SUPPLEMENTARY INFORMATION

Comparative whole-genome approach to identify traits underlying microbial interactions

Luca Zoccarato*, Daniel Sher*, Takeshi Miki, Daniel Segrè, Hans-Peter Grossart*

Corresponding authors(*)

Luca Zoccarato, zoccarato@igb-berlin.de; Daniel Sher, dsher@univ.haifa.ac.il;

Hans-Peter Grossart, hgrossart@igb-berlin.de

This PDF file includes:

Supplementary text

Supplementary figures 1 to 9

Other supplementary materials for this manuscript include the following:

Supplementary tables 1 to 7

Supplementary files 1 and 2

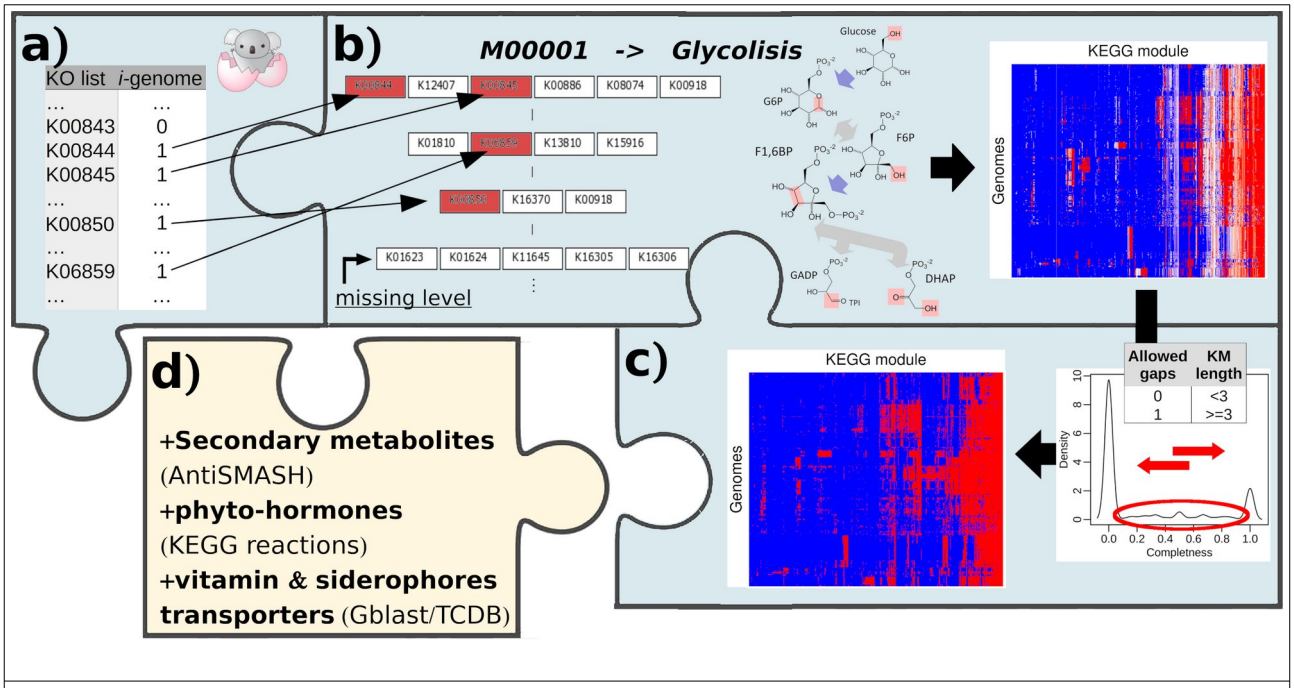
Remarks on the annotation pipeline

A relevant aspect in our analysis which needs to be kept in mind, we included only closed bacterial genomes (i.e. a single, high quality sequence of each DNA molecule such as chromosome or plasmid) or high-quality draft genomes (estimated using CheckM, see method section). The rational was to provide a comprehensive description of the full functional potential of pelagic marine bacteria which requires high-quality genomes to achieve the best information (i.e. functional genes; [1]. Gene annotation is *per-se* a challenging step, in particular, when it deals with environmental genomes for which many genes are still unknown and, therefore, cannot be properly annotated (~50% of the predicted coding sequences were annotated to KEGG orthologues (**KOs**) in our analysis).

A further step to improve cross-comparability among genomes was to re-annotate all of them with a standardized pipeline. For each genome, a list of genetic traits, termed as functional profile, was generated mainly recovering complete KEGG modules (**KMs**) from the annotated KOs (Supplementary Fig. 1a). We decided to focus our analysis on complete KMs instead of looking at the level of single annotated genes to seek for a more robust prediction of the genomic-inferred metabolic potential. To determine the presence of a KM, the detection of a single KO is not sufficient (unless the whole KEGG module is represented by a single KO), but the presence of a minimum complete set of related KOs is required. For example, for a KM with 5 required steps, and each step can be performed by several KOs, at least 1 KO per level has to be annotated in order to consider this genetic trait as complete (Supplementary Fig. 1b). This criterion was slightly relaxed by allowing for some gaps in bigger modules (i.e. 1 gap in KMs with > 3 steps) to account for possible annotation issues (i.e. missing KOs due to the absence of suitable reference genes in the database, unknown genes, miss-annotations, etc.; Supplementary Fig. 1c). The functional profiles were further enriched with the

annotation of other genetic traits using specific tools, e.g. for secondary metabolites (antiSMASH), phytohormones (KEGG pathway map01070), transporters for B vitamins and siderophore (BioV suite) (Supplementary Fig. 1d).

Additionally, the presence of a complete genetic trait does not necessarily translate into an expressed phenotype. Although several genetic traits are constitutively active, many of them can be under fine regulatory expression and be manifested only under specific environmental and physiological conditions.



Supplementary Fig. 1: Annotation workflow for the identification of genetic traits in genomes (a-c). Annotated KEGG orthologies (a) were recombined in all known KEGG modules (KM; b) and labeled as present or absent using a custom r script. The script, taking into account some completions rules, generates a presence/absence matrix (c). Further annotations were performed using antiSMASH for detection of secondary metabolites, KEGG reactions of the pathways map01070 for detection of phytohormones and Gblast against the Transporter Classification Database (TCDB) to

identify B vitamins and siderophore transporters (d).

16S rRNA gene classification

16S rRNA genes were identified using the RNAmmer option [2] during the prokka genome annotation step. The extracted sequences were classified using a local installation of SINA v1.6.1 [3] against the SILVA reference database SSU NR 99 v138 [4]. In case of multiple 16S rRNA gene copies in a genome we ensured that classification of each copy was coherent with all other copies. Taxonomy of *Cyanobacteria* genomes were manually curated using NCBI taxonomy associated with the NCBI Taxonomy IDs of the genomes.

GFCs and their phylogeny

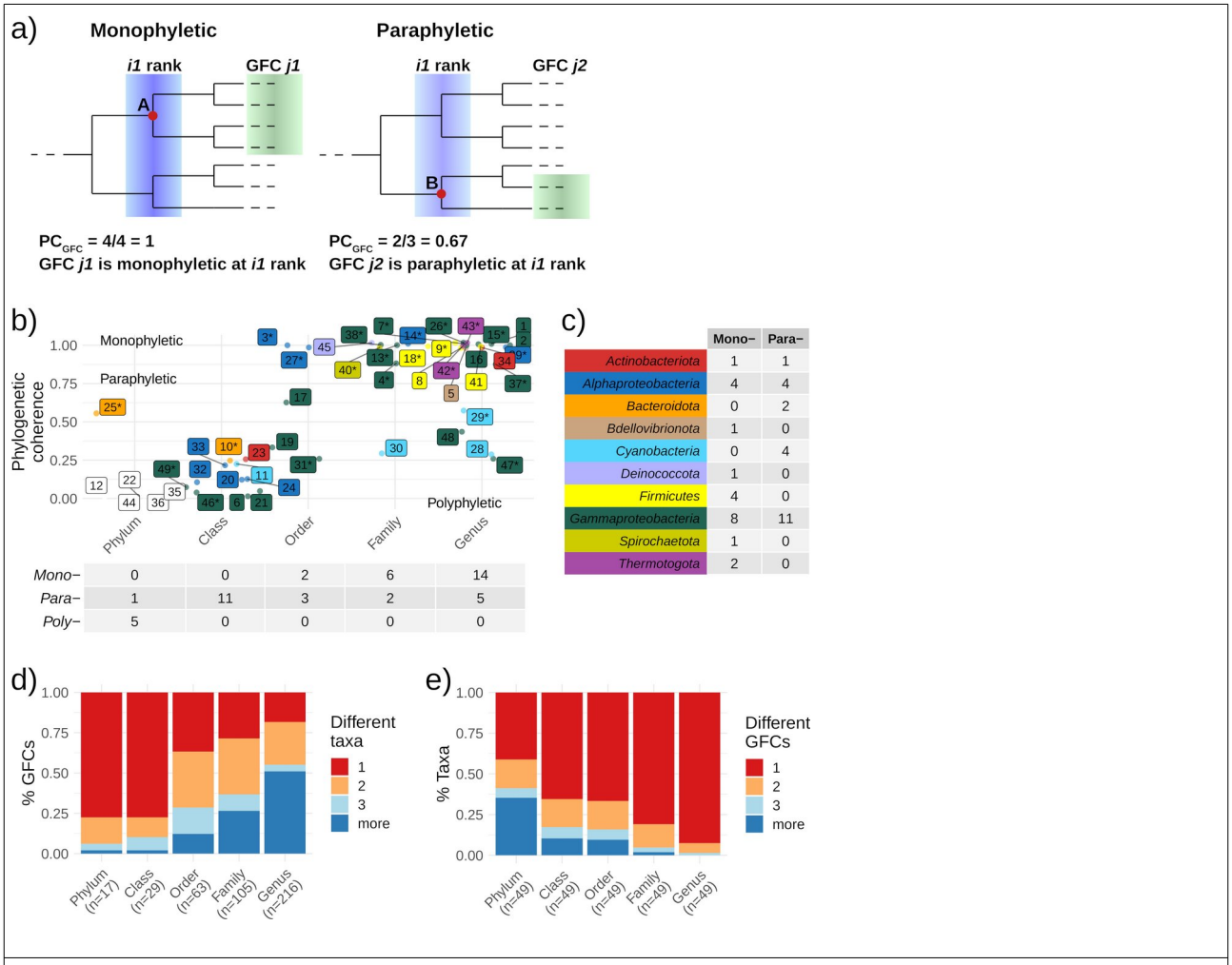
The genome clustering analysis retrieved a total of 49 genome functional clusters (GFCs). As shown in Supplementary Fig. 2d, most of these GFCs include only genomes of the same phylum (38), and fewer than 3 different families (14 with 1 and 17 with 2). At the genus levels more than half of the GFCs group together 3 or more different taxa. From the opposite perspective, the taxa level, more than half of the phyla were represented by 2 or more GFCs while nearly all genera by a single GFC (Supplementary Fig. 2e).

Similarly to [5], the phylogenetic coherence of each GFC was calculated based on the "local" phylogenetic coherence score (PC_{GFC}; Supplementary Fig. 2a):

$$PC_{GFC} = N_{GFC}/N_{taxon}$$

where N_{GFC} is the number of genomes grouped in a GFC, and N_{taxon} is the total number of genomes (included in our genome atlas) belonging to the last common ancestor of

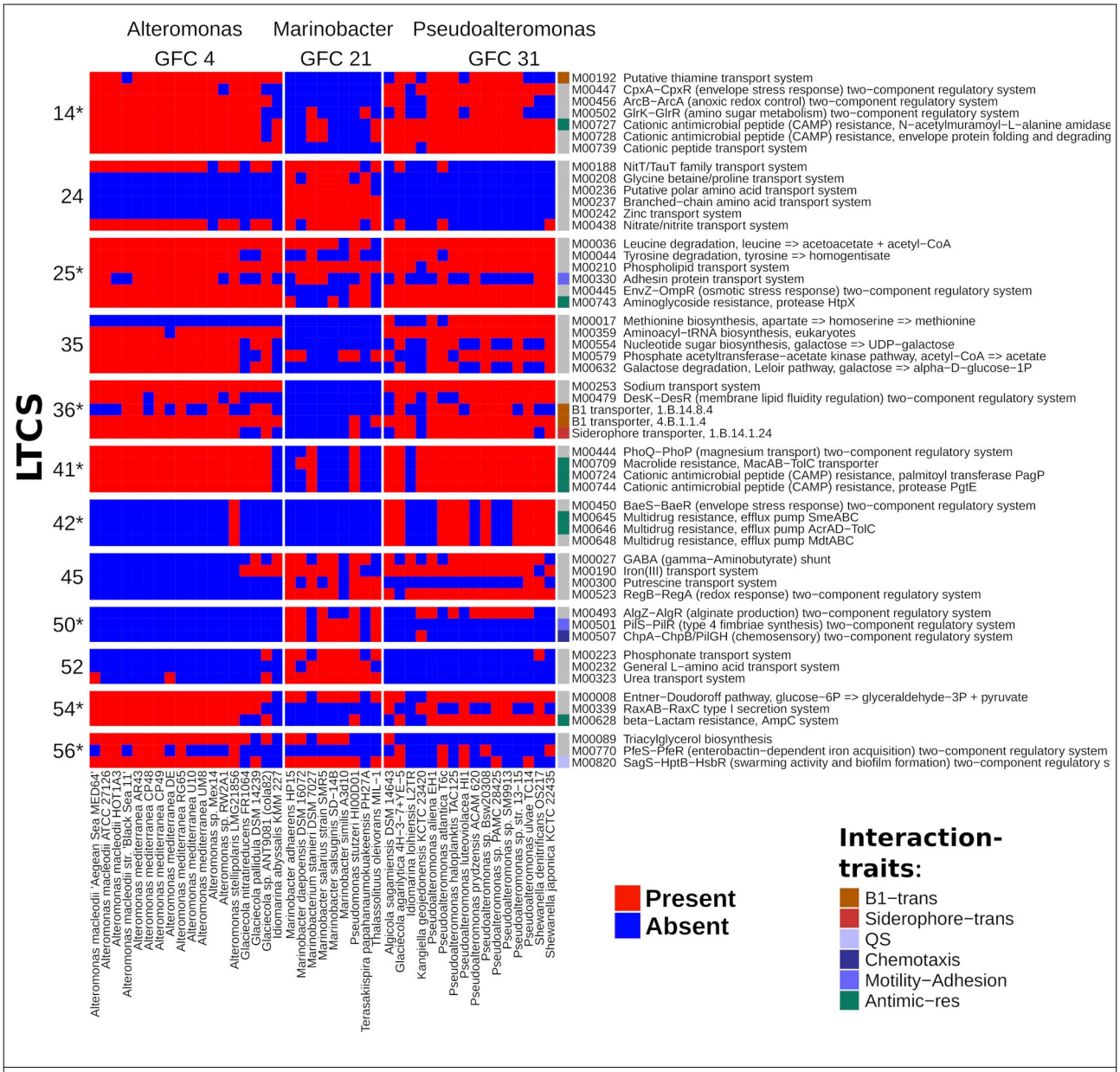
the GFC. The level of phylogenetic coherence of a GFC corresponds to the taxonomic rank of the last common ancestor. A PC_{GFC} of 1 indicates that all genomes belonging to the last common ancestor are included in the respective GFC which is considered to be monophyletic. A $PC_{GFC} < 1$ indicates instead a paraphyletic GFC. As the current genome availability doesn't allow for a uniform coverage of all bacterial taxa, a possible issue leading to taxonomic promiscuity of the paraphyletic GFCs might be related to the inclusion of "singleton" taxon (i.e. a single genome that represents a different taxon, as in GFCs 2, 7 or 29; see Supplementary Table 2 and Supplementary Fig. 2b). To avoid interpretation biases due to these singletons, we excluded a maximum of one singleton genome per GFC in the computation of the phylogenetic coherence. Nonetheless, ~30% of the paraphyletic GFCs included more evenly represented taxa as GFCs 14, 30 and 33, which grouped different genera with each genus represented by 3 or more genomes.



Supplementary Fig. 2: (a) Schematic example of calculating the phylogenetic coherence score in the genome functional clusters (GFCs) ‘j1’ and ‘j2’ carried out at the taxonomic rank ‘i1’. GFC ‘j1’ is monophyletic at that rank as it includes all genomes belonging to the related taxon ‘A’ (4/4) while GFC ‘j2’ is paraphyletic as it included only some genomes of the related taxon ‘B’ (2/3). (b) Highest phylogenetic coherence scores of each GFC separated by taxonomic ranks. Table underneath x-axis shows the amount of mono-, para- or polyphyletic GFCs for each rank; GFCs containing 1 “singleton” genome belonging to a different clade are marked with an asterisk; dots are color coded according to their taxonomy as shown in panel c. (c) Table summarizing the number of mono- and paraphyletic GFCs across different phyla. (d) Percentage of GFCs including 1, 2, 3 or more different taxa for each taxonomic rank. (e) Percentage of taxa included in 1, 2, 3 or more different GFCs for

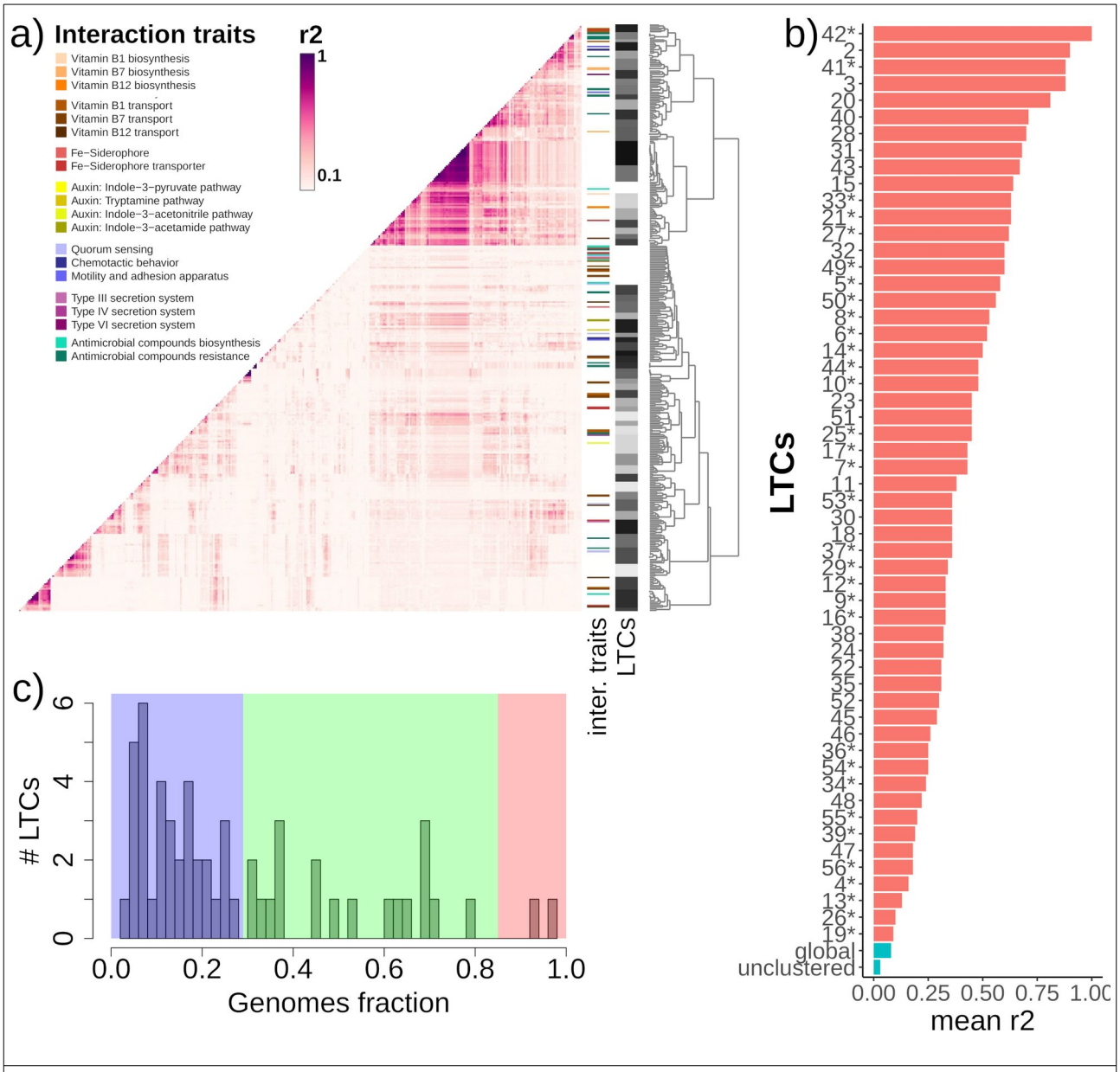
each taxonomic rank.

89



Supplementary Fig. 3: Insight analysis on the different linked trait clusters (LTCs) possessed by *Alteromonas*, *Marinobacter* and *Pseudoalteromonas* genome functional clusters (GFCs). The asterisk on the LTC names indicate that a LTC contains also one or more interaction traits. All tree GFCs share a set of complete LTCs, namely:2*, 3, 8*, 10*, 15, 17*, 21*, 23, 27*, 28, 33*, 40, 43 and 49*; check Supplementary Table 3

for more details abut these LTCs.



Supplementary Fig. 4: (a) Pairwise r^2 correlations among all genetic trait pairs; vertical color bar indicates the presence of different interaction traits while the vertical gray bar delineates specific linked trait clusters (LTCs). An interactive version of the same figure is available as Supplementary file 2. (b) Mean r^2 value of all pair of genetic traits included in each LTC; mean r^2 between all detected genetic traits ('global') and between all non-clustered genetic traits are shown as indication for

random correlation within the dataset. (c) Histogram of LTC occurrence across genomes showing the division in “core” (present in ≥ 90% of genomes; red), “common” (< 90% and ≥ 30%; green) and “ancillary” (≤ 30%; blue) LTC.

Absence-pattern among common LTCs

Some of the genetic traits included in the common LTCs 15, 23, 28 and 40 (found in 61-70% of the genomes) were consistently missing in some GFCs and/or taxonomic groups (Supplementary Fig. 5).

Broken TCA cycle

LTC 15 was absent in all *Cyanobacteria* (GFCs 11, 28, 29 and 30), *Thermotogota* (GFCs 42 and 43), *Spirochaetota* (GFC 40), *Deferribacterota* (GFC 12) and most GFCs of *Firmicutes* (8, 9 and 41). It included KEGG modules involved in the Citrate cycle (M00009, M00011, M00149), the biosynthesis of the amino acid serine and of phosphatidylethanolamine, the degradation oh histidine to glutamate and beta-oxidation (breakdown of fatty acids). Of note, LTC 15 is missing in ~30% of genomes and it encodes the complete TCA cycle (M00009) as well as the second carboxylation part (from 2-oxoglutarate to oxaloacetate; M00011). These genomes, however, still have the first carboxylation part of this cycle (from oxaloacetate to 2-oxoglutarate; M00010) which is included in the core LTC 2.

For almost four decades it was thought that *Cyanobacteria* indeed lack a full TCA cycle, until two new enzymes were discovered that, together catalyze the conversion of 2-oxoglutarate to succinate and thus functionally replace 2-oxoglutarate dehydrogenase and succinyl-CoA synthetase [6]. In our annotation, *Cyanobacteria*l

genomes lack these two ‘classical’ reactions of the TCA cycle (M00009), and therefore this pathway is flagged as incomplete, based on our rule of one gap for traits with up to 10 reactions. A manual search revealed that the 2-OGDC gene (K01652), which catalyzes the conversion of 2-oxoglutarate to succinic semialdehyde, is present in all *Cyanobacteria*l genomes, whereas the SSADH gene (K00135) that catalyzes the conversion of succinic semialdehyde to succinate is present only in the genomes of *Cyanobacteria* belonging to GFC 11. All of the pico-*Cyanobacteria* genomes lack the SSADH gene, consistent with the current view that these organisms lack a full TCA cycle [6].

In addition, our results show that other heterotrophic bacteria lack several or almost all of the reaction of this pathway. To the former case belongs genomes of *Spirochaetota* and *Thermotogota*, (grouped in GFCs 37 and 42, respectively) while to the latter case belong several *Firmicutes* (GFCs 9 and 38) and other *Thermotogota* (GFC 41) genomes. These findings confirm and broaden what has been previously reported for some genomes of these taxa [7–10].

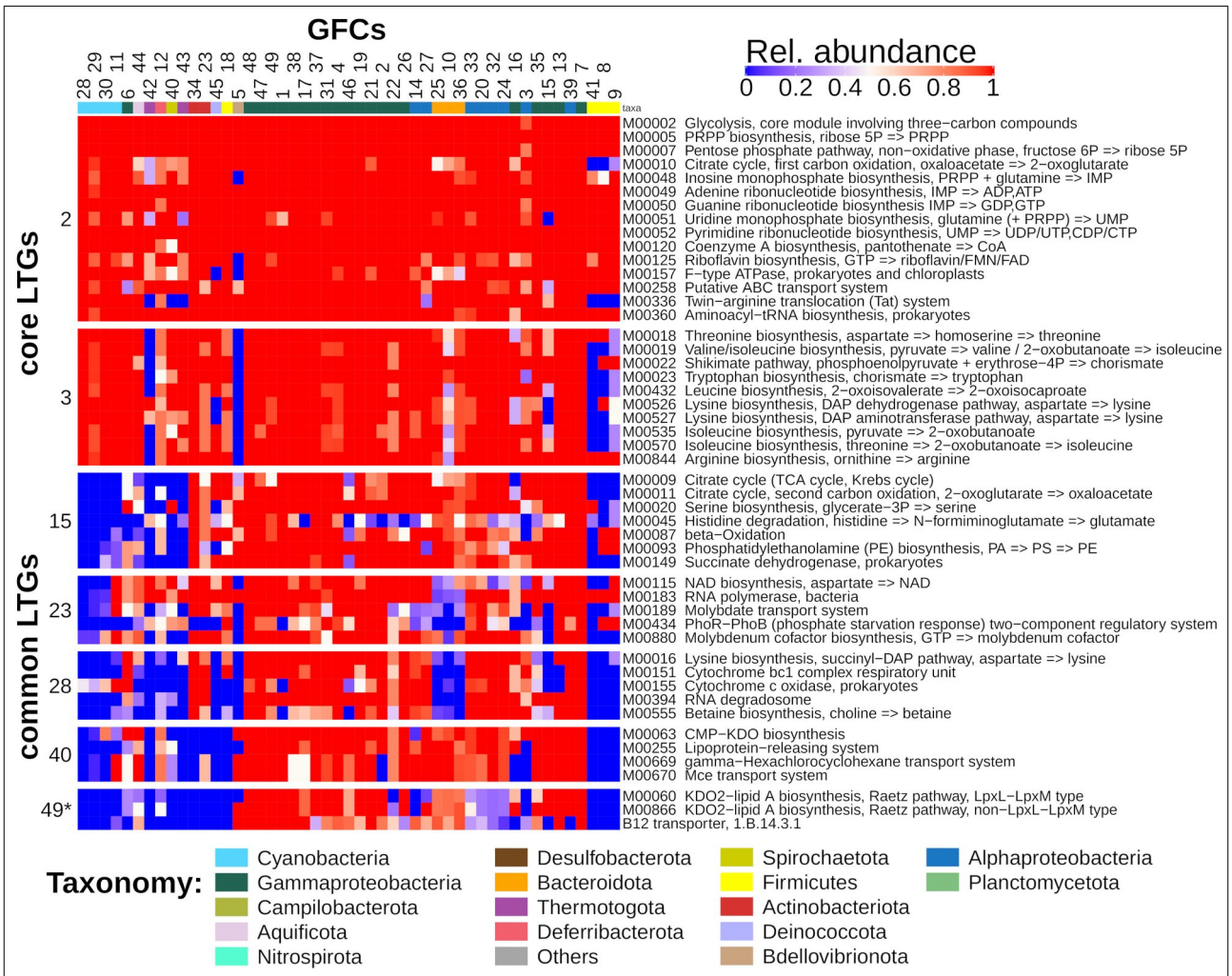
Differences in cell wall composition

In the LTC 40 and 49 (present in 70% and 65% of genomes), the absence of any pathway for the production of keto-deoxyoctulosonate (KDO, core constituent of lipopolysaccharide; M00060, M00063 and M00866) and of transporters for lipoprotein (M00255; Lol machinery), phospholipids (M00670; *mla* machinery) and gamma−Hexachlorocyclohexane (M00669; similar to *mla* machinery; Supplementary Fig. 5) provided another direct way to compare and validate our findings. The emerging incompleteness pattern of this LTC in GFCs 8, 9, 18, 41 (*Firmicutes*), 23, 34 (*Actinobacteriota*), and 45 (*Deinococcota*) could be largely explained by the bacterial cell wall type. In fact, all these GFCs grouped gram positive bacteria which are missing an outer membrane and, therefore, they don’t require any KDO (lipopolysaccharide

transporter (M00320; Lpt machinery) was an unclustered trait), nor export of any lipoprotein out of the cell membrane [11]. They also don’t need to exchange any phospholipids between the two membranes [12]. Despite being gram-negative, *Thermotogota* (GFCs 42 and 43), *Spirochaetota* (GFC 40) and most *Cyanobacteria* (GFCs 28, 29 and 30) were consistently missing some of the genetic traits of the LTC 40. Experimental evidence supports the lack of lipopolysaccharide in *Thermotogota* [13] and of KDO in the cell wall of *Cyanobacteria* [14] and *Spirochaetota* [15]. However the lack of any lipoprotein transporter in *Cyanobacteria* likely reflects a consistent gap in the annotation as these organisms are known to possess lipoproteins in the outer cell membrane [14].

Variants of RNA degradosome and polymerase

All *Bacteroidota* (GFCs 10, 25 and 36) and pico-*Cyanobacteria* (GFCs 10, 25 and 36) consistently lacked LTCs 23 and 28 grouping, among others, the biosynthesis of the RNA degradosome and polymerase. Absence of the two RNA related traits offered a good chance for validation. Indeed, *Bacteroidota* and pico-*Cyanobacteria* are known to possess different and unique types of σ factors and other subunits of the RNA polymerase [16, 17] and they do not possess the typical RNA degradosome complex [18–20]. The RNA degradosome was missing also in *Thermotogota* (GFCs 42 and 43), *Spirochaetota* (GFC 40), *Deferribacterota* (GFC 12) and in *Firmicutes* (GFCs 8, 9 and 41) for which similar conclusion could be drawn from the literature [18, 19].



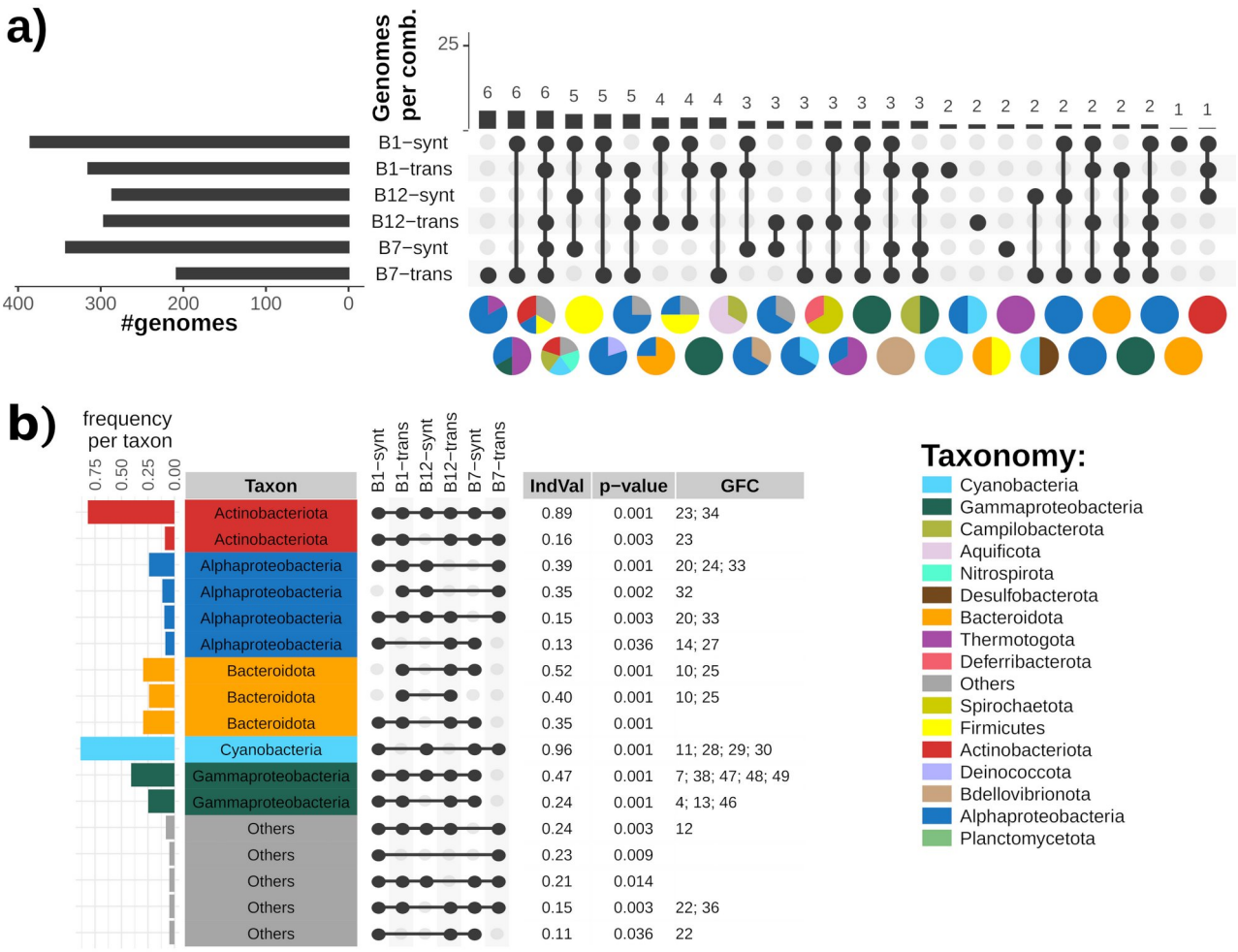
Supplementary Fig. 5: Relative abundance and absence-patterns of genetic traits grouped into core linked trait clusters (LTCs) 2 and 3, and in the common LTGs 15 (TCA cycle insight), 23, 28 (RNA polymerase and degradosome insight), 40 and 49 (cell wall insight) across all genome functional clusters (GFCs). The horizontal color bar indicate the most represented taxon in each GFC.

Combinations of vitamin traits’

As shown in Fig. 2c, some of the most frequent combinations of synthesis and uptake of vitamins B1, B12, and B7 appear more often in certain taxa than others. These patterns suggest the existence of taxa-specific evolutionary strategies for handling

these B vitamins. To test for this notion, we carried out an indicator species analysis as implemented in the function *multipatt* (func = "r.g", duleg = F, max.order = 2; package indicspecies) [21, 22]. The function is design to identify one or multiple species that can be used as indicators for certain habitats because of their strong species- habitat association. We performed the analysis using the most abundant B vitamins strategies (Fig. 2c) instead of species, while taxonomic groups and GFCs were used instead of habitats. We carried out the analysis mainly at the phyla level (except for Proteobacteria which we analyzed at the class level). Thereby, taxa represented by only 1-2 genomes were grouped together as ‘Others’. We also tested for specific associations of B vitamin strategies with combinations of site groups accounting for any possible taxa pair for combinations up to 5 different GFCs.

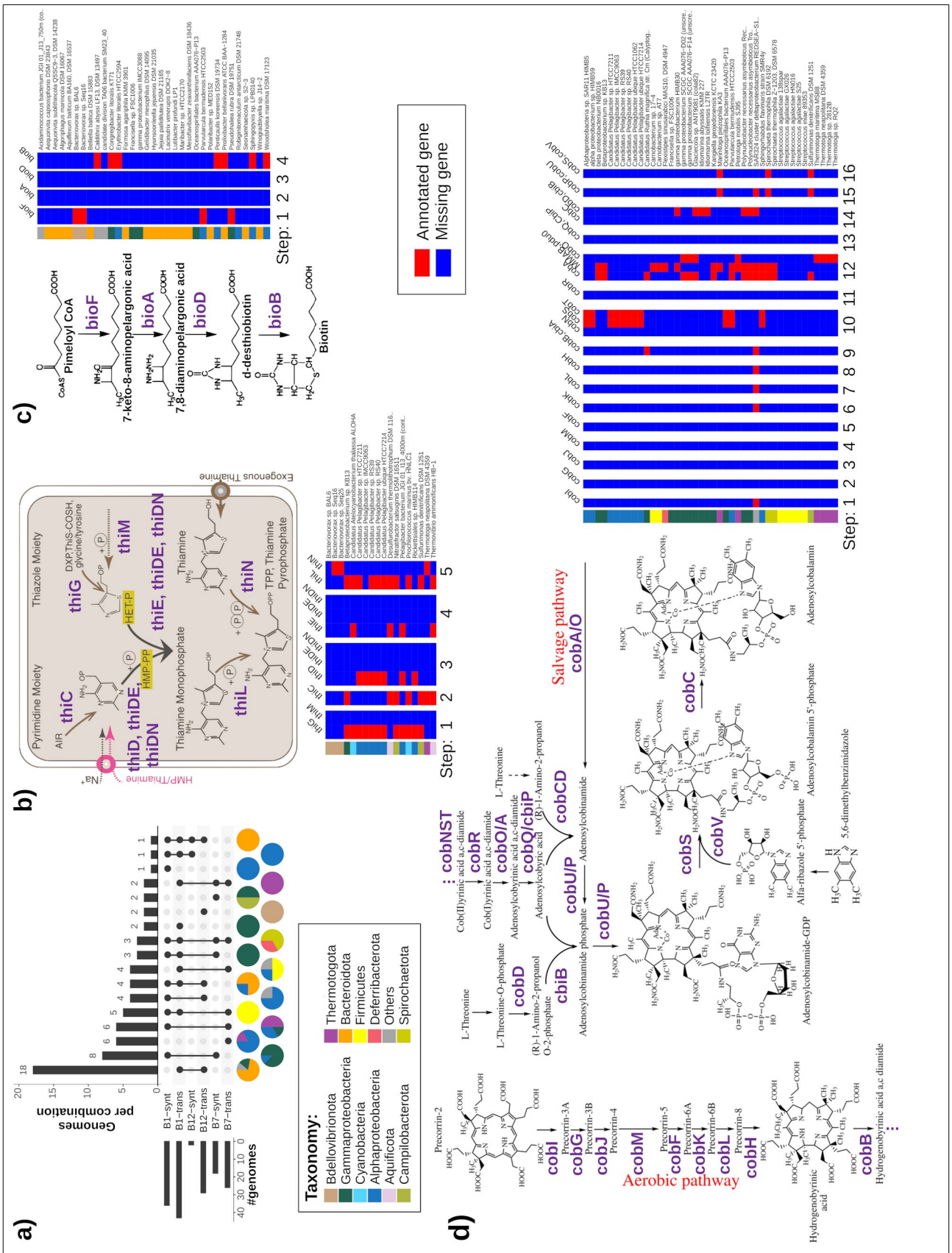
We identified a specific B vitamin strategy for 5 taxa (as ‘Others’ represent an artificial group containing Phyla with only a few representative genomes) and for 36 GFCs (Supplementary Fig. 6b). Except for *Cyanobacteria* and *Actinobacteriota*, all other taxa had more than one associated strategy which in some cases partitioned singularly into different GFCs, sometimes different GFCs of the same taxon possessed the same strategy, and in other cases the same strategy was present in GFCs of different taxa.



Supplementary Fig. 6: (a) UpSet plot showing the less common combinations of traits involved in the biosynthesis and transport of vitamins B1, B12, and B7. Horizontal bars indicate the total number of genomes for each trait, the dark connected dots indicate the different configurations of traits and vertical bars indicate the number of genomes possessing each configuration. Pie charts show the relative abundance of the taxonomic groups represented in each configuration. (b) Table with strategies for B vitamins uptake associated to specific taxa and genome functional clusters (GFCs). Horizontal bars show the frequency of strategies in each taxon and 'IndVal' expresses the strength of the strategy-taxon associations.

Problematic vitamin cases

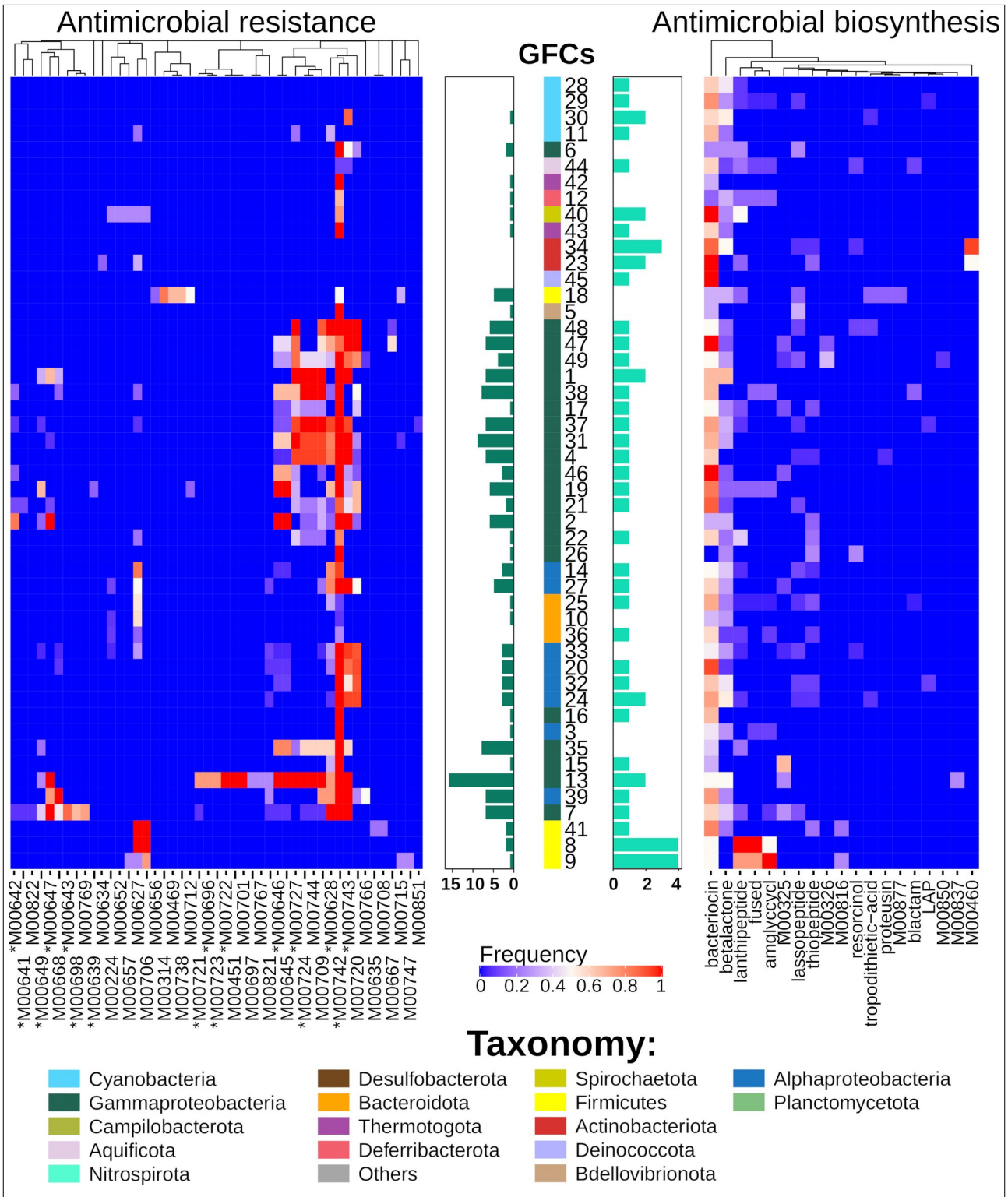
A small portion of genomes showed a fourth type of genetic configuration where neither the biosynthetic pathway nor the transporter for one (13%) or more (3%) vitamins have been detected (Supplementary Fig. 7a). These problematic cases could reflect limitations of our annotation process (e.g. unknown transporters or alternative genes/ pathways), however, there is growing evidence of organisms that even without complete biosynthetic pathways, but are able to grow on vitamin B intermediates [23, 24]. Therefore, we looked for the presence of possible fragmented biosynthetic pathways that could suggest forms of auxotrophy towards specific B-vitamin intermediates. We found that nearly all genomes with problematic vitamin B1 annotations possessed truncated biosynthetic pathways and need exogenous precursor supplies (Supplementary Fig. 7b). Contrary, for the problematic cases of vitamin B7 or B12, the related genomes owned only a few annotated genes or any at all (Supplementary Fig. 7c,d). This could point to an annotation issue of the related transporter, to yet unknown alternative biosynthetic pathways and reactions, or that these vitamins are not essential. Vitamin B12 is involved in amino acid and nucleotide synthesis, as well as in fatty- and amino acid breakdown; while vitamin B7 is involved in a “side” path of the TCA cycle and in the urea cycle. Most of these processes, however, have vitamin-independent routings (see in [25]) and genomes without biosynthetic capability nor vitamin B12 dependent genes have been reported [26] pointing that these vitamins, in some cases, might not be essential.



Supplementary Fig. 7: (a) UpSet plot showing incomplete combinations where both the biosynthetic pathway and the transporter for one (or more, 1% of genomes) of

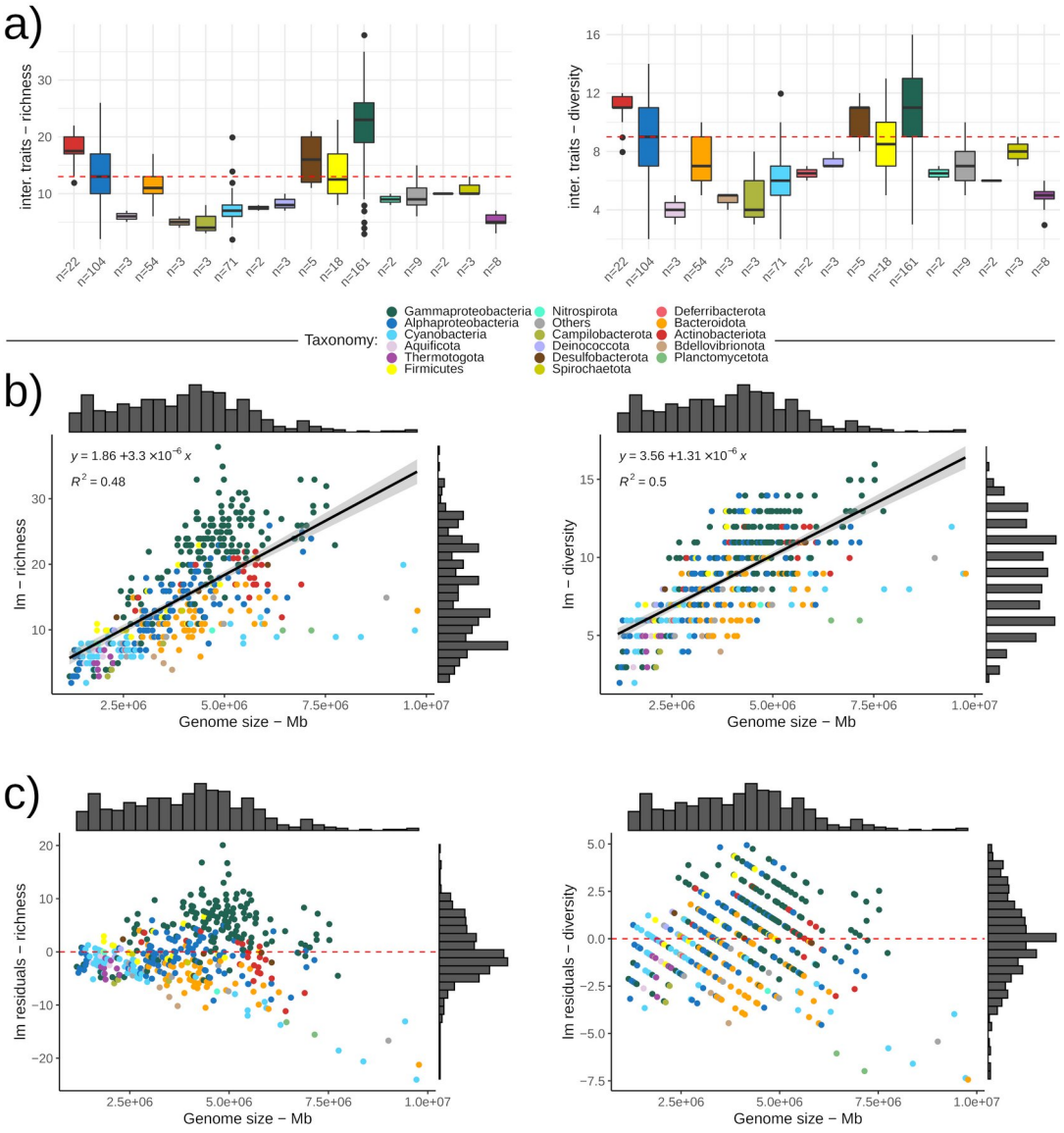
the B vitamin were missing. Horizontal bars indicate the total number of genomes for each trait, the dark connected dots indicate the different configurations of traits and vertical bars indicate the number of genomes possessing each configuration. Pie charts show the relative abundance of the taxonomic groups represented in each configuration. (b-d) Schematic overview of the biosynthetic pathways of (b) vitamin B1 (modified figure from [28], (c) B7 (modified figure from [27] and (d) B12 (modified figure from [29]. The related presence-absence maps show the annotated and missing genes involved in each pathway in genomes that were missing the genetic traits for both biosynthesis and transport of that vitamin. The genes are arranged as in the schematic view of the related pathway.

204



Supplementary Fig. 8: Distribution of antimicrobial resistance and biosynthetic traits in genome functional clusters (GFCs). Resistance traits marked with an asterisk are considered generic as they might be involved in other cellular metabolisms (Supplementary Table 4). Bar plots at the center shown the counts of traits with a

frequency > 0.5 in a GFC.



Supplementary Fig. 9: (a) Genome enrichment of interaction traits in terms of total number of interaction traits and diversity of interaction traits (as of different types e.g. vitamin B1 biosynthesis, type 4 secretion system,..). Box-plots show the distribution in different taxa and the dashed red lines represent the median value of trait richness and diversity across all genomes. (b) Linear regression between genomes size and interaction trait richness and diversity. (c) Plot of the residuals of the linear regression models. Genomes (i.e. dots) above the dashed red lines encode a higher number or a higher diversity of interaction traits that expected according to

their genome size, while genomes underneath the lines bear less interaction traits than expected.

References

1. Fadeev E, De Pascale F, Vezzi A, Hübner S, Aharonovich D, Sher D. Why close a bacterial genome? The plasmid of Alteromonas macleodii HOT1A3 is a vector for inter-specific transfer of a flexible genomic Island. Front Microbiol. 2016;7 MAR:1–13.

2. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8. doi:10.1093/nar/gkm160.

3. Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9. doi:10.1093/bioinformatics/bts252.

4. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41 Database issue:D590-6. doi:10.1093/nar/gks1219.

5. Koeppel AF, Wu M. Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. Nucleic Acids Res. 2013;41:5175–88. doi:10.1093/nar/gkt241.

6. Zhang S, Bryant DA. The tricarboxylic acid cycle in cyanobacteria. Science (80-). 2011;334:1551–3.

7. Neumann-Schaal M, Jahn D, Schmidt-Hohagen K. Metabolism the Difficile Way: The Key to the Success of the Pathogen Clostridioides difficile. Front Microbiol. 2019;10:219. doi:10.3389/fmicb.2019.00219.

8. Hoskins J, Alborn J, Arnold J, Blaszczak LC, Burgett S, Dehoff BS, et al. Genome of the bacterium Streptococcus pneumoniae strain R6. J Bacteriol. 2001;183:5709–17.

9. Wushke S, Fristensky B, Zhang XL, Spicer V, Krokhin O V, Levin DB, et al. A metabolic and genomic assessment of sugar fermentation profiles of the thermophilic Thermotogales, Fervidobacterium pennivorans. Extremophiles. 2018;22:965–74. doi:10.1007/s00792-018-1053-4.

10. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, et al. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. Nature. 1997;390:580–6.

11. Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harbor perspectives in biology. 2010;2:a000414. doi:10.1101/cshperspect.a000414.

12. Malinverni JC, Silhavy TJ. An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane. Proc Natl Acad Sci. 2009;106:8009–14. doi:10.1073/pnas.0903229106.

13. Plötz BM, Lindner B, Stetter KO, Holst O. Characterization of a novel lipid A containing D-galacturonic acid that replaces phosphate residues. The structure of the lipid A of the lipopolysaccharide from the hyperthermophilic bacterium Aquifex pyrophilus. J Biol Chem. 2000;275:11222–8. doi:10.1074/jbc.275.15.11222.

14. Durai P, Batool M, Choi S. Structure and effects of cyanobacterial lipopolysaccharides. Marine Drugs. 2015;13:4217–30. doi:10.3390/md13074217.

15. Vinogradov E, Paul CJ, Li J, Zhou Y, Lyle EA, Tapping RI, et al. The structure and biological characteristics of the Spirochaeta aurantia outer membrane glycolipid LGLB. Eur J Biochem. 2004;271:4685–95. doi:10.1111/j.1432-1033.2004.04433.x.

16. Vingadassalom D, Kolb A, Mayer C, Rybkine T, Collatz E, Podglajen I. An unusual primary sigma factor in the Bacteroidetes phylum. Mol Microbiol. 2005;56:888–902. doi:10.1111/j.1365-2958.2005.04590.x.

17. MacGregor BJ. Abundant intergenic TAACTGA direct repeats and putative alternate RNA polymerase β’ subunits in marine beggiatoaceae genomes: Possible regulatory roles and origins. Front Microbiol. 2015;6 DEC:1397. doi:10.3389/fmicb.2015.01397.

18. Kaberdin VR, Singh D, Lin-Chao S. Composition and conservation of the mRNA-degrading machinery in bacteria. Journal of Biomedical Science. 2011;18:23. doi:10.1186/1423-0127-18-23.

19. Aït-Bara S, Carpousis AJ. RNA degradosomes in bacteria and chloroplasts: Classification, distribution and evolution of RNase E homologs. Mol Microbiol. 2015;97:1021–35.

20. Zhang JY, Deng XM, Li FP, Wang L, Huang QY, Zhang CC, et al. RNase E forms a complex with polynucleotide phosphorylase in cyanobacteria via a cyanobacterialspecific nonapeptide in the noncatalytic region. RNA. 2014;20:568–79.

21. De Cáceres M, Legendre P. Associations between species and groups of sites: Indices and statistical inference. Ecology. 2009;90:3566–74. doi:10.1890/08-1823.1.

22. Cáceres DM, Legendre P, Moretti M. Improving indicator species analysis by combining groups of sites. Oikos. 2010;119:1674–84. doi:10.1111/j.1600-0706.2010.18334.x.

23. Paerl RW, Sundh J, Tan D, Svenningsen SL, Hylander S, Pinhassi J, et al. Prevalent reliance of bacterioplankton on exogenous vitamin B1 and precursor availability. Proc Natl Acad Sci. 2018;115:E10447–56.

24. Wienhausen G, Noriega-Ortega BE, Niggemann J, Dittmar T, Simon M. The exometabolome of two model strains of the Roseobacter group: A marketplace of microbial metabolites. Front Microbiol. 2017;8 OCT:1–15.

25. Romine MF, Rodionov DA, Maezato Y, Osterman AL, Nelson WC. Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors in microbial communities. ISME J. 2017;11:1434–46. doi:10.1038/ismej.2017.2.

26. Shelton AN, Seth EC, Mok KC, Han AW, Jackson SN, Haft DR, et al. Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. ISME J. 2019;13:789–804.

27. Croft MT, Warren MJ, Smith AG. Algae Need Their Vitamins. Eukaryot Cell. 2006;5:1175–83. doi:10.1128/EC.00097-06.

28. McRose D, Guo J, Monier A, Sudek S, Wilken S, Yan S, et al. Alternatives to vitamin B1 uptake revealed with discovery of riboswitches in multiple marine eukaryotic lineages. ISME J. 2014;8:2517–29.

29. Fang H, Kang J, Zhang D. Microbial production of vitamin B12: A review and future perspectives. Microb Cell Fact. 2017;16:1–14.