Title: Comparative metagenomics evidence distinct trends of genome evolution in 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.

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.

Abstract (168 words): 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. We compared metagenome-assembled genomes of sponge-dwelling and pelagic bacteria sourced from broad phylogenetic and geographic samplings. We discovered that sponge dwelling bacteria 1) have more energetically expensive genomes, reflecting a release from nutrient limitation, 2) share patterns of depletion and enrichment for functional categories of genes that evidence a non-pathogenic strategy to

virulence, and 3) harbor fewer prophages of smaller average lengths, although with higher densities of auxiliary genes that could be adaptive for the host bacteria. Given our observations, we propose two new evolutionary models, one that describes how alternatives to pathogenicity might develop in symbionts exhibiting mixed-mode transmission, and another that describes how density-dependent viral exposure can result in adaptive acquisition of novel genetic functions in sponge-dwelling bacteria. Collectively, these results illuminate the evolutionary pathways that might have first enabled microbes to stably associate with multicellular eukaryotes.

Main Text (3395 words):

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 the plummeting cost of high throughput sequencing, key statistical advancements, and ballooning CPU efficiencies have driven a renaissance in microbiology focused on the genetic underpinnings of microbial physiologies. Consequently, science has gained unprecedented insights into prokaryotic defense systems, molecular machinery related to virulence, 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 through contributions from both the environmental microbial pool as well as through parental inheritance. When both patterns of acquisition are observed the phenomenon is termed mixed-mode transmission; current evidence supports this mode of transmission as being the dominant form in SDPs (9). The prevalence of mixed-mode transmission in SDPs is at odds with the longevity of the sponge-microbial symbiosis, as conventional evolutionary theory states that any introduction of horizontal (environmental) transmission would undermine the stability of the symbiosis as these microbes would evolve towards higher levels of virulence, including pathogenicity (10–12). 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 the evolutionary paths of SDPs to explain what the consequences of sponge-inhabitation have been on SDP genome evolution, and how these patterns of evolution are involved in biological communication with the host. We use a comparative metagenomic analysis of SDPs and pelagic microbes that focuses on three 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 (13–16). Whether nutrient-rich sponge interiors release SDPs from the energetic shackles of oligotrophy remains undetermined. Second, we investigate the shared enrichment patterns of individual genes in SDPs with the goal of elucidating the forces of selection that are common across SDP taxa in multiple host species. While the presence of biotic and abiotic influences that are common across the interiors of sponge taxa is easy to conceptualize, it is largely unknown how they drive adaptation in SDPs at the genic level. Lastly, we examine the contributions of viruses to SDPs and how they help shape finesses of SDPs in the sponge interior. We focus on lysogenic bacteriophages since they are incorporated into the prokaryotic genome and thus likely code for a greater degree of evolutionary permanence relative to lytic phages. While bacteriophages have been identified as vectors of adaptive gene contents for prokaryotes in a handful of host sponge species (17), we have yet to tease apart the specific contributions of lysogenic phages and describe their contributions to SDP biology across global geographic scales and throughout the host and prokaryotic phylogenies.

Materials and Methods

DNA Extractions and Sequencing

DNA was extracted from two specimens that had been stored in 95% EtOH from each of the three nominal *Ircinia* species and seven growth forms (*Ircinia* spp. nov., Kelly et al. 2020 *in review*), resulting in a sample size of 20, using the Molzyme Molysis DNA isolation kit, which depletes host DNA in a sample resulting in an extraction that is enriched for microbial DNA.

DNA sequencing was performed 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 (18) 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 well as the control PhiX reads. Metagenomic contigs were assembled independently for each specimen using Megahit v1.2.9 (19) and subsequently binned using MetaBAT 2 v2:2.15 (20) 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 (21). The qualities of the refined bins were assessed using checkm v1.1.2 (22); only bins with qualities equal to or higher than 40, calculated as genome completeness – 5x contamination (21), 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 (23). 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 (23). Prior to quality control and dereplication, the pelagic dataset consisted of 2307 MAGs (NCBI bioproject PRJNA391943). The SDP MAG dataset was produced by combining our Caribbean *Ircinia* MAG dataset with 689 publicly available draft genomes of SDPs. Prior to dereplication and further QC, the SDP

dataset consisted of 1580 MAGs. The SDP and Tara Oceans datasets were independently dereplicated using drep v2.5.2 (24) with a cutoff of 96.5% identity, an estimate of maximum prokaryotic intraspecific genomic divergence backed by empirical evidence (25). 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 (26). Prophage genomes were identified from our dataset using ProphET v0.5.1 (27).

Protein predictions and subsequent KO annotations were performed on the MAGs (including the prophage sequences) and independently on the isolated prophage genomes using Enrichm v0.5.0 (28). Domain PFAM annotations were performed on the MAG and prophage proteins using Interproscan v5.39-77 (29). 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 MAGs in one category or the other (i.e. SDP vs free-living prokaryotes), 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 (30) 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.

We tested for statistical significance in the average number of prophages per MAG, average size of each prophage in bps, and the average number of auxiliary genes per prophage (normalized by the total number of genes predicted for each prophage) within functional categories using permutational t-tests, each with 10,000 iterations. We adjusted p-values for the auxiliary gene analysis using the FDR procedure as above.

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 ANOCOVA. We tested for significance in genome size, GC content, coding density, the total number of proteins, and the average number of prophages between SDPs and pelagic MAGs within a phylogenetic framework. Since the Archaea were represented by only 15 MAGs, we restricted this component of the analysis to bacteria. To perform this analysis, we estimated an ultrametric bacterial tree with the software PATHd8 (31) using the GTDB-Tk bacterial phylogeny as the input. Node dating was calibrated using the divergence times estimates of Battistuzzi et al. (32) 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 (33), 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, gene content, and prophage characteristics 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 (23). 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 (Tables 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 (Tables S3). We report our results below for analyses performed on the bacterial MAGs, as only 15 archaeal MAGs were recovered (Table S4). 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 of SDPs mirror those of free-living oceanic bacteria, 2) gene contents segregate SDP and pelagic bacteria, and 3) lysogenic phages confer adaptive advantages to their bacterial hosts that facilitate their residence in sponges.

Release from nitrogen limitation allows SDPs to evolve 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).

We hypothesize that differences in energy availability between the two life history strategies can 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 (13–16). 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. Excretion of nitrogenous compounds such as ammonium by the host sponge (34) could alleviate nitrogen limitation and enable the maintenance of larger genomes with higher GC contents in SDPs.

Gene enrichment patterns suggest SDPs have evolved towards lower pathogenicity and are equipped to tolerate and exploit the internal conditions of the sponge body

The contrasting selective regimes of pelagic and sponge-associated life history strategies are also evident at the level of individual genes. SDPs are depleted for cellular motility and lipopolysaccharide (LPS) biosynthesis/transport genes and are enriched for several eukaryotic-like domain-containing proteins (ELPs) (Fig. 3). These ELPs include the previously reported fibronectin type-III, ANK, LRR, TPR, NHL, WD40, cadherin, and PQQ domains (35–37) and two new ELPs: calx-beta motifs and eukaryotic-type carbonic anhydrase. Interestingly, some

TPRs, which are reported in other studies as being unanimously enriched in SDPs, were depleted in SDPs. Notably, SDPs were also enriched for genes involved in steroid biosynthesis. Steroid biogenesis genes have been observed before in SDPs of the sponge *Vaceletia sp*., where they represented an incomplete eukaryotic steroid biosynthesis pathway (KEGG pathway ko00100) that lacked the machinery to produce steroid precursor molecules. Instead, several complementary components of the pathways were discovered in the *Vaceletia sp*. genome, suggesting that steroid precursors may originate in the sponge host (38). In many prokaryotic clades in the present study, these genes are arranged in cassettes in SDPs (Fig. S1), suggesting that they might be part of a yet undescribed steroid biosynthesis pathway that is present across multiple prokaryotic phyla.

Given these observations, the patterns of virulence—which we define here as being composed of both pathogenicity and the promotion of residence in the host—are likely different between SDPs and pelagic bacteria. Motility and LPS genes are directly involved in pathogenesis (39); thus, a potential evolutionary outcome of the reduced frequency of motility and LPS genes might be one of reduced pathogenicity in SDPs. At face value, this observation appears to be at odds with conventional evolutionary theories that state 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 (10,11). 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 (12). Why then do we detect a depletion of pathogenic mechanisms in prokaryotes that are likely undergoing mixed-modes of

transmission in sponges (9)? We hypothesize that the answer might lie in the enrichment pattern of steroid biosynthesis genes in SDPs, which 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 (5,40) and hypoxic conditions (41,42) that are present in the interior of the host, as is the case in other symbiotic microbes such as the legume-inhabiting *Bradyrhizobium* (43–45). In SDPs where steroids have supplanted LPS in their role of maintaining the integrity of the cellular membrane, the growth rates of SDPs could be supplemented by vertical transmission through the larvae of their hosts. By circumventing damage to their hosts, SDPs could exhibit high rates of virulence by virtue of promoting their residence in the hosts while decreasing the pathogenic component.

Previous studies have concluded that molecular transporters (e.g. ABC transporters) are enriched in SDPs (35,37,46). Our results demonstrate that this finding is not universal (Fig. 3). Instead, we found that several classes of molecular transporters are enriched or depleted in SDPs, depending on what molecules are being transported. For example, in addition to the aforementioned LPS transporters, SDPs are depleted for Hg^{+}/heavy metal efflux proteins and the *pha* family of pH-dependent H^{+} cation transporters. Conversely, SDPs are enriched for amino acid transporters. Early work discovered that concentrations of free amino acids are relatively high in sponges, which is thought to play a role in osmoregulation (47). Microbial symbionts could be capitalizing on this pool of bioavailable nutrients to supplement their own metabolism. Additionally, SDPs are enriched for spermidine/putrescine transporters. Whether these compounds are supplementing the primary nutrition of the SDPs is difficult to discern. However, many of the spermidine/putrescine transporters were found in close proximity to steroid gene cassettes (data not shown), which might imply that they play a role in steroid biogenesis.

SDPs display patterns of prophage infection evidencing a capacity to select which viral elements are incorporated into their genomes

The nature of the associations of host prokaryotic cells with prophages differs between SDPs and pelagic microbes, as SDPs harbor fewer phages with smaller average sizes (Fig. 4). However, SDP prophages have higher densities of auxiliary genes that could be supplementing the functions of endogenous genes, including defense and resistance, primary and secondary metabolism, and genes involved in quorum sensing, cell growth, and biofilm formation (the 'other' category) (Fig. 4, auxiliary genes listed Table S7). We also found a non-significant enrichment for ELPs in SDP prophages that included fibronectin type III, TPRs, and ANKs. Jahn et al. (17) reported ANK-incorporated phages ('Ankyphages') in SDPs that contained auxiliary gene content thought to supplement prokaryotic host functioning (17). The same study also confirmed experimentally that Ankyphage-containing prokaryotes experience significantly less phagocytosis from their hosts. Here, we uncovered two new ELPs for which SDP phages are enriched and confirm that all three are incorporated into the host prokaryotic genome, potentially acting to bestow a higher degree of permanence to the fitness advantage resulting from ELPs in addition to other auxiliary gene content.

The combination of smaller average prophage sizes in SDPs and higher auxiliary gene content means that SDPs are likely having to spare less energy for a greater fitness return from viral genes that could supplement host metabolic functioning and host immune evasion relative to the prophages in pelagic MAGs. Additionally, the features of SDP prophages suggest that the incorporation of these viral elements into SDP genomes might undergo stronger host selectivity relative to the prophages of pelagic prokaryotes. We hypothesize that this selectivity by bacteria

on lysogenic phages, and ultimately the insertion of phages into SDP genomes that confer a higher fitness benefit, has evolved via a density-dependent relationship. Sponge microbiomes can have exceptionally high prokaryotic cellular densities that can reach 10^9 cells/cm^3, four orders of magnitude greater than the estimates of 0.5-5x10^5 cells/cm^3 in seawater (48,49). As such, social distancing may not be an option for SDPs. Under this scenario, cellular defense mechanisms that allow the cell to police incoming viral DNA might be evolutionarily favored. This expectation is met in our data, as SDPs are enriched for restriction-modification systems, CRISPR-CAS9, and DNA phosphothiolation genes (Fig. 3). However, higher exposure rates to viruses could also translate into higher rates of introduction to novel, potentially adaptive gene content (50–52). The cellular defenses of SDPs might allow them exclude prophages that are large and have a lower adaptive/non-adaptive ratio of gene content. As adaptive gene content in the form of ABC transporters, metabolism genes, or cellular defenses could mimic the prokaryotes' genetic composition, the higher ratio of these genes to non-adaptive viral content might result in a lower rate of destruction by endogenous prokaryotic defenses.

Conclusion

We discovered that SDP genomes are distinguished from the genomes of pelagic microbes across the categories of comparisons that we made, thus providing evidence that the selective pressures associated with living in sponges have driven shared genomic adaptations independently in multiple prokaryotic phyla. The differences in genomic features could, in part, be shaped by the alleviation of nutrient limitation associated with inhabiting the sponge mesohyl. The contrasting abiotic and biotic regimes between life history strategies have also resulted in the evolution of gene contents that help symbiotic bacteria optimize exploitation and tolerance of the

host environment. Additionally, the incorporation of prophages into SDP genomes has followed a different evolutionary path relative to pelagic prokaryotic genomes in that the SDP prophages are more restrictive with regard to size and their frequency of incorporation, suggesting that SDPs have evolved a capacity to select which lysogenic phages are allowed to persist in the prokaryotic genome. Altogether, these results provide insights into the shared evolutionary patterns associated with transitioning to inhabitation of multicellular eukaryotic hosts and elucidate how stability is maintained in aquatic host-microbial symbioses across geologic timescales.

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

11. 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

12. 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+.

13. 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

14. 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

15. 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

16. 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.

17. Jahn M, Arkhipova K, Markert S, Stigloher C, Lachnit T, Pita L, et al. A Phage Protein Aids Bacterial Symbionts in Eukaryote Immune Evasion. Cell Host Microbe. 2019 Sep 24;26.

18. 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

19. 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

20. 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

21. 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

22. 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

23. 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

24. 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

25. 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

26. 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

27. Reis-Cunha JL, Bartholomeu DC, Manson AL, Earl AM, Cerqueira GC. ProphET, prophage estimation tool: A stand-alone prophage sequence prediction tool with self-

updating reference database. PLoS One [Internet]. 2019 Oct 2;14(10):e0223364. Available from: https://doi.org/10.1371/journal.pone.0223364

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

29. 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

30. 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.

31. 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

32. 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

33. 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.

34. Brusca RC, Moore W, Shuster SM, Haver N. Invertebrates. 3rd ed. Sunderland, Massachusettes: Sinauer Associates, Inc.; 2016.

35. 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

36. 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

37. 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

38. 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

39. 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.

40. 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.

41. 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

42. 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.

43. 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

44. 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

45. 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

46. 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

47. 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

48. Hentschel U, Usher KM, Taylor MW. Marine sponges as microbial fermenters. FEMS Microbiol Ecol. 2006 Feb;55(2):167–77.

49. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A [Internet]. 1998 Jun 9;95(12):6578–83. Available from: https://pubmed.ncbi.nlm.nih.gov/9618454

50. Tsao Y-F, Taylor VL, Kala S, Bondy-Denomy J, Khan AN, Bona D, et al. Phage Morons Play an Important Role in Pseudomonas aeruginosa Phenotypes. J Bacteriol [Internet]. 2018 Oct 23;200(22):e00189-18. Available from: https://pubmed.ncbi.nlm.nih.gov/30150232

51. Fortier L-C, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence [Internet]. 2013/04/23. 2013 Jul 1;4(5):354–65. Available from: https://pubmed.ncbi.nlm.nih.gov/23611873

52. Brüssow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev [Internet]. 2004 Sep;68(3):560–602. Available from: https://pubmed.ncbi.nlm.nih.gov/15353570

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).

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. A and B: Bar charts depicting lower average number of prophages per MAG (Welsh t-statistic = 6.172, $p < 2.2e^{-16}$) and smaller average size of prophages in SDPs relative to pelagic MAGs (Welsh t-statistic = -3.6977, p = 0.0024). **C:** Bar charts depicting normalized abundances of genes on prophages within the categories of defense and resistance, ELP, primary and secondary metabolism, ABC transporters and membrane trafficking, others (see main text for descriptions), and total counts. Permutational t-tests for each category follow with fdr-adjusted p-values: defense and resistance (Welsh's t = 1.620, p = 0 0.0234), primary and secondary metabolism (Welsh's t = 2.0367, p = 0.0144), ABC transporters and membrane trafficking (Welsh's t = 0.748, p = 0.312), other (Welsh's t = 1.743, p = 0.0224), ELP Welsh's t = 1.234 p = 0.0934), total (Welsh's t = 2.801, p < 2.2e-16).

Figure S1. 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.

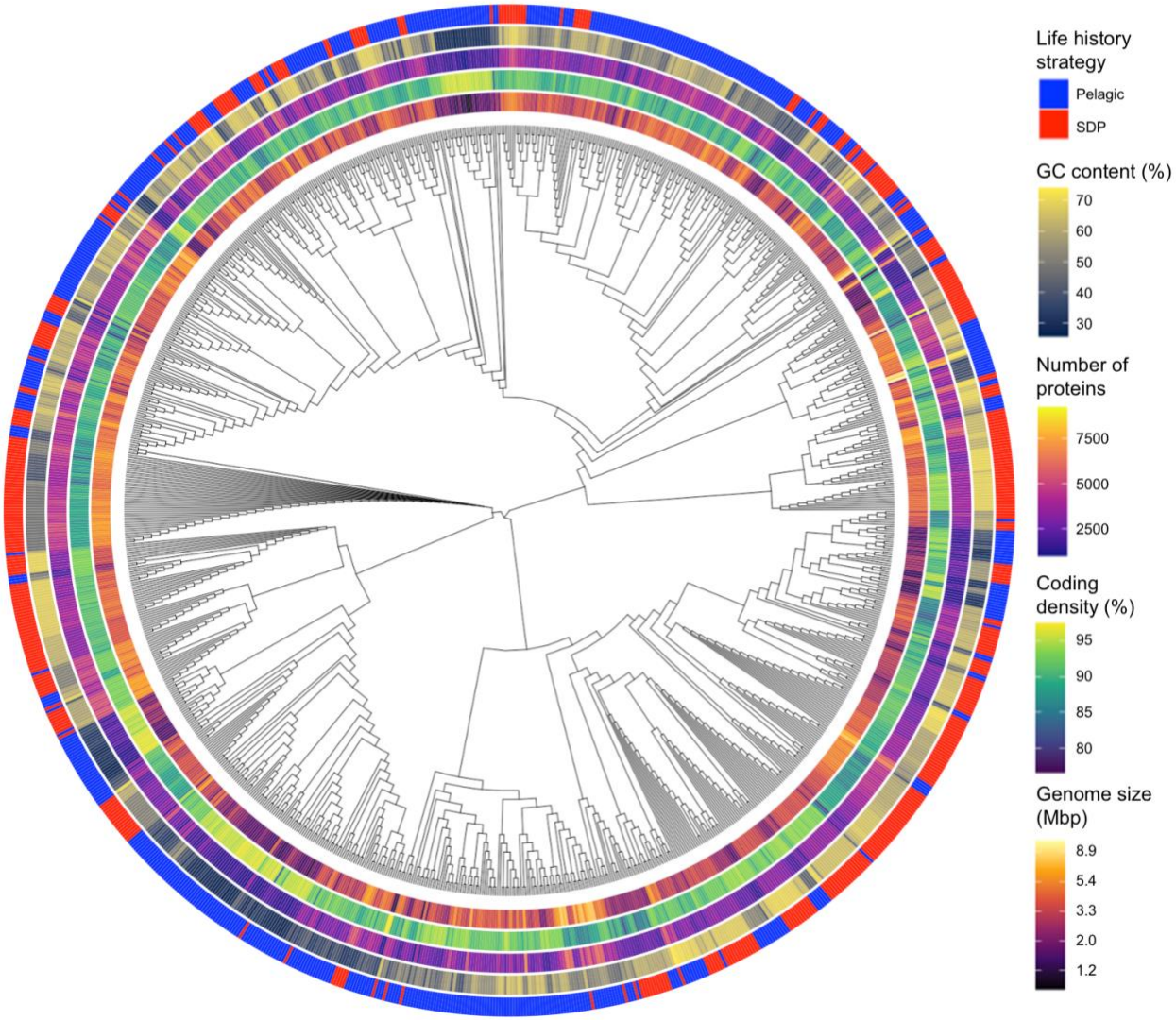

Figure 1. The phylogenetic distribution of endogenous genomic traits in bacteria is independent of life history strategy.

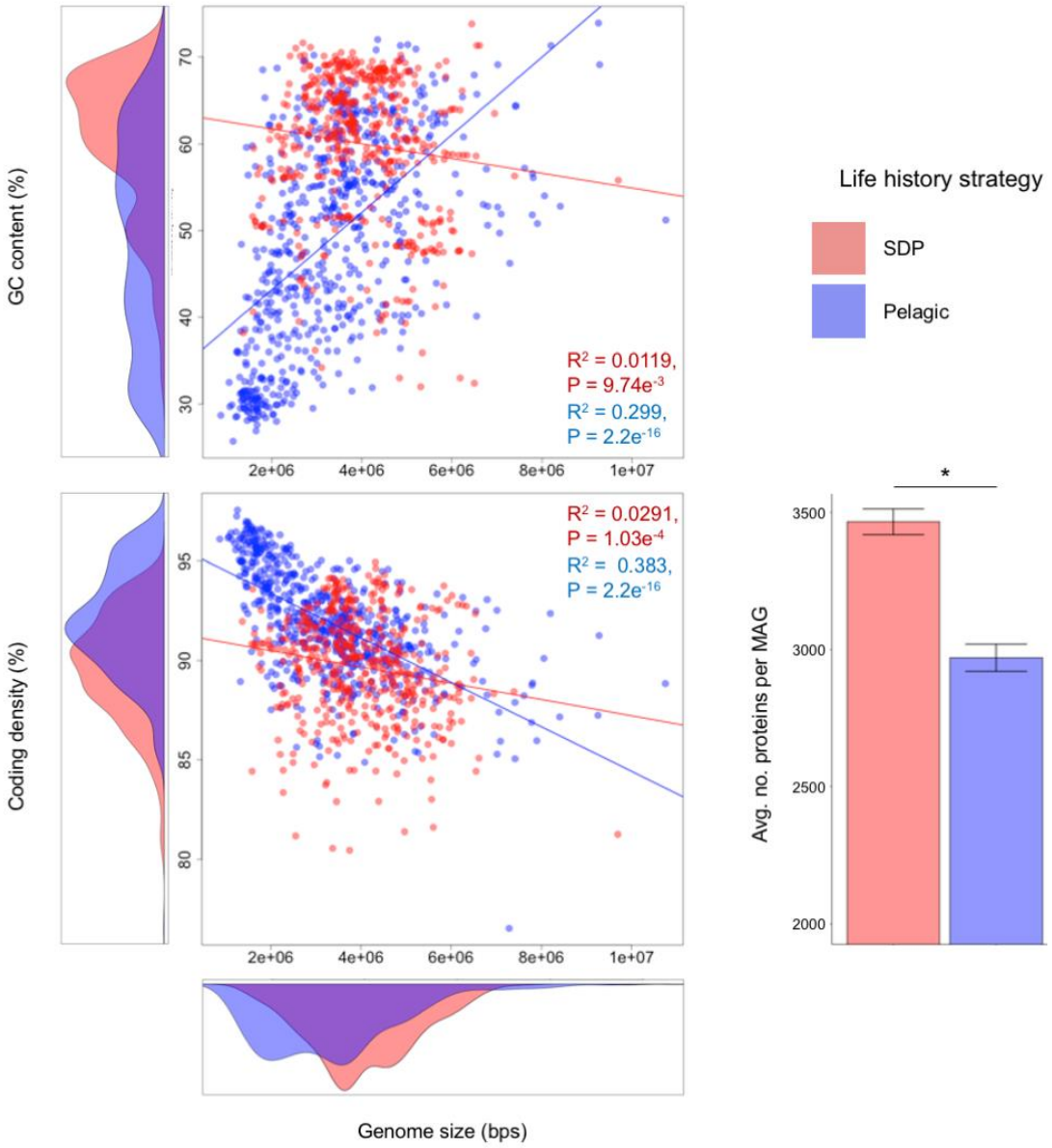

Figure 2. Genomes of SDPs are more energetically expensive than pelagic genomes.

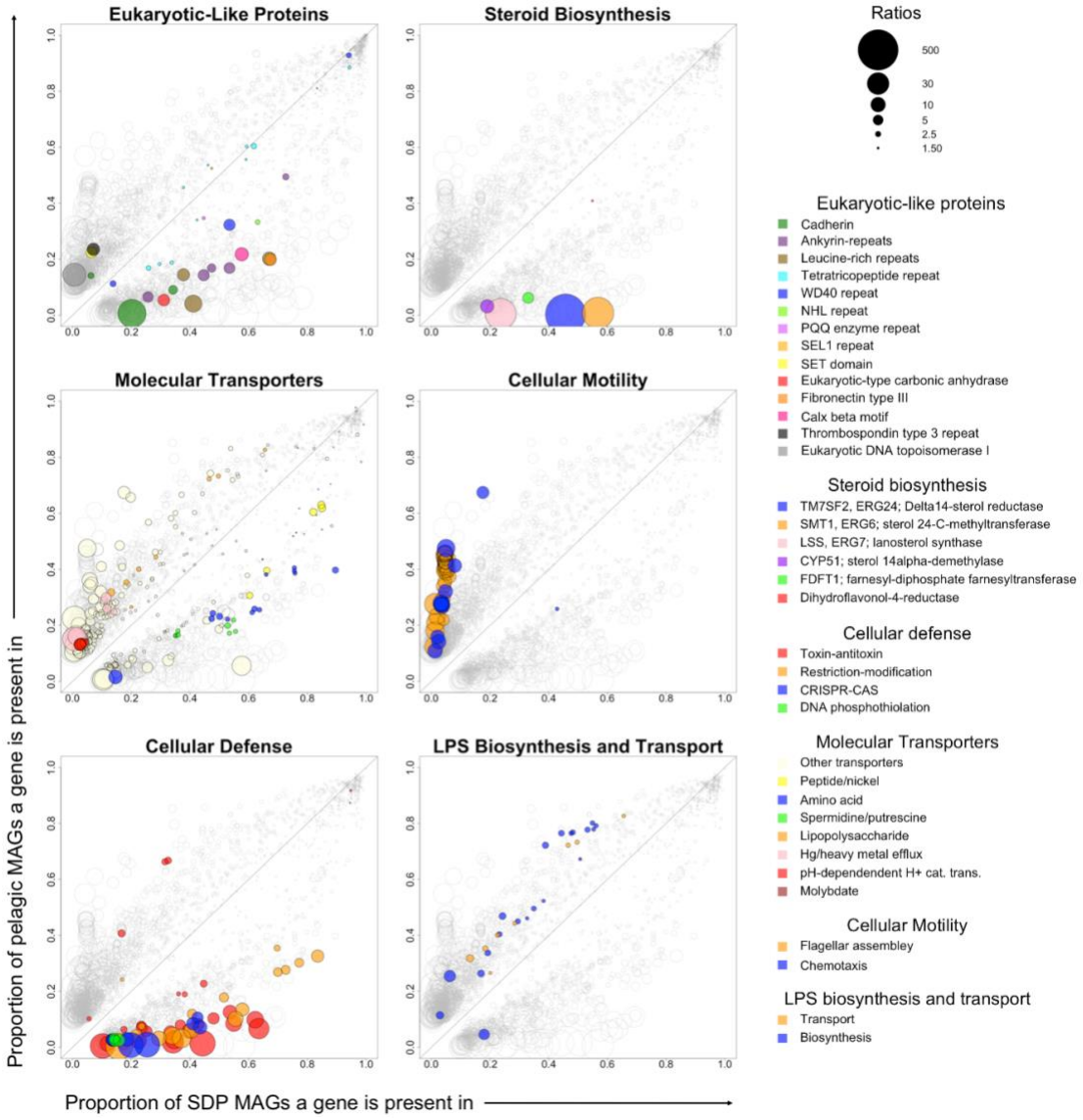

Figure 3. SDPs are well-defended against foreign DNA, are equipped to promote their residence in host sponges, and have reduced pathogenicity.

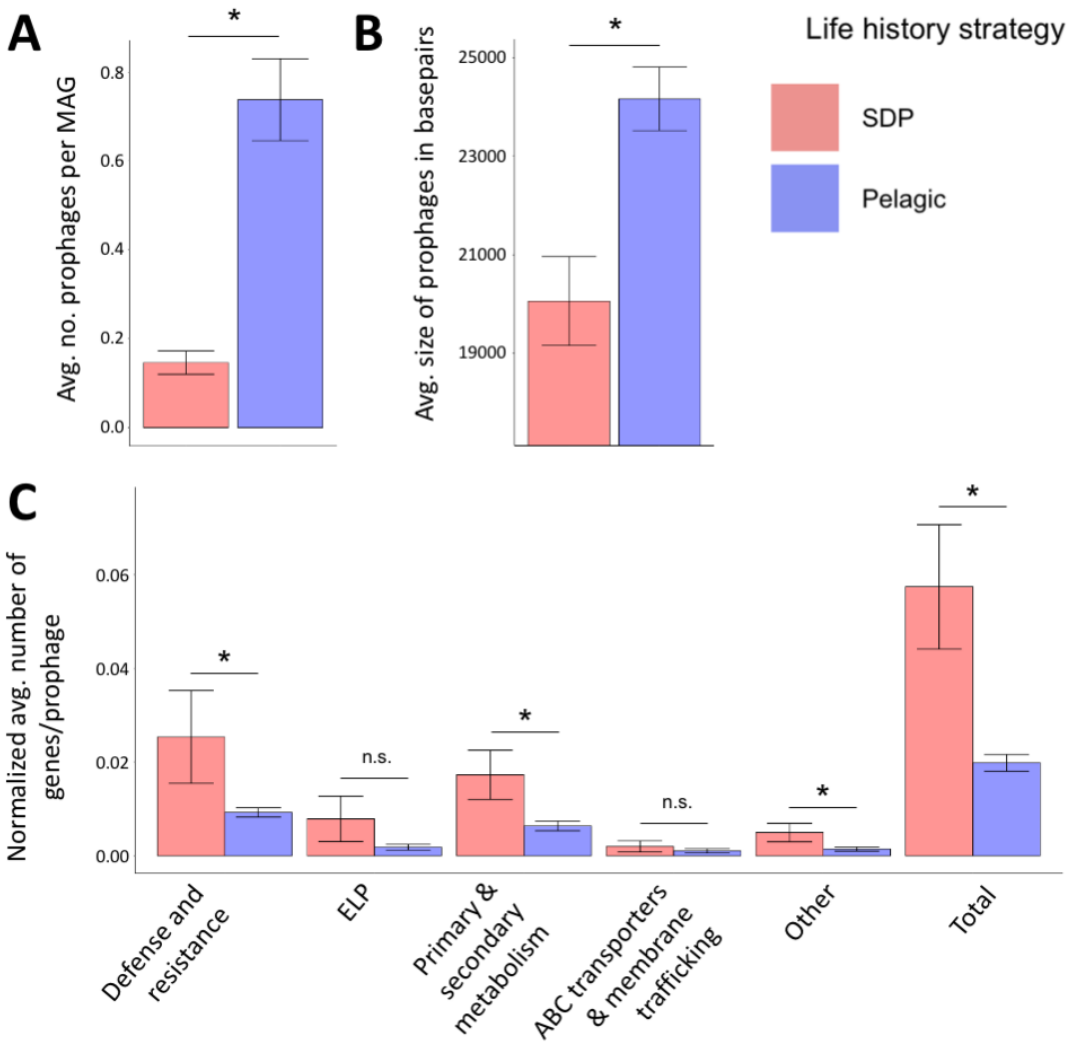

Figure 4. SDP prophages bestow a higher adaptive advantage to the host prokaryote.

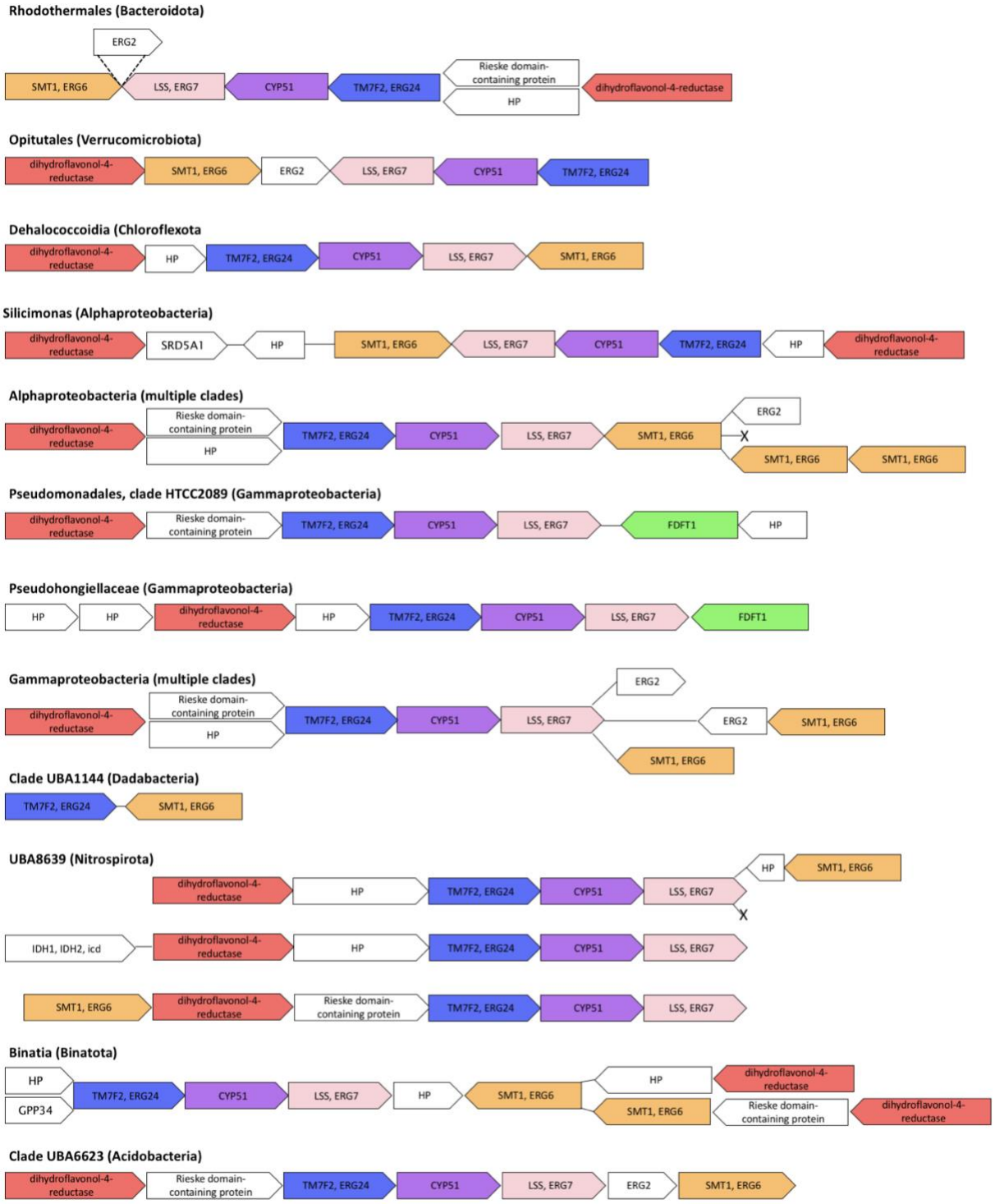

Figure S1. SDPs contain partial eukaryotic steroid biosynthesis pathways.