Comparative whole-genome approach to identify traits underlying microbial interactions

L. Zoccarato (1), D. Sher (2), T. Miki (3), D. Segrè (4,5), H.P. Grossart (1,6,7)

(1) Department Experimental Limnology, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany
(2) Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa, Israel
(3) Department of Environmental Solution Technology, Ryukoku University, Kyoto, Japan
(4) Departments of Biology, Biomedical Engineering, Physics, Boston University, Boston MA, USA
(5) Bioinformatics Program & Biological Design Center, Boston University, Boston MA, USA
(6) Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin, Germany
(7) Institute of Biochemistry and Biology, Potsdam University, Potsdam, Germany

ABSTRACT

Interactions among microorganisms affect the structure and function of microbial communities, with potentially far-reaching effects on ecosystem health and biogeochemical cycles. The functional traits mediating microbial interactions are known for several model organisms, but the prevalence of these traits across microbial diversity is unknown. We developed a new genomic approach to systematically explore the occurrence of metabolic functions and specific interaction traits (e.g. the production of vitamins, siderophores, antimicrobial compounds and phytohormones), and apply this approach to 473 sequenced genomes from marine bacteria. We identify 48 coherent genome functional clusters (GFCs), that are partly consistent with known bacterial ecotypes (e.g. within pico-Cyanobacteria and *Vibrio* taxa) and identify putative new ones (e.g. *Marinobacter*, *Alteromonas* and *Pseudoalteromonas*). Interaction traits such as the production of and resistance

towards antimicrobial compounds and the production of phytohormones are widely distributed among the GFCs, while other traits are less common (e.g. siderophores and secretion systems are found in 32% of genomes or less). Several GFCs lack the ability to produce B vitamins, suggesting that these metabolites represent essential trading goods for many bacteria. Alpha- and Gammaproteobacteria encode many interaction traits, and appear particularly poised to interact both synergistically and antagonistically with co-occurring bacteria and phytoplankton. Linked Trait Clusters (LTCs) group chemotaxis, motility and adhesion with regulatory systems involved in virulence and biofilm formation, and suggest that type-4 secretion systems may be used to inject the hormone indole acetic acid into target phytoplankton cells. Similar efficient processing and representation of multidimensional microbial functional information will be increasingly essential for translating genomes into ecosystem understanding across biomes.

Keywords: marine bacteria, phytoplankton, interactions, functional traits, vitamins, siderophores, phytohormones

Introduction

Interactions between microorganisms, such as symbiosis, competition and allelopathy, are a central feature of microbial communities [1]. In aquatic environments, heterotrophic bacteria interact with microbial primary producers (phytoplankton) in many ways, potentially affecting the growth of both organisms [2,3] and with consequences for ecosystem functioning and biogeochemical cycles [4,5]. For instance, heterotrophic bacteria consume up to 50% of the organic matter released by phytoplankton, significantly affecting the dynamics of the huge pool of dissolved organic carbon in the oceans [6]. Thus, if and how a bacterium can interact with other bacteria and eukaryotes may have important consequences for the biological carbon pump in the current and future oceans [7,8].

Recent studies, using specific model organisms in binary co-cultures, have started to elucidate mechanisms underlying marine microbial interactions (mostly between bacteria and phytoplankton). Many of these interactions are mediated by the exchange of metabolites used for growth or respiration. For example, bacteria associated to phytoplankton (*i.e.* within the phycosphere, [3,9]), gain access to labile organic carbon released by the primary producers, e.g. amino acids and small sulfur-containing compounds [10–15]. In return, phytoplankton benefit from an increased accessibility to nutrients via bacteria-mediated processes, e.g. nitrogen and phosphorus remineralization [16], vitamin supply [11,17] and iron scavenging via formation of siderophores [18,19]. In addition to such metabolic interactions, direct signaling may also occur between bacteria and phytoplankton, with heterotrophic bacteria directly controlling the phytoplankton cell cycle through phytohormones [10,20] or harming it using toxins [15,21]. Through such specific infochemical-mediated interactions, bacteria may also directly affect the rate of release of organic carbon from phytoplankton, as well as rates of mortality and aggregation [15,20,22].

While much is known about microbial interactions involving model organisms such as species of Roseobacter [10,15–17,21], *Alteromonas* [23–25], Cyanobacteria [26] or *Vibrio* [27,28], little is known to what extent

the potential for such interactions occurs in other species or microbial lineages. The few experimental studies that measure microbial interactions across diversity (e.g. [29–31]) are usually limited in their phylogenetic scope and are performed under conditions which are very different from those occurring in the natural marine environment. However, the knowledge obtained from model organisms on the molecular mechanisms underlying microbial interactions and the increasing availability of high-quality genomes present an opportunity to map known interaction mechanisms to a large set of bacterial species from various taxa. Here, we re-analyze the previously published genomes of 421 diverse marine bacteria, providing an "atlas" of their functional metabolic capacity. The atlas includes also 52 bacteria isolated from extreme marine habitats, human and plant roots that are meant to serve as functional out-groups and represent well known symbiotic plant bacteria (i.e. Rhizobacteria). In particular, we focus on genomic traits likely to be involved in mediating interactions between heterotrophic bacteria and other organisms. These traits are estimated based on the presence of KEGG modules or of genes encoding for transporters, phytohormones and secondary metabolite production. Trait-based approaches offer a new perspective to investigating microbial diversity with a more mechanistic understanding [32] and have been used in some specific cases to highlight putative bacterial interactions (e.g., reference [33]). As shown below, our results identify clusters of organisms whose genomes encode similar functional capacity (defined as genome functional clusters: GFC). We propose that organisms belonging to the same GFC are likely to interact in similar ways with other microorganisms. We also identify clusters of traits that are statistically linked, and propose that these linked trait clusters (LTCs) may have evolved to function together in microbial interactions. GFCs and LTCs provide a framework to extend the knowledge on microbial interactions gained from specific model systems, leading to testable hypotheses as to the prevalence of microbial interactions across bacterial diversity.

Results & Discussion (subheadings)

Genome functional clusters, a framework to capture potential new ecotypes

To obtain an overview on the functional capabilities of marine bacteria, we re-annotated a set of 473 high-quality genomes, and analyzed them using a trait-based workflow, which focuses on the detection of complete genetic traits rather than on the presence of individual genes (Supplementary Fig. 1, Supplementary information). Similar to the gene set enrichment analysis, this approach could provide a more robust interpretation of the genetic information by incorporating prior biological knowledge (e.g. biochemical or signaling pathways) [34]. Our analysis was based on metabolic KEGG modules, with the addition of specific traits which encode for mechanisms known to mediate interactions (Supplementary Table 1). These "interaction traits" include the production of certain secondary metabolites, secretion systems, vitamins and vitamin transporters, siderophores and phytohormones. As shown in Fig. 1, some traits were found across almost all genomes ("core" traits). These include basic cellular metabolisms (nucleotides, amino acids and carbohydrates), a few transport systems and cofactor biosynthesis. Other traits were found only in specific groups of organisms. Based on the patterns observed in Fig. 1, we could cluster the genomes into 48 Genome Functional Clusters (GFCs; see Supplementary Information for a more detailed description of the clustering method). Within each GFC, all genomes are inferred to encode similar traits, and thus these GFCs are expected to be coherent in terms of their functional and metabolic capacity, including the ways in which the related organisms interact with other microbes.

We next asked to what extent do the GFCs correlate with phylogeny? If all bacteria in a GFC belong to the same phylogenetic clade, and all bacteria in this phylogenetic clade are found together in the same GFC, this would imply that the phylogenetic affiliation of these bacteria can predict the

traits encoded in their genomes. We defined such GFCs as *phylogenetically coherent* and measure the coherence at multiple taxonomic ranks (i.e. genus, family, order, class or phylum; see Supplementary information and Supplementary Fig. 2a for more information). According to this metric, 22 out of the 49 GFCs were coherent (monophyletic), most of them at the genus level (Supplementary Fig. 2b,d,e). In these coherent GFCs, which include all Firmicutes and half of the Alpha- and Gammaproteobacteria GFCs (Supplementary Fig. 2b,c), the functional capacity (genome-encoded traits) seems to follow phylogeny closely. Of the remaining 29 GFCs, 22 were paraphyletic, i.e. GFCs contained genomes from multiple phylogenetic groups and some phylogenetic groups were partitioned among multiple GFCs (Supplementary Fig. 2b-e). The paraphyletic GFCs scored their highest phylogenetic coherence at the class level (the remaining half of the Alpha- and Gammaproteobacteria GFCs) and Genus level (all Cyanobacteria GFCs). In these GFCs a significant functional similarity (e.g. functional redundancy) existed between distantly related genomes. Finally, five GFCs were polyphyletic, i.e. they include organisms from multiple phyla. Some of these GFCs group organisms isolated from extreme environments (e.g. thermal vents or hyper saline environments; GFCs 12 and 44) and marine sediment (GFC 35). Such genomes were added in the analysis as outer groups and, therefore, their phylogenetic diversity was not adequately covered (Supplementary Table 2 and Supplementary information). Overall, the observation that functionality does not strictly follow phylogeny makes it difficult to infer the function of a community from its taxonomic structure, as previously shown across the global oceans [35,36].

Importantly, some of the GFCs correspond to previously defined ecotypes or to ecologically-defined species. For example, GFC 3 comprised all genomes of the order SAR11 (Pelagibacterales) (Supplementary Table 2), defining a group of highly abundant taxa with streamlined genomes adapted to thrive under oligotrophic conditions [37,38]. Similarly, pico-Cyanobacterial genomes clustered separately from all other Cyanobacteria, and, within the pico-Cyanobacterial group, GFC

29 consisted of exclusively high-light *Prochlorococcus* strains. GFC 28 comprised mostly low-light *Prochlorococcus* and GFC 30 comprised some low-light type IV *Prochlorococcus* and *Synechococcus* strains. Thus, while none of the Cyanobacterial GFCs were strictly coherent, they were consistent to a large extent with previous ecological and genomic studies (reviewed by [26]). Finally, genomes belonging to the family Vibrionaceae were clustered in three different GFCs (47, 48 and 49). GFC 48 mostly grouped *Vibrio alginolyticus*, which have been shown to prefer zooplankton hosts over other organic matter particles [39], whereas GFC 47 harbored free-leaving and non-pathogenic strains of *V. furnissii* and *V. natriegens* [40,41]. GFC 49 included several pathogenic strains of more generalist *Vibrio* species characterized by a wide range of aquatic hosts (e.g. *V. splendidus*; [42]), as well as a few human pathogens (*V. cholerae* and *V. vulnificus;* [40]). Along with *Vibrio* genomes, GFC 49 contained also genomes from additional taxa (e.g., *Photobacterium* (3 strains) and *Psychromonas* (2 strains)) which are also potential pathogens or gut endobionts of crustacean and marine snails [43,44].

The correspondence between the aforementioned GFCs and known bacterial ecotypes highlights that our analytical framework is able to categorize bacterial diversity into ecologically relevant units. Therefore, it is interesting that three groups of Gammaproteobacteria, i.e. *Alteromonas*, *Marinobacter* and *Pseudoalteromonas*, each formed their own GFC (4, 21 and 31, respectively). These organisms are all known as copiotrophs often associated with organic particles or phytoplankton [45–48]. However, these GFCs can be distinguished by the presence of different metabolic and interaction traits (see following chapters for the discussion on related biological implications; Supplementary Fig. 3). For example, *Pseudoalteromonas* and *Alteromonas* bear more traits involved in the resistance against antimicrobial compounds (especially *Pseudoalteromonas*), regulation for osmotic and redox stresses. They also have similar vitamin B1 and siderophore transporters, which are different from those encoded by *Marinobacter*. *Marinobacter* possess several more phosphonate and amino acid transporters, as well as specific regulatory systems for

164 adhesion (e.g. alginate and type 4 fimbriae production) and chemotaxis. These differences suggest
165 that these taxa form functionally coherent and separate ecological units that differ by their
166 interaction capabilities with other microorganisms. We note, however, that the GFCs do not resolve
167 all bacterial diversity, e.g. that found within the *Alteromonas* [49] or within the high-light
168 *Prochlorococcus* [26] clades.

169

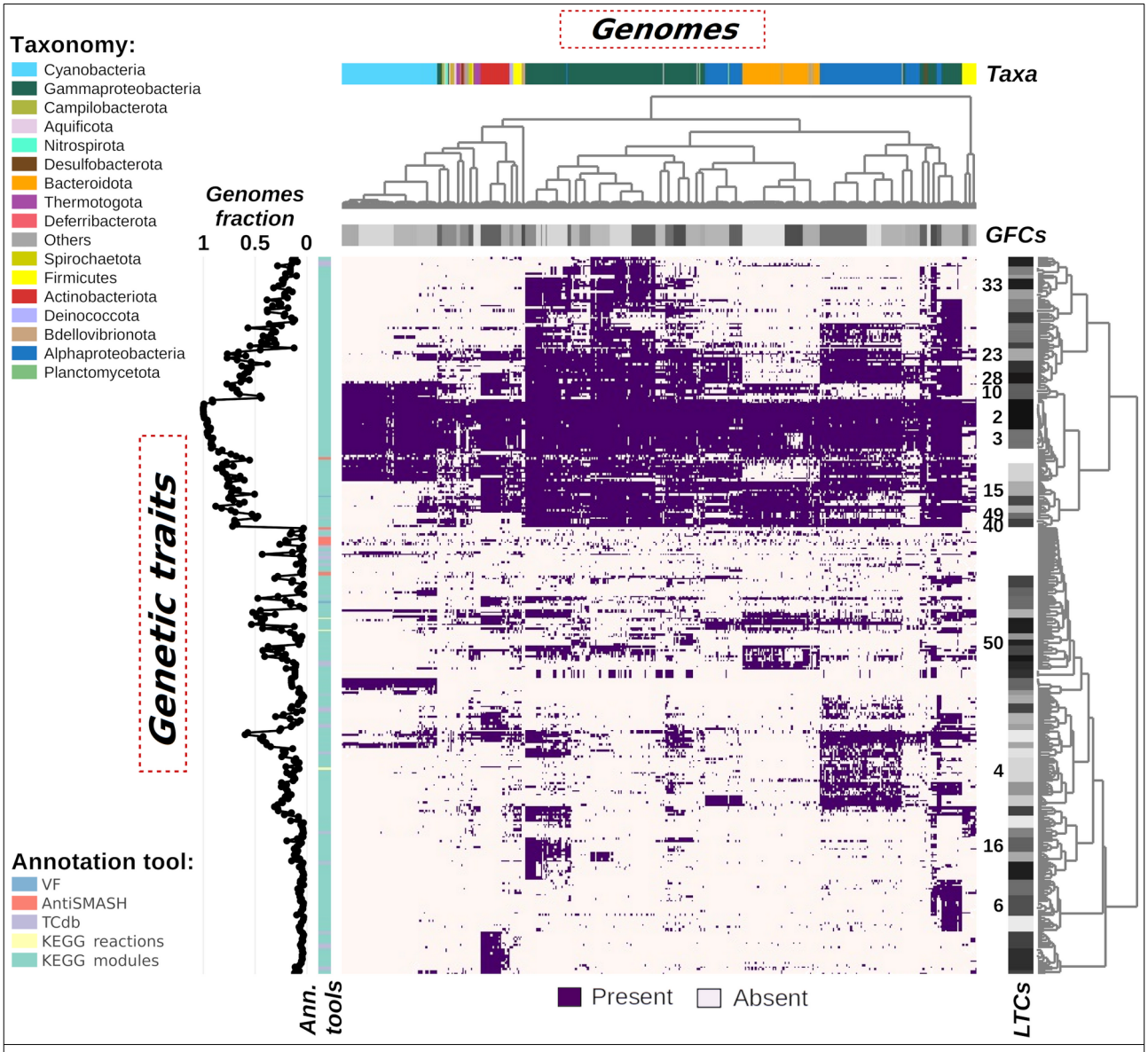

Fig. 1: ***Atlas of Marine Microbial Functional Traits*** *showing presence/absence of genetic traits across all analyzed genomes. Each column represents a genome and these are hierarchically clustered. The horizontal color bar represents the taxonomic affiliations of the genomes (mainly*

phyla, with the exception of Proteobacteria that are represented at the class level) and the horizontal grey bar delineates specific GFCs. Rows are the genetic traits clustered using the coefficient of disequilibrium into LTCs (vertical grey bar); relevant LTCs are marked. Left-side row annotations show the average abundance for the genetic traits and the annotation tool (color coded bar) with which they have been annotated. An interactive version of this figure is available as Supplementary media 1.

Missing common trait clusters highlight basic metabolic differences

Just as the genomes (represented by the columns in Fig. 1) could be grouped into GFCs, the genetic traits that drive this clustering could themselves be grouped into Linked Trait Clusters (LTCs). The traits within each LTC were found together in the genomes more often than expected by chance, and thus may be linked functionally (Supplementary Fig. 4). For example, LTC 10 includes pathways for assimilatory sulfate reduction, siroheme and heme biosynthesis, and biotin biosynthesis. The sulfate reduction and siroheme pathways are functionally linked, as siroheme is a prosthetic group for assimilatory sulfite reductases [50,51]. While heme and siroheme are different molecules and participate in different pathways, siroheme can be "hijacked" for the production of heme in sulfate-reducing bacteria [52]. Finally, once reduced, sulfur can be incorporated into essential molecules such as amino acids (methionine and cysteine), membrane lipids and the B vitamins thiamine and biotin [53]. The pathways for the biosynthesis of biotin and its precursor (pimeloyl-ACP) are indeed among the traits encoded in LTC 10. Moreover, every pair of traits in this LTC is much more likely to co-occur within a genome than random pairs of traits (the mean r^2 within this LTC is 0.48, compared with 0.08 among all traits and 0.03 among traits not clustered into any LTC; Supplementary Fig.

187 4b). Taken together, these results support an evolutionary relationship among the traits included in
188 LTC 10. Therefore, the LTC concept may be useful to identify traits that are likely to be functionally
189 connected and may have evolved together.

190 Based on their occurrence across genomes (Supplementary Fig. 4c), we divided the LTCs into three
191 subgroups: “core” (present in >90% of genomes), “common” (<90% and ≥30%) and “ancillary”
192 (≤30%). LTCs 2 and 3 (mean r^2 of 0.90 and 0.88) were the most commonly found across genomes
193 (>93%; Fig. 1) and represent the core LTCs. They included KEGG modules involved in the core
194 metabolism (glycolysis, pentose phosphate pathway and the first three reactions of the TCA cycle),
195 as well as pathways responsible for nucleotide, amino acid and cofactor metabolism. In addition,
196 LTC 2 also included an F-type ATPase, cofactor biosynthetic pathways (Coenzyme A and FAD) and
197 a few transporters, e.g. ABC type II (for a variety of small ions and macromolecules) and Tat
198 protein exporters (see Supplementary Table 3 for complete list of all LTCs and their respective
199 genetic traits).

200 Other common LTCs (15, 23, 28 and 40), which were found in 61-70% of the genomes, also
201 included traits involved in core metabolisms, e.g. parts of the TCA cycle in the LTC 15 (mean r^2 =
202 0.64). The lack of the full TCA cycle in some organisms, such as pico-Cyanobacteria, is consistent
203 with previous studies (Supplementary text and Supplementary Fig. 5) [54]. Moreover, our results show
204 that some heterotrophic bacteria also lack part of the TCA cycle (e.g. Spirochaetota, Thermotogota
205 and Firmicutes, corresponding to GFCs 40, 42, 43, 8, 9 and 41), confirming and broadening
206 previous studies [55–58]. However, other important cellular functions were also found in LTCs not
207 present in all genomes, for example genes involved in cell wall assembly (LTC 40 and 49, mean r^2
208 of 0.71 and 0.60) and variants of RNA degradosome and polymerase (LTCs 23 and 28, mean r^2 of
209 0.45 and 0.70; more details in the supplementary information). Additionally, as the linkage of the
210 traits inside each LTC is not complete, the presence or absence of a LTC is not a fully robust
211 indication for the presence or absence of each trait. Finally, a large fraction of the genes in each

genome was annotated as hypothetical or not annotated at all (range 22-66%), a strong reminder of the limitations of current genomic and physiological knowledge. We therefore interpreted the patterns of LTCs associated with microbial interactions, keeping in mind that these represent bioinformatics predictions requiring experimental validation.

Many bacteria need “to shop” for their vitamins

Vitamins B1, B7 and B12 are among the most essential cofactors for microbes [59,60]. They are metabolically expensive to produce as they often need several enzymes (e.g. about 20 for vitamin B12) [61], and their availability in the dissolved extracellular pools is extremely limited across all aquatic ecosystems [62,63]. Some phytoplankton are known to require exogenous vitamins from co-occurring heterotrophic bacteria, and indeed vitamin B12 may be a co-limiting micro-nutrient for primary productivity, e.g. in the Southern Ocean [64]. We thus analyzed the presence and absence of the pathways for the production of these vitamins, and of their transport across membranes, among our set of marine genomes. Less than half of all genomes are predicted to produce all these vitamins (~45%, including all pico-Cyanobacteria - GFC 27, 28 and 40 - and many Gammaproteobacteria; Fig. 2b). Of the rest, ~33% synthesize at least two B vitamins (e.g. Alphaproteobacteria, which produce mainly vitamin B1 and B12, and the rest of Gammaproteobacteria, which produce vitamin B1 and B7) and ~15% can produce only one type of B vitamin (or ~7% none at all). This suggests that there is a major “market” for B vitamins, and indeed almost all genomes encoded transporters for at least one vitamin.

A more detailed analysis of the genomes suggests that marine bacteria can be divided into three main groups based on their predicted strategy for B vitamins acquisition: (1) “consumers”, which lack the biosynthetic genes but harbor the vitamin transporters; (2) “independents”, which encode the biosynthetic pathways but not the relevant transporters; (3) “flexibles”, which encode both the

biosynthetic pathways and transporters for a specific vitamin (Fig. 2a). Bacteria possessing the latter strategy can potentially switch from being consumers to independent or vice-versa, according to what is more efficient given the surrounding conditions (e.g. availability of extracellular vitamins). The proportion of these three groups changes with the B vitamin studied and the taxonomy of the genomes. Very few genomes were “flexibles” for all three vitamins (~4%), and these were mostly Actinobacteriota (Fig. 3a). In contrast, the most common strategy for vitamin B1 was the “flexible” (just over half of the genomes), and almost half of these were Gammaproteobacteria (Fig. 2a). There were almost equal proportions of flexible, consumers and independents for vitamin B12, whereas the most common bacterial strategy for vitamin B7 was independent (Fig. 2a). Genomes with flexible strategy for vitamin B1 and B12 were quite common (52% and 32% of total genomes) and for several of them it was possible to speculate on their role as ‘providers’ for such vitamins. Indeed, ~20% of the genomes bearing both biosynthetic pathways and transporters for these two vitamins encoded a bidirectional transporter (Supplementary Table 4) that could allow to export and “share” the vitamin.

Notably, there were 64 possible combinations of traits for the synthesis and uptake of the three vitamins, and we could identify genomes corresponding to 45 of these possible combinations (Fig. 2c and Supplementary Fig. 6a). This plasticity is also reflected in the clustering of the related vitamin traits together with different metabolic traits in several different LTCs (Supplementary Table 3). Taken together, these results suggest a wide diversity of strategies to obtain vitamins, most of which require an exogenous uptake for a number of vitamins. We speculate that this represents a manifestation of the “Black Queen” hypothesis, where bacteria can “outsource” critical functions to the surrounding community, enabling a reduction of the metabolic cost [65]. Several of these strategies were specifically associated with one taxon and related GFCs (Supplementary text and Supplementary Fig. 6b). For example, all Cyanobacteria could produce all three vitamins, whereas their genomes encoded only vitamin B7 transporters, suggesting that if this group is a source (i.e.

provider) of vitamin B1 and B12, these vitamins become available to the rest of the community through cell death rather than metabolic exchange between living cells.

In our dataset, the highest fraction of B vitamins consumers, and hence putative auxotrophs, was observed for vitamin B12, followed by B7 and B1. This order may be related to the metabolic costs of producing each vitamin. About 20 genes are required for *de-novo* aerobic production of B12 [61], whereas only 4 genes are required to synthesize B7 [59] and 5 genes for B1 [66,67]. Notably, very few organisms were predicted to be auxotrophic for all three vitamins, suggesting that completely relying on exogenous sources for vitamins represent a risky strategy of difficult implementation in marine pelagic environments. Some of the putatively auxotrophic organisms actually encode parts of their vitamin biosynthesis pathways, and therefore may depend on the uptake of a precursor rather than of the vitamin itself (Supplementary text, Supplementary Fig.s 7a-d). Our results, together with experimental observations of vitamin limitation in laboratory cultures and in nature [59,62,68], and of shifts in the capacity of marine communities to produce vitamins [69], argue for an important role of vitamins or their precursors and their exchange between organisms in shaping marine microbial communities.

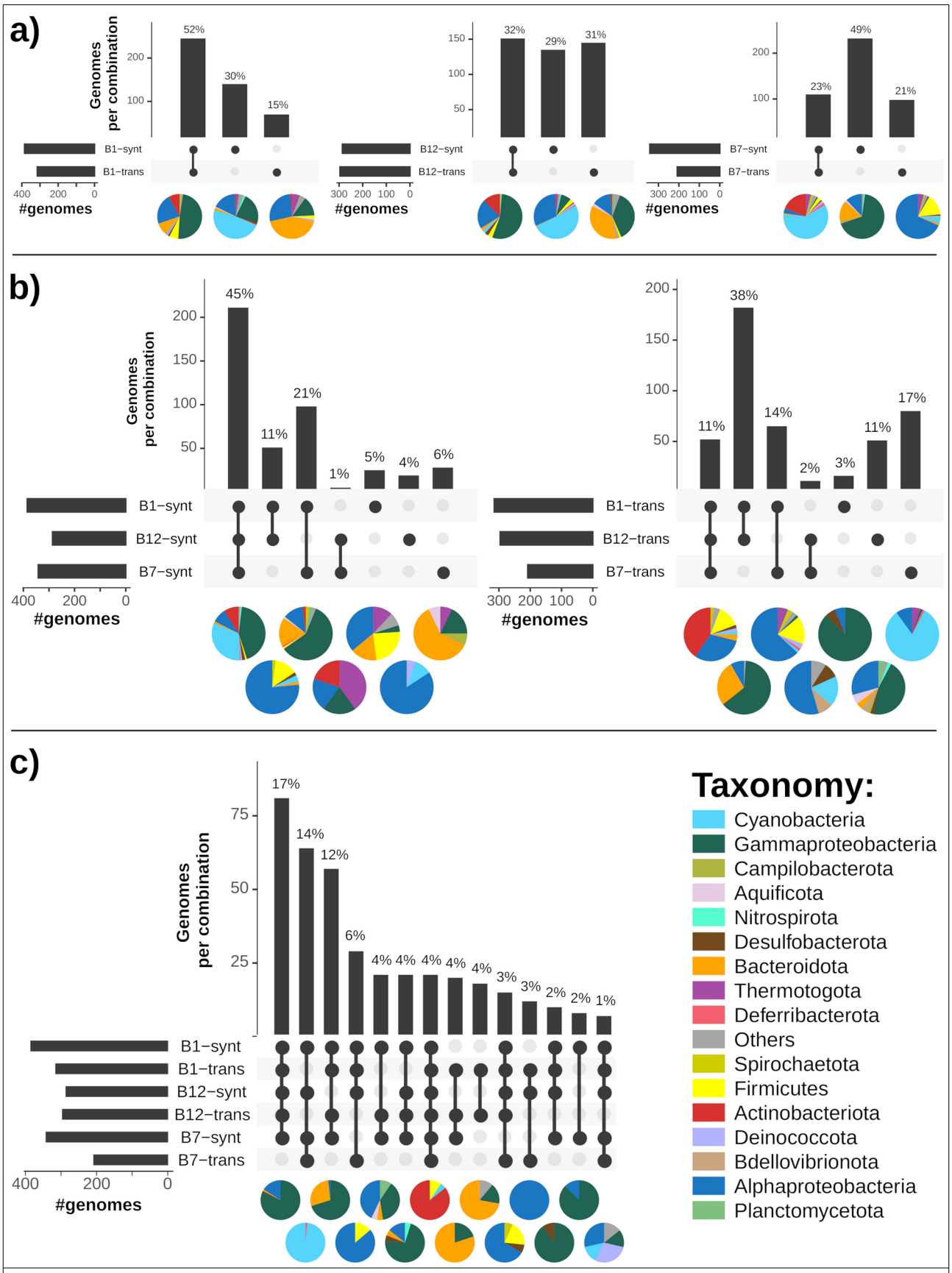

Fig. 2: UpSet plots exploring the different genomic configurations of traits involved in the biosynthesis and transport of vitamins B1, B12, and B7. (a) Different strategy of acquisition of B

vitamins among genomes in relation to their ability to produce or/and transport a certain vitamin. (b) Genome partitioning in relation to either production or transport of all selected B vitamins. (c) Most abundant combinations of all these traits across genomes; the remaining combinations are shown in Supplementary Fig. 6a. Overall, the left bar chart indicates the total number of genomes for each trait, the dark connected dots indicate the different configurations of traits and the upper bar chart indicates the number of genomes provided with such configuration. Pie charts show the relative abundance of the different taxa represented in each configuration.

Production of phytohormones and siderophores – common mechanisms of synergistic microbial interactions

We next focused on the traits that may mediate synergistic microbial interactions through the production and exchange of "common goods" such as siderophores [70], as well as of specific phytohormones like auxin [71]. Siderophores are organic molecules that bind iron, increasing its solubility and bioavailability [70]. Siderophore production by heterotrophic bacteria can stimulate the growth of phytoplankton, providing a potential trading good for synergistic interactions [18]. However, high affinity uptake of siderophores can also serve as a mechanism for competition for iron [72,73]. As shown in Fig. 3, approximately 32% of the genomes have the capacity to produce siderophores, and these genomes were primarily grouped in GFCs 5 (Bdellovibrionota), 9, 18 (Firmicutes), 34 (Actinobacteriota), as well as in GFCs 1, 13, 15 and 48 (Gammaproteobacteria). By contrast, many more genomes, from multiple phyla, encoded siderophore transporters (76% of the genomes, Fig. 3). Furthermore, microorganisms can utilize siderophore-bound iron also without the need for siderophore transporters, e.g. using ferric reductases located on the plasma membrane [74] or via direct endocytosis [75]. For example, about half of the genomes that produce siderophores can

293 produces vibrioferrin (~15% of total genomes), yet the vibrioferrin-bound iron is likely accessible
294 to many more organisms, including phytoplankton, upon photolysis [18]. Thus, in agreement with
295 recent considerations [76], we highlight the role of siderophores as "keystone molecules" [77] and take
296 this as an indication that the organisms producing them have an important role in the functionality
297 of microbial communities (e.g., references [78,79]).

298 Several recent studies have shown that bacteria can influence the growth of phytoplankton through
299 the production of phytohormones [10,15,21], and indeed one auxin hormone, indole-3-acetic acid (IAA),
300 has been identified in natural marine samples [10]. Almost half of the genomes (~49%) in our dataset
301 are predicted to produce IAA, including some Cyanobacteria. Four pathways for the production of
302 IAA were identified, with some organisms encoding more than one pathway. The indole-3-
303 acetamide pathway is the most common one and is present in nearly all GFCs comprising genomes
304 of Alphaproteobacteria and Actinobacteriota, as well as in some other taxa (Fig. 3). The second
305 most common is the tryptamide pathway, whereas the last two pathways are rarer and limited to
306 Alphaproteobacteria (indole-3-acetonitrile) and some genomes of Cyanobacteria and
307 Actinobacteriota (indole-3-pyruvate). It is tempting to speculate that the widespread distribution of
308 the capacity to produce IAA, and the diversity of biosynthetic pathways, suggest that many
309 heterotrophic bacteria can directly increase phytoplankton growth through specific signaling (e.g.
310 [10,15]). However, all pathways for IAA production are tightly intertwined with the metabolism of
311 tryptophan, either involved in tryptophan catabolism (to cleave the amino group for nitrogen
312 metabolism) or as a "release valve" to avoid the accumulation of toxic intermediates (e.g. α-keto
313 acid indolepyruvate and indoleacetaldehyde). Additionally, IAA can be catabolized as a carbon
314 source for growth (see [80] and references therein). Therefore, it is currently unclear whether IAA is
315 simply a byproduct of cellular metabolism or whether it is a key molecule utilized by diverse
316 bacteria to influence the growth of phytoplankton.

317

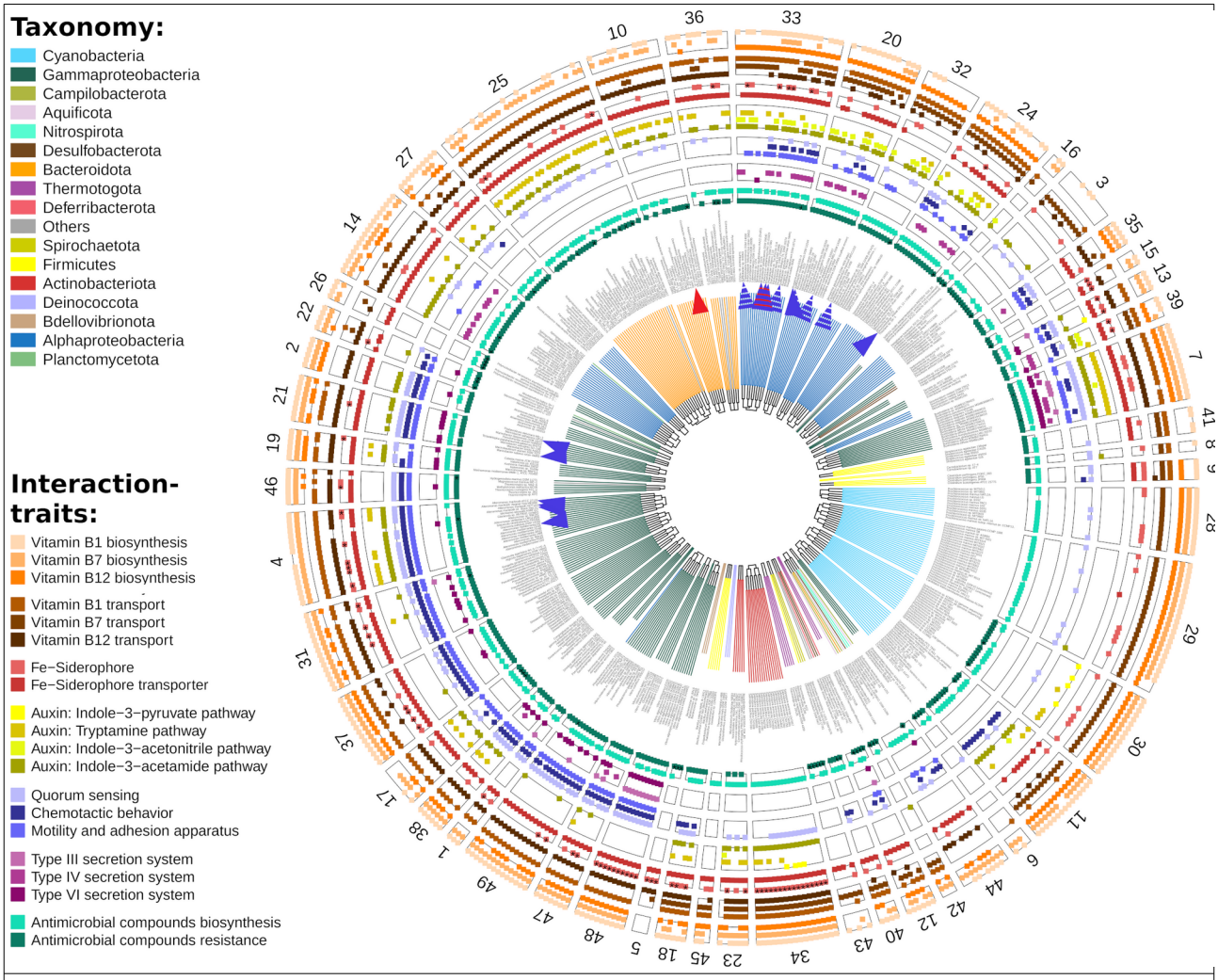

Fig. 3: Overview on interaction-traits across Genome Functional Clusters (GFCs). Each slice shows the interaction traits present in a GFC and, as dendrogram, the functional similarity of genetic traits of the grouped genomes. Known interactions (described in the referenced literature) between bacteria and phytoplankton are marked with arrows; a blue arrow indicates a positive interaction (an enhancing effect on phytoplankton growth), a red arrow a negative interaction, and a blue-red arrow a positive interaction that eventually becomes negative. A dashed arrow indicates a known interaction involving a bacterial genotype with a high similarity to one of the genomes included in the analysis. Asterisks in siderophores annotation indicate the presence of the specific vibrioferrin synthetic pathway along with the secondary metabolite pathway, while asterisks in the antimicrobial resistance annotation indicate that only generic resistance traits were annotated for that genome.

Traits underlying potential antagonistic interactions

Experimental measurements of interactions among marine bacteria suggest that antagonism is common (>50% of the tested isolates; [29,31]), but in most cases the mechanisms behind antagonistic interactions are unclear. Perhaps surprisingly, genes encoding for the production of putative antimicrobial compounds (as detected by antiSMASH [81] and KEGG modules; Supplementary Table 5 and 6) were found in almost all bacteria in our dataset, including GFCs poor in other interaction traits (77% of genomes, inner ring in Fig. 3). The most abundant traits across GFCs were bacteriocin and betalactone production (Supplementary Fig. 8). These two classes of compounds group several known molecules with potent bioactivity against bacteria and fungi [82,83], however their implication and role in natural environments are still understudied. Genes involved in the resistance to antimicrobial compounds were also relatively common (78% of genomes). We noted, however, that many pathways annotated in KEGG as resistance mechanisms against antimicrobial peptides have also other cellular functions (e.g. cell division, protein quality control and transport of other compounds; Supplementary Table 5) [84,85]. Cyanobacteria and Bacteroidota are notable examples of this as they don't possess any specific resistance traits other than the generic ones (Fig. 3 and Supplementary Fig. 8), suggesting they might be less efficient in resisting a chemical warfare. Some Cyanobacteria strains are indeed used as markers for antibiotic contamination because of their sensitivity (e.g., references [86,87]) and Bacteroidota often succumb when co-cultured with other bacteria that express antagonistic behaviour [29,31]. Overall, these genome-based predictions are in agreement with the experimental results of [29,31], which suggested that Alpha- and Gammaproteobacteria commonly inhibited other bacteria, whereas Bacteroidota showed the lowest inhibitory capacity and were the most sensitive to inhibition by other bacteria.

In contrast to the predicted widespread potential for allelopathy, the capacities to sense and move towards target organisms (quorum sensing, chemotaxis, motility and adhesion), and to directly

343 inject effector molecules (secretion systems), were more limited (respectively 61% and 24% of all
344 bacteria, Fig. 3). Chemotaxis, motility, adhesion and quorum sensing occurred together, primarily in
345 the GFCs that are rich in other interaction traits. These GFCs comprise Alpha- and
346 Gammaproteobacteria (Fig. 3). In a few cases these traits were grouped within the same LTC (e.g.
347 LTC 50, which contains traits for chemotaxis, flagellar motility and/or adhesion, or the LTC 6
348 which groups quorum sensing and potential resistance to antimicrobial compounds). Often, GFCs
349 that had these "antagonism LTCs" also encoded type IV or type VI secretion systems (T4SS and
350 T6SS, respectively). Notably, the T4SS and T6SS have different distributions among the GFCs,
351 with only GFC 7 and 47 (comprising *Burkholderia* and *Vibrio*, genera of Gammaproteobacteria)
352 bearing both systems. The T4SS system can perform multiple roles, including conjugation, DNA
353 exchange and toxin delivery in bacteria-bacteria or bacteria-eukaryote interactions [88]. T4SSs were
354 detected more frequently in Alphaproteobacteria (5 out of 8 GFCs). Similar to T4SSs, the T6SS
355 system can also deliver effector molecules into other bacterial or eukaryotic cells by using a
356 contractible sheath-like structure [89]. To date, T6SSs are known to be involved only in antagonistic
357 interactions, suggesting that the presence of this trait is a high-confidence predictor of the ability to
358 directly antagonize other cells ([90] and references therein). In our dataset, T6SSs occurred almost
359 exclusively in GFCs comprising Gammaproteobacteria, specifically in *Marinobacter* and *Vibrio*,
360 suggesting a strong capacity for contact-mediated antagonistic interactions in these taxa. Type III
361 secretion systems (T3SS) were found only in a few genomes (e.g. *Burkholderia* and *Aeromonas*),
362 which are often considered metazoan-associated and sometimes pathogenic bacteria [91]. T3SS allows
363 to delivering effector molecules that maintain the bacterial association with the host [92].

364

Gene functional clusters differ in their overall potential for interactions

When considering antagonistic and synergistic interaction traits together, it is clear that some GFCs encode significantly more interaction traits than others (Fig. 3), both in terms of diversity and richness (Supplementary Fig. 9a). One potential driver for the difference in richness and diversity of interaction traits could be the reduction in genome size associated with oligotrophic lineages such as pico-Cyanobacteria and SAR11, and indeed the number of interaction traits can partly be explained by genome size (Supplementary Fig. 9b-c) [93]. However, genome size could not explain by itself the number of different interaction traits, with Gammaproteobacteria and several Alphaproteobacteria encoding more interaction traits than predicted by genome size, and Bacteroidota encoding fewer than expected.

Conceptual Fig. 4 demonstrates four of the main auxiliary LTCs that we predict to link traits involved in microbial interactions. LTCs 33 and 50 show the linkage of traits for chemotaxis, motility and adhesion, a typical set of traits a bacterium would need to locate, reach and settle on an organic matter particle. LTC 33 (mean $r^2 = 0.63$) includes, in addition, a two-component regulatory system (BarA-UvrY), the biosyntesis of pyridoxal (one form of vitamin B6) and ubiquinone. BarA-UvrY genes are known to regulate virulence, metabolism, biofilm formation, stress resistance, quorum sensing and secretion systems (see [94] and references therein). This LTC is common in the GFCs grouping Gammaproteobacteria (10 out of 19) such as *Alteromonas* (GFC 4), *Marinobacter* (GFC 21), *Pseudoalteromonas* (GFC 31), *Shewanella* (GFC 37) and *Vibrio* (GFCs 47 and 48; Supplementary Table 3). All of these organisms are known as particle and phytoplankton associated bacteria (e.g. [23,39,45,46,95,96]). LTC 50 (mean r2 = 0.56) includes, in addition to traits for chemotaxis, motility and adhesion, also a two-component regulatory system (AlgZ-AlgR), which is a key element in the regulation of twitching motility, alginate production and biofilm formation during

389 *Pseudomonas aeruginosa* infections [97]. This LTC is found complete, for example, in GFC 21
390 representing mainly *Marinobacter* (Supplementary Table 3). A model system including a member of
391 this GFC, *M. adhaerens* HP15, has been shown to indeed interact with phytoplankton through
392 attachment (up-regulation of type-4 fimbriae synthesis which is controlled by the PilS-PilR
393 regulatory system included in LTC 50) and increasing host aggregation [98].

394 LTC 16 (mean r2 = 0.33) includes interaction traits for quorum sensing and vitamin B12 transport
395 along with amino acid and sugar transporters, and the RstB-RstA regulation system. Similar to the
396 other regulatory mechanisms described above, the RstB-RstA system is involved in adhesion,
397 biofilm formation, motility and hemolysis [99]. These effects have been shown in in *Vibirio*
398 *alginolyticus* and, not surprisingly, LTC 16 is found complete in the GFC 48 (including *V.*
399 *alginolyticus*) and in GFCs 1 and 47 (grouping *Aeromonas* and *V. natriegensis* genomes;
400 Supplementary Table 3). *V. alginolyticus* strains are found frequently associated with macro-algae
401 [100] supporting the role of this LTC in interactions with phytoplankton.

402 Finally, our results allow us to propose that the production of phytohormones may be linked with
403 other interaction traits. Specifically, LTC 4 (mean r^2 = 0.16), which contains the indole-3-
404 acetonitrile pathway, is found complete in GFCs 20, 33 and 39 (Supplementary Table 3). These
405 GFCs group genomes of bacteria known to interact via IAA-mediated mechanisms with
406 phytoplankton [10,15,21] and plants (GFC 39; [101]). LTC 4 also contains T4SS, which have been shown
407 to play a pivotal role in the delivery of another phytohormone, i.e. cytokinin, into host cells [102].
408 While the mean linkage within this LTC is relatively low (0.16), we speculate that, for some
409 bacteria, the T4SS may be used for injecting IAA into its target cells, as the phytohormone is
410 negatively charged under physiological conditions (pKa= 4.8) and thus not likely to cross
411 membranes passively [80]. Injection of IAA through T4SS into target cells may also explain why
412 addition of IAA to phytoplankton growth media did not produce the expected phenotype in a
413 phytoplankton model organism [10], and why physical connection between bacterial and host cells is

414 often observed in related model systems [15,21]. Moreover, other genetic traits found linked within
415 LTC 4 could add details to such mechanisms of interaction, like the manganese/iron transporter
416 (nutrient-dependent response), the transport of capsular polysaccharide (resistance to host defense
417 and pathogenicity; [103]) or the biosynthesis of putrescine (acts as a potential bio-stimulant of growth,
418 productivity, and stress tolerance in plants; [104]).

419

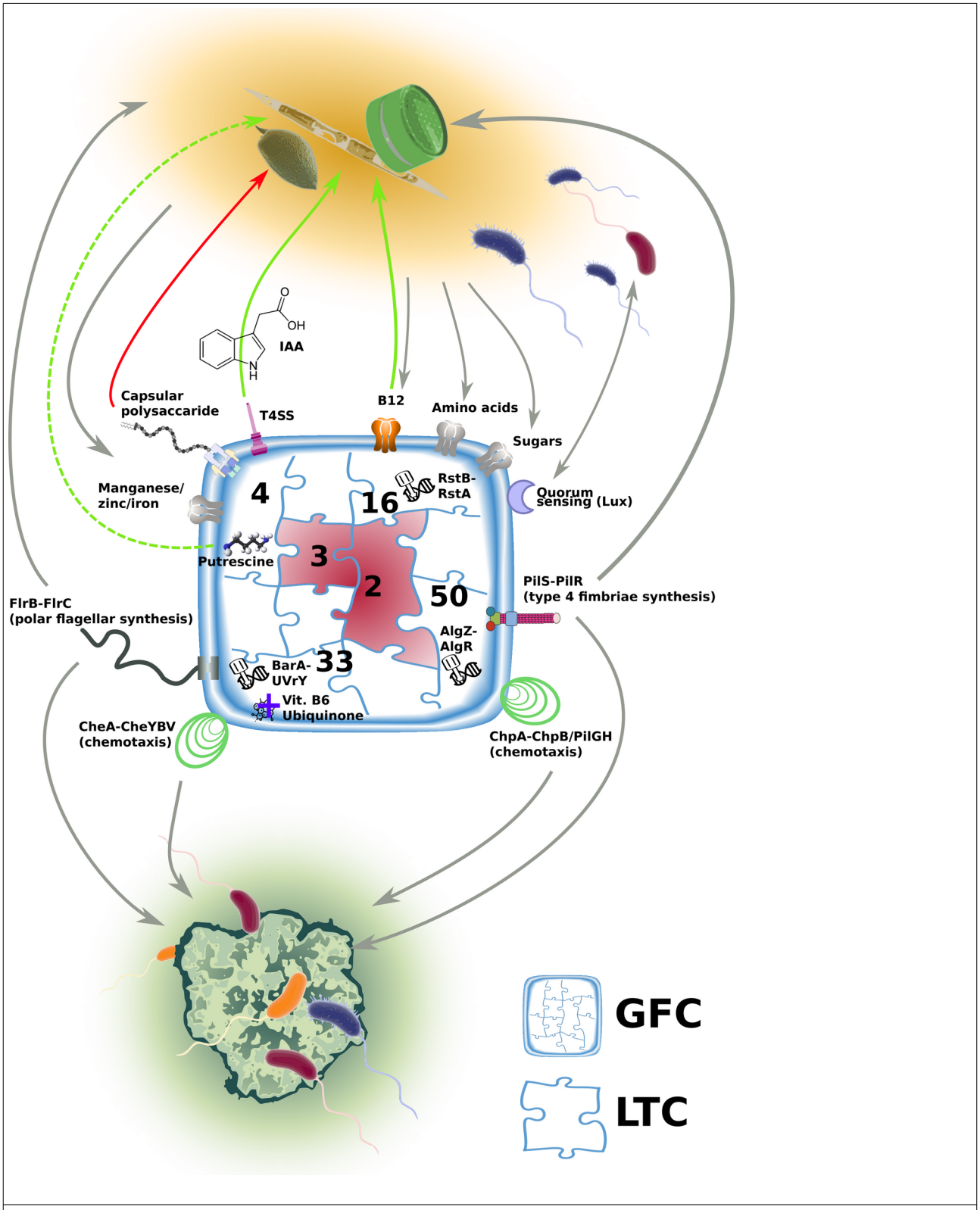

Fig. 4: Conceptual representation of a hypothetical bacterial cell characterized by the presence of the core LTCs (2 and 3, marked in red) and different common/ancillary LTCs mediating interactions with phytoplankton, other bacteria and particles. Green arrows indicate positive

effects (e.g. enhancing growth), grey arrows are for metabolites/chemical exchange, movement or attachment, and red arrows for negative effects (e.g. pathogenicity). LTCs 4 and 16 could be mainly involved in interactions with phytoplankton, while LTCs 33 and 50 in interactions with organic matter particles.

Summary and conclusions

We have presented two concepts, the GFCs and the LTCs, which provide an approach to extrapolate from studies of specific model organisms to the diversity of microbes, based on the traits encoded in their genomes. These concepts come with several caveats. First, both GFCs and LTCs are statistical in nature, representing the probability of bacteria having a similar functional capacity or of traits being linked. Second, the sequenced genomes represent only a part of the marine bacterial diversity, and bacterial taxa are unevenly represented (Figs. 1 and 3). Finally, a large fraction of the genes in each genome is currently uncharacterized while a few can be miss-annotated, and these instances may change our inferences (primarily those of a lack of a trait). These biases may be, in part, responsible for some of the "blurred" squares of the "checkerboard" that represent the Atlas of functional potential of marine bacteria (Fig. 1). Nevertheless, the GFC and LTC analyses enable us to identify prevalent patterns in the ability of specific bacterial groups to interact both physically (e.g. through motility and adhesion to substrates) and through the exchange of vitamins, toxins and other info-chemicals. These patterns provide hypotheses that can be tested experimentally. They can also provide an important tool for ecologists, in defining a quantitative probability of whether a trait is found, or an interaction may occur.

Importantly, the interaction traits alone cannot predict whether and under what conditions an interaction will indeed occur as these processes are often quite complex (e.g. [10,15,17,20,21]) and require functional and metabolic cooperation of several different genetic traits. Thus, each interaction trait refers to a specific mechanism of interaction that the bearing GFCs could exhibit, with the

441 combinations of these traits within a LTC or between different LTCs defining life-style plasticity of
442 each organism. Additional experimental work on established, as well as on new model organisms, is
443 required to test the many hypotheses raised by our analysis, to identify new mechanisms of
444 interaction and to understand how multiple interaction traits are combined when organisms grow
445 together in the oceans.
446 We believe that our work provides a remarkable headway in our knowledge on microbial functional
447 and interaction capacity, thus, addressing fundamental aspect ruling community dynamics and
448 assembling, with far-reaching consequences on ecosystem levels (e.g. biogeochemical cycles of C,
449 N, P and S). Moreover, we introduce a framework that can be easily scaled to different ecosystems
450 (e.g. freshwater or terrestrial) and expanded including information from additional model systems
451 of other environments (e.g. fishes, zooplankton, corals, sponges and humans). Finally, our
452 framework offers a new way to interpret amplicon and metagenome datasets, amenable to further
453 computational modeling and statistical analyses, as well as to experimental testing of specific
454 hypotheses on bacterial metabolism, behavior, and mechanisms of interactions.

Online Methods (subheadings)

Genome selection

A dataset of complete genomes of marine bacteria was compiled performing an extensive research on metadata available from NCBI (http://www.ncbi.nlm.nih.gov/genome), JGI (https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=FindGenomes&page=genomeSearch) and MegX (https://mb3is.megx.net/browse/genomes) websites. Although the focus of the analysis was on bacteria inhabiting the marine pelagic environment, some genomes from organisms isolated in extreme marine environments (i.e. thermal vents, saline and hypersaline environments, estuaries) and sediment, as well as from human and plant symbionts (*Sinorhizobium* and *Mesorhizobium*) were kept for comparison. All the genomes in the final list were downloaded from NCBI and JGI repositories and checked for completeness manually, and by mean of the software CheckM [105]. Only genomes whose chromosome was a continuous sequence in the fasta file (plus plasmids when present) or which met the criteria proposed for high-quality draft genomes (>90% of completeness, <10% of contaminations, >18 tRNA genes and all 3 rRNA genes present; [106]) were retained for downstream analyses. The final dataset included 473 complete genomes with 117 closed genomes and with >81% genomes that were >99% complete. Of these 473 genomes, 421 were isolated in marine pelagic and coastal zones, 34 in extreme environments (e.g. salt marsh or hydrothermal vent), 6 in marine sediment and, of the remaining, 8 were human associated and 4 plant roots associated (Supplementary Table 2).

Genome annotation

All retrieved genomes were re-annotated using a standardized pipeline. In brief, gene calling and first raw annotation steps were performed with Prokka [107]. The amino acid sequences translated from the identified coding DNA sequences of each genome were annotated against pre-computed hidden Markov model profiles of KEGG Ortholog (KEGG database v94.0) using kofamscan v1.2.0

[108]. Additional targeted analyses were performed to annotate secondary metabolites, phytohormones and specific transporters. The genbank files generated by Prokka were submitted to a local version of Anti-SMASH v5.1.2 (--clusterblast --subclusterblast --knownclusterblast --smcogs --inclusive --borderpredict --full-hmmer --asf --tta) which generated a list of predicted secondary metabolite biosynthesis gene clusters [81]. Phytohormones were manually identified mapping annotated KOs to the KEGG compounds belonging to the different phytohormones pathways present in the KEGG map01070. Only phytohormones that had at least the last 3 reactions present were considered in the analysis. Translated amino acid sequences were also used as input for a GBlast search (BioV suite; [109] to identify trans-membrane proteins, specifically vitamin B and siderophore transporters; as recommended by the authors, only the annotations with a trans-membrane alpha-helical overlap score >1 and a blast e-value $<1^{-6}$ were retained. Manual annotations were performed to specifically identify production of photoactive siderophores (i.e. vibrioferrin; [18], blasting predicted protein sequences (blastp; [110]) against reference dataset assembled using all available sequences of related genes (pvsABCDE operon; [111] available in UniProt.

KEGG module reconstruction

KEGG orthologies (**KO**s) annotations generated by kofamscan were recombined in KEGG modules (**KM**s) using an in-house R script. The KMs represent minimal functional units describing either the pathways, structural complexes (e.g. transmembrane pump or ribosome), functional sets (essential sets as Aminoacyl-tRNA synthases or nucleotide sugar biosynthesis) or signaling modules (phenotypic markers as pathogenicity). Briefly, using the R implementation of KEGG REST API [112], the script fetches the diagrams of all KMs from the KEGG website. Each diagram represents a pathway scheme of a KM listing all known KOs that can perform each of the reactions necessary to complete the pathway (Supplementary figure 1b). The completeness of a KM in a genome is calculated as the percentage of reactions for which at least one KOs was annotated over the total number of necessary reactions (e.g., a KM with 7 out of 8 reactions annotated is 87.5% complete).

KMs were considered complete based on the following rules: 1) KMs with fewer than 3 reactions required all reactions; 2) one gap was allowed in KMs with ≥3 reactions (i.e., a KM with 7 out of 8 annotated reactions was considered 100% complete).

Genetic and interaction traits identification

The annotated complete KMs, secondary metabolites, phytohormones and transporters represent the genetic traits identified in the genomes (556 genetic traits). From this list, the subset of interaction traits was manually extracted based on current knowledge about processes that likely play a role in microbial interactions (list of picked interaction traits in Supplementary Table 1). Within the KMs we identified traits related to vitamin biosynthetic pathways, quorum sensing, chemotaxis, antimicrobial resistance, motility and adhesion (Supplementary Table 5). Since the ecological role of most secondary metabolites is still unclear, a careful literature search was performed to identify and retain only the secondary metabolite clusters with a proposed function which can be linked to microbial interaction processes, such as siderophore production, quorum sensing and antimicrobial compound biosynthesis (Supplementary Table 6). The phytohormone annotations revealed the capability of producing indoleacetic acid (auxin), salicylic acid and ethylene however the last two were found in < 1% of genomes so only auxin production was included in the analysis (Supplementary Table 7). Vitamin and siderophore transporters were identified in the transporter annotations looking for the related transporter families (e.g. TonB, Btu) and the substrate information (Supplementary Table 4).

Statistical analyses

The presence/absence matrix of genetic traits in genomes served as basis to cluster the former into Linked Trait Clusters (LTCs) and the latter into Genome Functional Clusters (GFCs). The GFCs were generated feeding a genome functional similarity matrix calculated using Phi coefficient (i.e. Pearson correlation for binary variables; *phi* function, package sjstats) into the affinity propagation algorithm implemented in the *apcluster* function (q=0.5 and lam=0.5; r package apcluster; [113]. This

machine learning algorithm was chosen because it does not require the number of clusters to be determined *a priori*, allowing instead this feature to emerge from the data [114]. Briefly, the functional similarity matrix is used to construct a network where nodes and edges are known to be genomes and their pairwise Phi correlation, respectively. Starting from a random set of exemplar nodes, clusters are created by expansion towards the adjoining, most similar, nodes. Through iterations of this procedure, the algorithm tries to maximize the total similarity between nodes within each cluster eventually converging towards the best set of clusters.

For the LTC delineation, a similarity matrix of genetic traits was built calculating the square of Pearson's correlation coefficient (r^2; [115]). Out of a total of 556 genetic traits, 434 were retained for downstream analysis as they were found in >3% of the genomes (>14 genomes) and r^2 was computed using the *ld* function (r package snpStats; [116]. The LTCs were automatically extracted from the hierarchical clustering (*hclust* function in r, method = "ward.D2") of the dissimilarity matrix ($1-r^2$) using the function *cutreeDynamic* (method = "hybrid", deepSplit = 4 and minClusterSize = 3; r package dynamicTreeCut; [117]). Similar to the context of linkage disequilibrium (LD) [118], using r^2 as representative index [119] , r^2 have to be carefully interpreted as two genetic traits can be non-randomly associated because these traits are interactively linked to fitness or simply because they are closely located on the chromosome (i.e. lower chances of recombination). However, as we used r^2 to compare the genetic traits which commonly involve multiple genes, the second possibility is less likely. While exploring the functional potential, a LTC was considered complete in a genome when >60% of the grouped genetic traits were present and it was considered complete in a GFC when the average completeness of the included genomes was >60%.

All analyses were performed in R 3.6.0 [120].

References

1. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: Surviving and thriving in the microbial jungle. *Nature Reviews Microbiology* **8**, 15–25 (2010).

2. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 667–84 (2012).

3. Seymour, J. R., Amin, S. A., Raina, J.-B. & Stocker, R. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat. Microbiol.* **2**, 17065 (2017).

4. Farooq Azam and Francesca Malfatti. Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* **5**, 782–791 (2007).

5. Zoccarato, L. & Grossart, H.-P. Relationship Between Lifestyle and Structure of Bacterial Communities and Their Functionality in Aquatic Systems. in *Advances in Environmental Microbiology - The Structure and Function of Aquatic Microbial Communities* (ed. Hurst, C. J.) 13–52 (Springer Nature Switzerland AG 2019, 2019). doi:10.1007/978-3-030-16775-2_2

6. Kirchman, D. L. *Processes in Microbial Ecology*. (Oxford University Press, 2012). doi:10.1093/acprof:oso/9780199586936.001.0001

7. Worden, A. Z. *et al.* Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science (80-.).* **347**, 1257594 (2015).

8. Gibert, J. P. Temperature directly and indirectly influences food web structure. *Sci. Rep.* **9**, 5312 (2019).

9. Bell, W. & Mitchell, R. Chemotactic and Growth Responses of Marine Bacteria to Algal Extracellular Products. *Biol. Bull.* **143**, 265–277 (1972).

10. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**, 98–101 (2015).

11. Durham, B. P. *et al.* Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl. Acad. Sci.* **112**, 453–457 (2015).

12. Durham, B. P. *et al.* Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nat. Microbiol.* **4**, 1706–1715 (2019).

13. Moran, M. A. & Durham, B. P. Sulfur metabolites in the pelagic ocean. *Nature Reviews Microbiology* **17**, 665–678 (2019).

14. Paul, C., Mausz, M. A. & Pohnert, G. A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics* **9**, 349–359 (2013).

15. Segev, E. *et al.* Dynamic metabolic exchange governs a marine algal-bacterial interaction. *Elife* **5**, e17473 (2016).

16. Christie-Oleza, J. A., Sousoni, D., Lloyd, M., Armengaud, J. & Scanlan, D. J. Nutrient recycling facilitates long-term stability of marine microbial phototroph–heterotroph interactions. *Nat. Microbiol.* **2**, 17100 (2017).

17. Wang, H., Tomasch, J., Jarek, M. & Wagner-Döbler, I. A dual-species co-cultivation system to study the interactions between Roseobacters and dinoflagellates. *Front. Microbiol.* **5**, 311 (2014).

18. Amin, S. A. *et al.* Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17071–17076 (2009).

19. Keshtacher-Liebso, E., Hadar, Y. & Chen, Y. Oligotrophic Bacteria Enhance Algal Growth under Iron-Deficient Conditions. *Appl. Environ. Microbiol.* **61**, 2439–2441 (1995).

20. van Tol, H. M., Amin, S. A. & Armbrust, E. V. Ubiquitous marine bacterium inhibits diatom cell division. *ISME J.* 1–12 (2016). doi:10.1038/ismej.2016.112

21. Seyedsayamdost, M. R., Case, R. J., Kolter, R. & Clardy, J. The Jekyll-and-Hyde chemistry of Phaeobacter gallaeciensis. *Nat. Chem.* **3**, 331–335 (2011).

22. Grossart, H.-P. *et al.* Interactions between marine snow and heterotrophic bacteria: Aggregate formation and microbial dynamics. *Aquat. Microb. Ecol.* **42**, 19–26 (2006).

23. Aharonovich, D. & Sher, D. Transcriptional response of Prochlorococcus to co-culture with a marine Alteromonas: differences between strains and the involvement of putative infochemicals. *ISME J.* 1–15 (2016). doi:10.1038/ismej.2016.70

24. Coe, A. *et al.* Survival of Prochlorococcus in extended darkness. *Limnol. Oceanogr.* **61**, 1375–1388 (2016).

25. Hou, S. *et al.* Benefit from decline: The primary transcriptome of *Alteromonas macleodii str. Te101* during Trichodesmium demise. *ISME J.* **12**, 981–996 (2018).

26. Biller, S. J., Berube, P. M., Lindell, D. & Chisholm, S. W. Prochlorococcus: the structure and function of collective diversity. *Nat. Rev. Microbiol.* **13**, 13–27 (2014).

27. Cordero, O. X. *et al.* Population Genomics of Early Events in the Ecological Differentiation of Bacteria. *Science (80-.).* **336**, 48–51 (2012).

28. Tai, V., Paulsen, I. T., Phillippy, K., Johnson, D. A. & Palenik, B. Whole-genome microarray analyses of Synechococcus-Vibrio interactions. *Environ. Microbiol.* **11**, 2698–2709 (2009).

29. Long, R. A. & Azam, F. Antagonistic interactions among marine bacteria. *Appl. Environ. Microbiol.* **67**, 4875–4983 (2001).

30. Sher, D., Thompson, J. W., Kashtan, N., Croal, L. & Chisholm, S. W. Response of Prochlorococcus ecotypes to co-culture with diverse marine bacteria. *ISME J.* **5**, 1125–1132 (2011).

31. Grossart, H.-P., Schlingloff, A., Bernhard, M., Simon, M. & Brinkhoff, T. Antagonistic activity of bacteria isolated from organic aggregates of the German Wadden Sea. *FEMS Microbiol. Ecol.* **47**, 387–396 (2004).

32. Krause, S. *et al.* Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Frontiers in Microbiology* **5**, 251 (2014).

33. Bordron, P. *et al.* Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach. *Microbiologyopen* **5**, 106–117 (2016).

34. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

35. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science (80-.).* **353**, 1272–1277 (2016).

36. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).

37. Giovannoni, S. J. *et al.* Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science (80-.).* **309**, 1242–1245 (2005).

38. Grote, J. *et al.* Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio* **3**, e00252-12 (2012).

39. Hunt, D. E. *et al.* Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton. *Science (80-.).* **320**, (2008).

40. Darshanee Ruwandeepika, H. A. *et al.* Pathogenesis, virulence factors and virulence regulation of vibrios belonging to the Harveyi clade. *Rev. Aquac.* **4**, 59–74 (2012).

41. Lux, T. M., Lee, R. & Love, J. Complete genome sequence of a free-living Vibrio furnissii sp. nov. strain (NCTC 11218). *J. Bacteriol.* **193**, 1487–1488 (2011).

42. Preheim, S. P. *et al.* Metapopulation structure of Vibrionaceae among coastal marine invertebrates. *Environ. Microbiol.* **13**, 265–275 (2011).

43. Aronson, H. S., Zellmer, A. J. & Goffredi, S. K. The specific and exclusive microbiome of the deep-sea bone-eating snail, Rubyspira osteovora. *FEMS Microbiol. Ecol.* **93**, fiw250 (2017).

44. Prayitno, S. B. & Latchford, J. W. Experimental infections of crustaceans with luminous bacteria related to Photobacterium and Vibrio. Effect of salinity and pH on infectiosity. *Aquaculture* **132**, 105–112 (1995).

45. Holmström, C. & Kjelleberg, S. Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. *FEMS Microbiol. Ecol.* **30**, 285–293 (1999).

46. Eilers, H., Pernthaler, J., Glöckner, F. O. & Amann, R. Culturability and in situ abundance of pelagic Bacteria from the North Sea. *Appl. Environ. Microbiol.* **66**, 3044–3051 (2000).

47. Lupette, J. *et al.* Marinobacter Dominates the Bacterial Community of the Ostreococcus tauri Phycosphere in Culture. *Front. Microbiol.* **7**, 1–14 (2016).

48. Sonnenschein, E. C., Syit, D. A., Grossart, H.-P. & Ullrich, M. S. Chemotaxis of Marinobacter adhaerens and its impact on attachmentto the diatom Thalassiosira weissflogii. *Appl. Environ. Microbiol.* **78**, 6900–6907 (2012).

49. López-Pérez, M. & Rodriguez-Valera, F. Pangenome evolution in the marine bacterium *Alteromonas*. *Genome Biol. Evol.* **8**, evw098 (2016).

50. Siegel, L. M., Murphy, M. J. & Kamin, H. Reduced nicotinamide adenine dinucleotide phosphate-sulfite reductase of enterobacteria. I. The Escherichia coli hemoflavoprotein: molecular parameters and prosthetic groups. *J. Biol. Chem.* **248**, 251–264 (1973).

51. Scott, A. I., Irwin, A. J., Siegel, L. M. & Shoolery, J. N. Sirohydrochlorin. Prosthetic Group of a Sulfite Reductase Enzyme and Its Role in the Biosynthesis of Vitamin B12. *Journal of the American Chemical Society* **100**, 316–318 (1978).

52. Bali, S. *et al.* Molecular hijacking of siroheme for the synthesis of heme and d1 heme. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18260–18265 (2011).

53. Cook, A. M., Smits, T. H. M. & Denger, K. Sulfonates and Organotrophic Sulfite Metabolism. in *Microbial Sulfur Metabolism* 170–183 (Springer Berlin Heidelberg, 2008). doi:10.1007/978-3-540-72682-1_14

54. Zhang, S. & Bryant, D. A. The tricarboxylic acid cycle in cyanobacteria. *Science (80-.).* **334**, 1551–1553 (2011).

55. Neumann-Schaal, M., Jahn, D. & Schmidt-Hohagen, K. Metabolism the Difficile Way: The Key to the Success of the Pathogen Clostridioides difficile. *Front. Microbiol.* **10**, 219 (2019).

56. Hoskins, J. *et al.* Genome of the bacterium Streptococcus pneumoniae strain R6. *J. Bacteriol.* **183**, 5709–5717 (2001).

57. Wushke, S. *et al.* A metabolic and genomic assessment of sugar fermentation profiles of the thermophilic Thermotogales, Fervidobacterium pennivorans. *Extremophiles* **22**, 965–974 (2018).

58. Fraser, C. M. *et al.* Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. *Nature* **390**, 580–586 (1997).

59. Croft, M. T., Warren, M. J. & Smith, A. G. Algae Need Their Vitamins. *Eukaryot. Cell* **5**, 1175–1183 (2006).

60. Romine, M. F., Rodionov, D. A., Maezato, Y., Osterman, A. L. & Nelson, W. C. Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors in microbial communities. *ISME J.* **11**, 1434–1446 (2017).

694 61. Fang, H., Kang, J. & Zhang, D. Microbial production of vitamin B12: A review and future
695 perspectives. *Microb. Cell Fact.* **16**, 1–14 (2017).

696 62. Suffridge, C. P. *et al.* B Vitamins and Their Congeners as Potential Drivers of Microbial
697 Community Composition in an Oligotrophic Marine Ecosystem. *J. Geophys. Res.*
698 *Biogeosciences* **123**, 2890–2907 (2018).

699 63. Sañudo-Wilhelmy, S. A. *et al.* Multiple B-vitamin depletion in large areas of the coastal
700 ocean. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14041–14045 (2012).

701 64. Browning, T. J. *et al.* Nutrient co-limitation at the boundary of an oceanic gyre. *Nature* **551**,
702 242–246 (2017).

703 65. Morris, J. *et al.* The Black Queen Hypothesis: Evolution of Dependencies through Adaptive
704 Gene Loss. *MBio* **3**, e00036 (2012).

705 66. Jurgenson, C. T., Begley, T. P. & Ealick, S. E. The Structural and Biochemical Foundations of
706 Thiamin Biosynthesis. *Annu. Rev. Biochem.* **78**, 569–603 (2009).

707 67. McRose, D. *et al.* Alternatives to vitamin B1 uptake revealed with discovery of riboswitches
708 in multiple marine eukaryotic lineages. *ISME J.* **8**, 2517–2529 (2014).

709 68. Paerl, R. W. *et al.* Prevalent reliance of bacterioplankton on exogenous vitamin B1 and
710 precursor availability. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E10447–E10456 (2018).

711 69. Wienhausen, G., Noriega-Ortega, B. E., Niggemann, J., Dittmar, T. & Simon, M. The
712 exometabolome of two model strains of the Roseobacter group: A marketplace of microbial
713 metabolites. *Front. Microbiol.* **8**, 1–15 (2017).

714 70. Vraspir, J. M. & Butler, A. Chemistry of Marine Ligands and Siderophores. *Ann. Rev. Mar.*
715 *Sci.* **1**, 43–63 (2009).

716 71. De Smet, I. *et al.* Unraveling the evolution of auxin signaling. *Plant Physiol.* **155**, 209–21
717 (2011).

718 72. Hutchins, D. A., Witter, A. E., Butler, A. & Luther, G. W. Competition among marine
719 phytoplankton for different chelated iron species. *Nature* **400**, 858–861 (1999).

720 73. Joshi, F., Archana, G. & Desai, A. Siderophore cross-utilization amongst rhizospheric
721 bacteria and the role of their differential affinities for Fe3+ on growth stimulation under iron-
722 limited conditions. *Curr. Microbiol.* **53**, 141–147 (2006).

723 74. Morrissey, J. & Bowler, C. Iron utilization in marine cyanobacteria and eukaryotic algae.
724 *Front. Microbiol.* **3**, 43 (2012).

725 75. Kazamia, E. *et al.* Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition
726 in diatoms. *Sci. Adv.* **4**, eaar4536 (2018).

727 76. Kramer, J., Özkaya, Ö. & Kümmerli, R. Bacterial siderophores in community and host
728 interactions. *Nature Reviews Microbiology* **18**, 152–163 (2020).

77. Zimmer, R. K. & Ferrer, R. P. Neuroecology, chemical defense, and the keystone species concept. in *Biological Bulletin* **213**, 208–225 (2007).

78. Coale, T. H. *et al.* Reduction-dependent siderophore assimilation in a model pennate diatom. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23609–23617 (2019).

79. Basu, S., Gledhill, M., de Beer, D., Prabhu Matondkar, S. G. & Shaked, Y. Colonies of marine cyanobacteria Trichodesmium interact with associated bacteria to acquire iron from dust. *Commun. Biol.* **2**, 1–8 (2019).

80. Patten, C. L., Blakney, A. J. C. & Coulson, T. J. D. Activity, distribution and function of indole-3-acetic acid biosynthetic pathways in bacteria. *Crit. Rev. Microbiol.* **39**, 395–415 (2012).

81. Blin, K. *et al.* AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).

82. Robinson, S. L., Christenson, J. K. & Wackett, L. P. Biosynthesis and chemical diversity of β-lactone natural products. *Nat. Prod. Rep.* **36**, 458–475 (2019).

83. Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins-a viable alternative to antibiotics? *Nature Reviews Microbiology* **11**, 95–105 (2013).

84. Guo, L. *et al.* Lipid A Acylation and Bacterial Resistance against Vertebrate Antimicrobial Peptides systemic illnesses termed enteric fevers, which are char-acterized by microorganism colonization of the intes-tine, followed by systemic spread to tissues rich in. *Cell* **95**, 189–198 (1998).

85. Hinz, A., Lee, S., Jacoby, K. & Manoil, C. Membrane proteases and aminoglycoside antibiotic resistance. *J. Bacteriol.* **193**, 4790–4797 (2011).

86. Yasser, E.-N. & Adli, A. Toxicity of Single and Mixtures of Antibiotics to Cyanobacteria. *J. Environ. Anal. Toxicol.* **05**, 274 (2014).

87. EMEA. *European Medicines Agency. Doc ref. EMEA/CHMP/SWP/4447/00. http:// www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/ WC500003978.pdf*. (2006).

88. Grohmann, E., Christie, P. J., Waksman, G. & Backert, S. Type IV secretion in Gram-negative and Gram-positive bacteria. *Molecular Microbiology* **107**, 455–471 (2018).

89. Park, Y. J. *et al.* Structure of the type VI secretion system TssK–TssF–TssG baseplate subcomplex revealed by cryo-electron microscopy. *Nat. Commun.* **9**, 1–11 (2018).

90. Cianfanelli, F. R., Monlezun, L. & Coulthurst, S. J. Aim, Load, Fire: The Type VI Secretion System, a Bacterial Nanoweapon. *Trends in Microbiology* **24**, 51–62 (2016).

91. Silver, A. C. *et al.* Interaction between innate immune cells and a bacterial type III secretion system in mutualistic and pathogenic associations. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9481–9486 (2007).

765 92. Macho, A. P. & Zipfel, C. Targeting of plant pattern recognition receptor-triggered immunity
766 by bacterial type-III secretion system effectors. *Current Opinion in Microbiology* **23**, 14–22
767 (2015).

768 93. Shih, P. M. *et al.* Improving the coverage of the cyanobacterial phylum using diversity-driven
769 genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1053–1058 (2013).

770 94. Zere, T. R. *et al.* Genomic targets and features of BarA-UvrY (-SirA) signal transduction
771 systems. *PLoS One* **10**, e0145035 (2015).

772 95. Torres-Monroy, I. & Ullrich, M. S. Identification of Bacterial Genes Expressed During
773 Diatom-Bacteria Interactions Using an in Vivo Expression Technology Approach. *Front.*
774 *Mar. Sci.* **5**, 200 (2018).

775 96. Deng, J. *et al. Genomic Variations Underlying Speciation and Niche Specialization of*
776 *Shewanella baltica*. *mSystems* **4**, (2019).

777 97. Okkotsu, Y., Little, A. S. & Schurr, M. J. The pseudomonas aeruginosa AlgZR two-
778 component system coordinates multiple phenotypes. *Frontiers in Cellular and Infection*
779 *Microbiology* **4**, 82 (2014).

780 98. Gärdes, A., Iversen, M. H., Grossart, H.-P., Passow, U. & Ullrich, M. S. Diatom-associated
781 bacteria are required for aggregation of Thalassiosira weissflogii. *ISME J.* **5**, 436–445 (2011).

782 99. Huang, L., Xu, W., Su, Y., Zhao, L. & Yan, Q. Regulatory role of the RstB-RstA system in
783 adhesion, biofilm production, motility, and hemolysis. *Microbiologyopen* **7**, e00599 (2018).

784 100. Murthy, K. N., Mohanraju, R., Karthick, P. & Ramesh, C. Phenotypic and molecular
785 characterization of epiphytic vibrios from the marine macro algae of Andaman Islands, India.
786 *Indian J. Geo-Marine Sci.* **45**, 304–309 (2016).

787 101. Spaepen, S. & Vanderleyden, J. Auxin and plant-microbe interactions. *Cold Spring Harb.*
788 *Perspect. Biol.* **3**, 1–13 (2011).

789 102. Aly, K. A., Krall, L., Lottspeich, F. & Baron, C. The type IV secretion system component
790 VirV5 binds to the trans-zeatin biosynthetic enzyme Tzs and enables its translocation to the
791 cell surface of Agrobacterium tumefaciens. *J. Bacteriol.* **190**, 1595–1604 (2008).

792 103. Frosch, M., Edwards, U., Bousset, K., Krauße, B. & Weisgerber, C. Evidence for a common
793 molecular origin of the capsule gene loci in Gram negative bacteria expressing group II
794 capsular polysaccharides. *Mol. Microbiol.* **5**, 1251–1263 (1991).

795 104. Chen, D., Shao, Q., Yin, L., Younis, A. & Zheng, B. Polyamine function in plants:
796 Metabolism, regulation on development, and roles in abiotic stress responses. *Frontiers in*
797 *Plant Science* **9**, 1945 (2019).

798 105. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
799 assessing the quality of microbial genomes recovered from isolates, single cells, and
800 metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

106. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).

107. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

108. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).

109. Reddy, V. S. & Saier, M. H. BioV Suite - A collection of programs for the study of transport protein evolution. *FEBS J.* **279**, 2036–2046 (2012).

110. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

111. Tanabe, T. *et al.* Identification and Characterization of Genes Required for Biosynthesis and Transport of the Siderophore Vibrioferrin in Vibrio parahaemolyticus. *J. Bacteriol.* **185**, 6938–6949 (2003).

112. Tenenbaum, D. KEGGREST: Client-side REST access to KEGG. R package version 1.29.0. (2020).

113. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. Apcluster: an R package for affinity propagation clustering. *Bioinformatics* **27**, 2463–2464 (2011).

114. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science (80-.).* **315**, 972–976 (2007).

115. Hill, W. G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231 (1968).

116. Clayton, D. & Leung, H. T. An R package for analysis of whole-genome association studies. in *Human Heredity* **64**, 45–51 (2007).

117. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).

118. Kimura, M. Theoretical foundation of population genetics at the molecular level. *Theor. Popul. Biol.* **2**, 174–208 (1971).

119. VanLiere, J. M. & Rosenberg, N. A. Mathematical properties of the r2 measure of linkage disequilibrium. *Theor. Popul. Biol.* **74**, 130–137 (2008).

120. R Core Team. R: A Language and Environment for Statistical Computing. (2020).