Title: Comparative metagenomics evidence distinct evolutionary 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

47 ago (2). Sponge microbiomes can account for up to 40% of the sponge's volume (3) and
48 routinely comprise thousands of microbial taxa from dozens of eubacterial and archaeal phyla
49 (4). The taxonomic richness of sponge-dwelling prokaryotes (SDPs) is met by their physiological
50 diversity as numerous autotrophic and heterotrophic nutritional strategies have proliferated in
51 these microbes in addition to their capacity to synthesize a multitude of bioactive secondary
52 metabolites (5–8). Sponge microbiomes are assembled through contributions from both the
53 environmental microbial pool as well as through parental inheritance. When both patterns of
54 acquisition are observed the phenomenon is termed mixed-mode transmission; current evidence
55 supports this mode of transmission as being the dominant form in SDPs (9). The prevalence of
56 mixed-mode transmission in SDPs is at odds with the longevity of the sponge-microbial
57 symbiosis, as conventional evolutionary theory states that any introduction of horizontal
58 (environmental) transmission would undermine the stability of the symbiosis as these microbes
59 would evolve towards higher levels of virulence, including pathogenicity (10–12). The mystery
60 stands: among the distantly related SDPs that possess such metabolic diversity, and whose
61 transmission mode presents an evolutionary paradox, what features do they have in common that
62 account for their shared success within the environments of sponge bodies and the stability of the
63 symbiosis?

64 The current study seeks to illuminate the evolutionary paths of SDPs to explain what the
65 consequences of sponge-inhabitation have been on SDP genome evolution, and how these
66 patterns of evolution are involved in biological communication with the host. We use a
67 comparative metagenomic analysis of SDPs and pelagic microbes that focuses on three general
68 genomic facets. The first is the evolution of general genomic characteristics and encompasses
69 genome size and, particularly, how it scales with protein-coding and GC contents. In the open

70 ocean, these features are intrinsically linked to nutrient availability, which drives genome
71 streamlining of pelagic prokaryotic genomes and A-T biases in nucleotide composition (13–16).
72 Whether nutrient-rich sponge interiors release SDPs from the energetic shackles of oligotrophy
73 remains undetermined. Second, we investigate the shared enrichment patterns of individual
74 genes in SDPs with the goal of elucidating the forces of selection that are common across SDP
75 taxa in multiple host species. While the presence of biotic and abiotic influences that are
76 common across the interiors of sponge taxa is easy to conceptualize, it is largely unknown how
77 they drive adaptation in SDPs at the genic level. Lastly, we examine the contributions of viruses
78 to SDPs and how they help shape finesses of SDPs in the sponge interior. We focus on lysogenic
79 bacteriophages since they are incorporated into the prokaryotic genome and thus likely code for a
80 greater degree of evolutionary permanence relative to lytic phages. While bacteriophages have
81 been identified as vectors of adaptive gene contents for prokaryotes in a handful of host sponge
82 species (17), we have yet to tease apart the specific contributions of lysogenic phages and
83 describe their contributions to SDP biology across global geographic scales and throughout the
84 host and prokaryotic phylogenies.
85
86 **Materials and Methods**
87 *DNA Extractions and Sequencing*
88 DNA was extracted from two specimens that had been stored in 95% EtOH from each of
89 the three nominal *Ircinia* species and seven growth forms (*Ircinia* spp. nov., Kelly et al. 2020 *in*
90 *review*), resulting in a sample size of 20, using the Molzyme Molysis DNA isolation kit, which
91 depletes host DNA in a sample resulting in an extraction that is enriched for microbial DNA.

92 DNA sequencing was performed at the Yale Center for Genome Analysis on a NovaSeq6000
93 with a target of 30 million 2x150 reads per specimen.
94
95 *Metagenomic Analysis*
96 The reads were filtered and trimmed using FastP (18) with default parameters. Since the
97 DNA libraries were prepared in a facility that routinely handles vertebrate model organisms,
98 bbsplit.sh was used to remove contaminating reads mapping to the *Mus musculus* and *Homo*
99 *sapiens* genomes as well as the control PhiX reads. Metagenomic contigs were assembled
100 independently for each specimen using Megahit v1.2.9 (19) and subsequently binned using
101 MetaBAT 2 v2:2.15 (20) with default parameters. Contaminating contigs were identified on the
102 basis of being outliers with regard to genomic content (i.e., %GC and tetranucleotide content)
103 and removed from the bins using refinem v0.0.25 (21). The qualities of the refined bins were
104 assessed using checkm v1.1.2 (22); only bins with qualities equal to or higher than 40, calculated
105 as genome completeness – 5x contamination (21), were retained for downstream analysis. The
106 remaining bins were screened for uniqueness within host specimens using checkm.
107 To investigate the microbial mechanisms that promote their residence in sponge hosts, we
108 compared a dataset of SDP MAGs with a dataset of previously assembled MAGs from the Tara
109 Oceans expedition (23). The Tara Oceans dataset was produced from DNA isolated from filtered
110 oceanic seawater, and thus we assume that most of these microbes are pelagic; these are referred
111 to as the pelagic dataset throughout the manuscript (23). Prior to quality control and
112 dereplication, the pelagic dataset consisted of 2307 MAGs (NCBI bioproject PRJNA391943).
113 The SDP MAG dataset was produced by combining our Caribbean *Ircinia* MAG dataset with
114 689 publicly available draft genomes of SDPs. Prior to dereplication and further QC, the SDP

115 dataset consisted of 1580 MAGs. The SDP and Tara Oceans datasets were independently
116 dereplicated using drep v2.5.2 (24) with a cutoff of 96.5% identity, an estimate of maximum
117 prokaryotic intraspecific genomic divergence backed by empirical evidence (25). We retained
118 only MAGs with a quality equal to or greater than 40 and a completeness greater than 85% for
119 downstream analyses. Taxonomy was again assigned and a whole-genome phylogenetic tree was
120 produced for the combined datasets using GTDB-Tk v1.0.2 (26). Prophage genomes were
121 identified from our dataset using ProphET v0.5.1 (27).
122 Protein predictions and subsequent KO annotations were performed on the MAGs
123 (including the prophage sequences) and independently on the isolated prophage genomes using
124 Enrichm v0.5.0 (28). Domain PFAM annotations were performed on the MAG and prophage
125 proteins using Interproscan v5.39-77 (29). For proteins that contained duplicated identical
126 domain annotations, only one domain was retained so as to avoid double counting proteins.
127
128 *Statistical Analysis*
129 To identify features that were significantly enriched in MAGs in one category or the
130 other (i.e. SDP vs free-living prokaryotes), we performed pairwise permutational t-tests, each run
131 from 10,000 iterations, comparing the average abundance (e.g. copy number) of a feature in the
132 pelagic vs. the SDP dataset. To factor for differences in genome size, we normalized feature
133 counts by dividing them by the total number of predicted proteins prior to the t-tests. P-values
134 were corrected using the FDR adjustment of Benjamini-Hochberg (30) using the total number of
135 features within an annotation category (either KO or PFAM). Given the high number of multiple
136 comparisons, we further decided to assign a conservative significance cutoff of $p<0.01$.

137 Additionally, we omitted features that were present in less than 10% of both the SPD and pelagic
138 MAGs.
139 We tested for statistical significance in the average number of prophages per MAG,
140 average size of each prophage in bps, and the average number of auxiliary genes per prophage
141 (normalized by the total number of genes predicted for each prophage) within functional
142 categories using permutational t-tests, each with 10,000 iterations. We adjusted p-values for the
143 auxiliary gene analysis using the FDR procedure as above.
144 Linear models (linear regressions) were inferred to describe the relationship between
145 genome size vs. GC content and coding density. Statistical significance in differences between
146 the linear models of SDP and pelagic microbes was inferred using ANOCOVA. We tested for
147 significance in genome size, GC content, coding density, the total number of proteins, and the
148 average number of prophages between SDPs and pelagic MAGs within a phylogenetic
149 framework. Since the Archaea were represented by only 15 MAGs, we restricted this component
150 of the analysis to bacteria. To perform this analysis, we estimated an ultrametric bacterial tree
151 with the software PATHd8 (31) using the GTDB-Tk bacterial phylogeny as the input. Node
152 dating was calibrated using the divergence times estimates of Battistuzzi et al. (32) for the
153 cyanobacteria/actinobacteria, cyanobacteria/proteobacteria, and
154 alphaproteobacteria/gammaproteobacteria splits. We then performed phylogenetic ANOVAs for
155 each of the aforementioned genomic traits using the R package geiger (33), each for 10,000
156 iterations. All aforementioned tests of significance and regressions were performed in R v.3.6.0.
157
158 **Results and Discussion**

159 To investigate shared patterns of genome evolution in SDPs, we performed a comparative
160 metagenomic analysis evaluating nucleotide composition, genome size, gene content, and
161 prophage characteristics of SDPs from a broad sampling of SDP MAGs, sourced from multiple
162 host species and oceanic basins, against the MAG dataset of pelagic prokaryotes sequenced
163 during the Tara Oceans project (23). After screening for quality, completeness, and dereplication,
164 our final bacterial dataset was composed of 479 SDP MAGs sourced from 23 nominal and 10
165 undescribed host species and 629 pelagic MAGs; representatives from the Atlantic,
166 Mediterranean, Red Sea, and Pacific Ocean basins were present in both datasets (Tables S1).
167 Twenty-five bacterial phyla were recovered with a high degree of phylogenetic overlap between
168 the two datasets, with the majority of SDP MAGs (95.0%) and pelagic MAGs (92.7%) belonging
169 to phyla that contained both life history strategies (Tables S3). We report our results below for
170 analyses performed on the bacterial MAGs, as only 15 archaeal MAGs were recovered (Table
171 S4). We confirmed that the omission of Archaea did not change the interpretations of the results
172 by repeating our analyses on the total (Bacteria + Archaea) MAG dataset (data not shown). We
173 tested the hypotheses that: 1) general genomic characteristics of SDPs mirror those of free-living
174 oceanic bacteria, 2) gene contents segregate SDP and pelagic bacteria, and 3) lysogenic phages
175 confer adaptive advantages to their bacterial hosts that facilitate their residence in sponges.
176
177 *Release from nitrogen limitation allows SDPs to evolve larger genomes with higher GC contents*
178 The evolution of general genomic features across marine prokaryotic taxa has been
179 shaped by life history strategies. We found that SDPs have larger genomes, with both more
180 proteins and non-coding DNA, and higher GC content (Fig. 1, 2). The effect of life history
181 strategy remained significant after accounting for phylogenetic relatedness (Fig. 1). Life history

182 strategy also had a significant effect on the relationship between genome size and both GC
183 content (ANCOVA: $F_{1,1104} = 118.32$, $p < 2e^{-16}$) and coding density (ANCOVA: $F_{1,1104} = 104.3$, p
184 $< 2e^{-16}$) (Fig. 2). Based on comparisons among regression coefficients, we found that genome
185 size was a stronger predictor for GC content and coding density in pelagic MAGs relative to
186 SDPs (Fig. 2).

187 We hypothesize that differences in energy availability between the two life history
188 strategies can account for the contrasting genomic characteristics of SDP and pelagic MAGs.
189 Due to nitrogen limitation in the open ocean, pelagic prokaryotes tend to have streamlined
190 genomes and lower GC content; this strategy reduces the energetic costs of DNA replication
191 (13–16). Each of these expectations are met in our pelagic dataset, as pelagic MAGs have
192 smaller genomes with fewer genes, less non-coding DNA, and lower GC content. However,
193 opposite trends in genome content were discovered in SDPs. Excretion of nitrogenous
194 compounds such as ammonium by the host sponge (34) could alleviate nitrogen limitation and
195 enable the maintenance of larger genomes with higher GC contents in SDPs.

196

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

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

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

215 Given these observations, the patterns of virulence—which we define here as being
216 composed of both pathogenicity and the promotion of residence in the host—are likely different
217 between SDPs and pelagic bacteria. Motility and LPS genes are directly involved in pathogenesis
218 (39); thus, a potential evolutionary outcome of the reduced frequency of motility and LPS genes
219 might be one of reduced pathogenicity in SDPs. At face value, this observation appears to be at
220 odds with conventional evolutionary theories that state that lower levels of pathogenicity would
221 only be expected to confer a selective advantage to the microbe if the persistence of its genome
222 was completely dependent on survival of the host, i.e. under obligate vertical transmission
223 (10,11). On the other hand, higher rates of pathogenicity are expected to evolve in microbes
224 exhibiting mixed-transmission modes as pathogenic mechanisms act to protect the microbe, for
225 example through the strengthening of cell envelopes with LPS, while not damaging the microbial
226 population's growth rate since it is not linked to the host's (12). Why then do we detect a
227 depletion of pathogenic mechanisms in prokaryotes that are likely undergoing mixed-modes of

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

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

251

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

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

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

274 on lysogenic phages, and ultimately the insertion of phages into SDP genomes that confer a
275 higher fitness benefit, has evolved via a density-dependent relationship. Sponge microbiomes
276 can have exceptionally high prokaryotic cellular densities that can reach 10^9 cells/cm^3, four
277 orders of magnitude greater than the estimates of 0.5-5x10^5 cells/cm^3 in seawater (48,49). As
278 such, social distancing may not be an option for SDPs. Under this scenario, cellular defense
279 mechanisms that allow the cell to police incoming viral DNA might be evolutionarily favored.
280 This expectation is met in our data, as SDPs are enriched for restriction-modification systems,
281 CRISPR-CAS9, and DNA phosphothiolation genes (Fig. 3). However, higher exposure rates to
282 viruses could also translate into higher rates of introduction to novel, potentially adaptive gene
283 content (50–52). The cellular defenses of SDPs might allow them exclude prophages that are
284 large and have a lower adaptive/non-adaptive ratio of gene content. As adaptive gene content in
285 the form of ABC transporters, metabolism genes, or cellular defenses could mimic the
286 prokaryotes' genetic composition, the higher ratio of these genes to non-adaptive viral content
287 might result in a lower rate of destruction by endogenous prokaryotic defenses.
288

289 **Conclusion**

290 We discovered that SDP genomes are distinguished from the genomes of pelagic
291 microbes across the categories of comparisons that we made, thus providing evidence that the
292 selective pressures associated with living in sponges have driven shared genomic adaptations
293 independently in multiple prokaryotic phyla. The differences in genomic features could, in part,
294 be shaped by the alleviation of nutrient limitation associated with inhabiting the sponge mesohyl.
295 The contrasting abiotic and biotic regimes between life history strategies have also resulted in the
296 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

343 5. Thomas TRA, Kavlekar DP, LokaBharathi PA. Marine drugs from sponge-microbe
344 association--a review. Mar Drugs [Internet]. 2010 Apr 22;8(4):1417–68. Available from:
345 https://pubmed.ncbi.nlm.nih.gov/20479984
346 6. Hardoim CCP, Costa R. Microbial communities and bioactive compounds in marine
347 sponges of the family Irciniidae-a review. Mar Drugs. 2014;12(20):5089–122.
348 7. Pita L, Rix L, Slaby BM, Franke A, Hentschel U. The sponge holobiont in a changing
349 ocean: from microbes to ecosystems. Microbiome [Internet]. 2018;6(1):46. Available
350 from: https://doi.org/10.1186/s40168-018-0428-1
351 8. Freeman CJ, Thacker RW. Complex interactions between marine sponges and their
352 symbiotic microbial communities. Limnol Oceanogr. 2011;56(5):1577–86.
353 9. Thacker RW, Freeman CJ. Sponge-microbe symbioses: recent advances and new
354 directions. Adv Mar Biol. 2012;62:57–111.
355 10. Ferdy, Godelle B. Diversification of Transmission Modes and the Evolution of
356 Mutualism. Am Nat. 2005 Nov 1;166:613.
357 11. Ewald PW. The Evolution of Virulence: A Unifying Link between Parasitology and
358 Ecology. J Parasitol [Internet]. 1995 Apr 6;81(5):659–69. Available from:
359 http://www.jstor.org/stable/3283951
360 12. Lipsitch M, Siller S, Nowak MA. The evolution of virulence in pathogens with vertical
361 and horizontal transmission. Evolution (N Y). 1996 Apr 9;50:1729+.
362 13. Allen LZ, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LDH, et al. Influence
363 of nutrients and currents on the genomic composition of microbes across an upwelling
364 mosaic. ISME J [Internet]. 2012;6(7):1403–14. Available from:
365 https://doi.org/10.1038/ismej.2011.201

366 14. Mann S, Chen Y-PP. Bacterial genomic G+C composition-eliciting environmental
367 adaptation. Genomics [Internet]. 2010;95(1):7–15. Available from:
368 http://www.sciencedirect.com/science/article/pii/S0888754309002080
369 15. Luo H, Thompson LR, Stingl U, Hughes AL. Selection Maintains Low Genomic GC
370 Content in Marine SAR11 Lineages. Mol Biol Evol [Internet]. 2015 Jun 27;32(10):2738–
371 48. Available from: https://doi.org/10.1093/molbev/msv149
372 16. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome
373 streamlining in a cosmopolitan oceanic bacterium. Science. 2005 Aug;309(5738):1242–5.
374 17. Jahn M, Arkhipova K, Markert S, Stigloher C, Lachnit T, Pita L, et al. A Phage Protein
375 Aids Bacterial Symbionts in Eukaryote Immune Evasion. Cell Host Microbe. 2019 Sep
376 24;26.
377 18. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor.
378 Bioinformatics [Internet]. 2018 Sep 8;34(17):i884–90. Available from:
379 https://doi.org/10.1093/bioinformatics/bty560
380 19. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node
381 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
382 Bioinformatics [Internet]. 2015 Jan 20;31(10):1674–6. Available from:
383 https://doi.org/10.1093/bioinformatics/btv033
384 20. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive
385 binning algorithm for robust and efficient genome reconstruction from metagenome
386 assemblies. PeerJ [Internet]. 2019 Jul 26;7:e7359–e7359. Available from:
387 https://pubmed.ncbi.nlm.nih.gov/31388474
388 21. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al.

389 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree
390 of life. Nat Microbiol [Internet]. 2017;2(11):1533–42. Available from:
391 https://doi.org/10.1038/s41564-017-0012-7
392 22. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the
393 quality of microbial genomes recovered from isolates, single cells, and metagenomes.
394 Genome Res [Internet]. 2015/05/14. 2015 Jul;25(7):1043–55. Available from:
395 https://pubmed.ncbi.nlm.nih.gov/25977477
396 23. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-
397 assembled genomes from the global oceans. Sci Data [Internet]. 2018;5(1):170203.
398 Available from: https://doi.org/10.1038/sdata.2017.203
399 24. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic
400 comparisons that enables improved genome recovery from metagenomes through de-
401 replication. ISME J [Internet]. 2017;11(12):2864–8. Available from:
402 https://doi.org/10.1038/ismej.2017.126
403 25. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides
404 NC, et al. Microbial species delineation using whole genome sequences. Nucleic Acids
405 Res [Internet]. 2015/07/06. 2015 Aug 18;43(14):6761–71. Available from:
406 https://pubmed.ncbi.nlm.nih.gov/26150420
407 26. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify
408 genomes with the Genome Taxonomy Database. Bioinformatics [Internet]. 2019 Nov
409 15;36(6):1925–7. Available from: https://doi.org/10.1093/bioinformatics/btz848
410 27. Reis-Cunha JL, Bartholomeu DC, Manson AL, Earl AM, Cerqueira GC. ProphET,
411 prophage estimation tool: A stand-alone prophage sequence prediction tool with self-

412 updating reference database. PLoS One [Internet]. 2019 Oct 2;14(10):e0223364. Available
413 from: https://doi.org/10.1371/journal.pone.0223364
414 28. Boyd JA, Woodcroft BJ, Tyson GW. Comparative genomics using EnrichM. in prep.
415 2019;
416 29. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5:
417 genome-scale protein function classification. Bioinformatics [Internet]. 2014/01/21. 2014
418 May 1;30(9):1236–40. Available from: https://pubmed.ncbi.nlm.nih.gov/24451626
419 30. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful
420 approach to multiple testing. J R Stat Soc. 1995;57:289–300.
421 31. Britton T, Anderson CL, Jacquet D, Lundqvist S, Bremer K. Estimating Divergence
422 Times in Large Phylogenetic Trees. Syst Biol [Internet]. 2007 Oct 1;56(5):741–52.
423 Available from: https://doi.org/10.1080/10635150701613783
424 32. Battistuzzi FU, Feijao A, Hedges SB. A genomic timescale of prokaryote evolution:
425 insights into the origin of methanogenesis, phototrophy, and the colonization of land.
426 BMC Evol Biol [Internet]. 2004;4(1):44. Available from: https://doi.org/10.1186/1471-
427 2148-4-44
428 33. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, et al. Geiger
429 v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic
430 trees. Bioinformatics. 2014 Aug;30(15):2216–8.
431 34. Brusca RC, Moore W, Shuster SM, Haver N. Invertebrates. 3rd ed. Sunderland,
432 Massachusettes: Sinauer Associates, Inc.; 2016.
433 35. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, et al. Functional
434 equivalence and evolutionary convergence in complex communities of microbial sponge

435 symbionts. Proc Natl Acad Sci [Internet]. 2012 Jul 3;109(27):E1878 LP-E1887. Available
436 from: http://www.pnas.org/content/109/27/E1878.abstract
437 36. Reynolds D, Thomas T. Evolution and function of eukaryotic-like proteins from sponge
438 symbionts. Mol Ecol [Internet]. 2016;25(20):5242–53. Available from:
439 https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13812
440 37. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, et al. Genomic
441 blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable
442 Alphaproteobacteria. Sci Rep [Internet]. 2019;9(1):1999. Available from:
443 https://doi.org/10.1038/s41598-019-38737-x
444 38. Germer J, Cerveau N, Jackson DJ. The Holo-Transcriptome of a Calcified Early
445 Branching Metazoan. Front Mar Sci [Internet]. 2017;4:81. Available from:
446 https://www.frontiersin.org/article/10.3389/fmars.2017.00081
447 39. Chaban B, Hughes HV, Beeby M. The flagellum in bacterial pathogens: For motility and a
448 whole lot more. Semin Cell Dev Biol. 2015 Oct;46:91–103.
449 40. Kelly SR, Garo E, Jensen P, Fenical W, Pawlik J. Effects of Caribbean sponge secondary
450 metabolites on bacterial surface colonization. Aquat Microb Ecol. 2005 Sep 6;40:191–
451 203.
452 41. Lavy A, Keren R, Yahel G, Ilan M. Intermittent Hypoxia and Prolonged Suboxia
453 Measured In situ in a Marine Sponge. Front Mar Sci [Internet]. 2016;3:263. Available
454 from: https://www.frontiersin.org/article/10.3389/fmars.2016.00263
455 42. Weber M, Mendola D, Hoffmann F, Beer D De. Heterogeneous oxygenation resulting
456 from active and passive flow in two Mediterranean sponges , Dysidea avara and
457 Chondrosia reniformis. Limnol Oceanogr. 2010;55(3):1289–300.

458 43. Belin BJ, Busset N, Giraud E, Molinaro A, Silipo A, Newman DK. Hopanoid lipids: from
459 membranes to plant–bacteria interactions. Nat Rev Microbiol [Internet]. 2018;16(5):304–
460 15. Available from: https://doi.org/10.1038/nrmicro.2017.173
461 44. Silipo A, Vitiello G, Gully D, Sturiale L, Chaintreuil C, Fardoux J, et al. Covalently linked
462 hopanoid-lipid A improves outer-membrane resistance of a Bradyrhizobium symbiont of
463 legumes. Nat Commun [Internet]. 2014;5(1):5106. Available from:
464 https://doi.org/10.1038/ncomms6106
465 45. Kulkarni G, Busset N, Molinaro A, Gargani D, Chaintreuil C, Silipo A, et al. Specific
466 Hopanoid Classes Differentially Affect Free-Living and Symbiotic States of
467 Bradyrhizobium diazoefficiens. Ausubel FM, editor. MBio [Internet]. 2015 Oct
468 30;6(5):e01251-15. Available from: http://mbio.asm.org/content/6/5/e01251-15.abstract
469 46. Karimi E, Ramos M, Gonçalves JMS, Xavier JR, Reis MP, Costa R. Comparative
470 Metagenomics Reveals the Distinctive Adaptive Features of the Spongia officinalis
471 Endosymbiotic Consortium. Front Microbiol [Internet]. 2017 Dec 14;8:2499. Available
472 from: https://pubmed.ncbi.nlm.nih.gov/29312205
473 47. Sica D, Zollo F. Free amino acids in some sponges. Biochem Syst Ecol [Internet].
474 1977;5(2):129–31. Available from:
475 http://www.sciencedirect.com/science/article/pii/0305197877900412
476 48. Hentschel U, Usher KM, Taylor MW. Marine sponges as microbial fermenters. FEMS
477 Microbiol Ecol. 2006 Feb;55(2):167–77.
478 49. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl
479 Acad Sci U S A [Internet]. 1998 Jun 9;95(12):6578–83. Available from:
480 https://pubmed.ncbi.nlm.nih.gov/9618454

481 50. Tsao Y-F, Taylor VL, Kala S, Bondy-Denomy J, Khan AN, Bona D, et al. Phage Morons
482 Play an Important Role in Pseudomonas aeruginosa Phenotypes. J Bacteriol [Internet].
483 2018 Oct 23;200(22):e00189-18. Available from:
484 https://pubmed.ncbi.nlm.nih.gov/30150232
485 51. Fortier L-C, Sekulovic O. Importance of prophages to evolution and virulence of bacterial
486 pathogens. Virulence [Internet]. 2013/04/23. 2013 Jul 1;4(5):354–65. Available from:
487 https://pubmed.ncbi.nlm.nih.gov/23611873
488 52. Brüssow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens:
489 from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev [Internet].
490 2004 Sep;68(3):560–602. Available from: https://pubmed.ncbi.nlm.nih.gov/15353570
491

492 **List of Supplementary Materials:**

493 Tables S1-S7

494 **Figure captions:**

495 **Figure 1.** Phylogenetic ANOVAs of genomic characteristics show that they are significantly
496 different between SDPs and pelagic MAGs after correcting for phylogenetic relatedness and
497 multiple comparisons (fdr): genome size ($F_{1,1106}$=68.218, p=0.0488), coding density ($F_{1,1106}$=
498 188.62, p=0.0005), number of proteins ($F_{1,1106}$= 66.227, p= 0.0488), GC-content ($F_{1,1106}$= 292.44,
499 p= 0.00025).

500

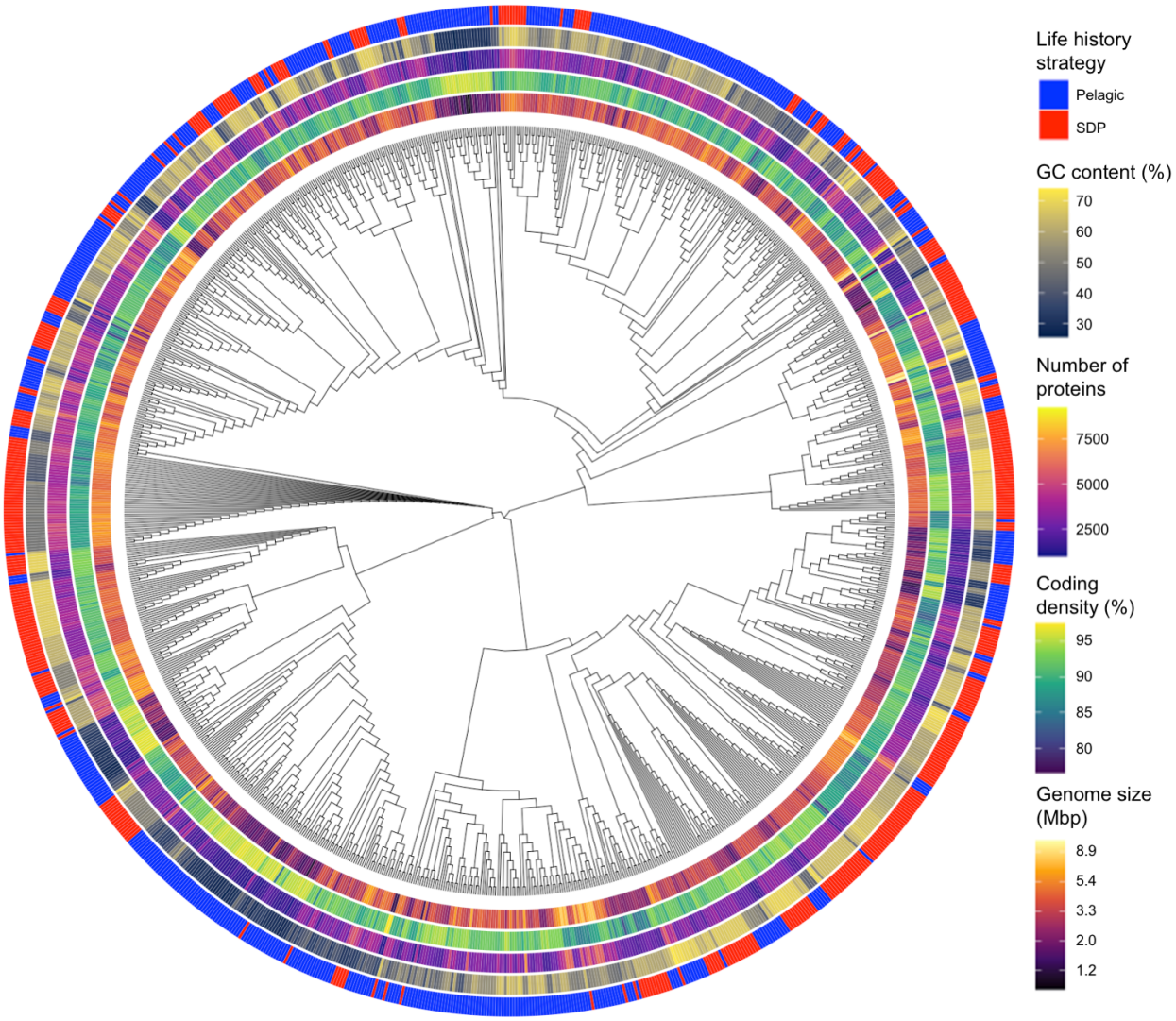
501 **Figure 2. A:** Scatterplot comparing relationships prokaryotic genome sizes to GC content and **B**,
502 genome size to coding density for SDPs and pelagic microbes. Coding density is reported as the
503 percentage of basepairs that are included in a coding region. R^2 and statistical significance is
504 reported for linear models inferred for the pelagic and SDP MAGs. Plots to the left of the y-axes
505 and below the x-axes are kernel density estimation plots. Degrees of freedom for SDP MAGs =
506 477, pelagic MAGs = 627. **C:** Bar chart comparing average number of proteins predicted from
507 SDP and pelagic MAGs (phylogenetic ANOVA reported in the caption of Figure 1). Asterisks
508 here and elsewhere in figures denote significance.

509

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

516 ratios are reported in the legend. The ELP graph was constructed from the Interproscan
517 annotations; all others were constructed using the KO annotations produced from enrichm.
518
519 **Figure 4. A and B:** Bar charts depicting lower average number of prophages per MAG (Welsh t-
520 statistic = 6.172, $p < 2.2e^{-16}$) and smaller average size of prophages in SDPs relative to pelagic
521 MAGs (Welsh t-statistic = -3.6977, p = 0.0024). **C:** Bar charts depicting normalized abundances
522 of genes on prophages within the categories of defense and resistance, ELP, primary and
523 secondary metabolism, ABC transporters and membrane trafficking, others (see main text for
524 descriptions), and total counts. Permutational t-tests for each category follow with fdr-adjusted p-
525 values: defense and resistance (Welsh's t = 1.620, p = 0 0.0234), primary and secondary
526 metabolism (Welsh's t = 2.0367, p = 0.0144), ABC transporters and membrane trafficking
527 (Welsh's t = 0.748, p = 0.312), other (Welsh's t = 1.743, p = 0.0224), ELP Welsh's t = 1.234 p =
528 0.0934), total (Welsh's t = 2.801, p < 2.2e-16).
529
530 **Figure S1.** Representative cassettes of steroid biosynthesis genes are depicted. Genes in parallel
531 denote common alternative combinations for a clade. TM7SF2, ERG24, Delta14-sterol reductase
532 [EC:1.3.1.70] (K00222); SMT1, ERG6, sterol 24-C-methyltransferase [EC:2.1.1.41] (K00559);
533 LSS, ERG7, lanosterol synthase [EC:5.4.99.7] (K01852); CYP51, sterol 14alpha-demethylase
534 [EC:1.14.14.154 1.14.15.36] (K05917); FDFT1, farnesyl-diphosphate farnesyltransferase
535 [EC:2.5.1.21] (K00801); and ERG2, C-8 sterol isomerase [EC:5.-.-.-] (K09829) are components
536 of the steroid biosynthesis pathway (ko00100). Dihydroflavonol-4-reductase [EC:1.1.1.219]
537 (K00091), an oxidoreductase acting on the CH-OH group of donors with NAD+ or NADP+ as
538 acceptor, is not assigned to a pathway and was included here and in Figure 2 due to its consistent

539 association with genes belonging to the steroid biosynthesis (ko00100) pathway and its likely
540 relevant biochemical activity. 'HP' abbreviates hypothetical protein. Arrows denote direction of
541 the coding sequence. Genes are colored to match the legend of the 'Steroid Biosynthesis' plot of
542 Figure 2. ERG2 is uncolored as it was removed from the data producing Figure 2 due to being in
543 low proportions (<10%) of both the SDP and pelagic MAG datasets. Note: genes are not drawn
544 to scale by length.
545
546

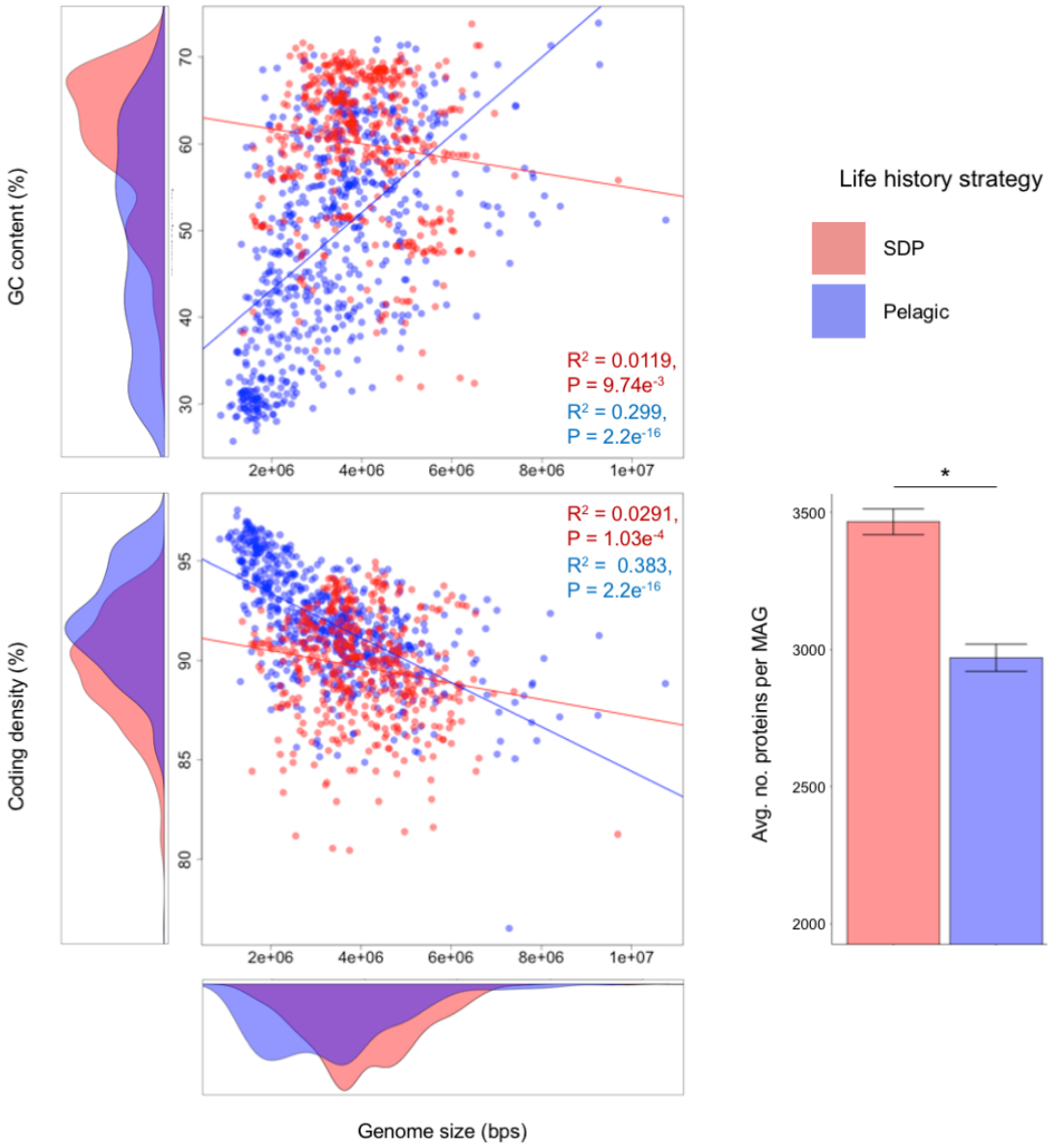


547

548 **Figure 1. The phylogenetic distribution of endogenous genomic traits in bacteria is**

549 **independent of life history strategy.**

550



551

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

553

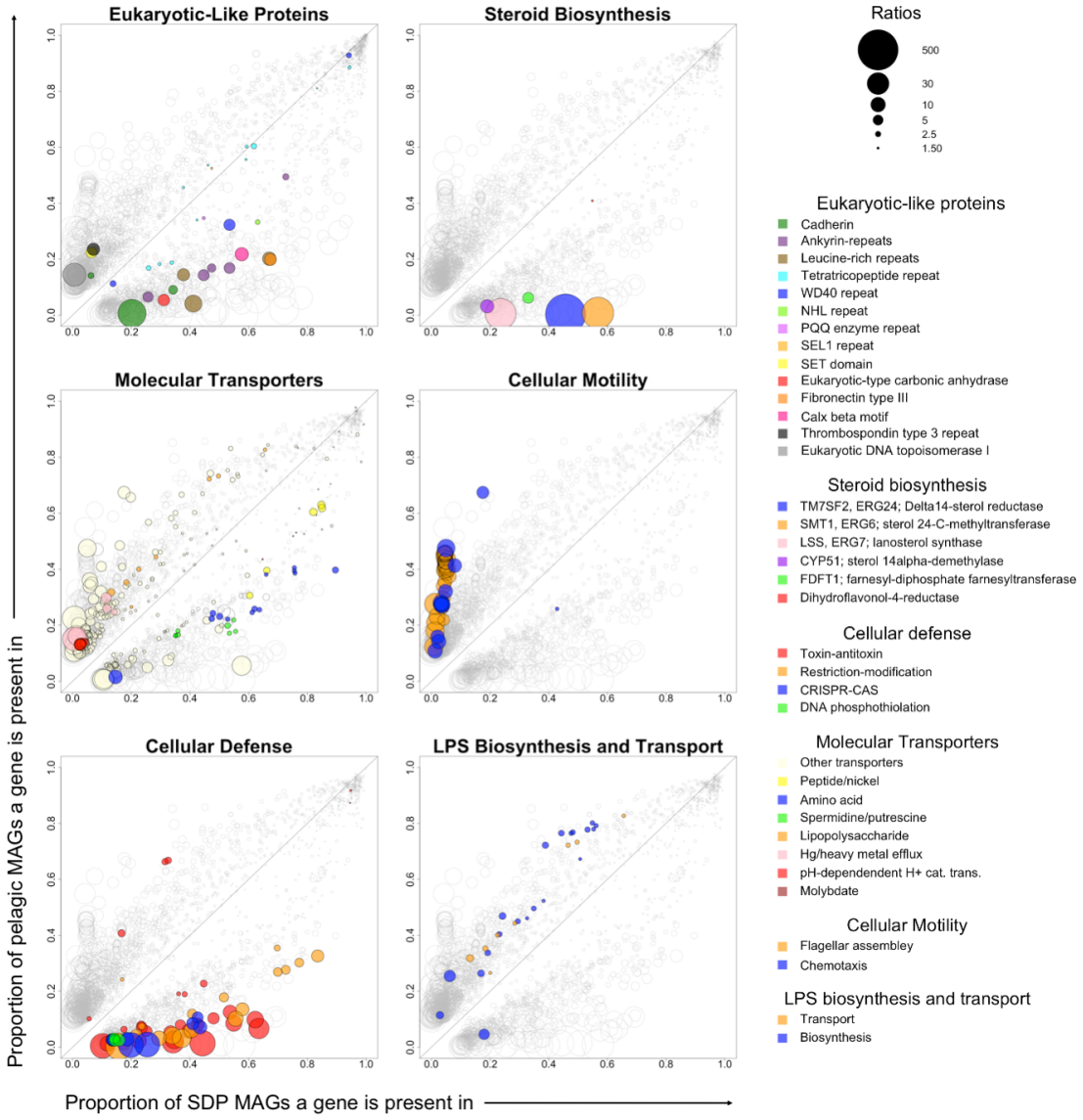
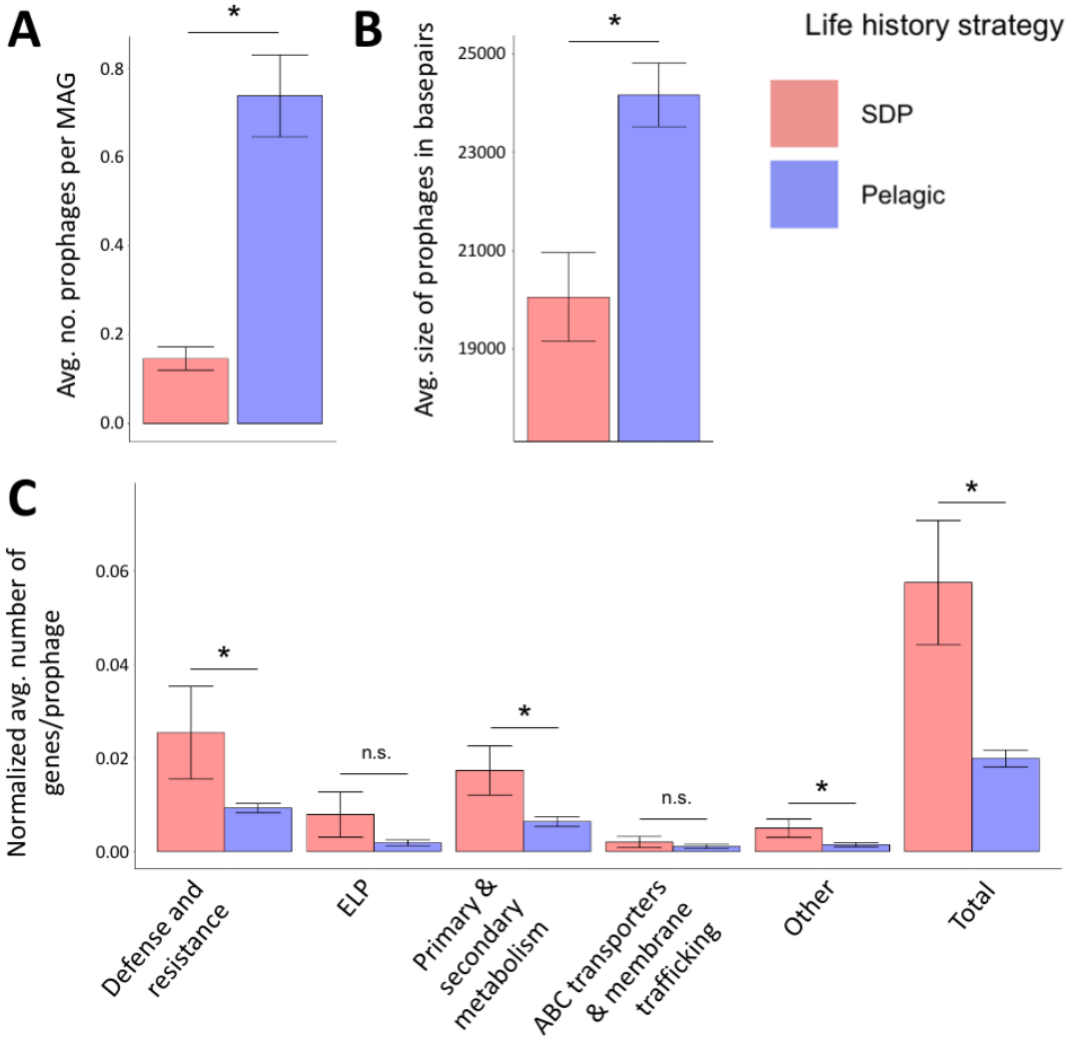


Figure 3. SDPs are well-defended against foreign DNA, are equipped to promote their residence in host sponges, and have reduced pathogenicity.



558

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

560

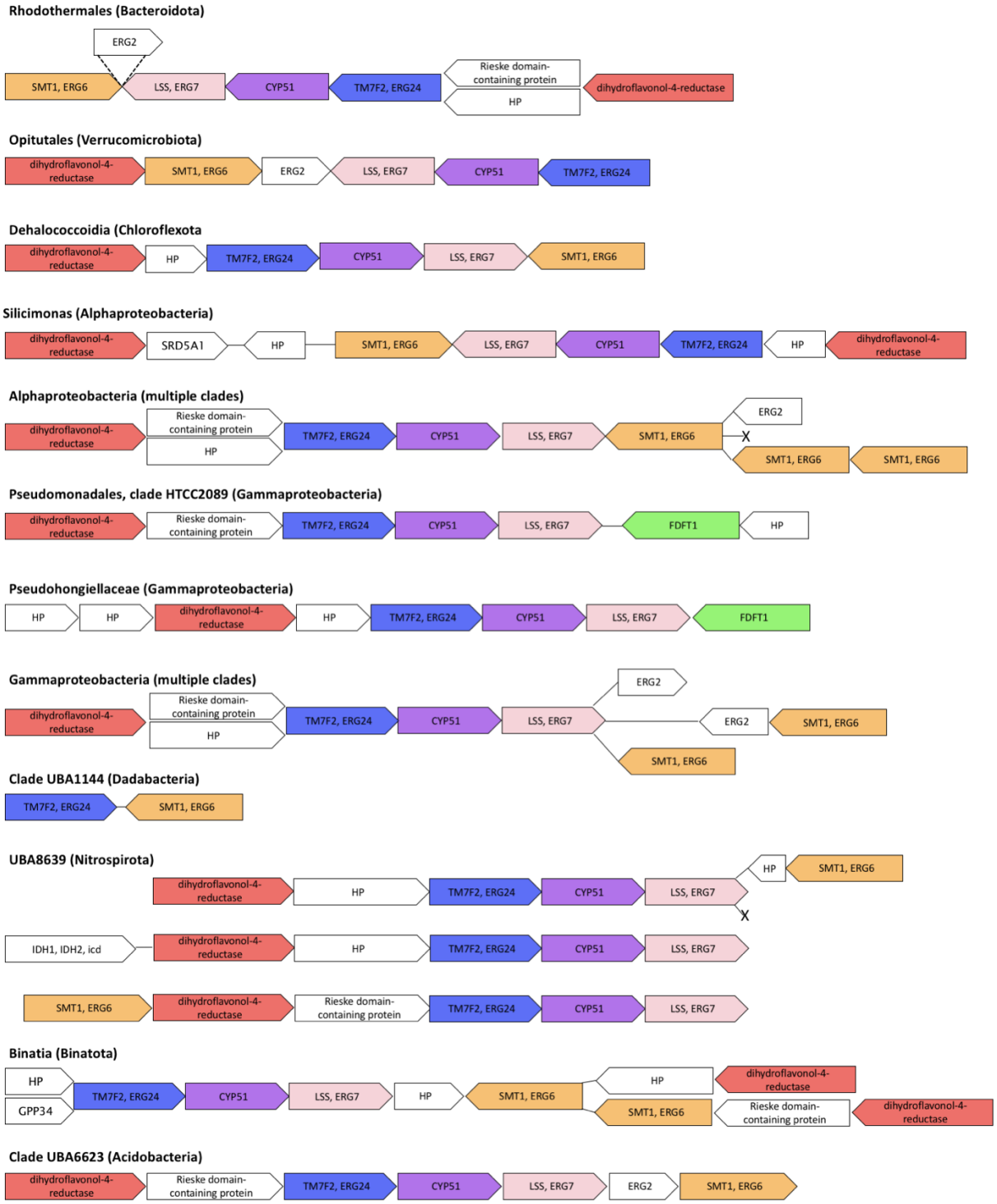


Figure S1. SDPs contain partial eukaryotic steroid biosynthesis pathways.