High transposase abundance in the deep ocean is linked to a particle-associated lifestyle

Juntao Zhong[1], Thais del Rosario Hernández[1,2], Oleksandr Kyrysyuk[1,3], Benjamin J. Tully[4], Rika E. Anderson[1*]

Author affiliations

[1]Department of Biology, Carleton College, Northfield, MN, USA

[2]Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA

[3]Department of Cancer Immunology and Virology, Dana Farber Cancer Institute, Boston, MA, USA

[4]Marine & Environmental Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA USA

*Corresponding author: randerson@carleton.edu

Abstract

Transposases are mobile genetic elements (MGEs) that move within and between genomes, enabling genomic plasticity within microbes. Transposases often carry functional genes, termed *cassette sequences*, when they migrate within and between genomes. In marine microbial communities, transposase abundance increases with depth, but the reasons for this trend are unclear. Through analysis of metagenomes from the *Tara* Oceans and Malaspina Expeditions, we identify the particle-associated lifestyle as a main covariate for the high occurrence of transposases in the deep ocean. The abundance of biofilm-associated genes correlates with transposases in a strong and depth-independent manner, suggesting that MGEs readily propagate among microbial communities in crowded biofilms. Genome size positively correlates with particle-associated populations and transposase abundance. This relationship may result from selection for larger genomes, in which essential genes are more sparse and less likely to be disrupted by the recombination activities of MGEs. Genes related to defense mechanisms are enriched in cassette sequences and upregulated in the deep sea. Further, defense mechanism genes in cassette sequences experience stronger selection than other cassette genes. Thus, while transposons can spread at the expense of their microbial hosts, they introduce novel genes and potentially benefit the hosts in accessing heavily contested resources. Together, our results reveal an emerging picture of deep ocean particles as gene sharing highways that result in expanded and defensively oriented microbial genomes.

Introduction

Mobile genetic elements (MGEs) are DNA segments that can mediate the movement of genetic sequences both within and between bacterial and archaeal genomes [1]. MGEs generally encode enzymes to mediate this process [1]: in transposons, the mediating enzymes are transposases,

through which the mechanism of transfer results in inverted repeats [2, 3]. When MGEs migrate from one genomic region to another, they frequently carry cassette sequences, which have been reported to express functional genes such as antibiotic resistance [4] and metal uptake [5, 6], and regulatory genes that affect gene expression in host cells [7, 8].

Transposases are the most abundant and ubiquitous genes in nature [9], despite their uneven distribution across taxa [10, 11]. In marine systems, one of the most striking – and as of yet unexplained – trends in microbial genomics is the distinct increase in transposase abundance with depth. Genomic analyses from station ALOHA in the North Pacific revealed a substantial increase in transposase abundance with depth from 500 m to their observed maximum at 4000m [12, 13]. As one of the most overrepresented COG-categories in ALOHA deep waters, transposases accounted for 1.2% of all fosmid sequences from 4000 m [12, 14]. Similarly, a study of hydrothermal chimneys showed that transposases were particularly abundant in biofilms and accounted for 8% of all metagenomic reads, a 10-fold increase compared to metagenomes in other habitats [15]. Based on these observations, our analyses focused on three central questions: (1) why is transposase abundance associated with greater depths and biofilms in marine systems, (2) are transposases selfish genes that spread widely with neutral or deleterious effects on host genomes, and (3) do transposases perform useful functions for microbial hosts that encode them?

To identify reasons for the high transposase abundance in the deep sea and to expand our knowledge of the role of MGEs in marine microbial communities, we analyzed 187 microbial metagenomes and 181 microbial metatranscriptomes from the *Tara* Oceans Expedition [16, 17]. The *Tara* Oceans samples were from 5-1000 m in depth; to represent bathypelagic microbial communities, we also incorporated 58 metagenomes, sampled between 2400-4000 m in depth, from the 2010 Malaspina Expedition [18]. To explore the role of transposases at a genome-resolved level, we analyzed 1562 metagenome assembled genomes (MAGs) previously generated from the *Tara*

Oceans [19] and Malaspina [18] metagenomes. Leveraging these publicly-available genomes from global oceans, we showed that the high abundance of transposases in deep ocean microbial communities was strongly linked to the shift towards a particle-associated lifestyle. We identified taxonomy, biofilm abundance, and genome size as key covariates to transposase abundance in MAGs. Further, we found that among the open reading frames (ORFs) in cassette sequences, the functional category "defense mechanism" was more abundant and under stronger selection, suggesting that transposons introduce beneficial genes to their microbial hosts inhabiting highly competitive habitats.

Results and Discussion

Previous studies have revealed that transposase abundance in microbial genomes is associated with increased ocean depth, a particle-associated lifestyle (as opposed to a planktonic lifestyle), and lower dissolved oxygen (DO) concentrations [10, 13, 20]. However, the reason and nature of these correlations has not been further explored. To confirm these findings on a broader scale, we screened for significant covariates for transposase abundance in the *Tara* Oceans and Malaspina metagenomes. We plotted the transposase abundance against depth, filter size fraction, and DO. In our metagenome analysis, transposase abundance was defined as the proportion of reads mapped back to all transposase ORFs (the database used to identify putative transposase ORFs is shown in **Table S1**). We confirmed that transposase abundance increases steadily as depth increases, despite large variations across samples (**Fig. S1**). In every ocean, the transposase abundances of the deep water samples were higher than those of the shallow water samples (**Table S2**). Here, *deep water* was defined as mesopelagic (depth 250-1000 m) and bathypelagic (2500-4000 m) waters, and *shallow water* represents surface (<10 m) and deep chlorophyll maximum (DCM, depth 17-120 m) waters.

To investigate how MGE abundances relate to the planktonic and particle-associated lifestyle, we calculated transposase abundance in metagenomes collected from different filter sizes. In the *Tara* Oceans samples, samples from the 0.22-3.0 μm filter size fraction were considered *particle-associated*, and samples from the 0.22-1.6 μm size fraction were *planktonic* [17]. In Malaspina samples, samples from the 0.8-5.0 μm size fraction were *particle-associated* and 0.2-0.8 μm were *planktonic* [18]. After controlling for depth, samples from particle-associated microbial communities were more enriched in transposases than their planktonic counterparts (**Fig. 1**; ANOVA *F-test P* = 0.0032).

Dissolved oxygen, however, was not a significant covariate for transposase abundance after controlling for depth (ANOVA *F-test P* = 0.71). In the surface and bathypelagic samples, DO positively correlated with transposase abundance, but this correlation was reversed in DCM and mesopelagic samples (**Fig. S2**).

Transposase and biofilm abundance were correlated in metagenomes and metatranscriptomes.

To further investigate the relationship between transposase abundance and particle-associated lifestyles, we examined the correlation between transposase and biofilm-associated ORFs in metatranscriptomic and metagenomic samples (the bathypelagic zone was excluded for all transcript analysis, because no metatranscriptomes were available for Malaspina samples). Biofilm formation is essential for surface attachment, so the abundance of biofilm-associated ORFs was used to quantify the degree to which a microbial community relied on a particle-associated lifestyle (the database of genes used to identify putative biofilm ORFs is shown in **Table S3**).

In samples from the mesopelagic zone, the average transcript abundance of biofilm-associated ORFs was doubled (95% CI: 1.70 to 3.13) compared to samples from the surface and

DCM (**Fig. S3**). As with metagenomic abundance, transcript abundance of a target gene was defined as the number of metatranscriptomic read hits divided by the total number of metatranscriptomic reads in a sample. The high transcript abundance of biofilm-associated ORFs in the mesopelagic zone suggested increased biofilm production and a shift toward a particle-associated lifestyle in the deep ocean. This shift also manifested at the population level: the proportion of particle-associated MAGs steadily increased with greater depth (**Fig. S4A**).

Biofilms on marine particles are thought to be hotspots for horizontal gene transfer and transposase propagation [21], and the prevalence of biofilm transcripts offers a promising explanation for the elevated transposase abundance in the deep ocean. In both metagenomic and metatranscriptomic samples, the abundance of transposases and biofilm-associated ORFs were strongly correlated (**Fig. 2A, 2B**). After controlling for depth and sample region, the abundance of biofilm-associated ORFs still explained a significant amount of the variation in transposase abundance: incorporating biofilm DNA abundance increased R^2 by 0.219 to 0.616 in metagenomes (ANOVA $P < 2.2\text{x}10^{-16}$); incorporating biofilm RNA abundance increased R^2 by 0.152 to 0.519 in metatranscriptomes (ANOVA $P < 1.9\text{x}10^{-12}$). Correlations in metagenomes and metatranscriptomes suggested that transposases spread more widely in microbial communities with a greater reliance on a particle-associated lifestyle. After observing such trends at the community level, we sought to determine if these trends hold true at a genome-resolved level to support our hypothesis that the particle-associated lifestyle is a main driver for the high transposase abundance in the deep ocean.

Genome size and taxonomy are key factors to transposase abundance in populations.

To identify factors affecting transposase abundance at a genome-resolved level, we analyzed 1307 MAGs recovered from the *Tara* Oceans metagenomes [19] and 255 MAGs from the Malaspina metagenomes [18]. All MAGs in our analysis had percent completeness >70% and redundancy

133 <10% (both criteria were calculated with CheckM [22]). The transposase abundance in MAGs was
134 quantified by the percentage of transposase ORFs among all ORFs in the MAG (abbreviated as *%-*
135 *transposase*). Similar to metagenomes, MAGs from deep waters had higher transposase abundance
136 than MAGs from shallow waters. Specifically, MAGs from the mesopelagic and bathypelagic zones
137 had a mean %-transposase 6 times (95% CI: 5.38 to 6.57) that of MAGs from the surface and DCM.
138 Given the enrichment of transposase ORFs in deep water MAGs, we ask 3 questions: (1)
139 whether the depth-independent correlation between transposase abundance and the particle-
140 associated lifestyle persists in MAGs; (2) whether specific taxa encode more transposases than
141 others, and thus whether community composition plays a role in transposase abundance; and (3)
142 since larger genomes provide more neutral insertion sites for transposases [13, 23], whether genome
143 size correlates to transposase abundance in MAGs.
144 To categorize the lifestyle of each MAG as *planktonic*, *particle-associated* or *mixed*, we compared
145 each MAG's relative mapping of metagenomic reads from the small and large filters (see Methods).
146 Taxonomy was analyzed at a *class* level, and the *complete genome size* was estimated using the genome
147 length of a MAG divided by its estimated percent completeness. Although MAG assembly might
148 bias against MGEs [24] and the completeness of MAGs might be underestimated for rare taxonomy
149 groups [25, 26], we can still gain valuable insights by analyzing the distribution of transposase ORFs
150 in MAGs of various genome sizes.
151 After controlling for depth, we observed a positive correlation between the estimated
152 complete genome size of a MAG and its transposase abundance (**Table 1**), confirming previous
153 speculations that larger genomes were linked to high transposase abundance [10, 13]. While deep
154 water genomes were bigger (**Fig. 3A**), the particle-associated lifestyle correlated with greater
155 complete genome size in every depth (**Fig. 3B**). Bacterial genomes typically contain around 500
156 essential genes, so as genomes get larger, the risk of inactivating essential genes quickly decreases,

157 and transpositions are more tolerated [23]. Thus, the particle-associated lifestyle might facilitate
158 transposase propagation by reducing the distance between microbes, and by providing more
159 neutral/non-fatal MGE insertion sites via larger genome size [23]. However, the causal relationship
160 between genome size, transposase abundance and particle association remains unclear. It is worth
161 noting that the transposase abundance in MAGs peaked when the complete genome size reached 5
162 Mbp, but it stabilized or declined beyond the peak (**Fig. 3C**). A previous study has also reported a
163 decreased proportion of transposase ORFs after bacterial genomes surpass 6 Mbp in size [11].
164 However, most (88.8%) MAGs recovered from the *Tara* Oceans and Malaspina metagenomes were
165 under 5 Mbp, so the association between large genomes and high transposase abundance generally
166 held true.

167 Taxonomic class was also a key predictor of transposase abundance in MAGs (**Table 1**),
168 after controlling for depth, genome size and lifestyle. Under the same environment, some taxonomic
169 classes were more enriched in transposases than others (**Fig. 3D, E**). MAGs of the taxonomic
170 classes *Alpha-*, *Beta-*, *Gammaproteobacteria*, and *Actinobacteria* were enriched in transposases compared
171 to other MAGs (for each class with ≥10 MAGs, The %-transposase of MAGs in that class were
172 compared against those of all other MAGs; *t*-test *P* cutoff: 0.05). MAGs of *Flavobacteria*,
173 *Acidimicrobidae*, *Gemmatimonadetes*, *Marinisomatia*, and *SAR202-2* had low transposase abundance. Taxa
174 that were enriched/low in transposases mostly matched with previous studies [10, 11], except that
175 *Actinobacteria* had previously been reported to be low in transposase abundance.

176 The link between taxonomy and transposase abundance could partially be attributed to
177 depth. MAGs of transposase-enriched classes were more abundant in deep waters, and MAGs from
178 low-transposase classes were more abundant in shallow waters (2 χ^2-tests, both $P < 1\text{x}10^{-6}$; **Fig.**
179 **S4B**). Thus, although community composition is a key covariate to transposase abundance, the
180 distribution of taxa is nonetheless linked to the depth of a microbial community.

Relaxed selection does not explain transposase abundance in genomes.

It is possible that transposases accumulate in deep ocean microbial populations as a result of genetic drift, because those populations tend to be smaller in size [27] and experience slow growth rate [28]. To examine this possibility, we tested whether the strength of selection on ORFs in metagenomes correlated with their transposase abundances. We calculated the *pN*/*pS* ratio for ORFs with ≥20x coverage in 50 randomly selected metagenomes from all four depths (see Methods). *pN*/*pS* represents the proportion of nonsynonymous mutations to the proportion of synonymous mutations, and characterizes selection at the level of the population, in contrast to *dN*/*dS*, which characterizes selection between individual species [29–31].

In the *Tara* Oceans metagenomes (surface, DCM, mesopelagic), the *pN*/*pS* of ORFs was similar between the planktonic and particle-associated size fractions. However, in the bathypelagic metagenomes, the *pN*/*pS* of ORFs was higher in particle-associated samples than planktonic samples (**Fig. S5**). Considering the planktonic filter size of 0.2-0.8 μm and large prefilter (20μm mesh) in the Malaspina Expedition samples compared to the Tara Ocean samples, it is possible that the high *pN*/*pS* in bathypelagic particle-associated samples is due in part to the inclusion of small eukaryotes, which are generally 1-3 μm in diameter [32], though these were not detected in high abundances in the Malaspina metagenomes.

In metagenomes from all depths, transposases had higher *pN*/*pS* than non-transposase ORFs (4 *t*-tests on log-transformed *pN*/*pS*, all $P < 10^{-11}$; **Fig. 4A**). The high *pN*/*pS* of transposases indicates relaxed selection pressure relative to the other ORFs on average. Moreover, ORFs from deeper ocean layers had higher *pN*/*pS* than layers above (*pN*/*pS* of ORFs from BAT > MES > DCM > SRF; 3 *t*-tests, all $P < 10^{-15}$; Fig. 4A). The high *pN*/*pS* of ORFs from the deeper layers suggested relaxed selection pressure relative to ORFs from shallower layers, which was in agreement with previous results [13]. The general explanation for such a trend is a slower replication rate, which

leads to smaller effective population sizes and a reduced selection effect in the deep ocean [12, 13]. However, further work is needed to substantiate this hypothesis.

Since transposases and deep-sea ORFs were both under relaxed selection, we tested whether high transposase abundance correlated with relaxed selection at a genome-resolved scale. Mapping of reads to each MAG represents variation within a microbial population, and the selection pressure was measured by the median *pN/pS* of all ORFs in a MAG (a MAG needed at least 100 ORFs to be included in the *pN/pS* calculation). After controlling for depth, we did not find any correlation between the median ORF *pNpS* of a MAG and its transposase abundance (**Fig. 4B**, ANOVA P = 0.85). Thus, although relaxed selection and transposase enrichment co-occur in deep oceans at a community scale, we did not observe this trend on a genome-by-genome basis, suggesting that there is not a direct connection between transposase abundance and relaxed selection at a genome-resolved scale, and thus within individual microbial populations.

Transposons/integrons carry a high proportion of ORFs related to defense mechanisms

One important question regarding high transposase abundance in the deep sea is whether these transposases are neutral or deleterious, or whether they perform important functions in the genomes encoding them. In some cases, transposases are deleterious and proliferate as a result of genetic drift [33], but transposases can also benefit the host by introducing functional and regulatory genes [7, 21, 34, 35].

Transposases mediate the migration and integration of cassette sequences into host genomes; a transposon or integron is the combination of a transposase and its cassettes. Thus, cassette sequences are functionally similar to auxiliary metabolic genes (AMGs) [36], in that they are carried by a selfish MGE (or viruses in the case of AMGs) to increase the fitness of the host and therefore increase the fitness of the MGE.

229 To determine whether transposons in marine habitats carry advantageous genes, we began
230 by querying the functional categories of cassette sequences. The software package *Integron Finder [11]*
231 was used to locate cassettes on contigs; the program searches for the two palindromic flanking sites
232 necessary for transposase-mediated recombination as well as a nearby transposase/integrase. We
233 treated transposons and integrons as equivalent, due to the sequence and function similarity between
234 integrases and transposases [11]. We found 8,519 cassette ORFs from the co-assembled Malaspina
235 [18] and *Tara* Oceans [17, 19] metagenomes (see Methods). Only 27% of the cassette sequences
236 were assigned with a Clusters of Orthologous Genes (COG) annotation, which was a lower
237 annotation rate compared to other ORFs in the *Tara* Oceans (55-60%) and Malaspina (68%)
238 metagenomes.

239 Compared to the rest of the metagenome, cassettes were enriched in ORFs of the COG
240 categories "replication, recombination, and repair", "defense mechanisms", and "mobilome:
241 prophages, transposons" (**Fig. 5A**). Transposons/integrons were expected to be enriched in ORFs
242 related to the mobilome and replication/recombination. The prevalence of defense mechanism
243 cassettes was noteworthy: defense mechanisms accounted for 14.3% of known function calls in
244 cassettes, but only accounted for 2.47% of known function calls in non-cassette ORFs (*known* was
245 defined as all functional groups except "function unknown" and "general function prediction"). Out
246 of the 276 defense mechanism cassettes,123 of them encoded toxin/antitoxin genes. Since
247 transposases are more pervasive on particles, toxin and defense genes they carry would be useful in
248 competition and protection on crowded biofilms [37–40]. Moreover, the abundance of biofilm-
249 associated ORFs – indicative of particle attachment – strongly correlated with the abundance of
250 defense mechanism ORFs (**Fig. 5B**), after controlling for depth. Incorporating biofilm abundance
251 increased R^2 by 0.351 to 0.801 (ANOVA $P < 2.2\text{x}10^{-16}$). Thus, if defense mechanism genes in

252 cassette sequences are conserved and expressed, this would substantiate the hypothesis that deep
253 ocean microbes benefit from novel genes introduced by integrons/transposons.
254
255 **Conservation of defense mechanism ORFs in cassette sequences**
256 To further investigate whether cassette sequences in transposons provide functional
257 advantages to the host, we examined selection pressure in cassette ORFs. We compared the *pN/pS*
258 of cassette ORFs to those of non-cassette ORFs in the metagenomes.
259 Out of 8519 identified cassette ORFs, 1286 and 477 ORFs from the *Tara* Oceans and
260 Malaspina metagenomes had sufficient coverage (≥20x) to calculate *pN/pS*, respectively. The 97
261 percentile *pN/pS* ratio of all cassette ORFs was 4, which was used as the upper cutoff for our
262 subsequent *pN/pS* analysis. Then, we calculated the *pN/pS* for 500000 randomly selected non-
263 cassette ORFs from the *Tara* Oceans and Malaspina metagenomes, according to the ratio of cassette
264 ORFs described above (477:1286).
265 Non-cassette ORFs had lower *pN/pS* than cassette ORFs (*t*-test, $P < 2.2\text{x}10^{-16}$, medians
266 were 0.137 and 0.283, respectively). Among cassette ORFs, those with COG function calls had
267 lower *pN/pS* than those without (**Fig. 6**). And cassette ORFs of the category defense mechanisms
268 had lower *pN/pS* than cassettes of other functional categories (**Fig. 6**). The low *pN/pS* of defense
269 mechanism cassettes suggested that they were under strong purifying selection, suggesting that they
270 provide a selective advantage to their hosts on densely populated biofilms.
271 Besides defense mechanisms, other cassette ORFs might also enhance host fitness. For
272 instance, cassettes related to "lipid transport and metabolism" and "amino acid transport and
273 metabolism" were under stronger selection than other ORFs carried by MGEs. The former had a
274 median *pN/pS* of 0.079 (n = 14), and the latter had a median *pN/pS* of 0.107 (n = 23); both were
275 lower than the median of 0.283 for cassette ORFs. Although these two categories were not as

abundant as defense mechanisms in cassettes (**Fig. 5**), their potential to exploit different organic carbon sources may be beneficial for hosts dwelling in the rapidly changing environment of a sinking particle.

Expression of transposase, biofilm, and defense mechanisms ORFs

In addition to examining the genetic potential of microbial communities through analysis of the abundance of transposase, biofilm, and defense mechanism genes in metagenomes, it is also helpful to examine the relative expression of these genes at different depths. Thus, we calculated the *RNA/DNA ratios* (defined as the transcript abundance of a target gene divided by its gene abundance) of biofilm, defense mechanisms, and transposase ORFs in each *Tara* Oceans sample with paired metagenome and metatranscriptome (n = 132). Malaspina samples were excluded from this analysis, because no metatranscriptomes were sequenced.

ORFs related to biofilms and defense mechanisms had greater RNA/DNA ratios in the mesopelagic zone than in the surface and DCM (**Fig. 7A, B, and D**), demonstrating a greater need for defensive genes as deep ocean microbial communities switch toward a particle-associated lifestyle. In contrast, although transposases were more abundant in the mesopelagic zone, their RNA/DNA ratios were similar across all depths (**Fig. 7C and D**). A possible explanation is that although transposases are not upregulated, they sometimes carry beneficial cassette genes, which tend to operate under purifying selection and are increasingly expressed in deep-sea microbial genomes.

Conclusion

Our analysis suggests a cyclic interaction between the particle-associated lifestyle, expanded genome size, and selection for defense mechanism genes to explain the increasing abundance of

transposases with depth (**Fig. 8**). From the surface to the bathypelagic zone, microbial communities shift from a predominantly planktonic lifestyle toward a predominantly particle-associated lifestyle, observed both in the literature [18, 41] and here via the increased transcript abundance of biofilm-associated ORFs in deeper oceans. On densely-populated biofilms, the close proximity between microbes leads to resource competition and allows transposases to spread more rampantly. Particle-associated microbes also tend to have larger genomes, which alleviate the danger from active transposition activities disrupting essential genes [23]. Thus, deep-sea prokaryotic genomes sustain high abundances of MGEs, allowing transposons to actively shape the ecology of microbial biofilm communities with the novel genes they introduce. Specifically, cassette sequences are enriched in defense mechanism genes, which are highly conserved and can benefit microbes engaging in the intense competition on biofilms [37, 39]. Altogether, transposons not only exploit the dense microbial population on particles to propagate but also contribute functional genes that help hosts compete in such an environment.

Investigating the elevated transposase abundance in the deep ocean is crucial for understanding gene flow, competition, and other ecological characteristics of the dark ocean, one of the largest habitats on Earth. Here, we show a strong association between transposase abundance and the particle-associated lifestyle, suggesting that transposons can enable microbial lineages to compete in crowded biofilm-associated habitats. We did not find correlation between the strength of selection and transposase abundance at a genome-resolved scale, so the relaxed selection pressure is unlikely to be the primary driver for the high transposase abundance in the deep ocean. Overall, our results reveal an emerging picture of the ocean as a stratified system in which the deep ocean acts as a gene sharing highway with increased networks of gene exchange, particularly on particles, resulting in large and defensively-oriented genomes in the deep-sea microbial communities, compared to the streamlined and specialized genomes that dominate the surface oceans.

Materials and Methods

Analysis of the Ocean Microbial Reference Catalog v2 (OM-RGC.v2) from the *Tara* Oceans Project. OM-RGC.v2 contained 47 million non-redundant open reading frames (ORFs) [16, 17], and the coverage of each ORF in every metagenomic (n = 187) and metatranscriptomic (n = 180) sample (website). We identified transposases and biofilm-associated ORFs in OM-RGC.v2 through TBLASTN (e-value cutoff of 10^{-5}). The transposase database contained "transposase" and "integrase" genes from the Pfam database [42] (Table S1), and the biofilm database contained representative genes for *lapABCE*, *csgABCD*, *pelABCD*, t*apAB, tasAB, esp,* and *FapC* [43] from the KEGG and UniProt database [44, 45] (Table S3). We identified ORFs of the "defense mechanisms" category through COG annotations in OM-RGC.v2. Finally, the coverage data in OM-RGC.v2 was used to determine the relative abundance of a target gene in a sample.

Mapping and calculating *pN/pS* ratios. To assess the strength of selection of target ORFs, we calculated the *pN/pS* ratio (the proportion of nonsynonymous mutations over the proportion of synonymous mutations) for all ORFs. To do this, we mapped the metagenomic reads against the assembled contigs for each metagenome. Metagenomic reads of the *Tara* Oceans Project and Malaspina Expedition were downloaded from the European Nucleotide Archive (ENA) under study accessions PRJEB402 and PRJEB44456, respectively. Methods for sample collection and Illumina sequencing were described by Sunagawa *et al.* [17] for the *Tara* Oceans Project and Acinas *et al.* [18] for the Malaspina Expedition. For the assembled contigs used as references for mapping of the *Tara* Oceans metagenomes, we downloaded (website) 10 existing co-assembled metagenomes from the 10 oceanic provinces [19]. For the reference contigs of the Malaspina metagenomes, we downloaded (website) the co-assembled metagenome of all 58 samples [18]. We mapped the raw reads of each of

348 the *Tara* Oceans metagenomes to the co-assembly from their corresponding provinces, and all raw
349 reads of each of the Malaspina metagenomes to one co-assembled metagenome using Bowtie2 [46]
350 (v2.2.9; paired-end alignment with default parameters). We calculated *pN/pS* using anvi'o [47] with
351 the script "anvi-script-calculate-pn-ps-ratio" with the "--min-coverage" flag set to 20.
352
353 **Calculating *pN/pS* ratios for individual metagenomes.** The co-assembled *Tara* Oceans
354 metagenomes merged samples of all depths in an oceanic province [19]. To analyze *pN/pS* at
355 different depths, we selected 10 individual *Tara* Oceans metagenomes from each of the surface,
356 DCM, and mesopelagic zone, and 30 individual metagenomic assemblies (Table S4) were
357 downloaded from ENA. We also separately assembled 20 individual metagenomes (Table S4) from
358 the Malaspina Expedition using MEGAHIT (v1.1.3; paired-end assembly with default parameters).
359 We doubled the number of samples because bathypelagic assemblies yielded fewer ORFs than those
360 from the *Tara* Oceans. Using the same procedures to calculate *pN/pS* in co-assembled
361 metagenomes, we calculated the *pN/pS* ratios in individual metagenomes, except that reads in each
362 sample mapped back only to its own assembly.
363
364 **Determination of MAG's depth and lifestyles.** MAGs that originated from co-assemblies in a
365 specific province in the *Tara* Oceans dataset (*e.g.*, Red Sea), as in Tully *et al.* [19], were used to recruit
366 reads from all *Tara* Oceans samples (n = 180) using Bowtie2 (default parameters). Each province
367 was performed separately. SAM files were converted to BAM files using samtools [48] and used to
368 determine RPKM (reads per kilobase pair MAG per Million pair metagenome) as in Graham et al.
369 [49].
370 To assign the depth origin of a MAG from the *Tara* Oceans metagenomes, we summed up
371 its RPKM in surface, DCM, and mesopelagic samples (3 groups, separately). The layer with the

372 highest sum RPKM was the depth of that MAG. All Malaspina MAGs belonged to the bathypelagic
373 layer.
374 To assign the lifestyle of a *Tara* Oceans MAG, we calculated its sum RPKM in samples of
375 0.2-1.6 μm and 0.2-3 μm size fractions (2 groups, separately). A MAG was designated as *particle-*
376 *associated* if its sum RPKM in 0.2-3 μm samples was >2x greater than that in 0.2-1.6 μm samples. A
377 MAG was *planktonic* if its sum RPKM in 0.2-1.6 μm samples was >2x greater than that in 0.2-3 μm
378 samples. MAGs that were not particle-associated or planktonic were designated as *mixed*. Malaspina
379 samples (n = 58) had size-fractions of 0.2-0.8 μm and 0.8-2 μm, but the same method was used to
380 determine the lifestyle of Malaspina MAGs.
381
382 **Integron and cassette sequences detection.** We used Integron Finder (v2) [11] to identify
383 cassette sequences from the *Tara* Ocean and Malaspina metagenomes. Integron Finder uses
384 HMMER to locate the integron-integrase *intI*, which is conserved for most integrons. Then, cassette
385 sequences are identified with the near-palindromic flaking regions. Anvi'o [47] was used to identify
386 and annotate ORFs on metagenomic contigs using the script 'anvi-run-ncbi-cogs.' If the anvi'o ORF
387 calls from Prodigal [50] were within 100bps (start and stop) of the Integron Finder's cassette calls,
388 the anvi'o COG annotations were used for those cassette ORFs.
389
390 **Statistical analysis**
391 We used two-tailed *t*-tests to compare the differences between means of two variables, and statistical
392 significance was assumed $P < 0.05$. The metagenomic/metatranscriptomic abundances, *pN*/*pS*
393 ratios, and RNA/DNA ratios were log transformed before any statistical analysis. All analyses were
394 performed using base R packages (v4.0.2).
395

Code and data availability.

All Python, R scripts, code explanation, and raw data for analysis are publicly accessible on GitHub at https://github.com/carleton-spacehogs/transposase-deep-ocean.

Acknowledgments

We would like to thank Dr. Shinichi Sunagawa for kindly providing information about the *Tara* Oceans datasets and the Ocean Microbial Reference Catalog, and Dr. Silvia G. Acinas for providing information about the Malaspina Deep Ocean dataset. Dr. Murat Eren provided help with anvi'o, and Mike Tie provided assistance with server administration. Funding for JZ and OK was provided by grants from the Towsley Endowment at Carleton College, and funding for TdRH was provided by the Summer Science Fellows program at Carleton College.

Competing Interest

The authors declare that they have no financial or non-financial competing interests.

References

1. Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005; **3**: 722–732.
2. Transposons. http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518. Accessed 30 Nov 2021.
3. Muñoz-López M, García-Pérez JL. DNA transposons: nature and applications in genomics. *Curr Genomics* 2010; **11**: 115–128.
4. Collis C M, Hall R M. Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrob Agents Chemother* 1995; **39**: 155–162.
5. Rankin DJ, Rocha EPC, Brown SP. What traits are carried on mobile genetic elements, and why? *Heredity* 2011; **106**: 1–10.
6. Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* 2001; **2**: 376–381.
7. Jones JM, Grinberg I, Eldar A, Grossman AD. A mobile genetic element increases bacterial host fitness by manipulating development. *Elife* 2021; **10**: e65924.
8. Casacuberta E, González J. The impact of transposable elements in environmental adaptation. *Mol Ecol* 2013; **22**: 1503–1517.
9. Aziz RK, Breitbart M, Edwards RA. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res* 2010; **38**: 4207–4217.
10. Vigil-Stenman T, Ininbergs K, Bergman B, Ekman M. High abundance and expression of transposases in bacteria from the Baltic Sea. *ISME J* 2017; **11**: 2611–2623.
11. Cury J, Jové T, Touchon M, Néron B, Rocha EP. Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res* 2016; **44**: 4539–4550.
12. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U, et al. Community

genomics among stratified microbial assemblages in the ocean's interior. *Science* 2006; **311**: 496–503.

13. Konstantinidis KT, Braff J, Karl DM, DeLong EF. Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl Environ Microbiol* 2009; **75**: 5345–5355.
14. Worden AZ, Cuvelier ML, Bartlett DH. In-depth analyses of marine microbial community genomics. *Trends Microbiol* 2006; **14**: 331–336.
15. Brazelton WJ, Baross JA. Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J* 2009; **3**: 1420–1424.
16. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh H-J, Cuenca M, et al. Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell* 2019; **179**: 1068–1083.e21.
17. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. *Science* 2015.
18. Acinas SG, Sánchez P, Salazar G, Cornejo-Castillo FM, Sebastián M, Logares R, et al. Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Communications Biology* 2021; **4**: 1–15.
19. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Scientific Data* 2018; **5**: 1–8.
20. Ganesh S, Parris DJ, DeLong EF, Stewart FJ. Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J* 2014; **8**: 187–211.
21. Stewart FJ. Where the genes flow. *Nat Geosci* 2013; **6**: 688–690.
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 2015; **25**: 1043–1055.

23. Touchon M, Rocha EPC. Causes of Insertion Sequences Abundance in Prokaryotic Genomes. *Mol Biol Evol* 2007; **24**: 969–981.

24. Meziti A, Rodriguez-R LM, Hatt JK, Peña-Gonzalez A, Levy K, Konstantinidis KT. The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample. *Appl Environ Microbiol* 2021; **87**.

25. Garcia SL, Buck M, McMahon KD, Grossart H-P, Eiler A, Warnecke F. Auxotrophy and intrapopulation complementary in the 'interactome' of a cultivated freshwater model community. *Mol Ecol* 2015; **24**: 4449–4459.

26. Rodríguez-Gijón A, Nuy JK, Mehrshad M, Buck M, Schulz F, Woyke T, et al. A Genomic Perspective Across Earth's Microbiomes Reveals That Genome Size in Archaea and Bacteria Is Linked to Ecosystem Type and Trophic Strategy. *Front Microbiol* 2021; **12**: 761869.

27. Costello MJ, Chaudhary C. Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation. *Curr Biol* 2017; **27**: R511–R527.

28. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci U S A* 2006; **103**: 12115–12120.

29. Simmons SL, Dibartolo G, Denef VJ, Goltsman DSA, Thelen MP, Banfield JF. Population genomic analysis of strain variation in Leptospirillum group II bacteria involved in acid mine drainage formation. *PLoS Biol* 2008; **6**: e177.

30. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 1991; **351**: 652–654.

31. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation

482 landscape of the human gut microbiome. *Nature* 2013; **493**: 45–50.

483 32. Massana R. Eukaryotic Picoplankton in Surface Oceans. *Annu Rev Microbiol* 2011; **65**: 91–110.

484 33. Escobar-Páramo P, Ghosh S, DiRuggiero J. Evidence for Genetic Drift in the Diversification

485 of a Geographically Isolated Population of the Hyperthermophilic Archaeon Pyrococcus.

486 *Molecular Biology and Evolution* . 2005. , **22**: 2297–2303

487 34. Flynn KJ, Swanson MS. Integrative conjugative element ICE-βox confers oxidative stress

488 resistance to Legionella pneumophila in vitro and in macrophages. *MBio* 2014; **5**: e01091–14.

489 35. Sullivan JT, Ronson CW. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that

490 integrates into a phe-tRNA gene. *Proc Natl Acad Sci U S A* 1998; **95**: 5145–5149.

491 36. Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial

492 realm. *Nat Microbiol* 2018; **3**: 754–766.

493 37. Nadell CD, Drescher K, Foster KR. Spatial structure, cooperation and competition in biofilms.

494 *Nat Rev Microbiol* 2016; **14**: 589–600.

495 38. Basler M, Ho BT, Mekalanos JJ. Tit-for-tat: type VI secretion system counterattack during

496 bacterial cell-cell interactions. *Cell* 2013; **152**: 884–894.

497 39. Hayes CS, Aoki SK, Low DA. Bacterial contact-dependent delivery systems. *Annu Rev Genet*

498 2010; **44**: 71–90.

499 40. Ho BT, Dong TG, Mekalanos JJ. A view to a kill: the bacterial type VI secretion system. *Cell*

500 *Host Microbe* 2014; **15**: 9–21.

501 41. Mestre M, Ruiz-González C, Logares R, Duarte CM, Gasol JM, Sala MM. Sinking particles

502 promote vertical connectivity in the ocean microbiome. *Proc Natl Acad Sci U S A* 2018; **115**:

503 E6799–E6807.

504 42. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam:

505 The protein families database in 2021. *Nucleic Acids Res* 2021; **49**: D412–D419.

43. Taglialegna A, Matilla-Cuenca L, Dorado-Morales P, Navarro S, Ventura S, Garnett JA, et al. The biofilm-associated surface protein Esp of Enterococcus faecalis forms amyloid-like fibers. *npj Biofilms and Microbiomes* 2020; **6**: 1–12.

44. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* 2021; **49**: D545–D551.

45. The UniProt Consortium, Bateman A, Martin M-J, Orchard S, Magrane M, Agivetova R, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2020; **49**: D480–D489.

46. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**: 357–359.

47. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 2021; **6**: 3–6.

48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.

49. Graham ED, Heidelberg JF, Tully BJ. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* 2018; **12**: 1861–1866.

50. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.

Figure Legends:

Fig. 1: Samples from the deep ocean and the particle-associated size fraction have higher transposase abundance. The abundance of transposase ORFs in metagenomic samples from the *Tara* Oceans and Malaspina Expedition, separated by depth and filter size fraction. SRF: surface; DCM: deep chlorophyll maximum; MES: mesopelagic zone; BAT: bathypelagic zone. For *Tara* Oceans metagenomes, the *particle-associated* size fraction was 0.22-3.0 μm, and *planktonic* was 0.22-1.6 μm. For Malaspina metagenomes, the particle-associated size fraction was 0.8-20 μm, and planktonic was 0.2-0.8 μm. The widths of the boxplots reflect the sample count in each category (counts were shown on the right).

Fig. 2: The gene potential (DNA) and transcript abundance (RNA) of transposases correlates with those of biofilm genes in all depths except the bathypelagic zone. **A** The correlation between the abundance of transposases and biofilm-associated ORFs in metagenomes (n = 236), separated by depths. **B** The same correlation in metatranscriptomes (n = 181).

Table 1: Multiple Stepwise Regression for log-transformed transposase abundance, *log(%-transposase)*, of 1562 previously generated MAGs. An ANOVA test was performed on each newly added covariate. ***depth:taxon**: the interaction term between depth and taxonomy. ***gsize***: the estimated complete genome size of a MAG.

Fig. 3: Deep-sea and particle-associated microbial populations tend to have larger genomes, which correlate with high transposase abundance in those populations. **A** The estimated complete genome size of MAGs (complete genome size = number of base pairs in a MAG / % completeness),

grouped by depth. Numbers on box plots indicated the count of MAGs in each group. **B** The estimated complete genome size of MAGs, grouped by depth and lifestyle. The boxplot's width reflects the number of MAGs in each group. **C** The transposase abundance in MAGs, grouped by depth and complete genome size. **D,E** Individual classes of MAGs grouped by depth and lifestyle. See Methods for the determination of depth and lifestyle of a MAG.

Fig. 4: The co-occurrence of relaxed selection pressure and high transposase abundance in deep ocean microbial communities does not support a correlation between the two at a genome-resolved basis. **A** The *pN/pS* ratio of ORFs in metagenomes, grouped by depth and whether this ORF was a transposase. The number on the right denotes the count of ORFs in each group. **B** The relationship between the selection pressure (median pN/pS) and the transposase abundance of a MAG, separated by depth. Only MAGs with ≥ 100 *pN/pS* were computed for median pN/pS.

Fig. 5: Cassette sequences include high proportions of defense mechanism genes, which are more abundant in microbial communities that rely heavily on biofilms. **A** The distribution of COG functional categories in cassette and metagenome ORFs. *TM: transport and metabolism. 1951 cassette ORFs with COG function calls were identified from the 10 co-assembled *Tara* metagenomes, and 359 from the co-assembled Malaspina metagenome. 5 million ORFs with COG function calls were then sampled from the *Tara* Oceans and Malaspina metagenomes according to the ratio of identified cassettes described above. **B** The correlation between the abundance of biofilm-associated and defense mechanism ORFs, separated by depth.

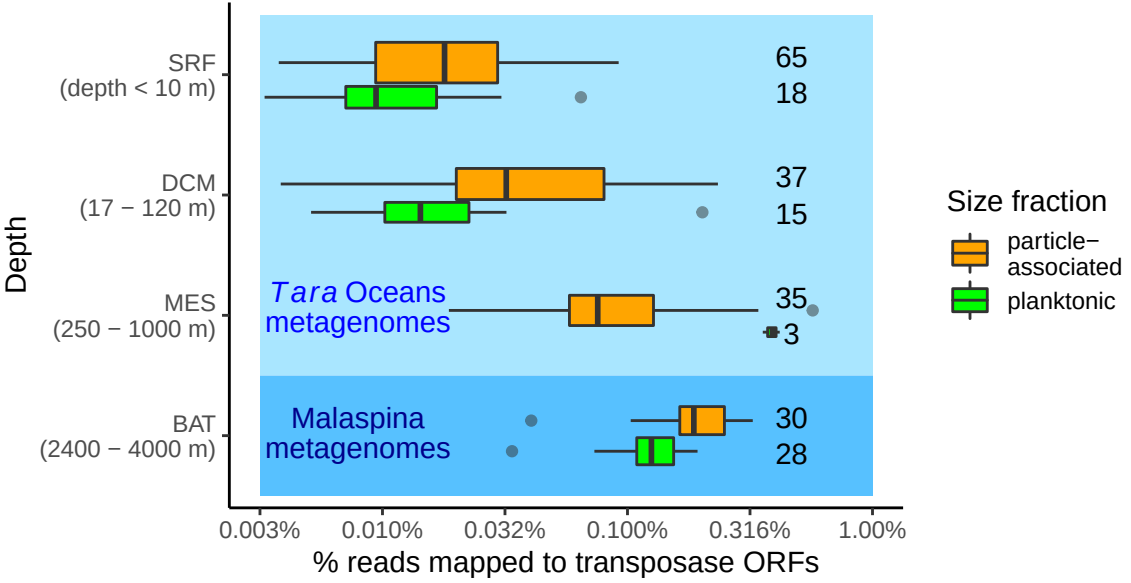
Fig. 6: The *pN/pS* for non-cassette ORFs and different categories of cassette ORFs. 2-tailed *t*-tests were performed on log-transformed *pN/pS*. *** $P < 0.001$, **** $P < 0.0001$.

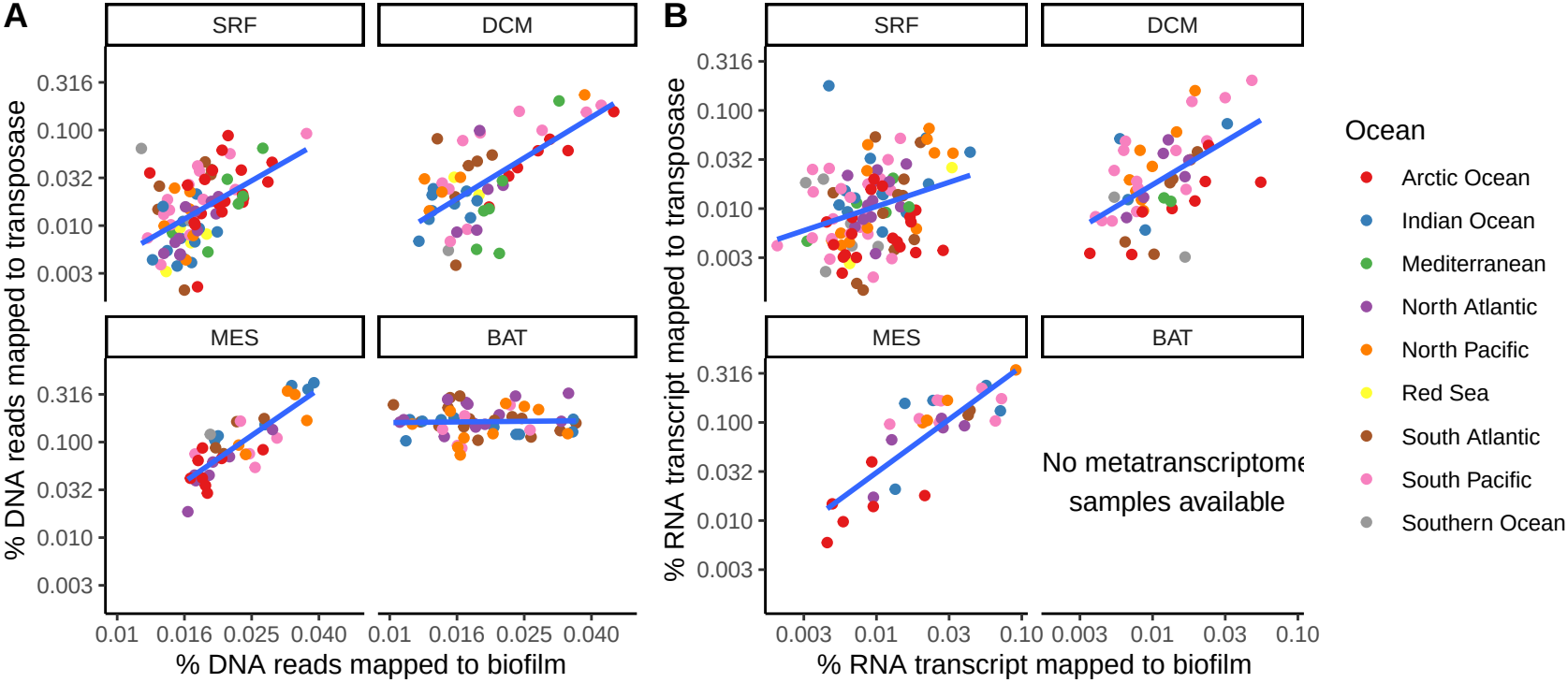
Fig. 7: Genes related to biofilm and defense mechanisms are upregulated in the mesopelagic zone, but transposases are not. **A, B, C** Median metagenomic and metatranscriptomic abundance of biofilm-associated, defense mechanism, and transposase ORFs in each depth. Error bars reflect the 25 and 75 percentiles. Each sample had its metagenome and metatranscriptome sequenced. **D** Log-transformed RNA/DNA ratios of target genes in each sample, separated by depth. 2-tailed *t*-tests were performed on log-transformed RNA/DNA ratios. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

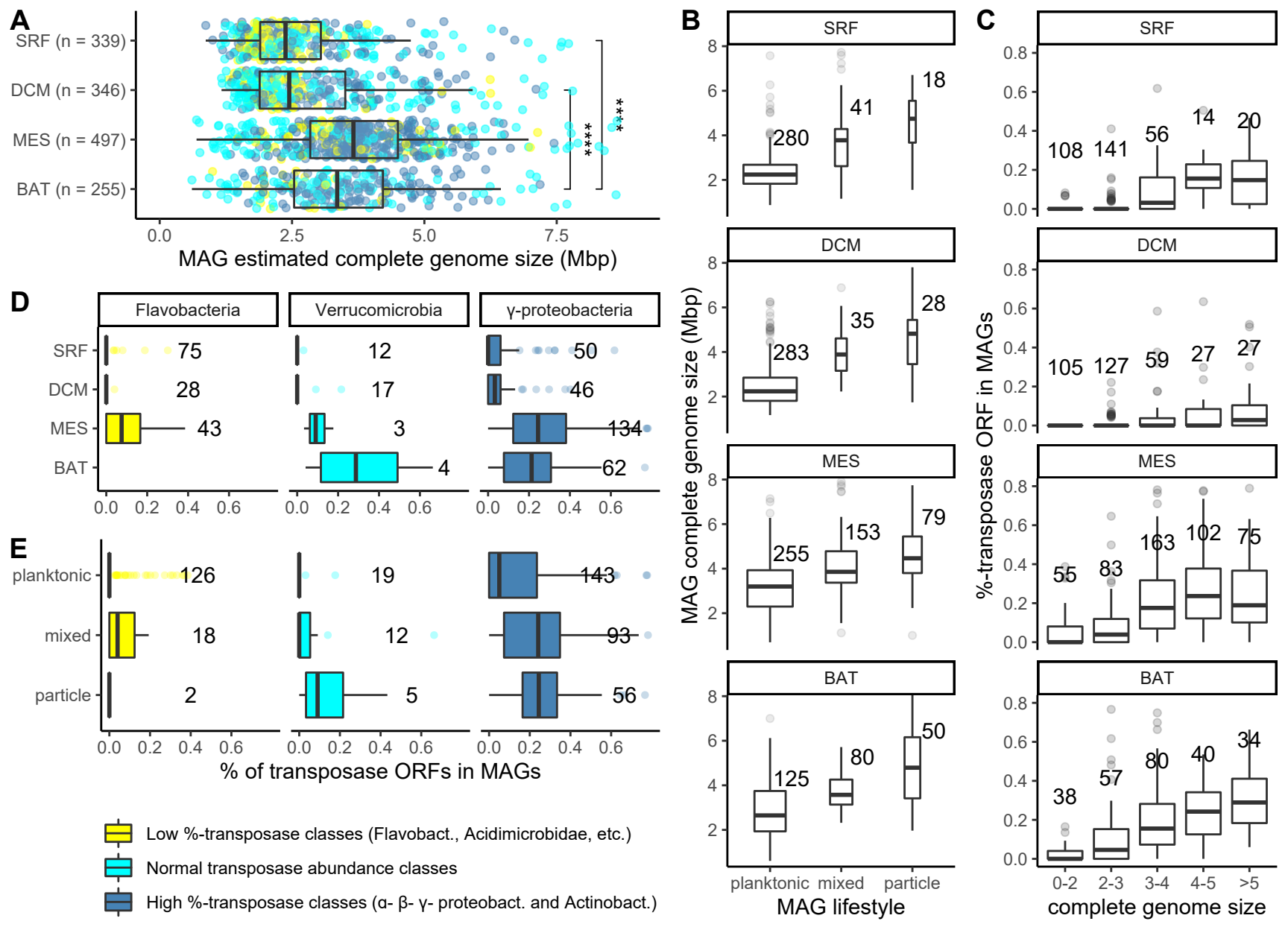
Fig. 8: Linking a particle-associated lifestyle to the high abundance of transposons in the deep ocean.

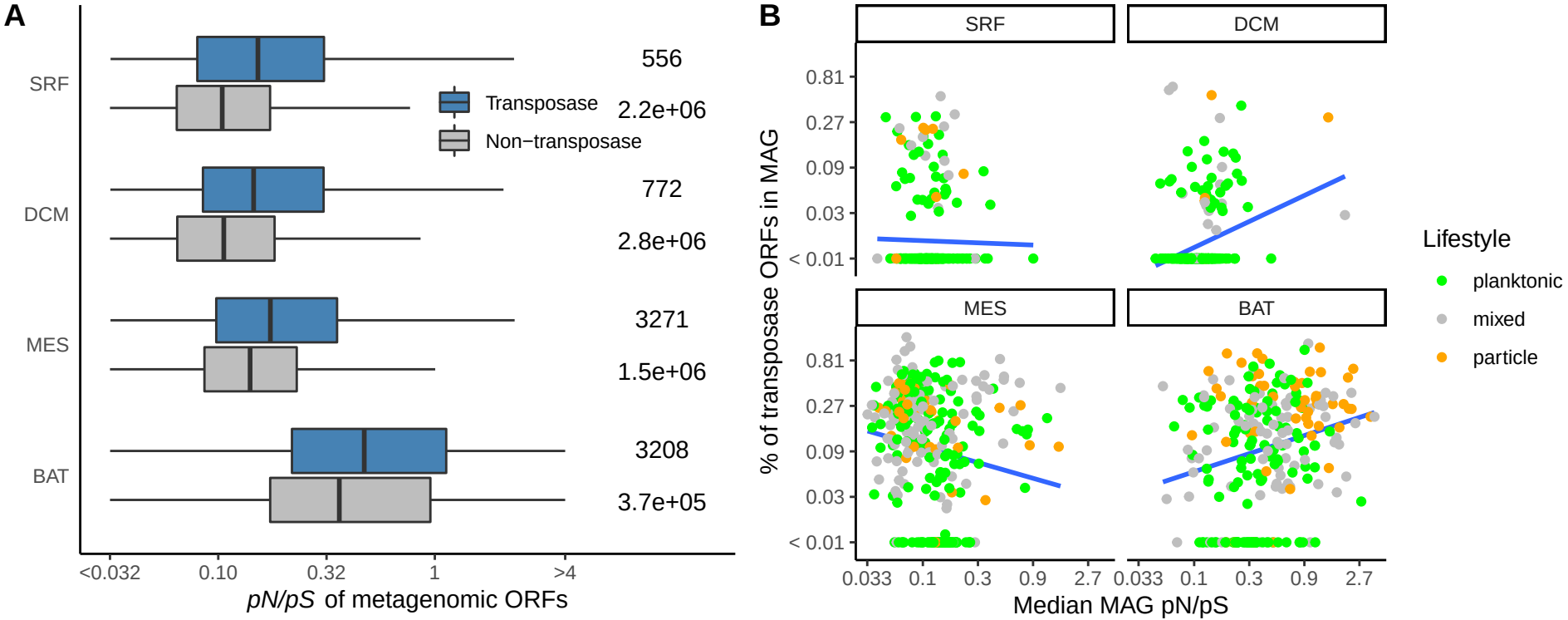
Table 1. Multiple Stepwise Regression for log-transformed transposase abundance in MAGs.

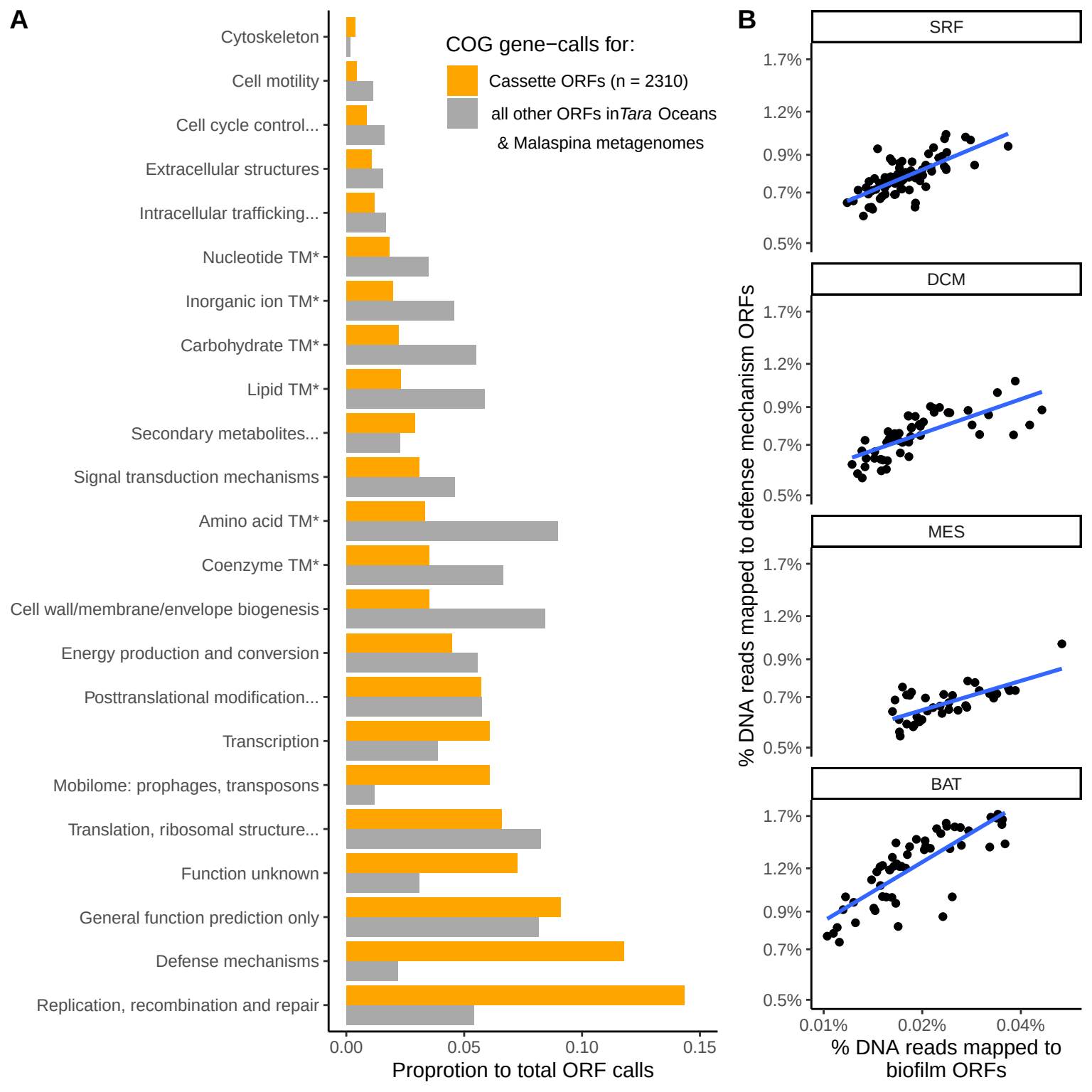
Explanatory Variable(s)	*P* for F-test	Cumulative R^2
depth	2.2×10^{-16}	0.291
depth + genome size (gsize)	2.2×10^{-16}	0.473
depth + lifestyle	2.2×10^{-16}	0.371
depth + gsize + lifestyle	6.7×10^{-8}	0.486
(as above) + taxon	2.2×10^{-16}	0.593
(as above) + depth:taxon	0.00016	0.622
(as above) + $gsize^2$	1.3×10^{-8}	0.633

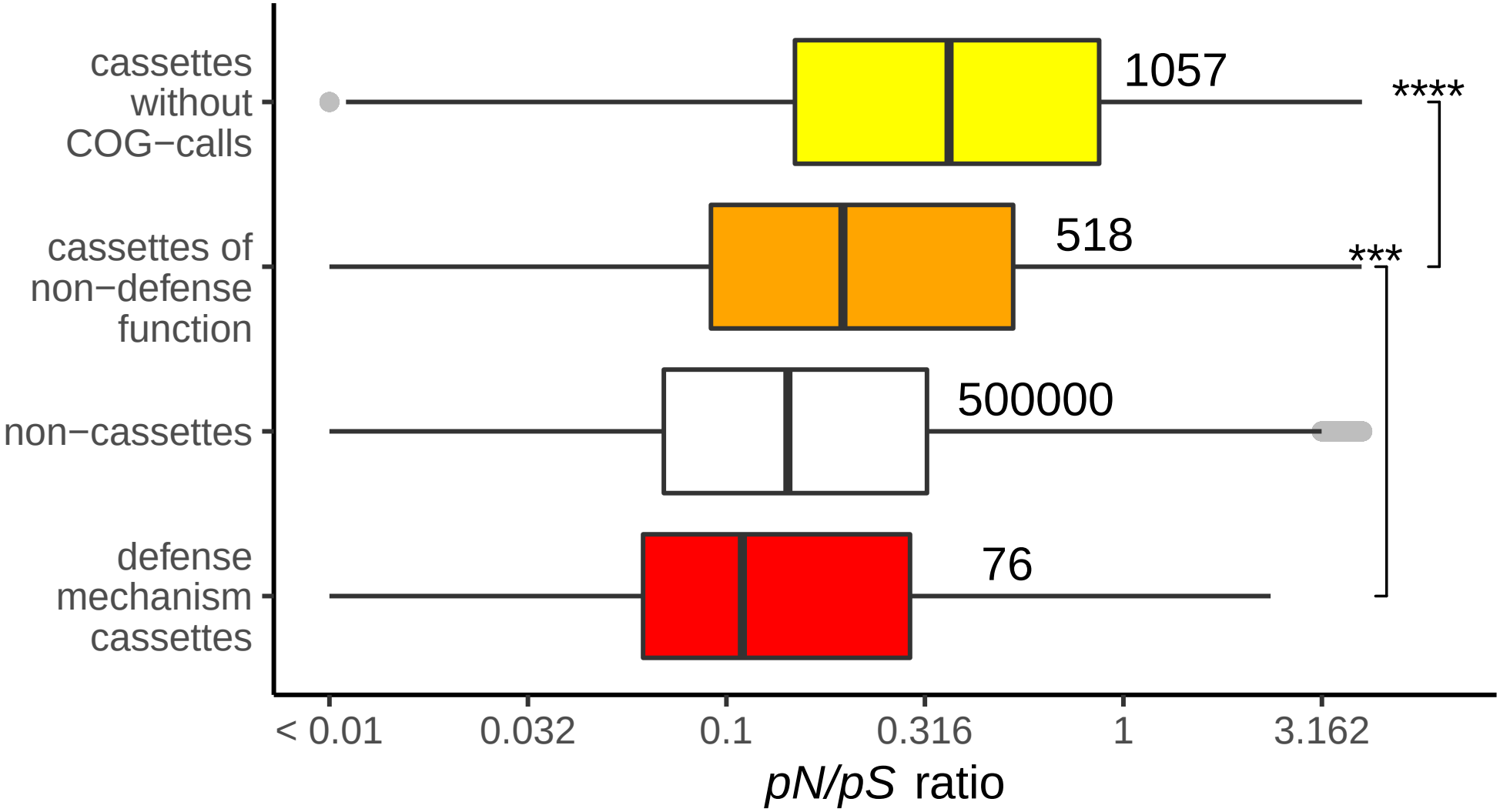

Depth
SRF
(depth < 10 m)
DCM
(17 − 120 m)
MES
(250 − 1000 m)
BAT
(2400 − 4000 m)
Tara Oceans
metagenomes
Malaspina
metagenomes
65
18
37
15
35
3
30
28
0.003%
0.010%
0.032%
0.100%
0.316%
1.00%
% reads mapped to transposase ORFs
Size fraction
particle−
associated
planktonic

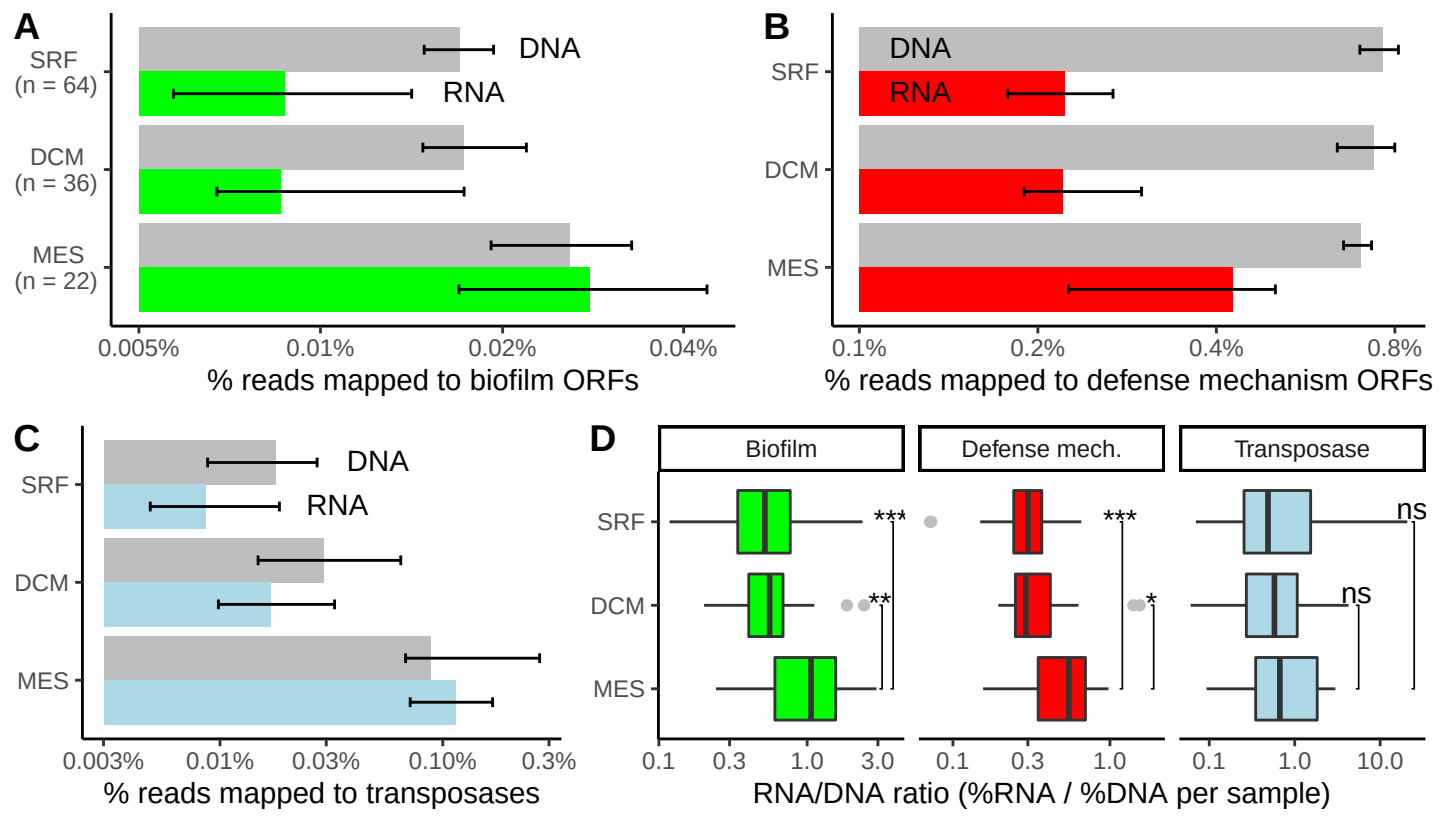

A
SRF
DCM
MES
BAT
% DNA reads mapped to transposase
0.316
0.100
0.032
0.010
0.003
0.01
0.016
0.025
0.040
% DNA reads mapped to biofilm
B
% RNA transcript mapped to transposase
0.003
0.01
0.03
0.10
% RNA transcript mapped to biofilm
No metatranscriptome samples available
Ocean
Arctic Ocean
Indian Ocean
Mediterranean
North Atlantic
North Pacific
Red Sea
South Atlantic
South Pacific
Southern Ocean

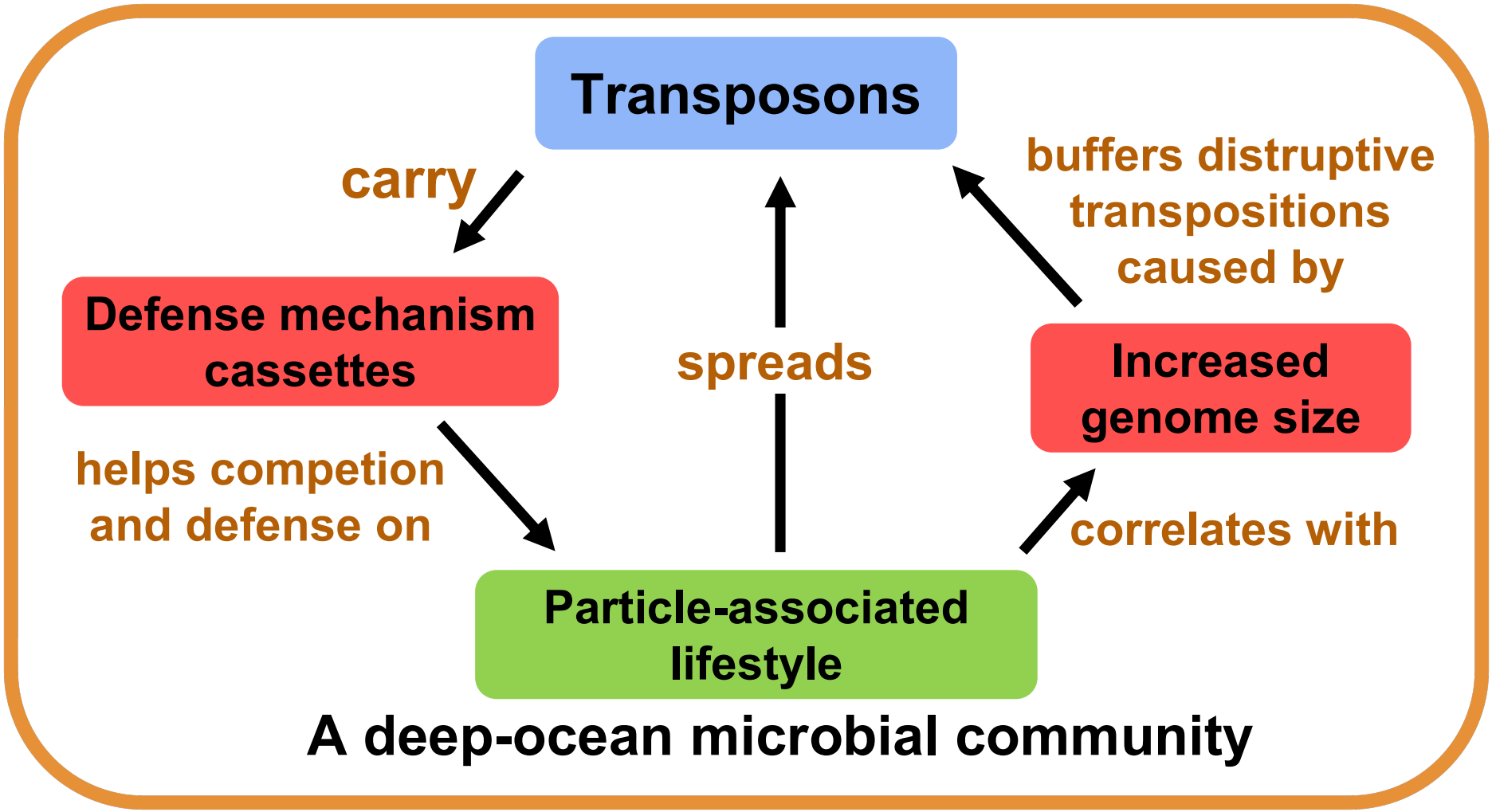

A
SRF (n = 339)
DCM (n = 346)
MES (n = 497)
BAT (n = 255)
****
****
0.0
2.5
5.0
7.5
MAG estimated complete genome size (Mbp)
B
SRF
DCM
MES
BAT
280
41
18
283
35
28
255
153
79
125
80
50
MAG complete genome size (Mbp)
planktonic
mixed
particle
MAG lifestyle
C
SRF
DCM
MES
BAT
108
141
56
14
20
105
127
59
27
27
55
83
163
102
75
38
57
80
40
34
%-transposase ORF in MAGs
0-2
2-3
3-4
4-5
>5
complete genome size
D
Flavobacteria
Verrucomicrobia
γ-proteobacteria
SRF
DCM
MES
BAT
75
28
43
12
17
3
4
50
46
134
62
E
planktonic
mixed
particle
126
18
2
19
12
5
143
93
56
% of transposase ORFs in MAGs
Low %-transposase classes (Flavobact., Acidimicrobidae, etc.)
Normal transposase abundance classes
High %-transposase classes (α- β- γ- proteobact. and Actinobact.)


A
SRF
DCM
MES
BAT
Transposase
Non−transposase
556
2.2e+06
772
2.8e+06
3271
1.5e+06
3208
3.7e+05
<0.032
0.10
0.32
1
>4
pN/pS of metagenomic ORFs
B
SRF
DCM
MES
BAT
% of transposase ORFs in MAG
0.81
0.27
0.09
0.03
< 0.01
0.033
0.1
0.3
0.9
2.7
Median MAG pN/pS
Lifestyle
planktonic
mixed
particle


A
COG gene−calls for:
Cassette ORFs (n = 2310)
all other ORFs in *Tara* Oceans & Malaspina metagenomes
Cytoskeleton
Cell motility
Cell cycle control...
Extracellular structures
Intracellular trafficking...
Nucleotide TM*
Inorganic ion TM*
Carbohydrate TM*
Lipid TM*
Secondary metabolites...
Signal transduction mechanisms
Amino acid TM*
Coenzyme TM*
Cell wall/membrane/envelope biogenesis
Energy production and conversion
Posttranslational modification...
Transcription
Mobilome: prophages, transposons
Translation, ribosomal structure...
Function unknown
General function prediction only
Defense mechanisms
Replication, recombination and repair
0.00
0.05
0.10
0.15
Proprotion to total ORF calls
B
SRF
DCM
MES
BAT
1.7%
1.2%
0.9%
0.7%
0.5%
% DNA reads mapped to defense mechanism ORFs
0.01%
0.02%
0.04%
% DNA reads mapped to biofilm ORFs


cassettes without COG−calls
cassettes of non−defense function
non−cassettes
defense mechanism cassettes
1057
518
500000
76
****
***
< 0.01
0.032
0.1
0.316
1
3.162
pN/pS ratio


A
SRF
(n = 64)
DCM
(n = 36)
MES
(n = 22)
DNA
RNA
0.005%
0.01%
0.02%
0.04%
% reads mapped to biofilm ORFs
B
SRF
DCM
MES
DNA
RNA
0.1%
0.2%
0.4%
0.8%
% reads mapped to defense mechanism ORFs
C
SRF
DCM
MES
DNA
RNA
0.003%
0.01%
0.03%
0.10%
0.3%
% reads mapped to transposases
D
Biofilm
Defense mech.
Transposase
SRF
DCM
MES
***
**
***
*
ns
ns
0.1
0.3
1.0
3.0
0.1
0.3
1.0
0.1
1.0
10.0
RNA/DNA ratio (%RNA / %DNA per sample)


Transposons
carry
buffers distruptive transpositions caused by
Defense mechanism cassettes
spreads
Increased genome size
helps competion and defense on
correlates with
Particle-associated lifestyle
A deep-ocean microbial community