Model cyanobacterial consortia reveal a consistent core microbiome independent of inoculation source or cyanobacterial host species

Andreja Kust[1,2], Jackie Zorz[1,2], Catalina Cruañas Paniker[2,#], Keith Bouma-Gregson[3], Netravathi Krishnappa[2], Wendy Liu[1], Jillian F. Banfield[1,2,4,5], Spencer Diamond[1,2,*]

[1]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, USA.
[2]Innovative Genomics Institute, University of California, Berkeley, CA, USA.
[3]U.S. Geological Survey, California Water Science Center, Sacramento, CA, USA.
[4]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA.
[5]School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia.

[#]Current Address: Department of Life Sciences, Imperial College London, London, United Kingdom.

***Corresponding author:**
Spencer Diamond
Innovative Genomics Institute
University of California - Berkeley
Innovative Genomics Institute Building
2151 Berkeley Way, Berkeley, CA 94720
Phone: (818) 268-5776
Email: sdiamond@berkeley.edu

Key Words: Cyanobacteria, Microbiology, Metagenomics, Microbial Ecology, Synthetic Communities.

Abstract

Cyanobacteria are integral to biogeochemical cycles, influence climate processes, and hold promise for commercial applications. In natural habitats, they form complex consortia with other microorganisms, where interspecies interactions shape their ecological roles. Although *in vitro* studies of these consortia have significantly advanced our understanding, they often lack the biological replication needed for robust statistical analysis of shared microbiome features and functions. Moreover, the microbiomes of many model cyanobacterial strains, which are central to our understanding of cyanobacterial biology, remain poorly characterized. Here, we expanded on existing *in vitro* approaches by co-culturing five well-characterized model cyanobacterial strains with microorganisms filtered from three distinct freshwater sources, generating 108 stable consortia. Metagenomic analyses revealed that, despite host and inoculum diversity, these consortia converged on a similar set of non-cyanobacterial taxa, forming a 25-species core microbiome. The large number of stable consortia in this study enabled statistical validation of both previously observed and newly identified core microbiome functionalities in micronutrient biosynthesis, metabolite transport, and anoxygenic photosynthesis. Furthermore, core species showed significant enrichment of plasmids, and functions encoded on plasmids suggested plasmid-mediated roles in symbiotic interactions. Overall, our findings uncover the potential microbiomes recruited by key model cyanobacteria, demonstrate that laboratory-enriched consortia retain many taxonomic and functional traits observed more broadly in phototroph-heterotroph assemblages, and show that model cyanobacteria can serve as robust hosts for uncovering functional roles underlying cyanobacterial community dynamics.

Introduction

In nature cyanobacteria form complex microbial communities [1] where inter-microbial interactions are critical for the resilience, stability, and environmental impact of cyanobacterial assemblages [2–7]. Cyanobacteria-associated microorganisms can act as fixed carbon sinks, modulate cyanobacterial productivity [8], and contribute to the biosynthesis of energetically costly compounds that benefit the entire microbial community [4]. These associated microbes may also play key roles in nutrient remineralization, enhancing the bioavailability of essential nutrients [6], and help mitigate oxidative stress [9]. Conversely, cyanobacteria-associated microorganisms may engage in competitive interactions, including competition with cyanobacteria for inorganic nutrients [9–12]. Cyanobacterial consortia share taxonomic and functional similarities with other phototroph-associated consortia [13] including those of the plant rhizosphere [14, 15], diatoms [16–18] and algae [19], suggesting that common ecological principles drive the assembly of these systems. However, a comprehensive understanding of the dynamics that drive cyanobacterial community assembly, and the roles of core microbial members within these communities, is still evolving [20–26].

Previous studies have provided important insights into cyanobacterial communities using cultivation-independent methods [6, 27–31], synthetic community assemblages [8, 32–35] and isolation and co-cultivation of cyanobacteria with their native microbial partners [21–23, 25, 26, 36–38]. However, *in vitro* consortia, which have made important contributions to our understanding of cyanobacterial communities [8, 21–25, 32–36], often contain few bacterial species and when assembled from isolates can lack taxonomic diversity that accurately reflects natural consortia. Alternatively, naturally sourced communities are not typically cultivated under standardized inoculation conditions or monitored longitudinally. Furthermore, the potential composition and function of microbiomes associated with cyanobacterial model organisms,

model species critical to our understanding of cyanobacterial biology, are largely uncharacterized. To address these limitations, we built on previous methodologies to establish a large number of *in vitro* communities grown under standardized laboratory conditions where model cyanobacterial species were used as hosts to enrich co-associated microbiomes from diverse natural inoculation sources. We chose well-characterized model cyanobacterial strains as community hosts including *Synechocystis* spp. [39, 40], *Synechococcus elongatus* [41, 42], and *Nostoc* spp. [43], where information about co-associated communities in natural environments is limited [32, 33, 44]. This system builds upon previous approaches by balancing the simplicity and genetic tractability of model cyanobacteria with the complexity of inoculating these hosts with species derived from environmental samples.

Ultimately this work resulted in the establishment of 108 stable *in vitro* consortia. Using 16S rRNA gene amplicon and genome-resolved metagenomic profiling, we observed that despite the taxonomic composition of the freshwater source microbiomes and diversity of cyanobacterial hosts, the resulting communities rapidly stabilized and became enriched with taxa commonly found in natural freshwater cyanobacterial and other phototrophic symbiotic assemblages [6, 25, 45, 46]. We monitored the stability of these communities over time, including post-cryopreservation revival, confirming their robustness, reproducibility, and ability to be transferred or re-utilized for additional study. The large degree of replication in this study enabled the identification of a 25 species core microbiome, which has not been previously evaluated using *in vitro* efforts. Comparison of core microbiome metagenome assembled genomes (MAGs) to other MAGs recovered in this study confirmed statistically significant enrichment of metabolic traits previously noted observationally in other *in vitro* studies [12, 16, 25, 28, 47] and identified additional functions in porphyrin metabolism and cyanobacterial alkane degradation associated with the core microbiome. Core microbiome members were also significantly enriched in putative plasmids, which encoded functionality potentially supportive of phototroph-heterotroph interactions [48]. Overall, our results uncover the diversity of microbiomes recruited by model cyanobacterial species, demonstrate that laboratory-enriched consortia using model strains retain many of the taxonomic and functional features seen in broader phototroph–heterotroph systems, and establish model cyanobacteria as effective platforms for uncovering new functional roles driving cyanobacterial community dynamics.

Materials and Methods

Cyanobacterial selection, inoculation, growth, and passaging

We selected five model cyanobacterial strains—*Synechococcus elongatus* PCC 7942 (7942), *Synechococcus elongatus* UTEX 3055 (3055), *Nostoc* spp., PCC 7120 (A7120), *Synechocystis* spp., PCC 6803 (6803), and *Leptolyngbya* spp**.,** BL0902 (L0902)—based on their genetic tractability, metabolic diversity, and relevance to cyanobacterial biology and biotechnology. Axenic model cyanobacterial strains were inoculated with bacterial biomass collected from three freshwater bodies (**Table S1**), and were grown in BG-11 [49] medium at 24°C under a 12:12 light cycle and passaged at both weekly and bi-weekly intervals for 12 weeks. Samples were collected at each passage and rigorous standardized assessments of culture contamination, and host dominance were applied to determine if a co-culture should be retained in the study (**Table S2**). For comprehensive details, see **Supplementary Methods** and **Fig. 1A**.

DNA extraction and sequencing

DNA for shotgun metagenomic sequencing was extracted from inoculation source material and biomass of co-culture samples at day 84 of passaging using a modified Qiagen DNeasy PowerSoil Pro Kit protocol. DNA from co-culture samples was sequenced with paired-end reads

of 150 bp, whereas DNA from the source material was sequenced with paired-end reads of 250 bp on a NovaSeq System (Illumina). We used longer sequencing reads (250 bp) and a larger library insert size for complex environmental source samples to improve assembly and contig recovery [50].

For 16S rRNA gene amplicon sequencing, genomic DNA was extracted [51] from co-culture samples of all passages. The extracted DNA was amplified for the V4–V5 regions of the 16S rRNA gene using Illumina-compatible primers and sequenced on a MiSeq System (Illumina) with a read depth of 50,000 reads per sample (mean = 46,098 reads per sample). This depth enabled frequent monitoring of community changes and complemented shotgun metagenomic sequencing, which was employed for species-level resolution and functional analysis of gene content. For comprehensive details, see **Supplementary Methods.**

16S rRNA gene data processing and statistical analysis

Amplicon sequence variants (ASVs) were generated from fastq files using the USEARCH-UNOISE3 pipeline [52], with modifications, to identify a total of 2,126 ASVs. Taxonomy was assigned using a naive Bayes classifier trained on the SILVA v138 SSU [53] reference database. We ultimately analyzed 869 rRNA gene amplicon samples (**Table S3**) that represented samples collected from all longitudinal passages of co-culture samples determined to be un-contaminanted, dominated by the expected host cyanobacterium, and also analyzed with shotgun genome-resolved metagenomics at day 84. Alpha-diversity (Richness, Shannon) was assessed with rarefied ASV counts (**Table S4**) using linear mixed-effects models. Beta-diversity was analyzed with additive log ratio (ALR) transformed ASVs and Aitchison distances (**Table S5**). Differential abundance was tested with Maaslin2 (**Tables S6**). For comprehensive details, see **Supplementary Methods** and **Code Availability.**

Assembly independent marker gene analysis of shotgun metagenomic data

As recovering MAGs from complex source samples was challenging, we also employed an analysis where marker genes were directly identified from shotgun metagenomic reads to more accurately compare diversity and taxonomic composition between source samples and co-cultures. This was conducted using ribosomal proteins as in prior work [54]. Using SingleM v0.13.2 [55], unique Ribosomal protein L6 (rpL6) were identified and clustered at 95% nucleotide identity to produce rpL6 OTUs that could be compared across samples (**Table S7**). Over/under-enrichment of bacterial orders in cyanobacterial communities relative to source samples was evaluated with a permutation-based method (**Table S8**). For comprehensive details, see **Supplementary Methods** and **Code Availability**.

Metagenome assembly, binning, bin de-replication

Metagenome assembly and binning were conducted using methods similar to [50]. However, scaffolds > 5000 bp were reassembled with COBRA v50 [56], and validated circular contigs were re-incorporated into the assembly. Genome binning was performed on a sample by sample basis and only used scaffolds > 2,500 bp using MetaBat2, maxbin2, CONCOCT, and vamb [57–60]. Best MAGs across the four methods for each sample were selected using DasTool [61] and evaluated with checkM [61, 62]. Only MAGs with ≥60% completeness and contamination of ≤10% were retained. MAGs were de-replicated at the species level and representative species MAGs were selected using dRep [63], considering MAGs as the same species if ANI was ≥ 95% across ≥ 10% of the MAG length. This resulted in a non-redundant set of 537 MAGs (**Table S9**). Taxonomic classification was conducted using GTDB-tk [64] classify workflow against the GTDB-R214 taxonomy. For comprehensive details, see **Supplementary Methods.**

Phylogenetic analysis
GToTree v1.6.34 [65] was employed to detect, align, trim, and concatenate 120 universal single copy genes [66] from non-redundant MAGs as well as an archaeal outgroup species *Acidianus hospitalis* W1 (GCF_000213215.1_2). MAGs with < 10% of markers were excluded. Maximum-likelihood phylogenetic reconstruction was conducted with IQ-TREE version 1.6.12 [67]. The best-fit evolutionary models for each marker protein partition were determined using ModelFinder Plus. Tree topology significance was assessed using ultrafast bootstrap analysis with 1,000 replicates and the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) with 1,000 replicates. The resulting phylogenetic tree was displayed using iTol v6.8.1 [68]. For comprehensive details, see **Supplementary Methods.**

MAG abundance mapping and diversity analysis
Shotgun metagenomic reads from each sample were mapped to species representative MAGs with Bowtie2 [69] and coverage quantified with CoverM [70]. MAGs were considered present in a sample if they met stringent coverage and breadth criteria (**Table S10**). Metagenomic samples were removed from analysis when the expected host cyanobacterium comprised ≤ 90 % of the cyanobacterial fraction resulting in 121 shotgun metagenomic samples (108 cyanobacterial cultures and 13 source environment samples) for downstream analysis.
Beta-diversity was assessed using Aitchison distance of MAG mapped read counts, and permutational analysis of variance (PERMANOVA) was used to estimate the influence of sample type (culture or source) and location (**Table S11**). For comprehensive details, see **Supplementary Methods** and **Code Availability.**

Core microbiome criteria
The concept of a core microbiome lacks a universally accepted definition [71, 72], so we developed an operational framework to identify core microbiome species (MAGs) across co-cultures based on two key criteria: (i) a core species must reproducibly co-occur (was considered present in a co-culture by MAG mapping criteria) with at least four cyanobacterial host species across three independent samples for each host (a minimum of 12 samples); and (ii) it must be consistently recruited (was considered present in a co-culture by MAG mapping criteria) from at least two distinct environmental sources across at least three independent samples per environmental inoculum source (a minimum of 6 samples). This operational definition ensures core species are consistently represented across both host diversity and environmental variability, reducing the likelihood of sampling artifacts or host-specific associations. To asses the relatedness of core microbiome MAGs assigned to the same species, we performed digital DNA-DNA hybridization (dDDH) analysis using the Genome-to-Genome Distance Calculator [73, 74](**Tables S12**).

Population level single nucleotide variant (SNV) analysis
Population-level SNV analysis was performed on 108 cyanobacterial co-culture samples using inStrain v1.6.4 [75], generating 3,088 pairwise comparisons for 133 MAGs, including 555 comparisons involving core microbiome species. No identical strain sharing was detected across 176 comparisons between geographically distinct inocula (popANI ≥ 99.999%; **Table S13**). Beta regression modeling showed significant effects of inoculum source and host strain on MAG microdiversity (adjusted P ≤ 0.05; **Table S14**). For comprehensive details, see **Supplementary Methods, Supplementary Results,** and **Code Availability.**

Assessment of co-culture community taxonomic overlap with other studies
We compiled a set of 1,843 MAGs that passed completeness and contamination criteria used in our study (920 distinct species) from 10 genome-resolved studies involving diverse cyanobacterial or plant rhizosphere samples (**Tables S15 and S16**). These included studies of

non-axenic cyanobacteria cultures [23, 76–78], environmental cyanobacterial consortia [27, 28, 30, 79], and rhizosphere derived bacteria [80, 81]. Identical species (≥ 95% genome ANI) between the 920 species-representative MAGs from other studies and the 319 MAGs observed in co-cultures of this study were identified using dRep [63]. Additionally we evaluated higher level taxonomic overlap between species-representative MAGs present in our co-cultures and those of other studies at the class and genus level using taxonomic classifications produced using GTDB-tk [64] classify workflow against the GTDB-R214 taxonomy.

Functional annotation and core microbiome functional enrichment analysis

Predicted proteins from all 537 MAGs were annotated using KOfam HMMs via kofamscan [82], and METABOLIC [83]. Alkane degradation genes were identified using hmmsearch [84] with HMMs from CANT-HYD [85]. To assess core microbiome functional enrichment (n = 25 MAGs) compared to non-core MAGs (n = 512 MAGs), Fisher tests were conducted on KEGG orthologs (KOs) with p-values adjusted for multiple testing (FDR ≤ 0.05). Two tests were performed: one using all MAGs (n = 537 MAGs) and another constrained to Pseudomonadota, Bacteroidota, and Spirochaetota MAGs (n = 391 MAGs) to account for taxonomic bias in the core microbiome MAG set. Broader functional categories were assigned, and differential representation was analyzed using Fisher tests with multiple testing correction. For comprehensive details, see **Tables S17-S20**, **Supplementary Methods,** and **Code Availability**.

Identification, curation, and analysis of putative mobile genetic elements

Contigs predicted to be mobile genetic elements (MGEs) were identified across 537 MAGs using geNomad [86] filtering for contigs ≥12 kb and FDR ≤ 0.05 (**Table S21**). We evaluated all contigs ≥12 kb that were confidently identified as putative plasmids, given the high frequency of circular element fragmentation in short-read metagenomic assemblies. Plasmid assignment to MAGs was based on a guilt-by-association approach, where plasmids co-binned with a MAG were considered associated. Plasmid content was compared between core, auxiliary, and source genomes using Dunn's test. All plasmid contigs were aligned to the IMG/PR database [87] using BLASTn for identification (**Table S22**). Genes related to conjugation, mobilization, and replication were identified using HMMs from Conjscan [88] and Pfam [89] via hmmsearch [84]. Plasmids were classified as mobilizable or conjugative based on the presence of key genes (e.g., MOBX, T4CP, virB4). General functional annotation was performed identically as for MAGs detailed above.

For MGE-associated functional enrichment, plasmid-contigs were removed from the 537 MAGs, and KEGG orthology (KO) enrichment analysis was repeated. Differential KO enrichment was compared between the full and MGE-removed analyses using Fisher's test, with FDR correction (FDR ≤ 0.05) (**Tables S18 and S19**). Fold changes in KO enrichment were calculated between the original and MGE-removed analyses. For comprehensive details, see **Supplementary Methods,** and **Code Availability**.

Results

Inoculation of model cyanobacterial strains with bacteria from diverse freshwater sources established stable *in vitro* communities

A total of 108 consortia were retained at the end of 12 weeks of passaging being free of contamination and dominated by the original cyanobacterial host **(Fig. S1 and Supplementary Methods)**. The retention of consortia in the experiment was influenced by the cyanobacterial host, inoculum source, and the interaction between host strain and passage rate. Among the

tested strains, 6803, L0902, and A7120 exhibited the highest retention rates (**Supplementary Results**).

Profiling of 16S rRNA gene amplicons showed that despite differences in inoculum source and host strain, communities rapidly stabilized within the first 28 days (**Figs. 1B-D and S2**). Cryopreservation and revival confirmed the long-term stability of the communities, with diversity metrics and taxonomic composition returning to pre-cryopreservation values **(Fig. 1B-D)**. The final consortia set evaluated for this study (84 days post inoculation) closely resembled natural cyanobacterial communities, displaying a consistent taxonomic structure (**Fig. 1, S3, and Supplementary Results**). These results confirm that model cyanobacterial species can form stable, reproducible consortia *in vitro*, simulating natural microbiomes [6, 9, 19, 90].

Genomic resolution enabled comprehensive characterization of *in vitro* communities

Shotgun sequencing of stable consortia (84 days) and source samples yielded 537 non-redundant species-level MAGs, with 59.8% classified as high-quality (completeness of ≥90%, contamination of ≤5%) and 40.2% as medium-quality (completeness of ≥60%, contamination of ≤10% (**Table S9**). These MAGs, spanning 17 phyla, primarily Bacteroidota and Pseudomonadota (**Fig. 2A**), included 84.2% of MAGs without species-level representatives in existing genomic databases, and59 MAGs unclassified beyond the family level (**Table S9**). The recovered MAGs effectively represented the *in vitro* communities, with 86.2% of co-culture reads mapping back to MAGs compared to only 9.3% of source sample reads mapping back to MAGs on average **(Fig. 2A, Table S2, Supplementary Results**).

Specific taxonomic groups become enriched in stable *in vitro* communities

We directly compared diversity between environmental source samples and co-cultures. This analysis was conducted both by comparing MAG diversity (**Fig. S5**) and performing an assembly-independent comparison of ribosomal protein L6 marker OTU diversity (L6OTU; **Fig. 2B and S4**), to overcome low MAG recovery from source samples. Both analyses revealed significantly lower alpha-diversity for *in vitro* consortia relative to source samples (**Fig. S4 and S5**). Using L6OTU analysis we found that 22 orders were overrepresented, whereas 59 were underrepresented within *in vitro* consortia relative to source environments. Bacterial orders from the Pseudomonadota and Bacteroidota were the most frequently enriched in co-cultures, and bacterial orders within the Candidate Phyla Radiation (CPR) and Omnitrophota were underrepresented (**Fig. 2B, Table S8**). Despite differences in inoculum sources, the *in vitro* environment consistently favored the enrichment of specific taxa (**Supplementary Results**).

Inoculum source and cyanobacterial host strain drive inter-community variation

Stable co-cultures were predominantly composed of their original cyanobacterial host strain, accompanied by prevalent Pseudomonadota and Bacteroidota species, along with less common phyla such as Spirochaetota, Actinomycetota, and Armatimonadota (**Fig. 3A-B and Table S10**). Between co-culture samples, we found that the source of inoculum (R^2 = 0.23, P < 0.0001; PERMANOVA), the cyanobacterial host strain (R^2 = 0.11, P < 0.0001; PERMANOVA), and the interaction between the inoculum source and host strain (R^2 = 0.08, P < 0.0001; PERMANOVA), all contributed to statistically significant variations in community composition, with inoculum source having the strongest influence (**Fig. S5G-I**). Despite these variations, communities converged toward a broadly similar taxonomic structure with enrichment of major taxonomic groups relative to source samples (**Fig. 2B and 3A-B**). Comparison of source samples and co-cultures revealed distinct clustering, with *in vitro* community samples tightly clustered relative to samples of source environments (**Fig. S4G, S5G-I, and Table S11**). These observations suggest that even though differences between *in vitro* communities exist, they are substantially smaller than the variations introduced during community formation from source inocula. Thus,

351 even though inoculum source and host strain shape community diversity, the resulting consortia
352 maintain high-level taxonomic coherence across different host strains and inocula.
353
354 **Core species identified across model cyanobacterial communities**
355 Although there is no universally accepted definition of a core microbiome, the consistent
356 taxonomic composition and developmental patterns in our *in vitro* cyanobacterial communities
357 (**Fig. 1B, 3A-B, and Fig. S3**) allowed us to define a 25 species core microbiome using genome-
358 resolved data (**See Methods**). These species were reproducibly detected across multiple
359 communities with different cyanobacterial hosts (>= 4 host species) and environmental inocula
360 (>= 2 source environments; **Fig. S6**). Hereinafter we refer to non-core species detected in
361 cultures as auxiliary species and species only detected in environmental samples as source
362 species (**Fig. 2A and Table S9**). Core microbiome species averaged 1.0 ± 1.3% (Range: 0 -
363 7.3%) of the total community abundance, increasing to 14.7 ± 19.4% (Range: 0 - 82.4%) when
364 considering only non-cyanobacterial taxa **(Fig. 3C**). Individual core species were present in
365 18.5% - 68.5% of all samples (**Table S9 and Fig. S6**). The majority of core species belonged to
366 taxonomic orders significantly enriched in the L6OTU analysis (**Fig. 2B**) with the most numerous
367 being Rhizobiales (n = 7 species), Rhodobacterales (n = 5 species) and Sphingomonadales (n
368 = 5 species) (**Fig. 3D**). Phylogenetically, core species were closely clustered, with significantly
369 lower phylogenetic distances between their MAGs on average relative to auxiliary or source
370 inoculum species (**Fig. 2A and S7**). Single nucleotide variant (SNV) analysis further confirmed
371 that identical species were independently selected into *in vitro* communities from distinct
372 geographic water sources, with no evidence of strain-sharing between environments (**Fig. S8,**
373 **Table S13-S14, Supplementary Methods and Results**). Eighteen core species lacked a
374 representative in public databases (**Table S9**), marking an expansion of genomic knowledge for
375 organisms that may occupy a generalized niche in cyanobacterial communities.
376
377 **Taxonomic similarity between *vitro* and natural phototrophic consortia**
378 We identified 29 (9.1%) shared species, **(Fig. 3E and S9**) between our *in vitro* communities and
379 MAGs compiled from other studies on cyanobacterial consortia (n = 1,843 MAGs). Among
380 these, three MAGs are core microbiome members identified in our study. In total, 22 (76%)
381 shared species also belonged to taxonomic orders enriched during *in vitro* community
382 establishment in this study (**Fig. 2B, 3E**). On average, this study shared more species with other
383 studies (3.2 ± 6.3 species) than the studies did with each other (0.7 ± 2.9 species). Most
384 pairwise comparisons between studies identified no shared species (42/55 comparisons). The
385 species shared with this study were primarily from freshwater-associated cyanobacterial
386 communities, with no species overlap in saltwater, soda lake, or plant rhizosphere communities.
387
388 At higher taxonomic levels we observed broad overlap between studies. Out of the 48 observed
389 orders in this study, 37 (77%) were shared with other studies (**Fig. S9B**), with the orders
390 enriched during *in vitro* community establishment in this study being the most commonly shared
391 (P = 0.018; Wilcoxon test, **Fig. S9C**). These findings support that the *in vitro* communities
392 developed here reasonably represent natural cyanobacterial consortia and share considerable
393 taxonomic overlap with other phototroph-heterotroph systems.
394
395 **Enriched metabolic functions in the core cyanobacterial microbiome**
396 We identified metabolic functions enriched in core microbiome MAGs (n = 25 MAGs) relative to
397 all non-core MAGs (n = 512 MAGs), and a constrained set of non-core MAGs from three phyla
398 in the core microbiome Bacteroidota, Pseudomonadota, and Spirocheaotota (n = 366 MAGs). In
399 both analyses, we found significant enrichment of 1073 and 1065 KEGG orthology groups
400 (KOs), respectively (FDR ≤ 0.05; Fisher test; **Tables S17-S19**). Enriched functional categories
401 included cofactor metabolism (Fisher test; OR_{All} = 1.6, FDR_{All} = 0.007; $OR_{Constrained}$ = 1.9,

402 $FDR_{Constrained}$ = 0.0002) and membrane transport (Fisher test; OR_{All} = 1.7, FDR_{All} = 1.9 e-6;
403 $OR_{Constrained}$ = 1.7, $FDR_{Constrained}$ = $1.1e^{-5}$). (**Fig. 4A, Table S19**). The constrained analysis
404 mirrored the full analysis and additionally identified signal transduction as (Fisher test;
405 $OR_{Constrained}$ = 1.7, $FDR_{Constrained}$ = 0.006) a significantly overrepresented functional category in
406 the core microbiome (**Fig. 4A**).
407
408 Within cofactor metabolism, 39 KOs in the porphyrin metabolic pathway were significantly
409 enriched in core microbiome MAGs including those involved in the production of vitamin B12,
410 chlorophylls, and hemes. The gene encoding 5-aminolevulinate synthase (K00643) which
411 facilitates one-step synthesis of 5-Aminolevulinate (ALA) from glycine was highly enriched in
412 core MAGs (OR_{All} = 9.9, FDR_{All} = 6.5 e-6; Fisher test), whereas non-core MAGs favored the
413 more complex three-step pathway from glutamate (K02492; Fisher test; OR_{All} = 24.6, FDR_{All} =
414 3.3 e-8). The gene *cobC* (K02225), involved in vitamin B12 synthesis, showed the strongest
415 core microbiome association (Fisher test; OR_{All} = 24.6, FDR_{All} = 3.3 e-8), alongside other
416 enzymes in the vitamin B12 and bacteriochlorophyll synthesis pathways (**Fig. 4B and Table**
417 **S18**).
418
419 We also found enrichment of KOs related to anoxygenic photosystems, with 80% (20/25) of
420 core microbiome members encoding these components (K08929/K08928/K13991; **Fig. 5, Table**
421 **S18 and S20**). This may indicate that light driven energy production is an important metabolic
422 strategy of cyanobacteria-associated bacteria. Electron donors for these anoxygenic
423 phototrophs are likely organic substrates [91], which agrees with functional predictions. Only
424 one member of the core microbiome, a *Georhizobium* species, encoded RuBisCO for carbon
425 fixation.
426
427 Genes for phosphonate degradation (K06162-6) and transport (K02041-4) were significantly
428 enriched in core MAGs, offering an alternate source of phosphorus to the community. KOs
429 involved in methylamine (K15228-9), nitrile (K01721/K20807), and urea (K01428-30) utilization
430 were also enriched in core MAGs, putatively expanding available carbon and nitrogen sources
431 (**Fig. 5 and Table S18 and S20**). Additionally, many core microbiome MAGs that encoded
432 methylamine to formaldehyde degradation also had genes to oxidize formaldehyde to formate
433 and CO_2. Finally, core microbiome members were enriched in several genes related to
434 degradation of cyanobacteria-derived long-chain alkanes. All cyanobacterial strains in our study
435 encoded KOs (K14330 and K14331) responsible for long-chain alkane biosynthesis [92, 93],
436 wheres core microbiome members were enriched in genes (*LadA*, *AhyA*, *AlkB*, and *Cyp153*) for
437 the degradation of these compounds [85] (**Fig. 5 and Table S18 and S20**). This suggests that
438 core microbiome species may be utilizing these cyanobacteria-produced alkanes as an energy
439 source.
440
441 **Core microbiome MAGs are enriched in putative plasmid elements**
442 Extrachromosomal replicons and plasmids have been shown to encode functions that mediate
443 phototroph-heterotroph symbioses [48]. Here we assessed if there were differences in the
444 frequency and distribution of putative plasmid elements associated with the MAGs resolved in
445 our study. We identified 850 putative plasmid contigs across 537 species-representative MAGs
446 (**Table S21**). Even though 98% of putative plasmids were not circularized, we found 15%
447 encoded plasmid-specific replication domains and 26% encoded proteins for plasmid
448 mobilization (**Fig. S10A**). In our comparison with the comprehensive IMG/PR database, 81
449 plasmids from our dataset matched entries in the database. Of these 81 matching plasmids,
450 80% were recovered from metagenomic datasets, primarily from aquatic environments, with
451 only 7 being assessed as putatively complete (**Table S22**).

We found that core microbiome MAGs had a significantly higher proportion of putative plasmid contigs than auxiliary (Dunn Test = 3.7, P = 0.0003) or source MAGs (Dunn Test = 6.5, P = 0) (**Fig 6A**). As bacteria from the order Rhizobiales are known to encode large numbers of extrachromosomal plasmids [94, 95] we also performed the comparison only between MAGs of this order. We found that core Rhizobiales MAGs contained a significantly larger proportion of putative plasmids than auxiliary (Dunn Test = 3.0, P = 0.0036) or source (Dunn Test = 2.4, P = 0.0253) MAGs **(Fig. S10D)**.

Putative plasmids of the core microbiome encode functions facilitating symbioses

To assess functional contribution of putative plasmids in core microbiome MAGs, we repeated our functional enrichment analysis, excluding plasmid-predicted contigs from all MAGs. We identified 197 KEGG Orthology (KO) groups that were no longer significantly enriched in the core microbiome (**Tables S18 and S19**), indicating that these KOs were enriched on putative plasmids. Functional category analysis of this KO set revealed significant enrichment of signal transduction (Fisher test; OR = 0.7, FDR = 0.038), membrane transport (Fisher test; OR = 0.6, FDR = 0.038), and carbohydrate metabolism (Fisher test; OR = 0.5, FDR = 0.049) on putative plasmids of the core microbiome (**Fig. 6B**).

Enriched plasmid-specific functions on core microbiome plasmids (**Fig. 6C and Table S18**) included those for plasmid replication (*repC*) and conjugal transfer (*TrbBCDFGI*). Carbohydrate metabolic functions included those mediating the interconversion and metabolism of sugars commonly found in cyanobacterial exopolysaccharides [96, 97], such as xylose (*XylABD*), rhamnose (*RhaAM*), and fucose (K18335, K07046, and *FucA*). Transporters for importing xylose (*XylG*) and simple sugars (*AbcSS* and *AbcMS*) were also enriched. Additionally, putative plasmids in the core microbiome may enable the utilization of peptides as alternative carbon and nitrogen sources as evidenced by the enrichment of dipeptide (*DppBCF*) and oligopeptide (*OppBC*) transport systems, and genes (*Mut*, *PccA*, and *BkdAB*) involved in converting 2-oxobutanoate to succinyl-CoA, linking amino acid degradation to the TCA cycle.

We identified plasmid-enriched genes that may mediate symbiotic interactions and biofilm formation. These genes include multiple flagellar structural proteins (FlgBDK and FlhAR), a flagellar regulatory protein (RpoD), and components of an Imp-like type VI secretion system (ImpBCFM) (**Fig. 6C and Table S18**). Furthermore, we identified a set of plasmid-enriched KOs involved in exopolysaccharide biosynthesis (*exoM* and *GumL*). We discovered comprehensive gene clusters for the biosynthesis of succinoglycan and alginate on a large (407 kb) putative plasmid (**Fig. 6D and Table S23**) associated with a Rhizobiales MAG from the core microbiome (StrawCreek_S_L0902_W_A_idba_concoct_4). Although this element was not fully circularized, it encoded plasmid-specific replication factors (*repABC*) and a complete conjugal transfer system, including a relaxase protein (TraI) and type IV secretion apparatus (*trb*).

Discussion

In this work, we built on existing approaches [21–23, 25, 26, 36–38] to generate and analyze 108 *in vitro* cyanobacterial consortia using well-characterized, model cyanobacterial hosts. This approach provides three core advances for the *in vitro* study of cyanobacterial communities. First, leveraging cyanobacterial host species that serve as key model systems in cyanobacterial research [43] significantly enhances our previously poor understanding of the microbiomes they recruit. Second, extensive replication enabled the identification of a core microbiome across multiple taxonomically diverse cyanobacterial hosts. Third, leveraging this core microbiome, we

confirmed, statistically rather than solely by observation, enrichment of functional traits associated with core species, and, furthermore, uncovered metabolic functions and the potential for plasmids to mediate phototroph-heterotroph symbioses in these systems.

The cyanobacterial hosts used in this study were selected for their diversity and importance as model organisms for advancing our understanding of cyanobacterial biology [41, 42, 98–103]. Despite their value as models, little is known about their behavior in ecological contexts. Whereas previous studies have utilized some of these species in two-member co-culture systems [8, 32, 33, 35], this study provides the insights into more complex microbial communities potentially recruited by these species from natural environments. We provide strong evidence that not only can *in vitro* communities rapidly stabilize around these model hosts (**Fig. 1B-D and S3**), but that these communities exhibit strong taxonomic overlaps with freshwater cyanobacteria in natural habitats [28, 30], cultivated in lower-throughput laboratory systems [23, 78], and share higher level taxonomic overlap with plant rhizosphere associated communities (**Fig. 3E, S9**). Furthermore, evidence supporting host-specific selective pressures was observed, as species co-occurring with the same host exhibited significantly higher strain-level ANI across co-cultures compared to the same species associated with different hosts **(Fig. S9)**.

By inducing co-culture diversity through variation of both cyanobacterial host species and environmental inoculum sources, consistently shared features could be robustly identified. This diversity was leveraged across consortia to define a 25-species core microbiome (**Fig. 3D**) identified across five model cyanobacterial host strains. Despite the limitation of using freshwater cyanobacterial species exclusively, the core microbiome identified here provides a foundation for understanding organisms that commonly co-occur across an array of taxonomically and functionally diverse cyanobacteria. Although some core species identified here were also present in cyanobacterial consortia from other studies **(Fig. 3E)**, our hypothesis is that the specific identities of these core microbiome members are less critical than the functional traits and “microbial archetypes” they represent. Therefore, comparing the genomic content of core and non-core species provides a framework for statistically identifying functional traits that play a broader role in supporting cyanobacterial ecosystems.

Functional genomic analysis of the core microbiome revealed a significant enrichment of metabolic pathways associated with the biosynthesis and potential cross-feeding of resource-intensive compounds (e.g., vitamin B12), the enhancement of nitrogen and phosphorus availability, and anoxygenic phototrophy (**Figs. 4 and 5**). Even though some of these pathways, such as anoxygenic photosynthesis [28, 30], have been previously reported [12, 16] this study provides robust statistical evidence for their specific enrichment within a core microbiome. Collectively, these findings support the hypothesis that core microbiome functions facilitate metabolic cooperation centered around phototrophy. This is further reinforced by the predominance of porphyrin cofactor metabolism among enriched functions, including numerous enzymes involved in vitamin B12 and bacteriochlorophyll biosynthesis pathways **(Fig. 4)**. The core microbiome exhibits enrichment in ALA synthase, which enables a more energetically efficient route for ALA production from glycine, potentially advantageous in communities where porphyrin compounds are critical resources. The importance of such exchanges is underscored by studies demonstrating enhanced cyanobacterial growth when supplemented with the porphyrin precursor ALA [104]. Although vitamin B12 cross-feeding in microbial communities has been studied intensively [105–107], the shared production and exchange of other energetically demanding and metabolically linked molecules, such as chlorophyll precursors, remain poorly characterized. The enrichment of anoxygenic phototrophy within the core microbiome, consistent with observations from other studies [91], may suggest that non-

553 competitive photosynthetic [108] partnerships are important in cyanobacterial microbiomes for
554 exchanging essential and costly metabolic precursors common to all organisms. Finally,
555 although multiple studies have posited the function of cyanobacterial long-chain alkanes [109],
556 this study shows the enrichment of long-chain alkane degradation pathways in core microbiome
557 organisms (**Fig. 5**). This capability was specifically prevalent in those core microbiome
558 organisms that also encoded phototropic capabilities. This observation would support that
559 cyanobacteria produce these compounds to specifically select and modulate their surrounding
560 communities by creating a metabolic niche.
561
562 Even though plasmids have been studied in microbial symbioses [110, 111], their roles within
563 cyanobacterial consortia remain largely unexplored. Here, core microbiome genomes exhibited
564 a higher prevalence of plasmids compared to non-core members (**Fig. 6**). Functionally core
565 microbiome plasmids encoded gene clusters linked to the biosynthesis of exopolysaccharides
566 such as succinoglycan, alginate, and xanthan gum-like compounds, polymers known to play
567 essential roles in plant-bacterial interactions [110, 112]. Succinoglycan is critical for nitrogen-
568 fixing Rhizobiales [113, 114], and alginate forms hydrogels that enhance biofilm formation [115–
569 117]. In particular, alginate encapsulation of *Synechococcus elongatus* has been shown to
570 improve both sucrose export and co-culture stability in laboratory settings [32]. Additionally, core
571 microbiome plasmids contained a two-component nitrate regulatory system (*NasST*), gene
572 clusters for mobility machinery, and Type VI secretion systems, features commonly found in
573 plant-associated bacteria [118], suggesting potential roles in colonization, competition, and
574 biofilm formation. Beyond colonization and biofilm forming functionality, plasmids in core
575 microbiome species encoded metabolic pathways for pentose and hexose sugar utilization,
576 including rhamnose and xylose metabolism, key components of cyanobacterial
577 exopolysaccharides [96, 97] (**Fig. 6, S10, Tables S17-18, S22**). Collectively, these findings
578 support the hypothesis that plasmids contribute to core microbiome establishment and
579 maintenance by enhancing biofilm formation and niche adaptation.The data further highlight
580 adaptive advantages plasmids may confer within cyanobacterial consortia, providing new
581 insights into how plasmids may shape these microbial interactions.
582
583 Even though compelling, several limitations of our approach warrant discussion. The absence of
584 cyanobacteria-free controls, limited our ability to disentangle effects solely attributable to media
585 or laboratory conditions. However, we observed substantial taxonomic overlap (**Fig. 3E and S9**)
586 between our co-cultures and other freshwater cyanobacterial and rhizosphere consortia [6, 28,
587 45, 46, 103]. Furthermore, the use of naturally sourced inocula carried the potential for
588 introducing diverse eukaryotic organisms, including flagellates, ciliates, fungi, and phototrophic
589 diatoms and algae. Although we rigorously screened cultures for readily detectable eukaryotes
590 (**See Methods**), we did not assess the diversity or function of eukaryotic species present within
591 final co-cultures, which on average constituted 0.64±1.89 % of assembled data on an
592 assembled contig size basis (**Table S2**). Our objective was to focus on the prokaryotic
593 interactions of model cyanobacteria, given the limited existing data on these hosts' recruited
594 microbiomes. Despite these constraints, our findings and related work [19] underscore the value
595 of *in vitro* cyanobacterial model systems for revealing insights into phototroph–heterotroph
596 interactions. Although methods for cultivating phytoplankton consortia are well-established [19,
597 23, 28, 38, 47], our framework and biobanked stable co-cultures offer a controlled yet
598 ecologically relevant platform to advance understanding of freshwater microbial community
599 dynamics. Future investigations using complementary methods such as long-read sequencing,
600 mesocosm experiments, cyanobacteria-free controls, and inclusion of eukaryotic organisms, will
601 help refine our understanding of aquatic microbiome functionality.
602

Data Availability

Raw sequencing read data generated as part of this project are available under the NCBI BioProject accession number PRJNA1271998. All metagenome bins and putative plasmid sequences recovered in this study have been deposited in FigShare under the following DOIs: 10.6084/m9.figshare.26148823 and 10.6084/m9.figshare.26148811

Code Availability

Code used in the analysis for this paper are available at the following GitHub repository: https://github.com/SDmetagenomics/Kust_Cyano_2023

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Acknowledgements

We would like to thank Prof. Susan Golden, Prof. James Golden, and Dr. Arnaud Taton for providing the cyanobacterial strains used in this study. Eel River sampling occurred at the UC Angelo Coast Range Reserve. Funding for this project was provided in part by the Gordon and Betty Moore Foundation (GBMF9321). Funding for this project was also provided in part by the Shurl and Kay Curci Foundation.

Author Contributions

AK, JFB, and SD conceived overall experimental concepts and design. AK and SD designed the sampling and cultivation strategy. AK, KBG, and SD collected water samples. AK, CCP, WL, and SD performed cultivation and sampling. AK, CCP, NK, and SD performed sequencing and sequence processing. AK and SD performed sequence assembly and genomic binning. AK, JZ, and SD performed annotation and statistical analysis on amplicon and metagenomic datasets. AK, JZ, and SD wrote the manuscript. AK, JZ, JFB, and SD edited the manuscript

Competing Interests

The authors declare no competing interests.

Corresponding Author

All correspondence should be addressed to Spencer Diamond (sdiamond@berkeley.edu)

References

1. Paerl HW, Pinckney JL. A mini-review of microbial consortia: Their roles in aquatic production and biogeochemical cycling. *Microb Ecol* 1996; **31**: 225–247.

2. Conover AE, Morando M, Zhao Y, Semones J, Hutchins DA, Webb EA. *Alphaproteobacteria* facilitate *Trichodesmium* community trimethylamine utilization. *Environ Microbiol* 2021; **23**: 6798–6810.

3. Zhang Z, Fan X, Peijnenburg WJGM, Zhang M, Sun L, Zhai Y, et al. Alteration of dominant cyanobacteria in different bloom periods caused by abiotic factors and species interactions. *J Environ Sci* 2021; **99**: 1–9.

4. Smith DJ, Kharbush JJ, Kersten RD, Dick GJ. Uptake of phytoplankton-derived carbon and cobalamins by novel Acidobacteria genera in *Microcystis* blooms inferred from metagenomic and metatranscriptomic evidence. *Appl Environ Microbiol* 2022; **88**: e01803-21.

5. Cai H, McLimans CJ, Jiang H, Chen F, Krumholz LR, Hambright KD. Aerobic anoxygenic phototrophs play important roles in nutrient cycling within cyanobacterial *Microcystis* bloom microbiomes. *Microbiome* 2024; **12**: 1–19.

6. Cook KV, Li C, Cai H, Krumholz LR, Hambright KD, Paerl HW, et al. The global interactome. *Limnol Oceanogr* 2020; **65**: S194–S207.

7. Li W, Baliu-Rodriguez D, Premathilaka SH, Thenuwara SI, Kimbrel JA, Samo TJ, et al. Microbiome processing of organic nitrogen input supports growth and cyanotoxin production of *Microcystis aeruginosa* cultures. *ISME J* 2024; **18**: wrae082.

8. Hays SG, Yan LLW, Silver PA, Ducat DC. Synthetic photosynthetic consortia define interactions leading to robustness and photoproduction. *J Biol Eng* 2017; **11**: 4.

9. Kim M, Shin B, Lee J, Park HY, Park W. Culture-independent and culture-dependent analyses of the bacterial community in the phycosphere of cyanobloom-forming *Microcystis aeruginosa*. *Sci Rep* 2019; **9**: 20416.

10. Dick GJ, Duhaime MB, Evans JT, Errera RM, Godwin CM, Kharbush JJ, et al. The genetic and ecophysiological diversity of *Microcystis*. *Environ Microbiol* 12 2021; **23**: 7278–7313.

11. Jackrel SL, Schmidt KC, Cardinale BJ, Denef VJ. Microbiomes reduce their host's sensitivity to interspecific interactions. *mBio* 2020; 11.

12. Baker D, Godwin CM, Khanam M, Burtner AM, Dick GJ, Denef VJ. Variation in resource competition traits among strains is affected by their microbiomes. *mLife* 2023; **2**: 401–415.

13. Durán P, Flores-Uribe J, Wippel K, Zhang P, Guan R, Melkonian B, et al. Shared features and reciprocal complementation of the *Chlamydomonas* and *Arabidopsis* microbiota. *Nat Commun* 2022; **13**: 1–14.

14. Ling N, Wang T, Kuzyakov Y. Rhizosphere bacteriome structure and functions. *Nat Commun* 2022; **13**: 1–13.

15. Morella NM, Weng FC-H, Joubert PM, Metcalf CJE, Lindow S, Koskella B. Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection. *Proc Natl Acad Sci USA* 2020; **117**: 1148–1159.

16. Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 2015; **522**: 98–101.

17. Kojadinovic-Sirinelli M, Villain A, Puppo C, Sing SF, Prioretti L, Hubert P, et al. Exploring the microbiome of the 'star' freshwater diatom *Asterionella formosa* in a laboratory context. *Environ Microbiol* 2018; **20**: 3601–3615.

18. Sison-Mangus MP, Jiang S, Tran KN, Kudela RM. Host-specific adaptation governs the interaction of the marine diatom, *Pseudo-nitzschia* and their microbiota. *ISME J* 2014; **8**: 63–76.

19. Jackrel SL, Yang JW, Schmidt KC, Denef VJ. Host specificity of microbiome assembly and its fitness effects in phytoplankton. *ISME J* 2020; **15**: 774–788.

20. Nelson C, Giraldo-Silva A, Garcia-Pichel F. A symbiotic nutrient exchange within the cyanosphere microbiome of the biocrust cyanobacterium, *Microcoleus vaginatus*. *ISME J* 2021; **15**: 282–292.

21. Zhu L, Zancarini A, Louati I, De Cesare S, Duval C, Tambosco K, et al. Bacterial communities associated with four cyanobacterial genera display structural and functional differences: evidence from an experimental approach. *Front Microbiol* 2016; **7**: 1662.

22. Cole JK, Hutchison JR, Renslow RS, Kim Y-M, Chrisler WB, Engelmann HE, et al. Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. *Front Microbiol* 2014; **5**: 109.

23. Zhao L, Lin L-Z, Zeng Y, Teng W-K, Chen M-Y, Brand JJ, et al. The facilitating role of phycospheric heterotrophic bacteria in cyanobacterial phosphonate availability and *Microcystis* bloom maintenance. *Microbiome* 2023; **11**: 142.

24. Halary S, Duperron S, Demay J, Duval C, Hamlaoui S, Piquet B, et al. Metagenome-based exploration of bacterial communities associated with cyanobacteria strains isolated from thermal muds. *Microorganisms* 2022; 10.

25. Duxbury SJN, Raguideau S, Rosko J, Cremin K, Coates M, Quince C, et al. Reproducible spatial structure formation and stable community composition in the cyanosphere predicts metabolic interactions. *bioRxiv* 2022.

26. Haines M, Vadlamani A, Daniel Loty Richardson W, Strous M. Pilot-scale outdoor trial of a cyanobacterial consortium at pH 11 in a photobioreactor at high latitude. *Bioresour Technol* 2022; **354**: 127173.

27. Bouma-Gregson K, Olm MR, Probst AJ, Anantharaman K, Power ME, Banfield JF. Impacts of microbial assemblage and environmental conditions on the distribution of anatoxin-a producing cyanobacteria within a river network. *ISME J* 2019; **13**: 1618–1634.

28. Pérez-Carrascal OM, Tromas N, Terrat Y, Moreno E, Giani A, Marques LCB, et al. Single-colony sequencing reveals microbe-by-microbiome phylosymbiosis between the cyanobacterium *Microcystis* and its associated bacteria. *Microbiome* 2021,**12**: 9.

29. Zhu C-M, Zhang J-Y, Guan R, Hale L, Chen N, Li M, et al. Alternate succession of aggregate-forming cyanobacterial genera correlated with their attached bacteria by co-pathways. *Sci Total Environ* 2019; **688**: 867–879.

30. Li H, Bhattarai B, Barber M, Goel R. Stringent response of cyanobacteria and other bacterioplankton during different stages of a harmful cyanobacterial bloom. *Environ Sci Technol* 2023; **57**: 16016–16032.

31. Smith DJ, Berry MA, Cory RM, Johengen TH, Kling GW, Davis TW, et al. Heterotrophic bacteria dominate catalase expression during *Microcystis* blooms. *Appl Environ Microbiol* 2022; **88**: e0254421.

32. Taylor L. Weiss, Eric J. Young, Daniel C. Ducat. A synthetic, light-driven consortium of cyanobacteria and heterotrophic bacteria enables stable polyhydroxybutyrate production. *Metabolic Engineering* 2017; **44**: 236–245.

33. Zuñiga C, Li T, Guarnieri MT, Jenkins JP, Li C-T, Bingol K, et al. Synthetic microbial communities of heterotrophs and phototrophs facilitate sustainable growth. *Nat Commun* 2020; **11**: 1–13.

34. Weissberg O, Aharonovich D, Sher D. Phototroph-heterotroph interactions during growth and long-term starvation across *Prochlorococcus* and *Alteromonas* diversity. *ISME J* 2022; 17: **227**–237.

35. Christie-Oleza JA, Sousoni D, Lloyd M, Armengaud J, Scanlan DJ. Nutrient recycling facilitates long-term stability of marine microbial phototroph–heterotroph interactions. *Nat Microbiol* 2017; **2**: 1–10.

36. Wu Q, Zhang Y, Li Y, Li J, Zhang X, Li P. Comparison of community composition between *Microcystis* colony-attached and free-living bacteria, and among bacteria attached with *Microcystis* colonies of various sizes in culture. *Aquat Ecol* 2019; **53**: 465–481.

37. Vardaka E, Kormas KA, Katsiapi M, Genitsaris S, Moustaka-Gouni M. Molecular diversity of bacteria in commercially available '*Spirulina'* food supplements. *PeerJ* 2016; 4.

38. Yancey CE, Kiledal EA, Chaganti SR, Denef VJ, Errera RM, Evans JT, et al. The Western Lake Erie culture collection: A promising resource for evaluating the physiological and genetic diversity of *Microcystis* and its associated microbiome. *Harmful Algae* 2023; **126**: 102440.

39. Ikeuchi M, Tabata S. *Synechocystis* sp. PCC 6803 - a useful tool in the study of the genetics of cyanobacteria. *Photosynth Res* 2001; **70**: 73–83.

40. Mehdizadeh Allaf M, Peerhossaini H. Cyanobacteria: model microorganisms and beyond. *Microorganisms* 2022; 10.

41. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, et al. The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci USA* 2015; **112**: E6634–43.

42. Diamond S, Jun D, Rubin BE, Golden SS. The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proc Natl Acad Sci USA* 2015; **112**: E1916–25.

43. Zeng X, Zhang C-C. The making of a heterocyst in cyanobacteria. *Annu Rev Microbiol* 2022; **76**: 597–618.

44. Cahill B, Straka L, Maldonado Ortiz J, Krajmalnik-Brown R, Rittmann BE. Effects of light intensity on soluble microbial products produced by *Synechocystis* sp. PCC 6803 and associated heterotrophic communities. *Algal Res* 2019; **38**: 101409.

45. Seymour JR, Amin SA, Raina J-B, Stocker R. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat Microbiol* 2017; **2**: 1–12.

46. Steinrücken P, Jackson S, Müller O, Puntervoll P, Kleinegris DMM. A closer look into the microbiome of microalgal cultures. *Front Microbiol* 2023; **14**: 1108018.

47. Jackrel SL, White JD, Evans JT, Buffin K, Hayden K, Sarnelle O, et al. Genome evolution and host-microbiome shifts correspond with intraspecific niche divergence within harmful algal bloom-forming *Microcystis aeruginosa*. *Mol Ecol* 2019; **28**: 3994–4011.

48. Wardell GE, Hynes MF, Young PJ, Harrison E. Why are rhizobial symbiosis genes mobile? *Philos Trans R Soc Lond B Biol Sci* 2022; **377**: 20200471.

49. Guillard RRL, Lorenzen CJ. Yellow-green algae with chlorophyllide c. *J Phycol* 1972; **8**: 10–14.

50. Diamond S, Andeer PF, Li Z, Crits-Christoph A, Burstein D, Anantharaman K, et al. Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat Microbiol* 2019; **4**: 1356–1367.

51. Clerico EM, Ditty JL, Golden SS. Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods Mol Biol* 2007; **362**: 155–171.

52. Edgar RC. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* 2016; 081257.

53. 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**: D590–6.

54. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. *mSystems* 2020; 5.

55. Woodcroft BJ, Aroney STN, Zhao R, Cunningham M, Mitchell JAM, Blackall L, et al. SingleM and Sandpiper: Robust microbial taxonomic profiles from metagenomic data. *bioRxiv* 2024; 2024.01.30.578060

56. Chen L, Banfield JF. COBRA improves the completeness and contiguity of viral genomes assembled from metagenomes. *Nat Microbiol* 2024; **9**: 737–750.

57. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016; **32**: 605–607.

58. Alneberg J, Bjarnason BS, Bruijn ID, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014; **11**: 1144–1146.

59. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015; **3**: e1165.

60. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, et al. Improved metagenome binning and assembly using deep variational autoencoders. *Nat Biotechnol* 2021; **39**: 555–560.

61. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 2018; 3: 836–843.

62. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**: 1043–1055.

63. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J 7* 2017; **11**: 2864–2868.

64. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2020; **36**: 1925–1927.

65. Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 2019; **35**: 4162–4164.

66. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2021; **50**: D785–D794.

67. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; **32**: 268–274.

68. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021; **49**: W293–W296.

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

70. Aroney STN, Newell RJP, Nissen JN, Camargo AP, Tyson GW, Woodcroft BJ. CoverM: read alignment statistics for metagenomics. *Bioinformatics* 2025; **41**:4.

71. Neu AT, Allen EE, Roy K. Defining and quantifying the core microbiome: Challenges and prospects. *Proc Natl Acad Sci USA* 2021; 118.

72. Dong C, Shao Q, Zhang Q, Yao T, Huang J, Liang Z, et al. Preferences for core microbiome composition and function by different definition methods: Evidence for the core microbiome of *Eucommia ulmoides* bark. *Sci Total Environ* 2021; **790**: 148091.

73. Auch AF, von Jan M, Klenk H-P, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2010; **2**: 117–134.

74. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013; **14**: 60.

75. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat Biotechnol* 2021; 39: 727–736.

76. Berthold M, Albrecht M, Campbell DA, Omar NM. Draft genomes of 3 cyanobacteria strains and 17 co-habiting proteobacteria assembled from metagenomes. *Microbiol Resour Announc* 2023; **12**: e0046023.

77. Cornet L, Bertrand AR, Hanikenne M, Javaux EJ, Wilmotte A, Baurain D. Metagenomic assembly of new (sub)polar Cyanobacteria and their associated microbiome from non-axenic cultures. *Microb Genom* 2018; 4.

78. Li Q, Lin F, Yang C, Wang J, Lin Y, Shen M, et al. A large-scale comparative metagenomic study reveals the functional interactions in six bloom-forming -epibiont communities. *Front Microbiol* 2018; **9**: 746.

79. Zorz JK, Sharp C, Kleiner M, Gordon PMK, Pon RT, Dong X, et al. A shared core microbiome in soda lakes separated by large distances. *Nat Commun* 2019; **10**: 4230.

80. Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, et al. Genomic features of bacterial adaptation to plants. *Nat Genet* 2017; **50**: 138–150.

81. Bai L, Cui J, Jie W, Cai B. Analysis of the community compositions of rhizosphere fungi in soybeans continuous cropping fields. *Microbiol Res* 2015; **180**: 49–56.

82. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 4 2020; **36**: 2251–2252.

83. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* 2022; **10**: 33.

84. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011; **7**: e1002195.

85. Khot V, Zorz J, Gittins DA, Chakraborty A, Bell E, Bautista MA, et al. CANT-HYD: A curated database of phylogeny-derived Hidden Markov Models for annotation of marker genes involved in hydrocarbon degradation. *Front Microbiol* 2021; **12**: 764058.

86. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. *Nat Biotechnol* 2024; **42**: 1303–1312.

87. Camargo AP, Call L, Roux S, Nayfach S, Huntemann M, Palaniappan K, et al. IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. *Nucleic Acids Res* 2024; **52**: D164–D173.

88. Cury J, Abby SS, Doppelt-Azeroual O, Néron B, Rocha EPC. Identifying conjugative plasmids and integrative conjugative elements with CONJscan. *Methods Mol Biol* 2020; **2075**: 265–283.

89. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res* 2021; **49**: D412–D419.

90. Smith DJ, Tan JY, Powers MA, Lin XN, Davis TW, Dick GJ. Individual *Microcystis* colonies harbour distinct bacterial communities that differ by *Microcystis* oligotype and with time. *Environ Microbiol* 2021; **23**: 3020–3036.

91. Yurkov VV, Beatty JT. Aerobic anoxygenic phototrophic bacteria. *Microbiol Mol Biol Rev* 1998; **62**: 695–724.

92. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB. Microbial biosynthesis of alkanes. *Science* 2010; **329**: 559–562.

93. Lea-Smith DJ, Biller SJ, Davey MP, Cotton CAR, Perez Sepulveda BM, Turchyn AV, et al. Contribution of cyanobacterial alkane production to the ocean hydrocarbon cycle. *Proc Natl Acad Sci USA* 2015; **112**: 13591–13596.

94. Leinberger J, Holste J, Bunk B, Freese HM, Spröer C, Dlugosch L, et al. High potential for secondary metabolite production of CP157, isolated from the crustacean. *Front Microbiol* 2021; **12**: 688754.

95. Czarnecki J, Chapkauskaitse E, Bos J, Sentkowska D, Wawrzyniak P, Wyszyńska A, et al. Differential localization and functional specialization of centromere-like sites in replicons of Alphaproteobacteria. *Appl Environ Microbiol* 2022; **88:** e0020722.

96. De Philippis R, Vincenzini M. Exocellular polysaccharides from cyanobacteria and their possible applications. *FEMS Microbiol Rev* 1998; **22**: 151–175.

97. Mota R, Flores C, Tamagnini P. Cyanobacterial Extracellular Polymeric Substances (EPS). In: Oliveira JM, Radhouani H, Reis RL (eds). *Polysaccharides of Microbial Origin*. 2021. Springer, Cham, pp 1–28.

98. Yang Y, Lam V, Adomako M, Simkovsky R, Jakob A, Rockwell NC, et al. Phototaxis in a wild isolate of the cyanobacterium. *Proc Natl Acad Sci USA* 2018; **115**: E12378–E12387.

99. Schatz D, Nagar E, Sendersky E, Parnasa R, Zilberman S, Carmeli S, et al. Self-suppression of biofilm formation in the cyanobacterium *Synechococcus elongatus*. *Environ Microbiol* 2013; **15**: 1786–1794.

100. Cohen SE, Golden SS. Circadian rhythms in cyanobacteria. *Microbiol Mol Biol Rev* 2015; **79**: 373–385.

101. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement*). DNA Res* 1996; **3**: 185–209.

102. Lázaro S, Fernández-Piñas F, Fernández-Valiente E, Blanco-Rivero A, Leganés F. pbpB, a gene coding for a putative penicillin-binding protein, is required for aerobic nitrogen fixation in the cyanobacterium *Anabaena* sp. strain PCC7120. *J Bacteriol* 2001; **183**: 628–636.

103. Taton A, Lis E, Adin DM, Dong G, Cookson S, Kay SA, et al. Gene transfer in *Leptolyngbya* sp. strain BL0902, a cyanobacterium suitable for production of biomass and bioproducts*. PLoS One* 2012; **7**: e30901.

104. Promotive effect of 5-aminolevulinic acid on the growth and photosynthesis of *Spirulina platensis*. Journal of Fermentation and Bioengineering 1995; **79**: 453–457.

105. Grant MAA, Kazamia E, Cicuta P, Smith AG. Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures. *ISME J* 2014; **8**: 1418–1427.

106. Sultana S, Bruns S, Wilkes H, Simon M, Wienhausen G. Vitamin B is not shared by all marine prototrophic bacteria with their environment. *ISME J* 2023; **17**: 836–845.

107. Hessler T, Huddy RJ, Sachdeva R, Lei S, Harrison STL, Diamond S, et al. Vitamin interdependencies predicted by metagenomics-informed network analyses and validated in microbial community microcosms. *Nat Commun* 2023; **14**: 4768.

108. Koblížek M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol Rev* 2015; **39**: 854–870.

109. Parveen H, Yazdani SS. Insights into cyanobacterial alkane biosynthesis. *J Ind Microbiol Biotechnol* 2022; 49.

110. Finks SS, Martiny JBH. Plasmid-encoded traits vary across environments*. mBio* 2023; **14**: e0319122.

111. Finks SS, Moudgalya P, Weihe C, Martiny JBH. The contribution of plasmids to trait diversity in a soil bacterium. *ISME Commun* 2024; **4**: ycae025.

112. Acosta-Jurado S, Fuentes-Romero F, Ruiz-Sainz J-E, Janczarek M, Vinardell J-M. Rhizobial exopolysaccharides: genetic regulation of their synthesis and relevance in symbiosis with

legumes. *Int J Mol Sci* 2021; 22.

113. Skorupska A, Janczarek M, Marczak M, Mazur A, Król J. Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microb Cell Fact* 2006; **5**: 1–19.

114. Reuber TL, Walker GC. Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Cell* 1993; **74**: 269–280.

115. Allison DG, Sutherland IW. The role of exopolysaccharides in adhesion of freshwater bacteria. *Microbiology* 1987; **133**: 1319–1327.

116. Urtuvia V, Maturana N, Acevedo F, Peña C, Díaz-Barrera A. Bacterial alginate production: an overview of its biosynthesis and potential industrial production. *World J Microbiol Biotechnol* 2017; **33**: 198.

117. Boyd A, Chakrabarty AM. *Pseudomonas aeruginosa* biofilms: role of the alginate exopolysaccharide. *J Ind Microbiol* 1995; **15**: 162–168.

118. Bernal P, Llamas MA, Filloux A. Type VI secretion systems in plant-associated bacteria. *Environ Microbiol* 2018; **20**: 1–15.

Figure Legends

Figure 1 | Assessment of *in vitro* community stability over the 119 d passaging experiment. (A) Experimental design for cyanobacterial *in vitro* community cultivation. Water samples were collected from three freshwater locations (two sites per location) resulting in six source inocula. Bacterial fractions were prefiltered, then concentrated using 0.2 μm and 0.1 μm filters, combined, and used to inoculate five cyanobacterial host strains in triplicate. Each host was paired with six source inocula twice for passaging at both 7- and 14-day intervals over 12 weeks, resulting in 180 initial cultures. At each passage, samples were collected for microscopic analysis, OD measurements, and 16S rRNA gene amplicon sequencing. After the 84 d passage, samples were collected for metagenomic sequencing and cryopreservation. The frozen consortia were then regrown, passaged, and harvested as before. **(B)** Order-level community composition for *in vitro* communities at 14 d intervals over 119 d, based on 16S rRNA gene amplicon sequencing averaged across all samples. Orders found in the core microbiome are indicated with an asterisk (*). In all plots the vertical red line separates pre-cryopreservation passaging (left) and post-cryopreservation passaging (right). **(C)** Change in Aitchinson distance between previous passage time points for *in vitro* communities at 14 d intervals over 119 d. Gray dots give change for each community compared to itself in the preceding time point, purple squares indicate estimated marginal means, and error bars represent 95% CI of marginal mean. Statistical significance between comparisons estimated using linear mixed effects models with comparisons not sharing the same letter being significantly different (FDR ≤ 0.05). **(D)** Observed richness of *in vitro* communities at 14 d intervals over 119 d (total ASV counts). Gray dots give values for each community at a given time point, blue squares indicate estimated marginal means, and error bars represent 95% CI of marginal mean. Statistical significance between time points estimated using linear mixed effects models with time points not sharing the same letter being significantly different (FDR ≤ 0.05).

1097
1098 **Figure 2 | Maximum likelihood phylogenetic tree and enrichment of microbial orders. (A)**
1099 A phylogenetic tree of 537 MAGs in our study constructed using a concatenated alignment of
1100 120 bacterial-specific marker genes. An archaeal reference genome (GCF_000213215.1)
1101 served as an outgroup for rooting the tree. Nodes marked with black dots represent a high
1102 bootstrap support (≥ 95%, calculated using ufboot with n = 1000 iterations). The outer ring
1103 highlights core microbiome MAGs (in blue), and the inner ring differentiates MAGs observed in
1104 *in vitro* cultures (orange) and those exclusively observed in source samples (purple). Branch
1105 tips with cyanobacterial host species are denoted by yellow stars and have no outer ring colors.
1106 Scale bar indicates the number of nucleotide substitutions per site. **(B)** Microbial orders showing
1107 significant over or under enrichment when comparing ribosomal protein L6 marker gene OTU
1108 (rpL6 OTU) diversity between true *in vitro* community compositions and randomly sampled *in*
1109 *vitro* communities (n = 10,000 permutations) where all observed rpL6 OTUs were sampled. The
1110 Z-score indicates the level of enrichment of an order in observed communities relative to
1111 communities generated from random sampling. Only orders with statistically significant
1112 enrichment values are shown (FDR ≤ 0.05; permutation test). Orders positioned on the left of
1113 the red line are underrepresented in true consortia, and those on the right of the line are over-
1114 represented in true consortia. Blue squares identify taxa found in the core microbiome. Bars are
1115 colored by phylum as indicated in the key of Fig. 2A with additional colors added and noted in
1116 the key of Fig. 2B for minor phyla (Other) and Omnitrophota. Also see Tables S8 and S9.
1117
1118 **Figure 3 | Microbial community composition, core microbiome, and cross community**
1119 **overlap. (A)** Phylum level relative abundance of *in vitro* communities based on mapping
1120 samples to a database of 537 species-level MAGs. Sample compositions are averaged within
1121 cyanobacterial strain for each environmental inoculum. **(B)** Phylum level taxonomic composition
1122 of *in vitro* communities with cyanobacteria removed. **(C)** Composition of *in vitro* communities
1123 with cyanobacteria removed depicting the relative fraction assigned to core and auxiliary
1124 microbiome MAGs. **(D)** Cyanobacterial host association and order-level taxonomy of the 25
1125 genomes identified as core microbiome members. The numbers above the colored stacked bars
1126 indicate the total number of core species shared. The colored stacked bars represent the order-
1127 level taxonomy of these species. Connected dots below each stacked bar indicate the set of
1128 cyanobacterial hosts that share the species above. The numbers in parentheses on the right of
1129 the plot indicate the total number of species detected in at least three cultures for each host
1130 strain. The legend on the right of the figure provides the color code for each order in the stacked
1131 bar plot. An asterisk (*) indicates orders that were significantly enriched in co-cultured
1132 communities relative to source environment samples (Also See Fig. 2B). **(E)** The number and
1133 taxonomy of species shared between the communities of our study and those of other studies
1134 where genome-resolved analysis of cyanobacterial consortia was conducted. The numbers
1135 above the colored stacked bars indicate the total number species shared. The colored stacked
1136 bars indicate the order-level taxonomy of these species. Blue asterisks above the stacked bar
1137 plots indicate that a core microbiome species identified in our study was shared. Connected
1138 dots below each stacked bar indicate the set of studies that share the species above. Numbers
1139 in parentheses on the right of the plot indicate the total number of species recovered in each
1140 study. The green and blue color shading under the connected dot plots indicates the type and

origin of cyanobacterial community for the compared studies. The coloring of orders in the stacked bar plot of this figure are the same as in Figure 3D.

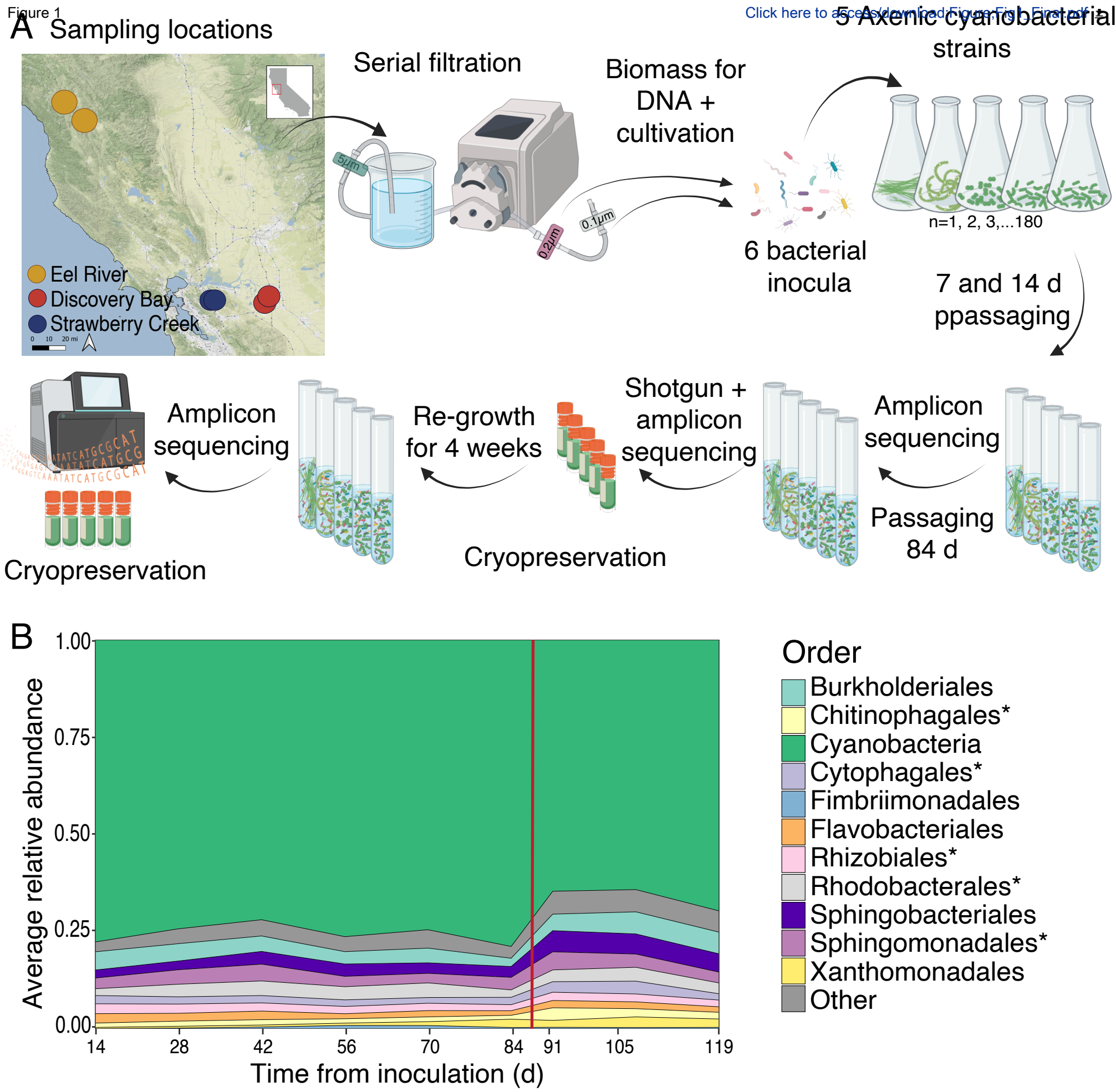
Figure 4 | Metabolic functions enriched in the core microbiome. (A) Functional categories of statistically over or under enriched KEGG functions in the core microbiome. The enrichment odds-ratio for each functional category is shown for the set of KEGG functions with differential enrichment when the core microbiome MAGs were compared to all 512 non-core MAGs (red) and the 366 non-core MAGs in the phyla *Bacteroidota*, *Proteobacteria*, and *Spirocheaotota* (blue). Bars are solid colors for categories with statistically significant over or under representation (FDR ≤ 0.05; Fisher test) and transparent if not significant. **(B)** Metabolic reaction map for porphyrin compound biosynthetic pathways. Numbers indicate the KEGG KO associated with a reaction. This map shows enrichment data from the comparison of core MAGs to all 512 non-core MAGs. The size of colored circles indicates the percentage of MAGs in core (green; n = 25) or non-core (blue; n = 512) sets that encode the respective reaction. Circles are solid colors if a KEGG KO was detected as significantly over or under enriched in the core microbiome (FDR ≤ 0.05; Fisher test), and transparent if not significant. Also see Tables S16-S18.

Figure 5 | Comparison of metabolic functions in cyanobacterial hosts and core microbiome species. Encoded metabolic potential for specific biogeochemical turnover reactions identified in cyanobacterial host species MAGs and the 25 core microbiome MAGs. Specific metabolic functions (left) are separated by higher level metabolic categories (right). Functions marked with and asterisk (*) had at least one associated KO with significant enrichment in the core microbiome relative to all 512 non-core MAGs. Each column represents an individual MAG in the core microbiome, MAGs are grouped at different phylum levels, and colored by order level taxonomy. Identification of the potential for a specific function in each MAG is indicated by a colored box, and boxes share order and color. Criteria for identifying specific functions can be found in Table S19.

Figure 6 | Predicted mobile genetic element distribution and functional potential. (A) Proportion of contigs predicted to be mobile genetic elements in each MAG of the core, auxiliary, and source microbiome. Points are colored by phylum level taxonomy of the MAG. Box plots behind points indicate median, first, and third quartiles of data. Statistical differences between groups are indicated by letters, and groups not sharing the same letter are significantly different ($Statistic_{Core\text{-}v\text{-}Aux}$ = 3.7, FDR < 0.001; $Statistic_{Core\text{-}v\text{-}Source}$ = 6.5, FDR < 0.0001; $Statistic_{Aux\text{-}v\text{-}Source}$ = 6.6, FDR < 0.0001; Dunn Test). **(B)** Functional category analysis for the set of KEGG KOs that becomes non-significant when all MGE contigs are removed from MAGs. Functional categories that are significantly over or underrepresented in this set of KOs are colored accordingly (FDR ≤ 0.05; Fisher test). X-axis indicates the natural logarithm of the Fisher test enrichment score. Also see Tables S15, S16 and S20. **(C)** Individual KEGG KOs from membrane transport, carbohydrate metabolism, and signal transduction categories that become non-significant when the enrichment analysis for core microbiome functions is repeated with predicted MGE contigs removed from all MAGs. The x-axis indicates the -log10(FDR) for each KO in the original enrichment analysis. The y-axis indicates the fold change in the odds-

1185 ratio for a KO between the original analysis and the analysis conducted with all MGE contigs
1186 removed. **(D)** A representative plasmid contig (407,520 bp) co-binned with an *Allorhizobium*
1187 MAG from the core microbiome (StrawCreek_S_L0902_W_A_idba_concoct_4). Genes are
1188 colored by functional category. Dashed lines indicate a break in the contig for visualization
1189 purposes. Also see Table S23.

Click here to access/download;Figure Fig1_Final.pdf



Click here to access/download;Figure;Fig2_Final.pdf

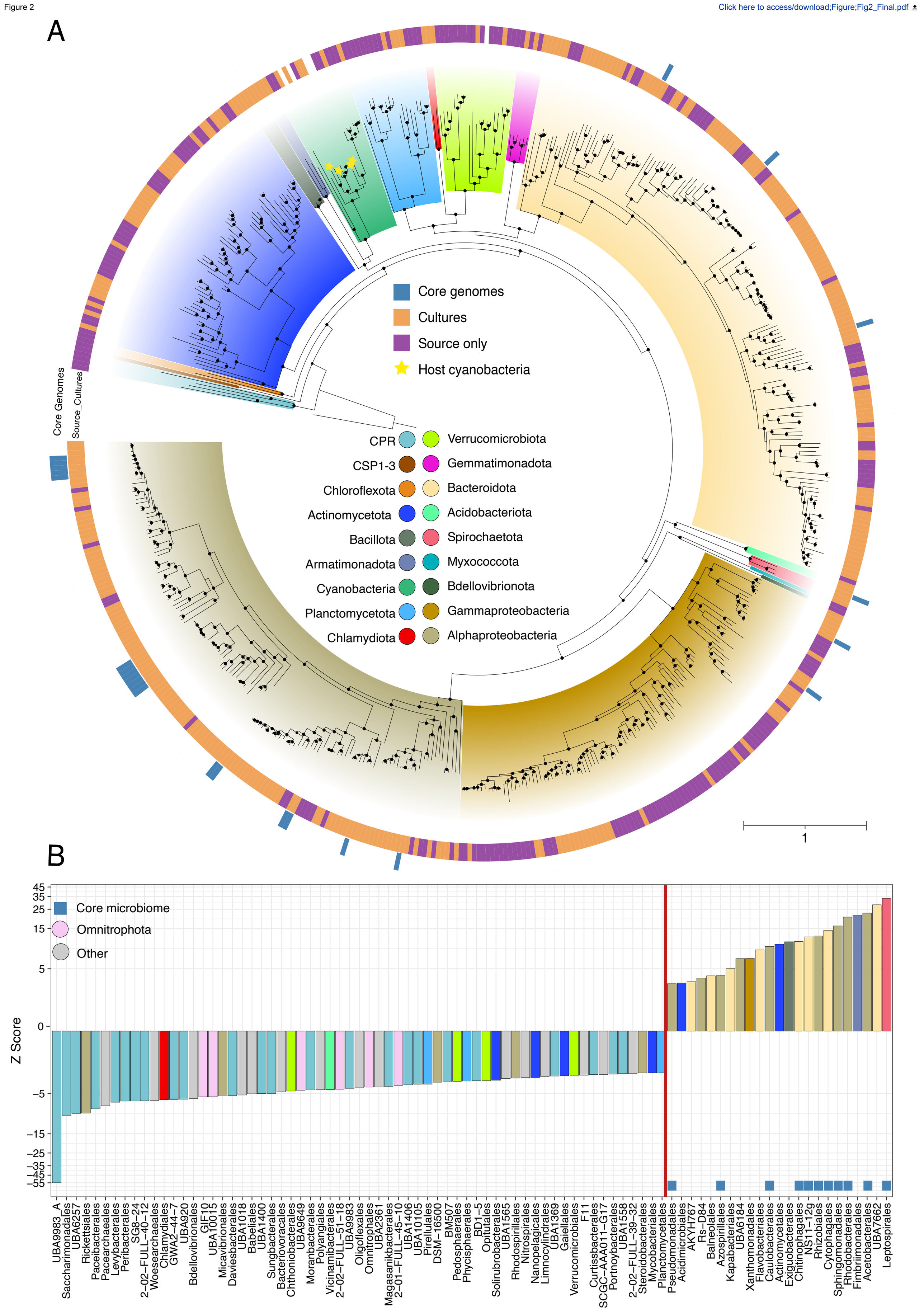
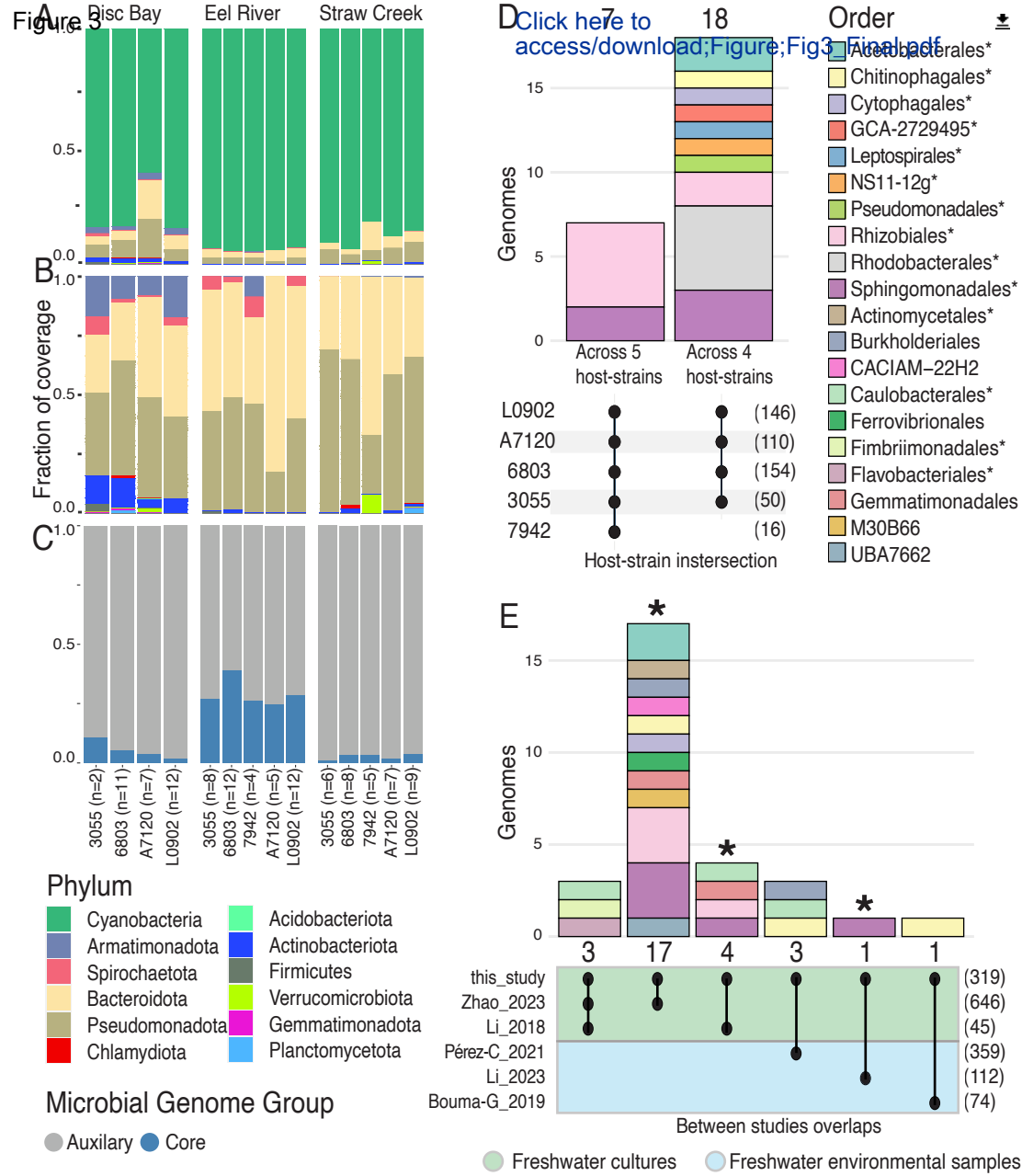


Figure 3

Click here to access/download;Figure;Fig3_Final.pdf



Click here to access/download;Figure;Fig4_Final.pdf

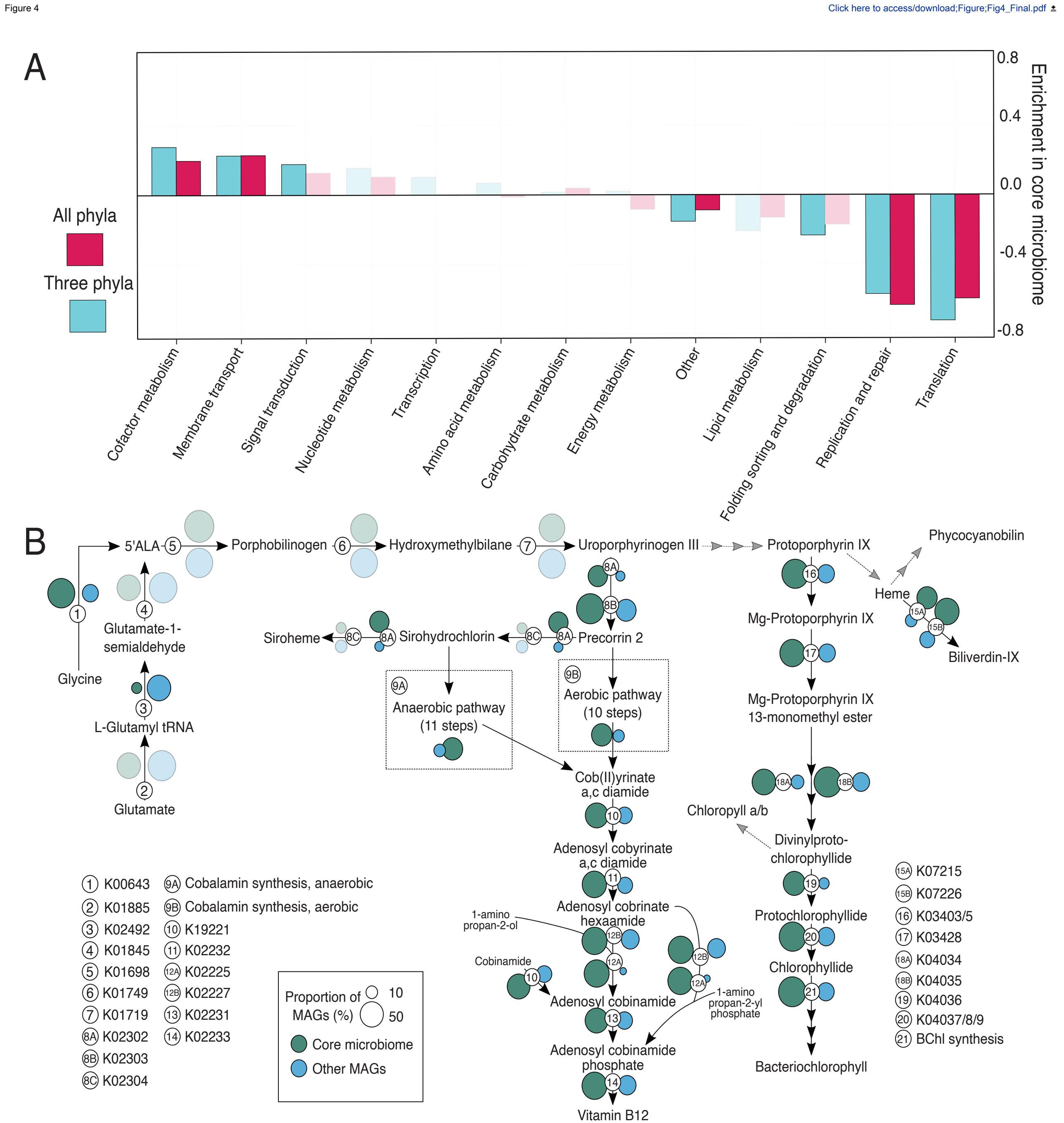
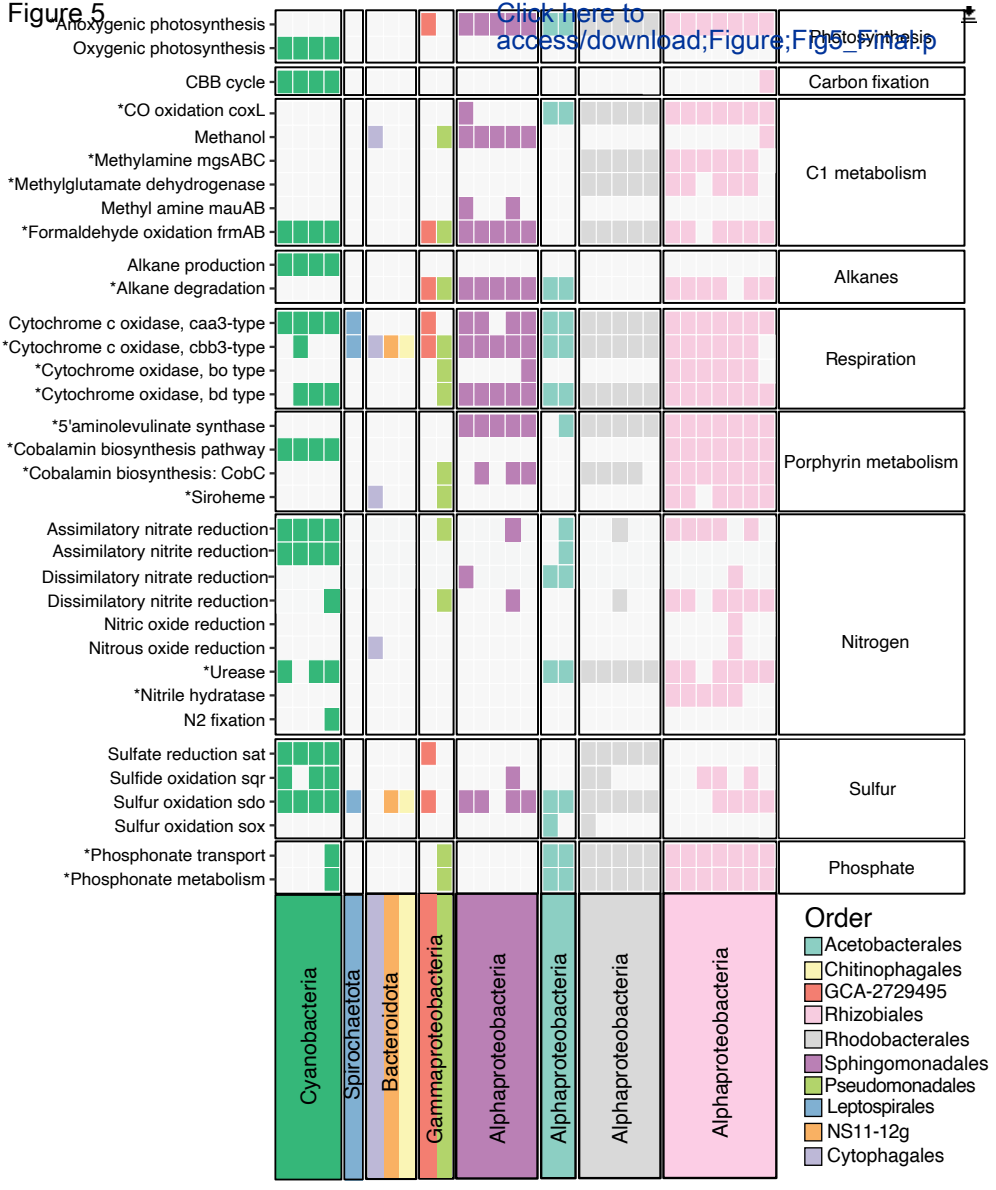
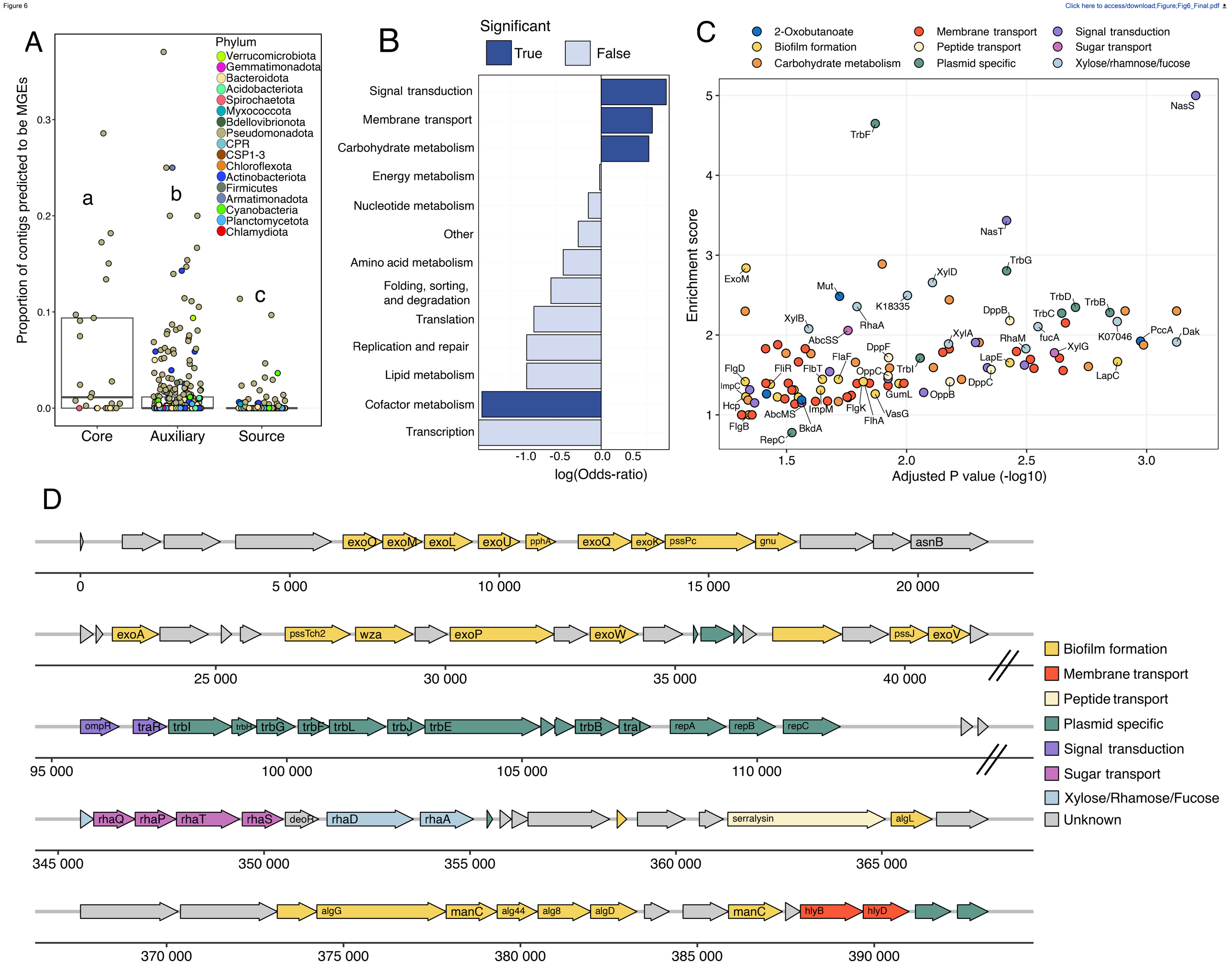


Figure 5

Click here to access/download;Figure;Fig5_Final.p



Click here to access/download;Figure;Fig6_Final.pdf