomic traits using the R package geiger (45), each for 10,000 iterations. All aforementioned tests of significance and regressions were performed in R v.3.6.0.

Results and Discussion

To investigate shared patterns of genome evolution in SDPs, we performed a comparative metagenomic analysis evaluating nucleotide composition, genome size, and gene content of SDPs from a broad sampling of SDP MAGs sourced from multiple host species and oceanic basins against the MAG dataset of pelagic prokaryotes sequenced during the Tara Oceans project (15). After screening for quality, completeness, and dereplication, our final bacterial dataset was composed of 479 SDP MAGs sourced from 23 nominal and 10 undescribed host species and 629 pelagic MAGs; representatives from the Atlantic, Mediterranean, Red Sea, and Pacific Ocean basins were present in both datasets (Table S1). Twenty-five bacterial phyla were recovered with a high degree of phylogenetic overlap between the two datasets, with the majority of SDP MAGs (95.0%) and pelagic MAGs (92.7%) belonging to phyla that contained both life history strategies (Table S4). We report our results below for analyses performed on the bacterial MAGs, as only 15 archaeal MAGs were recovered (Table S5). We confirmed that the omission of Archaea did not change the interpretations of the results by repeating our analyses on the total (Bacteria + Archaea) MAG dataset (data not shown). We tested the hypotheses that 1) general genomic characteristics distinguish SDPs from free-living oceanic bacteria and 2) enrichment patterns in functional categories of genes that are relevant to host-symbiont crosstalk, including genes that facilitate horizontal transmission (e.g. motility and chemotaxis), are different in SDPs relative to pelagic bacteria.

SDPs have evolved larger genomes with higher GC contents

The evolution of general genomic features across marine prokaryotic taxa has been shaped by life history strategies. We found that SDPs have larger genomes, with both more proteins and non-coding DNA, and higher GC content (Fig. 1, 2). The effect of life history strategy remained significant after accounting for phylogenetic relatedness (Fig. 1). Life history strategy also had a significant effect on the relationship between genome size and both GC content (ANCOVA: $F_{1,1104} = 118.32$, $p < 2e^{-16}$) and coding density (ANCOVA: $F_{1,1104} = 104.3$, $p < 2e^{-16}$) (Fig. 2). Based on comparisons among regression coefficients, we found that genome size was a stronger predictor for GC content and coding density in pelagic MAGs relative to SDPs (Fig. 2). Our finding of larger genome sizes and higher GC contents in SDPs relative to seawater bacteria is consistent with previous observations made of SDPs inhabiting three Mediterranean sponge species, *Petrosia ficiformis*, *Sarcotragus foetidus*, and *Aplysina aerophoba*, suggesting that these are geographically widespread genomic characteristics found in SDPs across host taxa (24).

We hypothesize that differences in energy availability between the two life history strategies may account for the contrasting genomic characteristics of SDP and pelagic MAGs. Due to nitrogen limitation in the open ocean, pelagic prokaryotes tend to have streamlined genomes and lower GC content; this strategy reduces the energetic costs of DNA replication (16–19). Each of these expectations are met in our pelagic dataset, as pelagic MAGs have smaller genomes with fewer genes, less non-coding DNA, and lower GC content. However, opposite trends in genome content were discovered in SDPs, suggesting that the sponge host might be providing the symbionts with nitrogen. This hypothesis is corroborated by the observation that two sponge species, *Xestospongia muta* (46) and *Haliclona cymiformis* (47), are

documented sources of inorganic nitrogen (NH_3) for SDPs. Further research could investigate whether the exploitation of nitrogenous waste by SDPs is widespread throughout sponge host taxa to elucidate whether a sponge-dwelling lifestyle alleviates SDPs of nitrogen limitation and enables the maintenance of larger genomes with higher GC contents.

Gene enrichment patterns suggest SDPs are equipped to exploit the organic molecules present in the sponge body and have evolved towards lower pathogenicity

Previous studies have concluded that several classes of molecular transporters (e.g. ABC transporters) are enriched in SDPs (22,25,48). Our results are consistent with this finding (Fig. 3) and provide further resolution into which classes of ABC transporters are enriched or depleted in SDPs. For example, SDPs are depleted for lipopolysaccharide (LPS) transporters, Hg^+/heavy metal efflux proteins, and the *pha* family of pH-dependent H^+ cation transporters. Conversely, and consistent with observations made in previous studies (22,28,48–50), SDPs are enriched for amino acid and spermidine/putrescine transporters. Early work discovered that concentrations of free amino acids are relatively high in sponges, which are thought to play a role in host osmoregulation (51). Microbial symbionts could be capitalizing on this pool of bioavailable nutrients to supplement their own metabolism, possibly to provide cheap sources of amino acids for protein biosynthesis. Likewise, spermidine and putrescine are polyamines that are essential for cell growth (52) and perform protective functions for host-associated prokaryotes (53). Recent genomic analysis evidenced a capacity for members of at least one clade, the gammaproteobacterial order Tethybacterales, to metabolize and derive energy from spermidine (54). Collectively, our results suggest that SDPs have evolved a capacity to utilize the organic molecules in sponge bodies to meet their energetic needs.

Consistent with the observations made for gammaproteobacterial SDPs made by Tian et al. (2017) and Karimi et al. (2018), we detected a near complete absence of motility and chemotaxis genes across SDPs. Additionally, we detected a depletion of lipopolysaccharide (LPS) biosynthesis genes and the aforementioned LPS transport genes in SDPs (Fig. 3). We also recovered enrichment patterns of ELPs including the previously reported fibronectin type-III, ANK, LRR, TPR, NHL, WD40, cadherin, and PQQ domains (22,25,26,48,55) and two new ELPs: calx-beta motifs and eukaryotic-type carbonic anhydrase. Most notably, we discovered that SDPs were also enriched for genes involved in steroid biosynthesis that are localized on the bacterial chromosomes in cassettes (Fig. 4) that constitute an incomplete eukaryotic steroid biosynthesis pathway (KEGG pathway ko00100). Other genes present in this pathway were found to be missing from the SDP genomes. Our results are similar to prior observations made of SDPs inhabiting the sponge *Vaceletia sp*. where, again, these genes represented an incomplete ko00100 pathway that lacked the machinery to produce steroid precursor molecules. Instead, it was found that several complementary components of the pathways were discovered in the host *Vaceletia* genome, suggesting that steroid precursors may originate in the sponge itself (56).

Proteins involved in motility, such as flagellin, and bacterial lipopolysaccharides are directly involved in eliciting the immune response from the host that causes pathogenicity (57–60); thus, a potential evolutionary outcome of the depletion of these genes in SDPs might be one of reduced symbiont pathogenicity. At face value, this hypothesis appears to be at odds with the theory that lower levels of pathogenicity would only be expected to confer a selective advantage to the microbe if the persistence of its genome was completely dependent on survival of the host, i.e. under obligate vertical transmission (11,12). On the other hand, higher rates of pathogenicity are expected to evolve in microbes exhibiting mixed-transmission modes as pathogenic

mechanisms act to protect the microbe, for example through the strengthening of cell envelopes with LPS, while not damaging the microbial population's growth rate since it is not linked to the host's (13). Why then do we detect a depletion of pathogenic mechanisms in prokaryotes that are likely undergoing mixed-modes of transmission in sponges (9,10)? One possibility is that the products of the steroid biosynthesis gene cassettes could be playing an analogous role to LPS in strengthening the cellular envelope and contributing to the resistance of stressors such as antimicrobial secondary metabolites (5,61) and hypoxic conditions (62,63) that are present in the interior of the host, as is the case in other symbiotic microbes such as the legume-inhabiting *Bradyrhizobium* (64–66). An alternative role for the steroid genes might be that their products have immunosuppressive properties, similar to the endogenous steroid hormones of eukaryotes, which could help prevent removal of the SDPs by the hosts' immune system (67,68). Under this scenario, SDPs could still experience selection against LPS, cellular motility, and chemotaxis genes if the dependence of SDPs on their hosts for survival outweighs the fitness benefits that these genes confer outside of the host. Experimental determination of the gene products of the cassettes containing steroid biosynthesis genes and functional assays are required to delineate the roles that the steroid genes play in the sponge-SDP symbiosis.

Conclusion

Our results provide new insights into the characteristics that distinguish SDP from those of free-living microbes and help elucidate the metabolic and defensive strategies adopted by SDPs that translate to their success in sponge hosts. We discover greater genome sizes and GC contents of SDPs could that could, in part, be enabled by allocation of nitrogen from the host sponge to SDPs via nitrogenous waste products. Additionally, we discovered the near complete

absence of genes related to cellular motility chemotaxis and a depletion of LPS biosynthesis and transport genes in SDPs, evidence that inhabitation of the sponge interior involves an evolution towards decreased pathogenicity. Finally, we discovered a novel defining characteristic of SDPs: the presence of steroid gene cassettes. In particular, we advocate for future research to focus on identifying the products of these steroid-like pathways and identifying the roles that they play in mediating the symbiosis between sponges and SDPs.

Acknowledgements. We thank Dr. Jackie L. Collier and Dr. Liliana Davalos-Alvarez for their input and comments on the manuscript; the staffs of the Smithsonian Tropical Research Institute's Bocas Research Station, the Smithsonian's Carrie Bow Cay Field Station, and the MOTE Marine Laboratory and Aquarium's Elizabeth Moore International Center for Coral Reef Research & Restoration for their support on logistical aspects of the field work; and Barrett Brooks (Smithsonian Museum of Natural History) and Karen Koltes (form. Office of the Secretary of the Interior) for help with specimen collections in Carrie Bow Cay. We thank the Yale Center for Genome Analysis for their attentiveness in the metagenomic DNA library preparation and sequencing; the staff of the SeaWulf HPC at SBU, which we used to perform the majority of our analyses; and our funders at Experiement.com. **Funding:** Fieldwork and DNA sequencing for this study was supported by the following grants awarded to JBK: Fellowship of Graduate Student Travel (Society for Integrative and Comparative Biology), Dr. David F. Ludwig Memorial Student Travel Scholarship (Association for Environmental Health and Sciences Foundation), and through crowdfunding via Experiment.com. Additional funding was provided by Stony Brook University. This work was also partially supported by grants to RWT from the U.S. National Science Foundation (DEB-1622398, DEB-1623837, OCE-1756249).

Author contributions: study conceptualization, methodology: JBK, JSL, DC; formal analysis and investigation: JBK; resources and funding acquisition: JBK and RWT; writing (original draft preparation): JBK; writing (review and editing): JBK, JSL, DC, RWT; supervision: RWT. **Data and materials availability:** Raw data that was produced during this study are deposited under GenBank Accession XXX. Assembled MAGs are deposited under NCBI BioProject XXX. Accessions for publicly available data are announced in the main text and supplement.

Competing interests statement: The authors declare that no competing interests exist. Funding sources are listed in the acknowledgements section following the main text of the manuscript.

References:

1. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;
2. Brunton FR, Dixon OA. Siliceous sponge-microbe biotic associations and their recurrence through the Phanerozoic as reef mound constructors. Palaios. 1994;9:370–87.
3. Vacelet J. Etude en microscopie electronique de l'association entre bacteries et spongiaires du genre Verongia (Dictyoceratida). J Microsci Biol Cell. 1975;23(3):271–88.
4. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, et al. Diversity, structure and convergent evolution of the global sponge microbiome. Nat Commun [Internet]. 2016 Jun 16;7:11870. Available from: https://doi.org/10.1038/ncomms11870
5. Thomas TRA, Kavlekar DP, LokaBharathi PA. Marine drugs from sponge-microbe association--a review. Mar Drugs [Internet]. 2010 Apr 22;8(4):1417–68. Available from:

https://pubmed.ncbi.nlm.nih.gov/20479984

6. Hardoim CCP, Costa R. Microbial communities and bioactive compounds in marine sponges of the family Irciniidae-a review. Mar Drugs. 2014;12(20):5089–122.

7. Pita L, Rix L, Slaby BM, Franke A, Hentschel U. The sponge holobiont in a changing ocean: from microbes to ecosystems. Microbiome [Internet]. 2018;6(1):46. Available from: https://doi.org/10.1186/s40168-018-0428-1

8. Freeman CJ, Thacker RW. Complex interactions between marine sponges and their symbiotic microbial communities. Limnol Oceanogr. 2011;56(5):1577–86.

9. Thacker RW, Freeman CJ. Sponge-microbe symbioses: recent advances and new directions. Adv Mar Biol. 2012;62:57–111.

10. Björk JR, Díez-Vives C, Astudillo-García C, Archie EA, Montoya JM. Vertical transmission of sponge microbiota is inconsistent and unfaithful. Nat Ecol Evol [Internet]. 2019;3(8):1172–83. Available from: https://doi.org/10.1038/s41559-019-0935-x

11. Ferdy, Godelle B. Diversification of Transmission Modes and the Evolution of Mutualism. Am Nat. 2005 Nov 1;166:613.

12. Ewald PW. The Evolution of Virulence: A Unifying Link between Parasitology and Ecology. J Parasitol [Internet]. 1995 Apr 6;81(5):659–69. Available from: http://www.jstor.org/stable/3283951

13. Lipsitch M, Siller S, Nowak MA. The evolution of virulence in pathogens with vertical and horizontal transmission. Evolution (N Y). 1996 Apr 9;50:1729+.

14. Frank SA. Models of parasite virulence. Q Rev Biol. 1996 Mar;71(1):37–78.

15. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data [Internet]. 2018;5(1):170203.

Available from: https://doi.org/10.1038/sdata.2017.203

16. Allen LZ, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LDH, et al. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. ISME J [Internet]. 2012;6(7):1403–14. Available from: https://doi.org/10.1038/ismej.2011.201

17. Mann S, Chen Y-PP. Bacterial genomic G+C composition-eliciting environmental adaptation. Genomics [Internet]. 2010;95(1):7–15. Available from: http://www.sciencedirect.com/science/article/pii/S0888754309002080

18. Luo H, Thompson LR, Stingl U, Hughes AL. Selection Maintains Low Genomic GC Content in Marine SAR11 Lineages. Mol Biol Evol [Internet]. 2015 Jun 27;32(10):2738–48. Available from: https://doi.org/10.1093/molbev/msv149

19. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. Science. 2005 Aug;309(5738):1242–5.

20. Gao Z-M, Wang Y, Ren-Mao T, Wong YH, Batang ZB, Al-Suwailem AM, et al. Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont “Candidatus Synechococcus spongiarum". MBio. 2014;5(2):e00079–e14.

21. Burgsdorf I, Slaby BM, Handley KM, Haber M, Blom J, Marshall CW, et al. Lifestyle evolution in cyanobacterial symbionts of sponges. MBio. 2015;6(3):e00391-15.

22. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, et al. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. Proc Natl Acad Sci [Internet]. 2012 Jul 3;109(27):E1878 LP-E1887. Available from: http://www.pnas.org/content/109/27/E1878.abstract

23. Slaby BM, Hackl T, Horn H, Bayer K, Hentschel U. Metagenomic binning of a marine

sponge microbiome reveals unity in defense but metabolic specialization. Nat Publ Gr [Internet]. 2017;11(11):2465–78. Available from: http://dx.doi.org/10.1038/ismej.2017.101

24. Horn H, Slaby BM, Jahn MT, Bayer K, Moitinho-Silva L, Forster F, et al. An Enrichment of CRISPR and Other Defense-Related Features in Marine Sponge-Associated Microbial Metagenomes. Front Microbiol. 2016;7:1751.

25. Karimi E, Ramos M, Gonçalves JMS, Xavier JR, Reis MP, Costa R. Comparative Metagenomics Reveals the Distinctive Adaptive Features of the Spongia officinalis Endosymbiotic Consortium. Front Microbiol [Internet]. 2017 Dec 14;8:2499. Available from: https://pubmed.ncbi.nlm.nih.gov/29312205

26. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, et al. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. ISME J. 2010 Dec;4(12):1557–67.

27. Tian R-M, Zhang W, Cai L, Wong Y-H, Ding W, Qian P-Y. Genome Reduction and Microbe-Host Interactions Drive Adaptation of a Sulfur-Oxidizing Bacterium Associated with a Cold Seep Sponge. Gilbert JA, editor. mSystems [Internet]. 2017 Apr 21;2(2):e00184-16. Available from: http://msystems.asm.org/content/2/2/e00184-16.abstract

28. Karimi E, Slaby BM, Soares AR, Blom J, Hentschel U, Costa R. Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. FEMS Microbiol Ecol [Internet]. 2018 Apr 25;94(6). Available from: https://doi.org/10.1093/femsec/fiy074

29. Raina J-B, Fernandez V, Lambert B, Stocker R, Seymour JR. The role of microbial

motility and chemotaxis in symbiosis. Nat Rev Microbiol [Internet]. 2019;17(5):284–94. Available from: https://doi.org/10.1038/s41579-019-0182-9

30. Kelly JB, Carlson DE, Low JS, Rice T, Thacker RW. The relationship between microbiomes and selective regimes in the sponge genus Ircinia. bioRxiv [Internet]. 2020 Jan 1;2020.08.31.275412. Available from: http://biorxiv.org/content/early/2020/09/01/2020.08.31.275412.1.abstract

31. Kelly JB, Thacker RW. Microbiomes and host genetics provide evidence for ecological diversification among Caribbean members of the sponge genus Ircinia Nardo, 1833. bioRxiv [Internet]. 2020 Jan 1;2020.09.04.282673. Available from: http://biorxiv.org/content/early/2020/09/04/2020.09.04.282673.abstract

32. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics [Internet]. 2018 Sep 8;34(17):i884–90. Available from: https://doi.org/10.1093/bioinformatics/bty560

33. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics [Internet]. 2015 Jan 20;31(10):1674–6. Available from: https://doi.org/10.1093/bioinformatics/btv033

34. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ [Internet]. 2019 Jul 26;7:e7359–e7359. Available from: https://pubmed.ncbi.nlm.nih.gov/31388474

35. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al.

Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol [Internet]. 2017;2(11):1533–42. Available from: https://doi.org/10.1038/s41564-017-0012-7

36. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res [Internet]. 2015/05/14. 2015 Jul;25(7):1043–55. Available from: https://pubmed.ncbi.nlm.nih.gov/25977477

37. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J [Internet]. 2017;11(12):2864–8. Available from: https://doi.org/10.1038/ismej.2017.126

38. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, et al. Microbial species delineation using whole genome sequences. Nucleic Acids Res [Internet]. 2015/07/06. 2015 Aug 18;43(14):6761–71. Available from: https://pubmed.ncbi.nlm.nih.gov/26150420

39. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics [Internet]. 2019 Nov 15;36(6):1925–7. Available from: https://doi.org/10.1093/bioinformatics/btz848

40. Boyd JA, Woodcroft BJ, Tyson GW. Comparative genomics using EnrichM. in prep. 2019;

41. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics [Internet]. 2014/01/21. 2014 May 1;30(9):1236–40. Available from: https://pubmed.ncbi.nlm.nih.gov/24451626

42. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57:289–300.

43. Britton T, Anderson CL, Jacquet D, Lundqvist S, Bremer K. Estimating Divergence Times in Large Phylogenetic Trees. Syst Biol [Internet]. 2007 Oct 1;56(5):741–52. Available from: https://doi.org/10.1080/10635150701613783

44. Battistuzzi FU, Feijao A, Hedges SB. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol Biol [Internet]. 2004;4(1):44. Available from: https://doi.org/10.1186/1471-2148-4-44

45. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, et al. Geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. 2014 Aug;30(15):2216–8.

46. Fiore CL, Baker DM, Lesser MP. Nitrogen Biogeochemistry in the Caribbean Sponge, Xestospongia muta: A Source or Sink of Dissolved Inorganic Nitrogen? PLoS One [Internet]. 2013 Aug 26;8(8):e72961. Available from: https://doi.org/10.1371/journal.pone.0072961

47. Davy SK, Allemand D, Weis VM. Cell Biology of Cnidarian-Dinoflagellate Symbiosis. Microbiol Mol Biol Rev [Internet]. 2012 Jun 1;76(2):229 LP – 261. Available from: http://mmbr.asm.org/content/76/2/229.abstract

48. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, et al. Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable Alphaproteobacteria. Sci Rep [Internet]. 2019;9(1):1999. Available from: https://doi.org/10.1038/s41598-019-38737-x

49. Gauthier M-EA, Watson JR, Degnan SM. Draft Genomes Shed Light on the Dual Bacterial Symbiosis that Dominates the Microbiome of the Coral Reef Sponge Amphimedon queenslandica. Front Mar Sci [Internet]. 2016;3:196. Available from: https://www.frontiersin.org/article/10.3389/fmars.2016.00196

50. Moitinho-Silva L, Díez-Vives C, Batani G, Esteves AIS, Jahn MT, Thomas T. Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered metatranscriptomics. Isme J [Internet]. 2017 Mar 24;11:1651. Available from: https://doi.org/10.1038/ismej.2017.25

51. Sica D, Zollo F. Free amino acids in some sponges. Biochem Syst Ecol [Internet]. 1977;5(2):129–31. Available from: http://www.sciencedirect.com/science/article/pii/0305197877900412

52. Cohen SS (Seymour S, 1917-. guide to the polyamines. UN FAO of the, editor. Oxford University Press; 1998. p. : ill.; 26 cm.

53. Shah P, Swiatlo E. A multifaceted role for polyamines in bacterial pathogens. Mol Microbiol. 2008 Apr;68(1):4–16.

54. Taylor JA, Palladino G, Wemheuer B, Steinert G, Sipkema D, Williams TJ, et al. Phylogeny resolved, metabolism revealed: functional radiation within a widespread and divergent clade of sponge symbionts. ISME J [Internet]. 2020; Available from: https://doi.org/10.1038/s41396-020-00791-z

55. Reynolds D, Thomas T. Evolution and function of eukaryotic-like proteins from sponge symbionts. Mol Ecol [Internet]. 2016;25(20):5242–53. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13812

56. Germer J, Cerveau N, Jackson DJ. The Holo-Transcriptome of a Calcified Early

Branching Metazoan. Front Mar Sci [Internet]. 2017;4:81. Available from: https://www.frontiersin.org/article/10.3389/fmars.2017.00081

57. Chaban B, Hughes HV, Beeby M. The flagellum in bacterial pathogens: For motility and a whole lot more. Semin Cell Dev Biol. 2015 Oct;46:91–103.

58. Ramos HC, Rumbo M, Sirard J-C. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol. 2004 Nov;12(11):509–17.

59. Miyake K. Innate recognition of lipopolysaccharide by Toll-like receptor 4-MD-2. Trends Microbiol. 2004 Apr;12(4):186–92.

60. Medzhitov R, Preston-Hurlburt P, Janeway CA. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature [Internet]. 1997;388(6640):394–7. Available from: https://doi.org/10.1038/41131

61. Kelly SR, Garo E, Jensen P, Fenical W, Pawlik J. Effects of Caribbean sponge secondary metabolites on bacterial surface colonization. Aquat Microb Ecol. 2005 Sep 6;40:191–203.

62. Lavy A, Keren R, Yahel G, Ilan M. Intermittent Hypoxia and Prolonged Suboxia Measured In situ in a Marine Sponge. Front Mar Sci [Internet]. 2016;3:263. Available from: https://www.frontiersin.org/article/10.3389/fmars.2016.00263

63. Weber M, Mendola D, Hoffmann F, Beer D De. Heterogeneous oxygenation resulting from active and passive flow in two Mediterranean sponges , Dysidea avara and Chondrosia reniformis. Limnol Oceanogr. 2010;55(3):1289–300.

64. Belin BJ, Busset N, Giraud E, Molinaro A, Silipo A, Newman DK. Hopanoid lipids: from membranes to plant–bacteria interactions. Nat Rev Microbiol [Internet]. 2018;16(5):304–15. Available from: https://doi.org/10.1038/nrmicro.2017.173

65. Silipo A, Vitiello G, Gully D, Sturiale L, Chaintreuil C, Fardoux J, et al. Covalently linked hopanoid-lipid A improves outer-membrane resistance of a Bradyrhizobium symbiont of legumes. Nat Commun [Internet]. 2014;5(1):5106. Available from: https://doi.org/10.1038/ncomms6106

66. Kulkarni G, Busset N, Molinaro A, Gargani D, Chaintreuil C, Silipo A, et al. Specific Hopanoid Classes Differentially Affect Free-Living and Symbiotic States of Bradyrhizobium diazoefficiens. Ausubel FM, editor. MBio [Internet]. 2015 Oct 30;6(5):e01251-15. Available from: http://mbio.asm.org/content/6/5/e01251-15.abstract

67. Coutinho AE, Chapman KE. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. Mol Cell Endocrinol [Internet]. 2010/04/14. 2011 Mar 15;335(1):2–13. Available from: https://pubmed.ncbi.nlm.nih.gov/20398732

68. Wyle FA, Kent JR. Immunosuppression by sex steroid hormones. The effect upon PHA- and PPD-stimulated lymphocytes. Clin Exp Immunol [Internet]. 1977 Mar;27(3):407–15. Available from: https://pubmed.ncbi.nlm.nih.gov/862230

List of Supplementary Materials:

Tables S1-S7

Figure captions:

Figure 1. Phylogenetic ANOVAs of genomic characteristics show that they are significantly different between SDPs and pelagic MAGs after correcting for phylogenetic relatedness and multiple comparisons (fdr): genome size ($F_{1,1106}$=68.218, p=0.0488), coding density ($F_{1,1106}$= 188.62, p=0.0005), number of proteins ($F_{1,1106}$= 66.227, p= 0.0488), GC-content ($F_{1,1106}$= 292.44, p= 0.00025). Bacteria phyla are annotated by the alternating gray and black sections of the outer circle.

Figure 2. A: Scatterplot comparing relationships prokaryotic genome sizes to GC content and **B**, genome size to coding density for SDPs and pelagic microbes. Coding density is reported as the percentage of basepairs that are included in a coding region. R^2 and statistical significance is reported for linear models inferred for the pelagic and SDP MAGs. Plots to the left of the y-axes and below the x-axes are kernel density estimation plots. Degrees of freedom for SDP MAGs = 477, pelagic MAGs = 627. **C:** Bar chart comparing average number of proteins predicted from SDP and pelagic MAGs (phylogenetic ANOVA reported in the caption of Figure 1). Asterisks here and elsewhere in figures denote significance.

Figure 3. Plots depicting genes that are enriched in pelagic (i.e. depleted in SDPs) or enriched in SDPs. Each circle corresponds to a gene that was determined as being differentially abundant in either SDPs or pelagic MAGs in terms of normalized average gene copy number (see methods). The size of each circle is scaled to the log-transformed ratio of the gene calculated as (avg. copy number SDP) / (avg. copy number pelagic) for genes under the y=x diagonal and (avg. copy number pelagic) / (avg. copy number SDP) for genes above the y=x diagonal. Non-transformed

ratios are reported in the legend. The ELP graph was constructed from the Interproscan annotations; all others were constructed using the KO annotations produced from enrichm.

Figure 4. Representative cassettes of steroid biosynthesis genes are depicted. Genes in parallel denote common alternative combinations for a clade. TM7SF2, ERG24, Delta14-sterol reductase [EC:1.3.1.70] (K00222); SMT1, ERG6, sterol 24-C-methyltransferase [EC:2.1.1.41] (K00559); LSS, ERG7, lanosterol synthase [EC:5.4.99.7] (K01852); CYP51, sterol 14alpha-demethylase [EC:1.14.14.154 1.14.15.36] (K05917); FDFT1, farnesyl-diphosphate farnesyltransferase [EC:2.5.1.21] (K00801); and ERG2, C-8 sterol isomerase [EC:5.-.-.-] (K09829) are components of the steroid biosynthesis pathway (ko00100). Dihydroflavonol-4-reductase [EC:1.1.1.219] (K00091), an oxidoreductase acting on the CH-OH group of donors with NAD+ or NADP+ as acceptor, is not assigned to a pathway and was included here and in Figure 2 due to its consistent association with genes belonging to the steroid biosynthesis (ko00100) pathway and its likely relevant biochemical activity. 'HP' abbreviates hypothetical protein. Arrows denote direction of the coding sequence. Genes are colored to match the legend of the 'Steroid Biosynthesis' plot of Figure 2. ERG2 is uncolored as it was removed from the data producing Figure 2 due to being in low proportions (<10%) of both the SDP and pelagic MAG datasets. Note: genes are not drawn to scale by length.

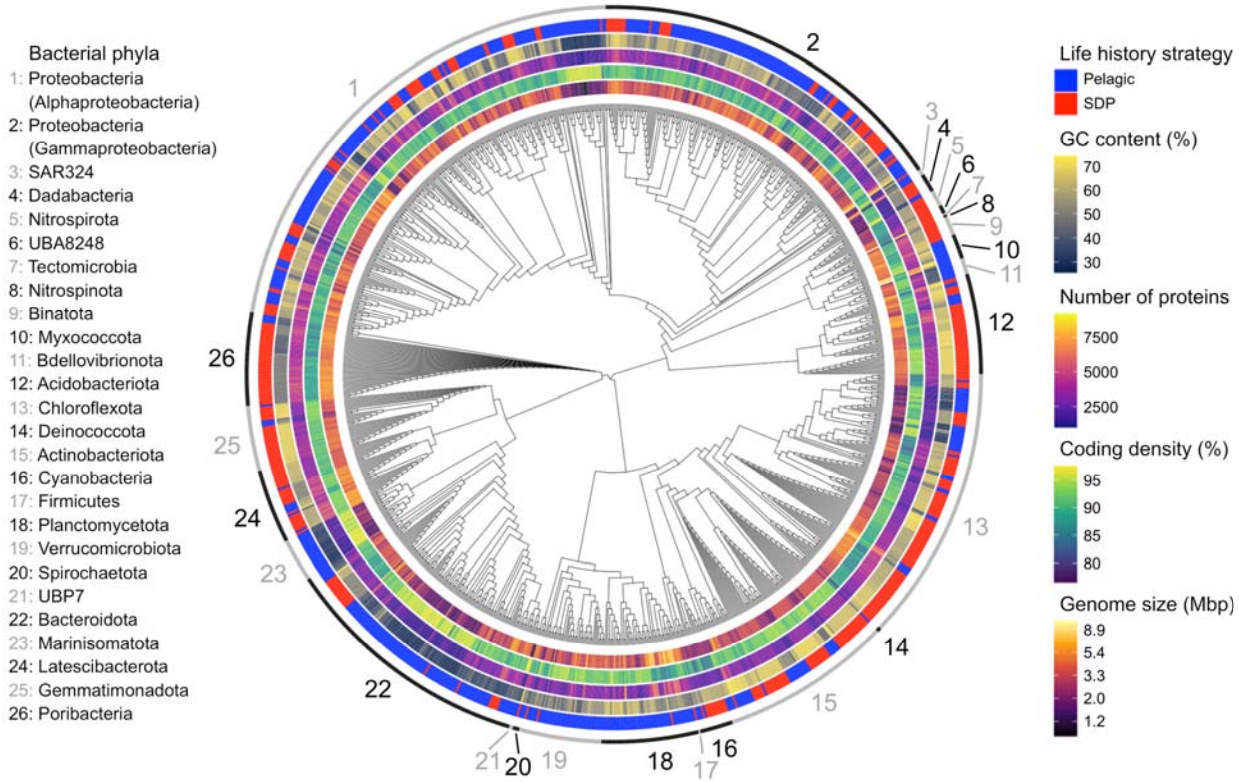
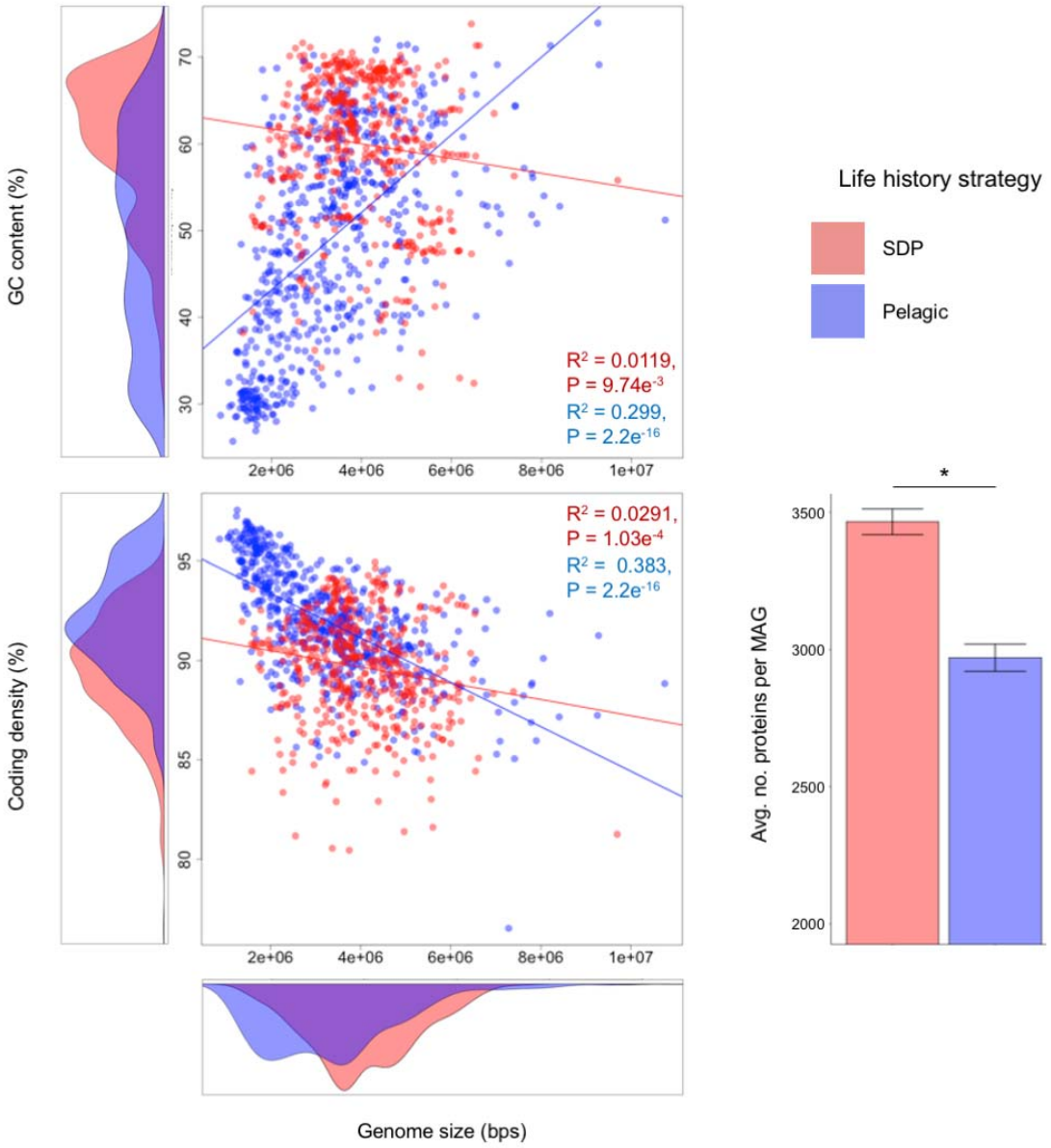


Figure 1. The phylogenetic distribution of endogenous genomic traits in bacteria is independent of life history strategy.



592

593 **Figure 2. Genomes of SDPs are more energetically expensive than pelagic genomes.**

594

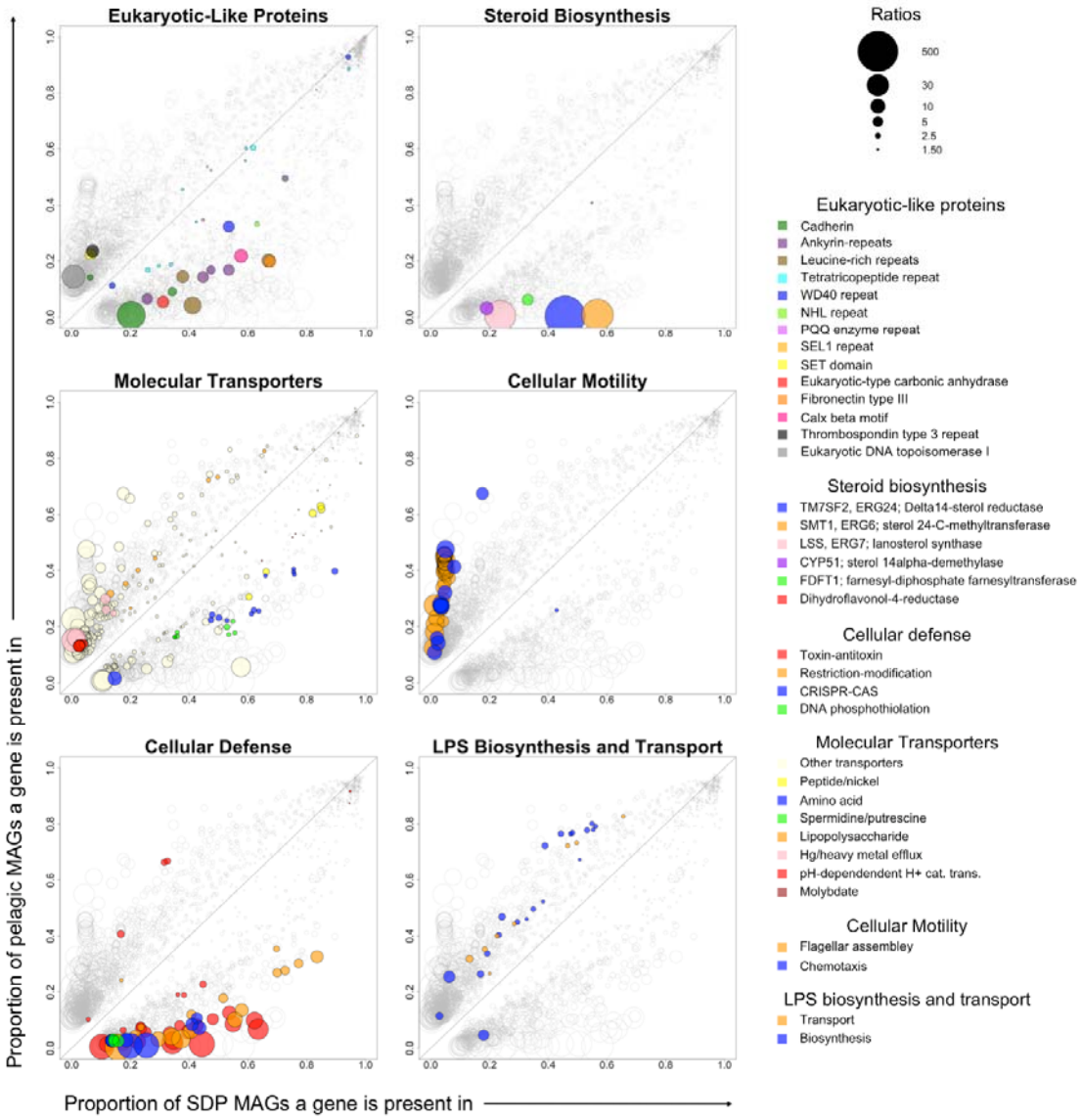


Figure 3. SDPs are well-defended against foreign DNA, are equipped to promote their residence in host sponges, and have reduced pathogenicity.

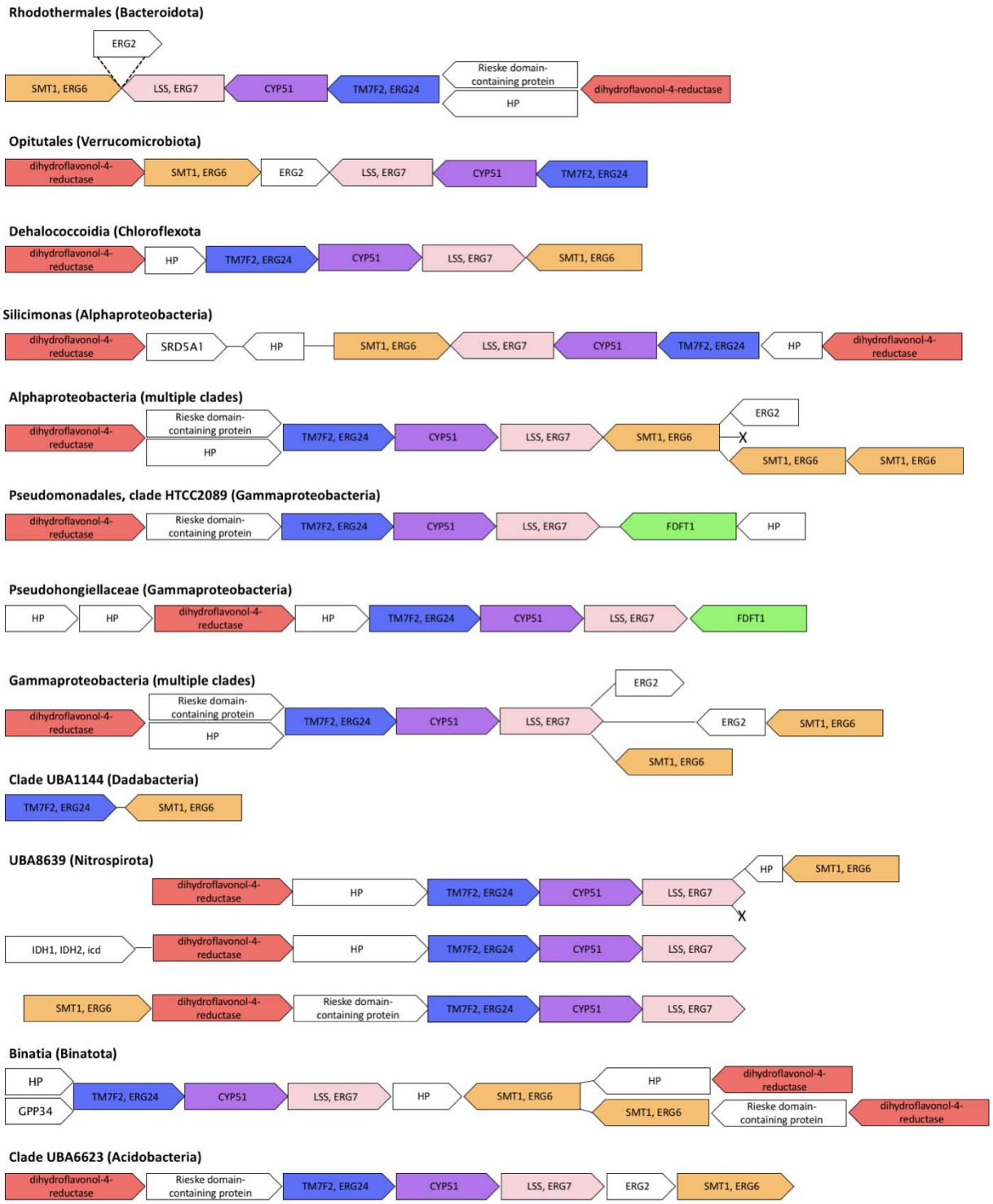


Figure 4. SDPs contain partial eukaryotic steroid biosynthesis pathways.