Microbial model communities reveal widespread auxotrophies in abundant bacteria

Armando Pacheco-Valenciana[1*], Anna Tausch[2], Jennah E. Dharamshi[1], Fabian Bergland[1], Luis Fernando Delgado-Zambrano[3], Alejandro Rodríguez-Gijón[1], Anders F. Andersson[3], Sarahi L. Garcia[1,2,4*].

[1] Department of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden.

[2] Institute for Chemistry and Biology of the Marine Environment (ICBM), School of Mathematics and Science, Carl von Ossietzky Universität Oldenburg, 26129, Oldenburg, Germany

[3] Department of Gene Technology, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden.

[4] Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), Oldenburg, Germany

Correspondence: armando.pacheco@su.se, sarahi.garcia@su.se

KEYWORDS: microbial model communities, metagenomics, auxotrophs, microbial interdependencies.

ABSTRACT

Microorganisms exist in complex communities shaped by intricate interactions. Here, we examined auxotrophies and metabolic interdependencies in microbial model communities established from Baltic Sea pelagic sample dilutions. We recovered 527 metagenome-assembled genomes (MAGs) from 204 microbial model communities, including some of the most abundant taxa such as Pelagibacterales. Combined with 1073 environmental MAGs, we created a genomic catalog of 701 species-clusters (95% average nucleotide identity - ANI). Our results show that more complex communities included microorganisms with smaller genome sizes and communities with more than three species exhibited lower average biosynthetic potential for amino acids and B vitamins. Furthermore, microorganisms with lower biosynthetic potential for those compounds are more abundant in the environment and have smaller estimated genome sizes. Overall, our results highlight the value of cultivating microorganisms in groups of naturally co-occurring microorganisms for unraveling ecological strategies such as auxotrophies and genomic streamlining which are widespread features of successful microorganisms.

INTRODUCTION

Microbial communities in diverse environments operate as complex systems driven by multi-species interactions[1]. Understanding such complex interactions is essential because microorganisms play key roles in the biogeochemical cycles on Earth[2]. To unravel microbial interactions, we need to investigate microbial communities at various levels of biological organization[3,4], ranging from one-to-one species interactions, to model communities[5], to naturally occurring communities.

At the community level, metagenomics has become a powerful tool for uncovering the genetic potential of microbial communities without the need for isolating individual species[6–8]. Analyzing metagenomic data reveals not only the vast diversity of microbial species[9] but also the metabolic potential and co-occurrence networks, which are important for ecosystem functioning[10]. To bridge the gap between broad metagenomic insights and detailed ecological understanding, a few studies have explored genome-specific traits and their ecological implications. Some of these studies have found a link between small genome sizes and a higher number of auxotrophies for both amino acids and vitamins[11–14]. Specifically, auxotrophs are often deficient in biosynthetic pathways required for producing energetically costly but essential compounds such as amino acids[15–17] and certain vitamins (e.g., B_1, B_7, B_{12})[18–20]. Auxotrophic organisms can acquire the essential metabolites from other community members to meet their nutritional requirements[21–23]. Beyond individual interactions, modeling work has shown that microbial communities enriched in auxotrophies can exhibit greater robustness under ecological disturbances, suggesting that these metabolic interdependencies may contribute to overall community stability[24]. While metagenomics offers a broad understanding of microbial communities, interactions cannot easily be derived from co-occurrences within natural complex ecosystems.

Experimental systems are needed to observe microbial interaction dynamics under controlled conditions. Studies have increasingly turned to simplified systems[25–27], such as co-cultures[28–30] and mixed cultures[16,31,32], in order to identify specific types of interactions. To further contextualize these findings, our literature review in **Table S1** provides a comprehensive overview of publications where microbial interaction patterns were observed in different experimental settings, including isolation attempts, co-cultures, and different versions of microbial model communities. Across these studies, cross-feeding mechanisms and mutualistic

interactions are mostly studied in cultures with two different populations or species[29,33,34]. To a lesser extent, more complex metabolic interactions have also been studied by mixing different isolated species together in model ecosystems, with the goal of increasing complexity to more closely resemble natural environments[31,35,36]. However, these methods focus on cultured isolates, and the vast majority of microorganisms remain uncultivated[37]. An alternative, yet underutilized, method of establishing model ecosystems composed of previously uncultivated microorganisms is through dilution cultivation. Dilution to mixed-cultures of naturally co-occurring microorganisms has the potential to allow observation of natural microbial interactions[37]. Such cultures, also known as microbial model communities, represent a small subset of the many interactions likely occurring in the natural system[35]. By studying a larger number of microbial model communities, we can gain a more comprehensive understanding of microbial interactions occurring in a natural environment.

In our study, we have focused on studying auxotrophies and potential interdependencies at two levels of biological organization by using high-throughput dilution cultivation of model communities together with genome-resolved metagenomics to unravel ecological strategies of microorganisms in the Baltic Sea. The 204 obtained microbial model communities revealed insights into community complexity. Moreover, we identified correlations between genome size, potential biosynthesis of essential metabolites, relative abundance, and prevalence using genomes obtained from both microbial model communities and metagenomic data from Baltic Sea pelagic samples. Our findings demonstrate that microbial model communities are an effective cultivation technique for identifying putative microbial interactions, including auxotrophies that lead to metabolic interdependencies and thereby promote stable microbial community structure and function.

RESULTS

A Baltic Sea MAG catalog

To generate the microbial model communities, we used the dilution-to-extinction cultivation technique in two formats. The first type, low inoculum size (LS), involved inoculating between approximately 2 and 100 cells per well in 1 ml 96-well plates used for each of the inoculum sizes. The second type, high inoculum size (HS), ranged from approximately 200 to $1x10^6$ cells inoculated per microbial model community in 100 ml volumes. After a 4-week incubation period, we sent an aliquot of all cultures for lysis and DNA amplification. Out of 801 cultures,

315 passed the amplification threshold and were sent to sequencing. In total, 204 microbial model communities together yielded 527 MAGs. Moreover, from the original sample used to establish the microbial model communities, we generated two metagenomes (each from a distinct DNA extraction method) that yielded 305 MAGs. To create a comprehensive genomic catalog, we added 771 MAGs deriving from 110 Baltic Sea metagenomes (**Figure 1A**, **Table S2**)[39–41].

Examining this comprehensive MAG catalog allowed us to assess microbial genomic characteristics of all microorganisms found both in the environmental samples and our 204 microbial model communities (**Table S3**). Altogether, the 1603 MAGs were dereplicated into 701 species-clusters (ANI > 95%), which form the Baltic Sea genomic catalog (BalticMAG catalog)[42] used in this study (**Table S4**). The average completeness of the 701 species-cluster representative MAGs is 88%, and they were all used to analyze taxonomy, abundance, prevalence, and estimated genome size (**Figure S1**). The varying completeness of the MAGs has a very minor effect on estimated genome size and relative abundance, just as observed in other studies[14,43]. However, only 450 species-cluster representative MAGs are of high-quality (completeness >90% and contamination <5%) and were used to reliably investigate metabolic potential (**Figure S1D)**.

Examining the source of genomes in the BalticMAG catalog, we found that 33 of the species-clusters included MAGs exclusively from the microbial model communities, 629 included MAGs exclusively from the environmental metagenomes, and 39 (54% of all cultured species) included MAGs from both sources (**Figure 1C, Table S4**, **Figure S2).**

To investigate the relative abundance of the BalticMAG catalog, we mapped all the environmental metagenomics reads against the genome catalog. We observed that salinity significantly co-varied with the proportion of metagenomic reads that mapped to all species representative genomes (**Figure 1B**). While on average, 38.63% of the metagenomic reads per sample mapped to the BalticMAG catalog, the highest mapping percentage (63.99% and 62.89%) was observed at salinity concentrations 11.28‰ and 7.65‰, respectively. Notably, the two metagenomes from this study (sample with salinity concentration 7.12‰) displayed some of the highest mapping rates at 51.77% and 60.16%, reflecting that the BalticMAG catalog is most complete for salinities between 6 and 11‰.

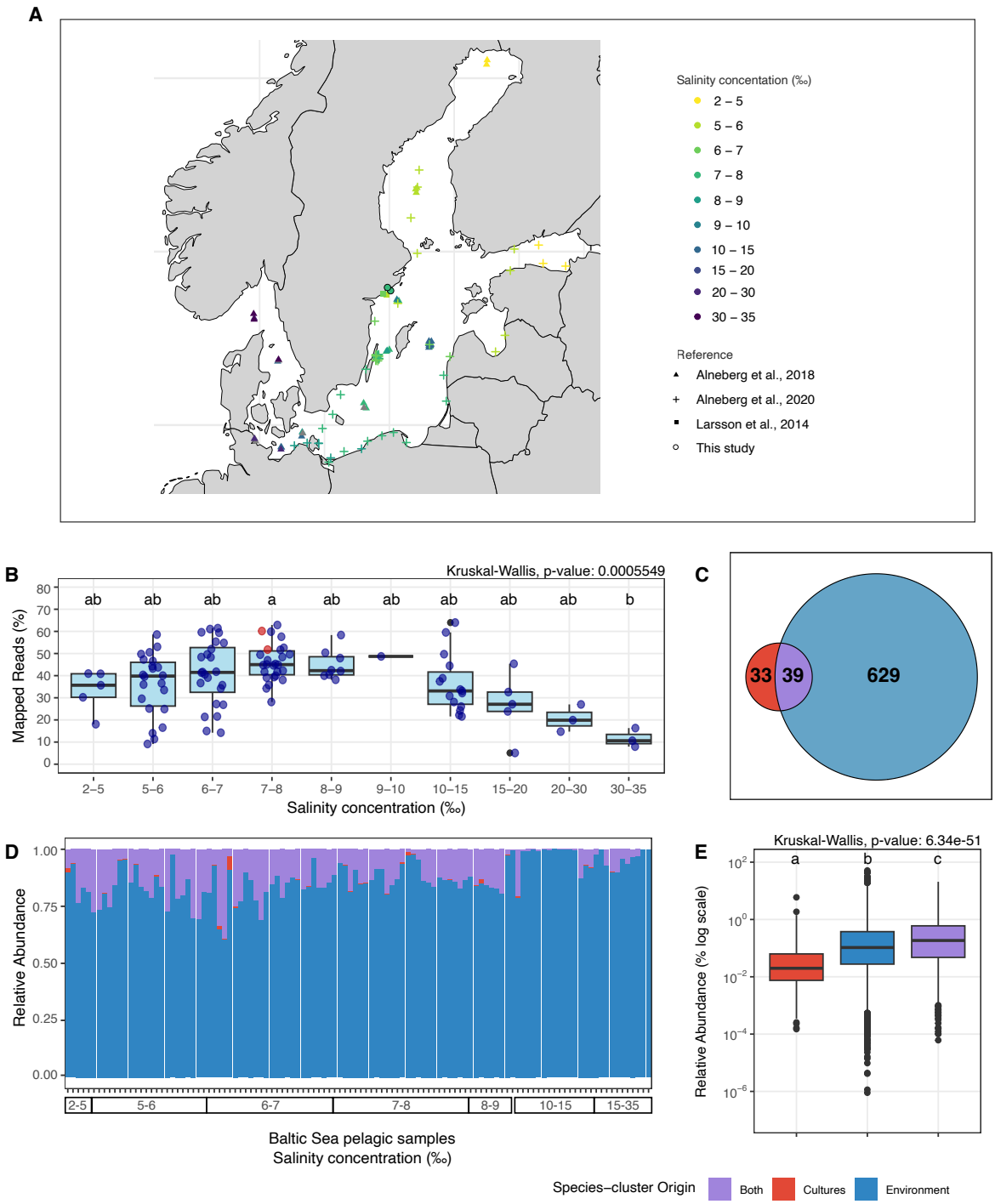

Figure 1. The BalticMAG catalogue. (A) Map of the Baltic Sea showing the geographic location of all the metagenomes analyzed in this study, including our sampling site (n =112). Sampling sites are color-coded according to their salinity gradient (measured in ‰, equivalent to PSU or parts per thousand). The shapes correspond to the different reference sources. (B) Boxplot showing the percentage of mapped reads from all metagenomes to our species-cluster collection, categorized by salinity concentration (Kruskal-Wallis test). (C) Venn diagram showing the overlap of species-clusters presence among cultures (red), the environment (blue), and found in both (purple). (D) Relative abundance of species-clusters across the collection of metagenomes, ordered by salinity gradient from left to right (corresponding to the order in panel B). (E) Boxplot comparing the relative abundance of species-clusters categorized by their presence in cultures, the environment, or both (Kruskal-Wallis test).

Kruskal-Wallis tests were followed by Dunn's post hoc test for pairwise comparisons. Letters indicate groups with significant differences ($p < 0.05$).

Despite salinity differences (**Figure 1D**), the average relative abundance of the 39 species-clusters that included MAGs from both microbial model communities and environmental metagenomes was significantly higher in the whole dataset (**Figure 1E),** but also when comparing only the metagenomes from the location and salinity from which we sampled **(Figure S3**). Moreover, the 33 species-clusters with MAGs sourced exclusively from microbial model communities were detected across environmental metagenomes, albeit at significantly lower abundances. Interestingly, 26 of these 33 species-clusters grew in model communities with more than two members and were less abundant in the environment (**Figure 1E**), indicating that the cultivation approach not only captures abundant taxa but also enables the recovery of species that may be missed in metagenomic surveys.

Model communities with more than three species harbor the most diverse and prevalent taxa

In the LS microbial model communities, the more cells we inoculated, the higher the number of cultures that yielded MAGs (**Figure 2A**). In total, 94 LS model communities yielded only one MAG each, while 110 model communities resulted in two or more MAGs. In model communities with more than one MAG, each MAG belonged to a different species-cluster in our analysis. Therefore, for clarity, we refer to different MAGs within a model community as different species. On the one hand, microbial model communities with an inoculum size of less than approximately 30 cells primarily resulted in single-species cultures (**Figure 2B**). On the other hand, starting from an inoculum size of approximately 30 cells, microbial model communities with more than two species appear more often (**Figure 2B** and **C**). Nearly 82% ($n = 433$) of the microbial model community MAGs were obtained from multi-species cultures, and the highest number of co-occurring species were found in the HS microbial model communities inoculated with 5000 cells (**Figure 2C**). Despite using a complex inoculum from the Baltic Sea, observing a maximum of 13 co-occurring species suggests that our cultivation conditions may impose a threshold on model community complexity. Alternatively, since sequencing followed MDA, there is also the possibility that some model communities included more species that were not amplified or assembled and binned. Nevertheless, the increased growth success at higher inoculum sizes may be attributed to a greater chance of inoculating cells that will grow either alone or together.

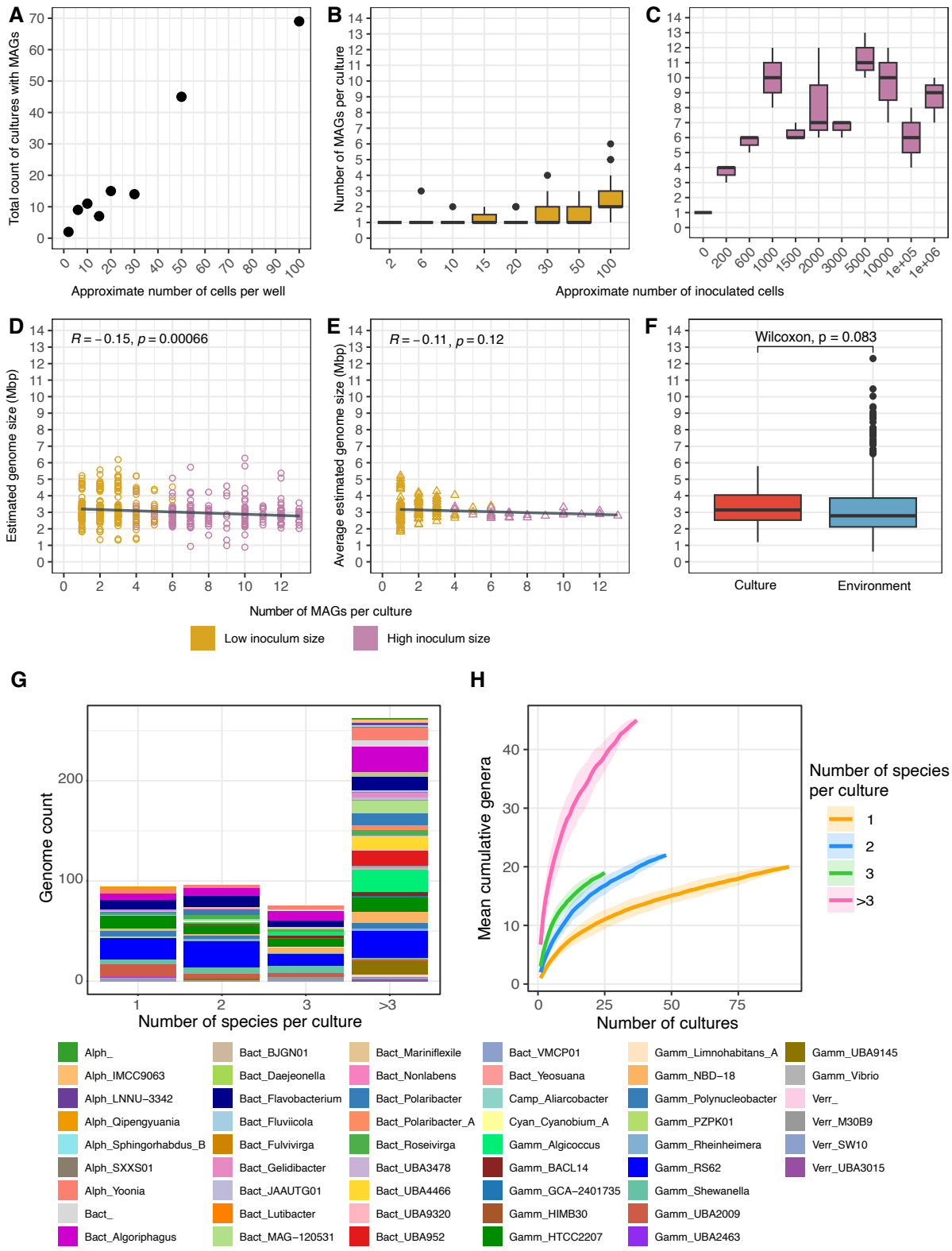

Figure 2. Impact of inoculum size on microbial richness, genome size, and taxonomic diversity of cultures. (A) Relationship between the number of inoculated cells (low inoculum size = 0 to 100 cells/well) and the total number of cultures from which MAGs were assembled and binned. (B) Number of MAGs per culture for the low inoculum size (yellow) and (C) high inoculum size (pink) cultures. (D) Estimated genome size (circles) of MAGs from cultures (n = 527) in relation to the number of

species growing in the same culture, and (E) average estimated genome size (triangles) of MAGs growing in the same culture. (F) Boxplot comparing the average estimated genome size of species-clusters from cultures and those from the environment (Wilcoxon rank-sum test). (G) Barplot showing the total genome count in relation to the number of genomes recovered per culture. Genomes are color-coded by genera, and the legend shows abbreviated prefixed indicating the class of each genus (e.g., Alph_ for Alphaproteobacteria, Bact_ for Bacteroidia, Camp_ for Campylobacteria, Cyan_ for Cyanobacteria, Gamm_ for Gammaproteobacteria, and Verr_ for Verrucomicrobiae. (H) Rarefaction curves of cultured microbial genera by species-per-culture category. Each curve represents the mean cumulative number of unique genera recovered as cultured samples are progressively added. Cultures were categorized into four groups based on the number of species they contained: 1, 2, 3, or >3 species per culture. For each category, genus accumulation was calculated across 100 random permutations (bootstrapped iterations), and shaded ribbons represent ±1 standard deviation from the mean.

Among the 527 MAGs obtained from the microbial model communities, the estimated genome sizes ranged from 0.88 to 6.27 Mbp, with an average of 3.05 Mbp (**Figure 2D** and **F**). For the MAGs derived from Baltic Sea metagenomes, estimated genome sizes ranged from 0.61 to 12.32 Mbp, with an average of 3.00 Mbp. While there was no significant difference between the average estimated genome sizes of species-clusters from microbial model communities and Baltic Sea metagenomes, we found that species-clusters exclusively composed of genomes from microbial model communities had significantly larger genome sizes than species-clusters including genomes from both sources (**Figure S4**). Moreover, we found that as the number of retrieved MAGs per microbial model community increased, more microorganisms with smaller genomes emerged, as indicated by a slight but significant correlation between estimated genome size and the number of MAGs per microbial model community (**Figure 2D**). Further analysis reinforced the observation that genomes from HS microbial model communities and from model communities with more than 3 species had on average, smaller genome sizes (2.89 and 2.92 Mbp, respectively - **Figure S5** and **S6**). Genomes from HS microbial model communities were also on average more prevalent and had higher average relative abundance in the investigated Baltic Sea metagenomes. Additionally, the average genome size per microbial model community stabilized at around 3 Mbp in cultures containing more than three species (**Figure 2E**).

Furthermore, the taxonomic diversity in our microbial model communities spanned 51 genera across 5 of the most abundant phyla (*Pseudomonadota*, *Bacteroidota*, *Campylobacterota*, *Cyanobacteriota*, and *Verrucomicrobiota*) present in the Baltic Sea environmental metagenomic sample[44,45], with microbial model communities with more than three species exhibiting a higher cumulative and rarefied diversity at the genus level (**Figure 2G** and **H**).

The analysis illustrates how more diversity growing together allows for the cultivability of different taxa that would normally not grow in axenic cultures under the conditions tested.

The diverse taxa cultivated in model communities together accounted for a relative abundance of approximately 20% in the original environmental sample (**Figure S7**). Notably, one of our HS model communities yielded a MAG that was found to be very abundant in the environment (~4.7%) classified in the order Pelagibacterales (SAR11) but from an unnamed genus according to GTDB-tk v2.1.1[46]. The order Pelagibacterales is widely recognized as one of the most abundant yet challenging to cultivate microorganisms in the ocean surface waters due to their nutritional requirements, which has been attributed to their small genome size and low biosynthetic potential[47,48]. In our study, the Pelagibacterales MAG possessed an estimated genome size of 0.93 Mbp and grew in a community of at least 10 species which had larger genome sizes, some up to 6 Mbp. Both the large novel diversity cultivated and their high abundance in nature highlight the potential of model communities to capture hard-to-cultivate yet abundant microbes from the environment, offering a promising future for microbial cultivation.

Species strictly found growing in groups are found more frequently in the environment.

In our microbial model communities, a total of 72 species were identified. The prevalence of species among the model communities varied greatly, from some detected only once (e.g., sc_576 and sc_639, names of species-clusters are found in **Table S4** and **Figure 3**) to some found in over 40 model communities (e.g., sc_011 and sc_014). Fifteen species (21%) were present in more than 10 model communities, collectively accounting for approximately 63% of the total MAGs (330 out of 527). Distinct patterns were observed across species, with 4 species found exclusively on their own (e.g., sc_035 and sc_281) and 48 species found only in groups (e.g., sc_102 and sc_177). Notably, 20 species were found both independently and in groups (e.g., sc_011 and sc_014). Most of the frequently retrieved MAGs showed this flexibility, growing independently or in groups (**Figure 3A** and **B**). However, our experiment did not systematically test if species only found on their own, could also grow in groups or if species found in groups only, could also grow alone. Nevertheless, when we examined presence patterns in the microbial model communities in relation to relative abundance across the Baltic Sea environmental sample used as inoculum, we see a trend for the species found both independently and in groups (**Figure 3C**). The total number of genomes found in the cultures per species increases as their relative abundance in the environment of origin increases.

On the other hand, the two most abundant species identified in the environment of origin belong to the order Pelagibacterales (sc_139) and Methylacidiphilales (sc_121) and were found exclusively in groups and cultivated only a few times. Moreover, for species that demonstrated the ability to grow both independently and in groups, as well as those that exclusively grew in groups, a significant positive correlation emerged between the total number of genomes per species and their prevalence in the environment (**Figure 3D**). The relationship indicates that species with higher environmental prevalence are more frequently cultured in laboratory settings. Moreover, the correlation is further enhanced when species grow more easily, meaning those found growing both independently and in groups. Nevertheless, the species with the highest prevalence across all environmental samples, as well as the highest relative abundance in the sample of origin, were found exclusively growing in groups.

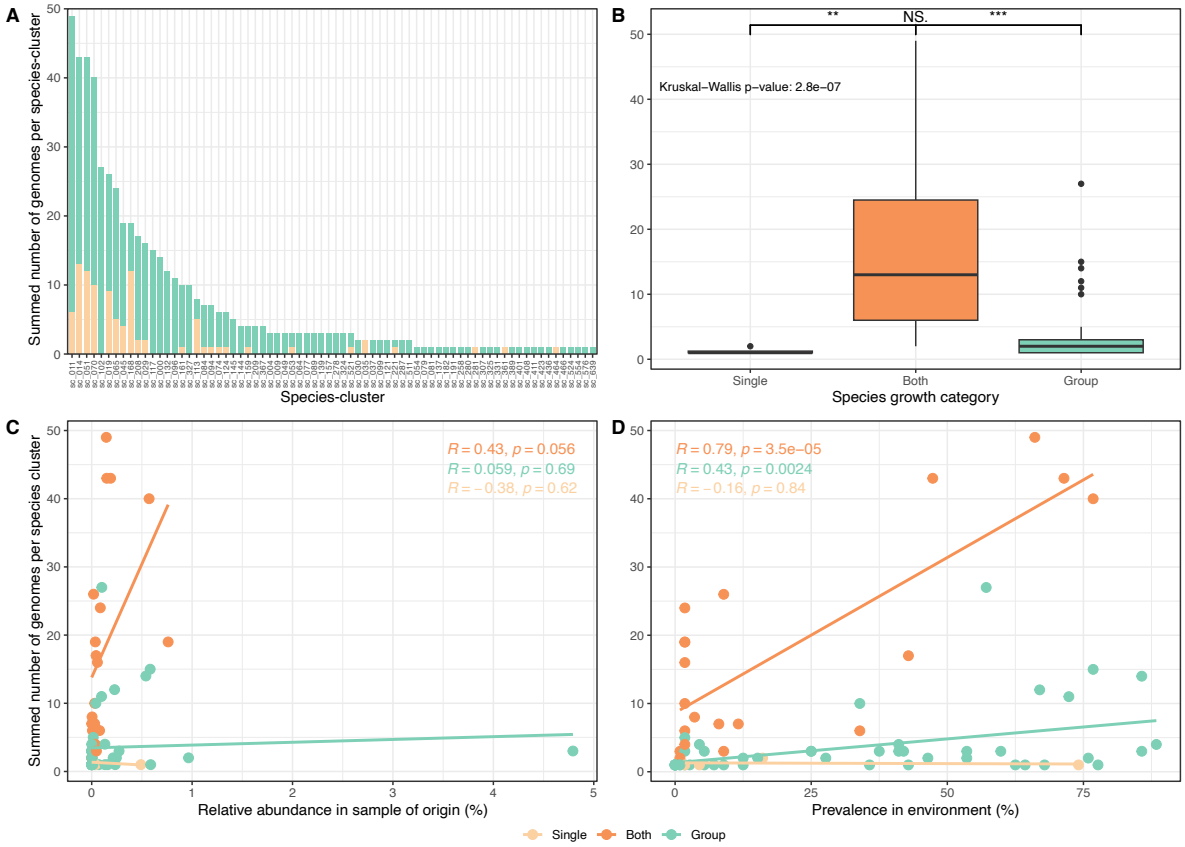

Figure 3. Cultured genomes per species-cluster in relation to their observed growth condition (alone, group, or both), relative abundance in the sample of origin, and prevalence in the Baltic Sea. (A) Barplot displaying the total number of cultured genomes for each species-cluster. Color coding indicates whether the genomes originate from single-genome-cultures (light orange) or multi-genome-cultures (green). Species-clusters were further differentiated according to whether they grew exclusively on their own ("Single" = light orange), exclusively in groups ("Group" = green), or both on their own as well as in groups ("Both" = orange). (B) Boxplot showing the total number of cultured genomes per species-

cluster for these three growth categories. Letters indicate groups with significant differences ($p < 0.05$) (Kruskall-Wallis-Test followed by a Dunn's post-hoc test with Bonferroni correction). (C) The dot plots display the relationship between the total number of cultured genomes per species-cluster and relative abundance in the sample of origin (D) as well as the prevalence of these species-clusters in all environmental samples. Each data point represents one species-cluster. Linear regression models were fitted to the data.

Auxotrophies and interdependencies are more common in MAGs from model communities with more than three species

We compared estimated genome sizes and biosynthetic potential among species found on their own sometimes or always (singleton-capable) or strictly in groups (group-only). Although there was no significant difference in estimated genome sizes between singleton-capable and group-only species (**Figure 4A**), we observed a significant difference in average pathway completeness for both amino acids and vitamin modules, with species found exclusively in groups exhibiting a lower average pathway completeness (**Figure 4B**). We next examined biosynthetic potential across all microbial model communities, categorizing them by their number of species (one, two, three, or more than three). Here, we found that average pathway completeness for amino acids and vitamins was significantly lower for model communities with more than 3 species found together (**Figure 4C** and **D**). Amino acid biosynthesis potential is in particular lower for species that exclusively grow in groups and are found in model communities that contain more than 3 species (**Figure S8**).

We explored interdependencies among species growing together. For this, we used the stepwise module framework in all MAGs from the 204 model communities, regardless of genome completeness, as partial genomes still provide informative evidence of biosynthetic potential. Model communities with more species have cumulatively all the necessary steps for the biosynthesis of essential metabolites (**Figure 5**).

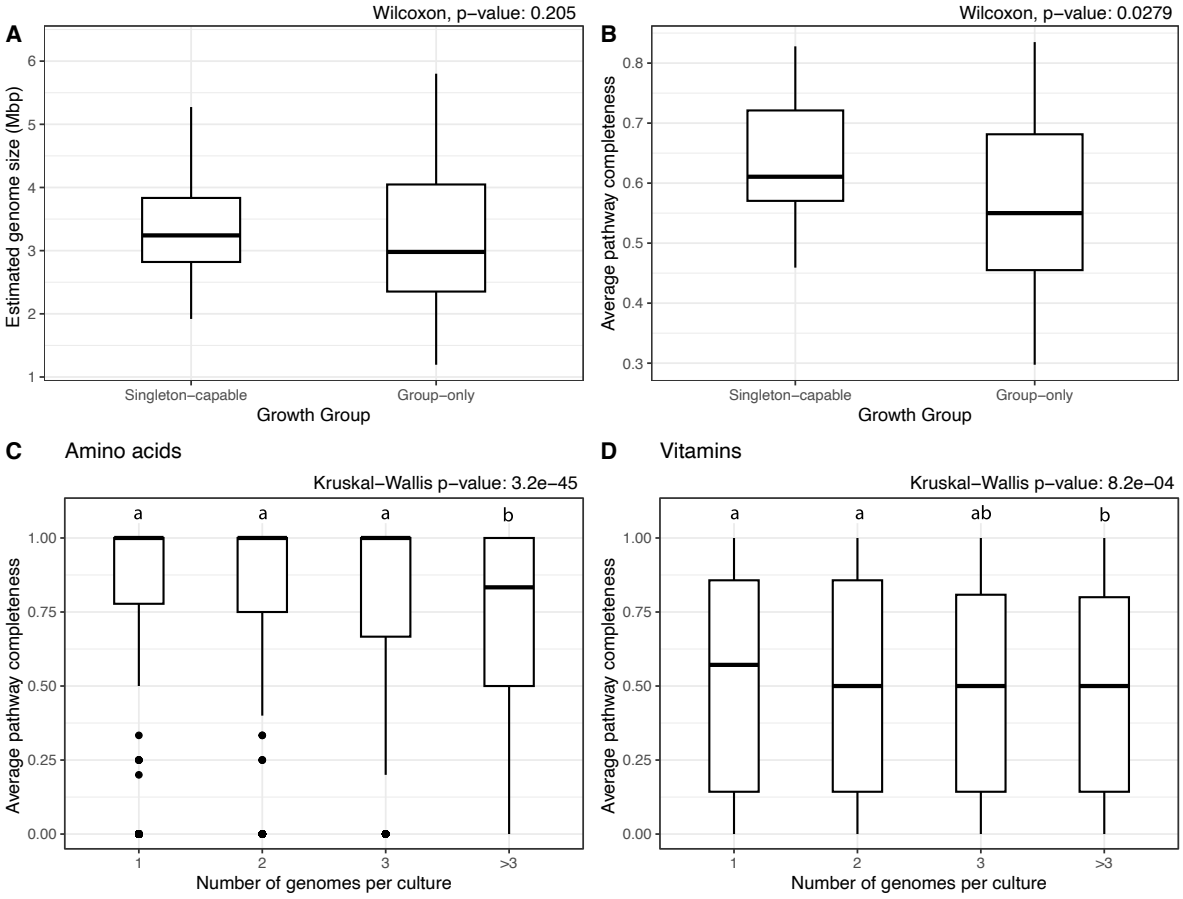

Figure 4. Relationship between genome size, biosynthesis pathway completeness, and growth strategy. (A) Boxplot comparing the estimated genome size of species-cluster representative genomes (n = 72) growing exclusively in groups (mixed-only) and capable of growing alone (singleton-capable). (B) Boxplot comparing the average pathway completeness for amino acid and vitamin biosynthesis modules between species-cluster representative genomes (n = 57, completeness >90%) growing always in groups and those capable of growing alone (Wilcoxon rank-sum test). (C) Boxplot displaying the average pathway completeness for amino acid biosynthesis and (D) vitamin biosynthesis across genomes (n = 305, completeness >90%) in cultures containing different numbers of species. Letters indicate groups with significant differences ($p < 0.05$) (Kruskal-Wallis test followed by Dunn's post-hoc test with Bonferroni correction). (E) Heatmap showing the average biosynthetic module completeness for amino acid and B vitamin modules across all 204 microbial model communities. Values reflect the average presence of module steps (genes) across all genomes per culture (n = 527), regardless of genome completeness. Modules are grouped by compound type, and microbial model communities are ordered by richness. Gradient scale ranges from white (0 = incomplete) to black (1 = complete).

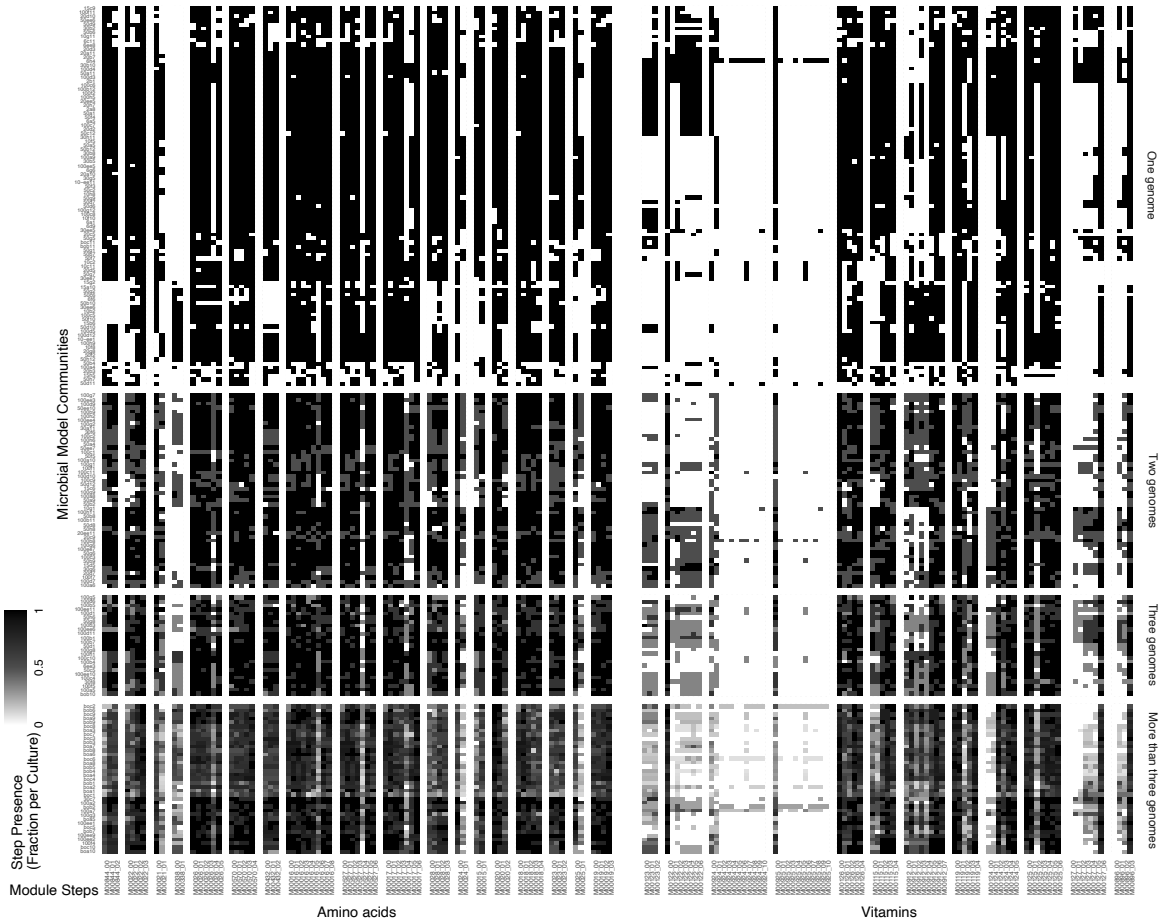

Figure 5. Stepwise module framework analysis of amino acid and B vitamin biosynthetic pathways across microbial model communities. Heatmap shows the distribution of individual biosynthetic module steps of all genomes in the 204 microbial model communities. Each column represents a specific KEGG module step (e.g., M000020_01) grouped by compound and module type. Rows correspond to individual model communities, organized by the number of species they contain (one, two, three, or more than three). Values indicate the proportion of genomes in each culture that encode a specific module step, ranging from 0 (white) to 1 (black).

Ubiquitous and highly relatively abundant species tend to have smaller estimated genome sizes and more auxotrophies.

Overall, we observed a negative correlation between amino acid and vitamin average pathway completeness and average relative abundance for species growing in groups as well as for species growing alone (**Figure 6A** and **Figure S9**). When considering vitamin biosynthesis alone, we found a significant positive correlation between estimated genome size and average pathway completeness for those species growing exclusively in groups (**Figure 6B**). Our results suggest that species found exclusively in groups have higher average relative abundances and smaller genomes with reduced biosynthetic capacity, particularly for vitamins. Scaling the observations from model communities to all species from the BalticMAG catalog, we found that higher average relative abundances (**Figure 6C**) and smaller estimated genome sizes (**Figure 6D**) were linked to lower average completeness of biosynthetic pathways for amino acids and vitamins. Overall, prokaryotes with smaller estimated genome sizes, are more prevalent and abundant in the Baltic Sea and possess lower average biosynthetic potential for amino acids and vitamins (**Figure 6** and **Figure S10)**. Specifically, we found that biosynthetic pathways for cysteine, proline, arginine, tryptophan, histidine, leucine, tyrosine, and phenylalanine, vitamin B_1, B_7, and B_{12} were more frequently absent in many species (**Figure S11**).

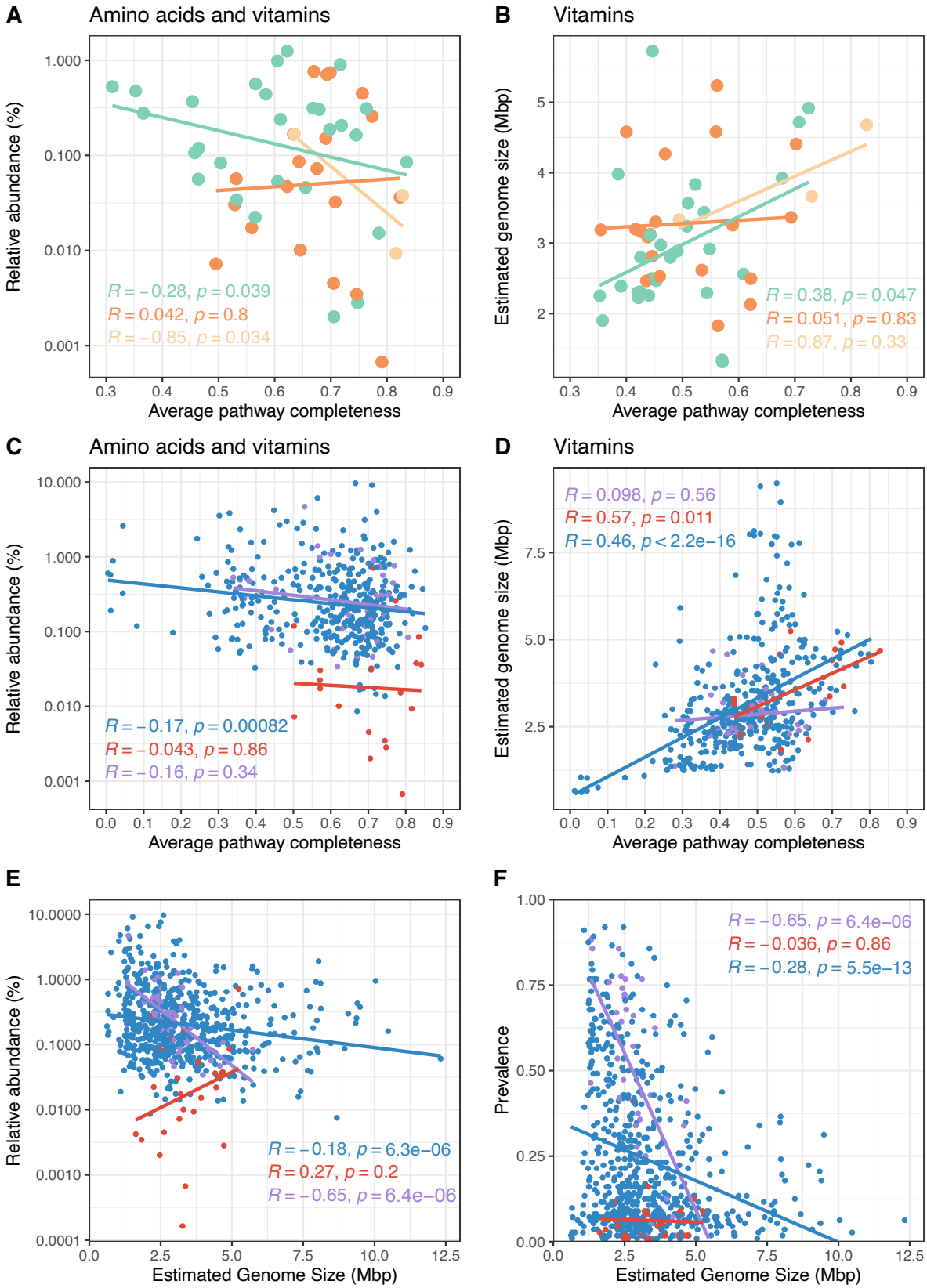

Figure 6. Link between pathway completeness, relative abundance, and estimated genome size of cultured species and the BalticMAG species catalog. (A) Correlation between average biosynthesis pathway completeness of amino acids and vitamin modules and their relative abundance of cultured species (n = 57, completeness >90%). (B) Correlation between the average biosynthesis completeness for vitamin modules and the estimated genome size of cultured species-clusters. Each data point

represents one species cluster (n = 57, completeness >90%), which either grew exclusively on their own (light orange), exclusively in groups (green), or both on their own as well as in groups (orange). Panels (C) and (D) replicate the analysis from panels (A) and (B) but include all species-clusters (n = 450, completeness >90%) from both cultures and environmental samples. Correlation between estimated genome size and average relative abundance and (F) correlation between estimated genome size and prevalence of all species-clusters (n = 701). Data points are color-coded to represent species clusters from cultures only (red), environments only (blue), and found in both (purple).

DISCUSSION

The field of microbiology has traditionally relied on the isolation of microorganisms from nature in pure culture[49–54]. The practice has yielded foundational insights into microbial physiology[55–57], metabolism[58–61], and ecological interactions such as mutualism[30,62,63], and competition[64–66]. However, this reductionistic approach has limitations, particularly a strong cultivation bias favoring microorganisms with larger genome sizes[43,67], with broader biosynthetic potential[68], and greater independence. As a result, a substantial portion of the most abundant microbes in natural environments remains uncultivated[69].

Here, we demonstrate the utility of microbial model communities established through dilution cultivation from a Baltic Sea pelagic sample for obtaining a wide range of abundant and biosynthetically limited microorganisms. Despite not extensively varying physicochemical parameters, our method has provided principles that allow for the cultivation of groups of organisms with the potential to showcase important ecological principles governing microorganisms in nature. Our results suggest that by using high-throughput cultivation of microbial model communities, relevant ecological principles can be better examined and that by varying cultivation physicochemical parameters, such as adding catalase[70], changing light regimes[71], or temperature[72], we will be able to cultivate a greater proportion of the abundant microorganisms found in nature.

In particular, we found that species growing in microbial model communities composed of more than three species exhibited lower biosynthetic potential for both amino acids and B vitamins compared to species found growing in smaller groups or independently. The threshold for species growing together (n > 3) indicates that at higher community complexities, microorganisms with auxotrophies can form metabolic networks to support their nutritional requirements[73]. Moreover, a previous study has also found a threshold of microbial diversity at which competition and complementation saturate[32]. Specifically, they observed that beyond

approximately 26 taxa, further increases in diversity had little detectable impact on an overall community function (respiration), suggesting a saturation of functional capacity. Although our system looks at different types of functions, both studies observe a diversity threshold for functional saturation, emphasizing that microbial communities likely achieve a balanced state of interaction complexity beyond certain diversity levels, thereby optimizing ecological efficiency.

Our findings also align with the Black Queen Hypothesis, which posits that certain functions, particularly costly biosynthetic pathways, can be lost by some community members as long as the production of the metabolite at the community level is sufficient to sustain the individual[74]. In this context, the reduced genome sizes observed in the abundant species from the model communities and the environment likely reflect a genome streamlining process[75]. The findings further align with the hypothesis that obligate co-existing microbes have evolved to rely on their community for essential nutrients[76], potentially creating social networks that might bring microbial community stability[11,77–79]. To support this, our stepwise module framework analysis revealed that microbial model communities with more than three species collectively maintain the biosynthetic potential for amino acids and vitamins.

Importantly, our findings also complement large-scale studies investigating the ecological distribution of auxotrophies. A recent analysis of over 26000 representative bacterial genomes across diverse environments found that auxotrophy is more common in host-associated environments while relatively rare in aquatic and soil ecosystems. However, their analysis draws a robust but strict boundary on how to categorize microorganisms in a binary model of prototrophs and auxotrophs (more than 40% genes missing per pathway for auxotrophy)[13]. While this categorization has been widely used, it might also be hiding potential interdependencies. In our study, we draw no boundary for defining auxotrophy. Instead, we studied the pathways by observing their completeness and compared them across the different species in the dataset. We believe this has the potential to reveal a wide spectrum of possibilities where microorganisms might just need a precursor to complete biosynthesis[16].

Interestingly, we observe that in the Baltic Sea, a brackish environment where salinity is a driver for diversity, ecology, and evolution of microorganisms[80], the most abundant and prevalent taxa possess streamlined genomes with low biosynthetic pathway completeness, as previously observed in another study from the Baltic Sea[81] . This pattern has also been reported

432 in freshwater[14] and marine[11] ecosystems, reinforcing the idea that streamlined, auxotrophic
433 microorganisms are widespread and ecologically successful when they reside in communities
434 capable of fulfilling their metabolic needs.
435
436 In summary, our study highlights the larger potential of microbial model communities in
437 bridging the gap between laboratory cultivation and ecological representation. By allowing
438 naturally co-occurring microorganisms to grow together as communities, we can capture some
439 of the most abundant taxa with the potential for biosynthetic interdependence. These findings
440 provide evidence that increasing community complexity is associated with reduced metabolic
441 autonomy, suggesting that microbial interdependencies are common and can emerge as a
442 natural consequence of a shared metabolic lifestyle. By cultivating groups rather than
443 individuals, we can access a more ecologically relevant understanding of how microbes
444 survive, interact, and evolve in nature.
445
446 **MATERIALS AND METHODS**
447
448 **Sampling and sample processing**
449 We collected an environmental sample from the surface layer of the Baltic Sea close to Askö
450 in the Trosa archipelago (Lat 58°48.20'N, Lon 17°37.42'E) in October 2021 (**Figure 1A**). The
451 sample was subsequently processed in the laboratory for DNA extraction and metagenomic
452 sequencing. Briefly, the water sample was filtered through a 0.1 µm membrane and used to
453 extract environmental DNA using either the FastDNA® SPIN Kit for Soil (MP Bio) and the
454 Dneasy PowerWater kit (Qiagen). Additionally, we filtered water through a 0.1 µm hollow fiber
455 cartridge (Cytiva) and used the filtrate as media to establish cultures with different starting
456 inoculum sizes.
457
458 **Establishment of microbial model communities.**
459 We used the dilution-to-extinction technique, as previously described[76], to achieve
460 approximately the desired starting number of cells in our cultures. The cultures were designed
461 in two ways: low inoculum size (LS) and high inoculum size (HS). The LS cultures included
462 hundreds of individual cultures in 96-well plates with a starting number of cells of 0 (control),
463 2, 6, 10, 15, 20, 30, 50, and 100 cells/well. The HS cultures were prepared in bottles with
464 starting numbers of cells of 0 (control), 200, 600, 1000, 1500, 2000, 3000, 5000, 10000,
465 100000, and 1000000 cells/bottle. All cultures were incubated with light/dark cycles (light

466 6:42, dark 18:32 each day) and 12.2 °C at light and 11.8 °C at dark for four weeks before further
467 processing. This regime was aimed to emulate the time of dark and light and temperature that
468 the cells experienced in nature. However, after the experiment had already run and samples
469 were sequenced, we learned that our light intensity was about five times higher than in nature.
470
471 **Flow cytometry.**
472 We used a CytoFLEX instrument manufactured by Beckman Coulter to process the
473 environmental sample and to calculate the cell count as the number of events/mL. Briefly, we
474 stained 50 µl of the samples with Syto13 at a final concentration of 0.025 mM. During the flow
475 cytometer acquisition process, we set the following parameters: FSC 2500, SSC 2500, and
476 FITC 800 with a flow rate of 60 µl/min.
477
478 **MDA, library preparation, and sequencing.**
479 Before sending our samples for metagenomic sequencing, we performed multiple displacement
480 amplification (MDA) on all cultures to increase the concentration of DNA[83]. The MDA
481 reaction consisted of 0.6 µL of culture and 4.4 µL of reaction mix using the Repli-g Single Cell
482 kit (Qiagen). After DNA amplification, we found that 315 of the cultures passed the
483 amplification threshold of the negative controls set in the MDA reaction. These cultures were
484 deemed positive and were sent for sequencing. The selected cultures and our environmental
485 sample extracted with two different DNA extraction methods were subjected to library
486 preparation using TruSeq PCRfree DNA library preparation kit (Illumina Inc.) followed by
487 metagenomic Illumina sequencing at the SNP&SEQ Platform at Uppsala University. This
488 sequencing technology utilized cluster generation and 150 cycles of paired-end sequencing on
489 an SP flow cell, employing the NovaSeq 6000 system with v1.5 sequencing chemistry
490 (Illumina Inc).
491
492 **Genome resolved metagenomics pipeline.**
493 We removed low-quality reads from the raw sequences using the software Trimmomatic
494 (v0.36)[84] with the following options: ILLUMINACLIP:TruSeq3-PE-
495 2.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
496 MINLEN:50. We employed the MetaWRAP pipeline (v1.3.2)[85] to process our clean
497 metagenomic reads. First, clean reads were assembled in a single-sample assembly style using
498 the "metaWRAP_assembly" module with MegaHit (v1.1.3) for the environmental samples[86]
499 and with SPAdes (v3.15.3)[87] for the culture samples. The quality of the assemblies was assessed

500 with QUAST (v.5.0.2)[88]. Since the culture DNA was amplified using MDA, the read
501 information for coverage could not be used. For this reason, we included background
502 metagenomic data from previous projects in the Baltic Sea (**Figure 1A** and **Table S2**) for
503 binning. Subsequently, we used multiple-sample coverage binning to decrease the
504 contamination and increase the completeness of bins[89]. The reads were mapped against all
505 assemblies using the Input_POGENOM pipeline[90], which uses Bowtie2[91] with default
506 parameters. After mapping and obtaining the BAM files, the min coverage was calculated using
507 samtools (v1.9)[92]. Input_POGENOM estimates the coverage of a genome in a sample by only
508 mapping a subset of the reads, which includes the ones that reached a mean coverage of 20 and
509 mean breadth of 40. These coverage values for each sample were combined and processed with
510 the "metaWRAP_bining" module, which uses three metagenomic binning tools: metaBAT2[93],
511 maxBIN2[94], and CONCOCT[95]. We consolidated all the bins generated by these different tools
512 using the "metaWRAP_bin_refinement" module. Finally, we first classified the resulting bins
513 taxonomically with GTDB-tk v2.1.1[46]. Subsequently, we assessed the quality of the bins using
514 CheckM (v1.1.3)[96]. We considered bins as MAGs when they had a completeness of >45% and
515 a contamination of <10%, and these MAGs were included for further analysis.
516
517 **Complementing with previously published MAGs.**
518 The de-replicated MAGs obtained here were supplemented with 771 MAGs from an earlier
519 study[39] that were based on metagenomics data from three studies[39–41]. We de-replicated the
520 collection of MAGs to obtain species-cluster representatives using ANI > 95% with mOTUpan
521 (v0.3.2)[97], and selected the genome with the highest quality as the species-cluster
522 representative genome.
523
524 **Relative abundance analysis.**
525 To calculate the relative abundance of our 701 species-clusters in the BalticMAG catalog, we
526 employed the mapping tool Strobealign (v0.14.0)[98] , which aligns the short metagenomic reads
527 to our species-clusters collection using a high-speed indexing method (**Table S5**). Briefly, we
528 created three different indexes with different lengths (100, 125, and 150) for our 701 species-
529 clusters. We filtered out low-quality reads from our 112 environmental samples with
530 Trimmonatic (v0.36)[84] with the following options: ILLUMINACLIP:TruSeq3-PE-
531 2.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
532 MINLEN:50. After this, we did a competitive mapping of all reads against our created index
533 to obtain the corresponding BAM file for each sample. We sorted the BAM files with the

534 Anvi'o platform (v7.1)[99] using the anvi-init-bam function, and we calculated the coverage of
535 each genome per sample with the function anvi-profile-blitz. We calculated each species-
536 cluster's relative abundance with the previously obtained output by dividing each genome's
537 mean coverage inner quartiles (i.e., q2q3_cov) by the overall sample mean coverage.
538 Moreover, we measured the "prevalence", defined as the frequency of each species-cluster
539 across samples. Specifically, it represents the proportion of samples in which a species-cluster
540 was detected with a relative abundance > 0.
541
542 **Functional annotation of KEGG pathways (KOs).**
543 We used the Anvi'o platform (v7.1)[100] to perform functional annotation of KEGG pathways.
544 Initially, we created a contigs database for each genome, which served as the basis for the
545 subsequent functional annotation steps. To annotate each genome, we ran the "anvi-run-kegg-
546 kofams" program, which searches for functions and metabolic pathways based on the KEGG
547 orthologs (KOs) database[101]. We then predicted the metabolic capabilities of each genome by
548 running the "anvi-estimate-metabolism" program[102]. We evaluated the completeness of various
549 metabolic modules in each genome, focusing on the biosynthetic modules for amino acids and
550 B vitamins (**Table S6**).
551
552 **Statistical analyses.**
553 All statistical analyses from this study were performed with R version (v4.4.0)[103] and
554 RStudio[104]. We used the Shapiro test to assess the normality of our data to be compared; if $p <$
555 0.05, we interpret this as not normally distributed[105]. Since our data was not normally
556 distributed, we employed a non-parametric test, such as the Wilcoxon test, to find differences
557 between pairs of groups (e.g., culture vs environment)[106]. We also performed a Kruskal-Wallis
558 test to compare different groups, and the p-values were corrected using the Benjamin-Hochberg
559 procedure[107]. Kruskal-Wallis tests were followed by Dunn's post hoc test for pairwise
560 comparisons[108], and significant differences are indicated by the different letters when $p < 0.05$.
561 All comparisons were reported to be significant if the p-values were < 0.05.
562
563 **Data availability.** The paired-end sequences of both environmental (n =2) and culture (n =
564 204) metagenomic samples from the Baltic Sea samples from this study, along with the
565 corresponding 827 MAGs (>45% completeness and <10% contamination), have been
566 deposited in NCBI under the BioProject ID PRJNA1134408. These data are not yet publicly
567 available but can be accessed during peer review via the following temporary link:

https://dataview.ncbi.nlm.nih.gov/object/PRJNA1134408?reviewer=neftpak1i4ufv6i9i7qj3o697e. All other metagenomes from the Baltic Sea (n = 110) were downloaded from public repositories, and the metadata is included in **Table S2**, including their publication reference.

ACKNOWLEDGEMENTS

This work was funded by SciLifeLab and by the Swedish Research Council VR (grant 2022-03077). The authors would like to acknowledge support from the Genomics infrastructure services at Science for Life Laboratory in Uppsala. The MDA was done by Claudia Bergin at the Microbial Single Cell facility, and sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facilities are part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The authors acknowledge support from SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for access to the UPPMAX computational infrastructure as well as the National Academic Infrastructure for Supercomputing in Sweden (NAISS). Computational work and data handling were enabled by resources in the projects SNIC 2022/5-392, 2023/5-126, and naiss2023-5-379 provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through a grant agreement no. 2018-05973 and 2022-06725.

SLG conceptualized, designed, and supervised the research. FB and SLG performed the cultivation work. AFA and LFD-Z provided the Baltic Sea MAGs that contributed to the genomic catalog. AP-V, JED, LFD-Z, and SLG conducted bioinformatics analysis. AR-G supported bioinformatic analysis. AP-V, AT, and SLG performed the data interpretation and visualization. AP-V led the writing of the manuscript with input from SLG. All co-authors contributed to the literature searches, participated in editing and reviewing the manuscript, and approved the final version.

CONFLICTS OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Konopka, A. What is microbial community ecology? *The ISME Journal* **3**, 1223–1230 (2009).
2. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320**, 1034–1039 (2008).
3. Gralka, M. Searching for Principles of Microbial Ecology Across Levels of Biological Organization. *Integrative And Comparative Biology* **63**, 1520–1531 (2023).
4. Herman, M. A. *et al.* A Unifying Framework for Understanding Biological Structures and Functions Across Levels of Biological Organization. *Integrative and Comparative Biology* **61**, 2038–2047 (2022).
5. O'Toole, G. A. We have a community problem. *J Bacteriol* e00073-24 (2024) doi:10.1128/jb.00073-24.
6. Benoit, G. *et al.* High-quality metagenome assembly from long accurate reads with metaMDBG. *Nat Biotechnol* **42**, 1378–1383 (2024).
7. Eren, A. M. & Banfield, J. F. Modern microbiology: Embracing complexity through integration across scales. *Cell* **187**, 5151–5170 (2024).
8. Méheust, R., Castelle, C. J., Jaffe, A. L. & Banfield, J. F. Conserved and lineage-specific hypothetical proteins may have played a central role in the rise and diversification of major archaeal groups. *BMC Biol* **20**, 154 (2022).
9. Hug, L. A. *et al.* A new view of the tree of life. *Nat Microbiol* **1**, 16048 (2016).
10. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, 13219 (2016).
11. Giordano, N. *et al.* Genome-scale community modelling reveals conserved metabolic cross-feedings in epipelagic bacterioplankton communities. *Nat Commun* **15**, 2721 (2024).
12. Hessler, T. *et al.* Vitamin interdependencies predicted by metagenomics-informed network analyses and validated in microbial community microcosms. *Nat Commun* **14**, (2023).
13. Ramoneda, J., Jensen, T. B. N., Price, M. N., Casamayor, E. O. & Fierer, N. Taxonomic and environmental distribution of bacterial amino acid auxotrophies. *Nat Commun* **14**, (2023).

14. Rodríguez-Gijón, A. *et al.* The ecological success of freshwater microorganisms is mediated by streamlining and biotic interactions. Preprint at https://doi.org/10.1101/2025.03.24.644981 (2025).
15. Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3695–3700 (2002).
16. Garcia, S. L. *et al.* Auxotrophy and intrapopulation complementary in the 'interactome' of a cultivated freshwater model community. *Mol Ecol* **24**, 4449–4459 (2015).
17. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* **111**, (2014).
18. Paerl, R. W. *et al.* Prevalent reliance of bacterioplankton on exogenous vitamin B1 and precursor availability. *Proc. Natl. Acad. Sci. U.S.A.* **115**, (2018).
19. Ryback, B., Bortfeld-Miller, M. & Vorholt, J. A. Metabolic adaptation to vitamin auxotrophy by leaf-associated bacteria. *The ISME Journal* **16**, 2712–2724 (2022).
20. Sirithanakorn, C. & Cronan, J. E. Biotin, a universal and essential cofactor: synthesis, ligation and regulation. *FEMS Microbiology Reviews* fuab003 (2021) doi:10.1093/femsre/fuab003.
21. Embree, M., Liu, J. K., Al-Bassam, M. M. & Zengler, K. Networks of energetic and metabolic interactions define dynamics in microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15450–15455 (2015).
22. Gómez-Consarnau, L. *et al.* Mosaic patterns of B-vitamin synthesis and utilization in a natural marine microbial community. *Environmental Microbiology* **20**, 2809–2823 (2018).
23. Wienhausen, G. *et al.* Ligand cross-feeding resolves bacterial vitamin B12 auxotrophies. *Nature* (2024) doi:10.1038/s41586-024-07396-y.
24. Oña, L. & Kost, C. Cooperation increases robustness to ecological disturbance in microbial cross-feeding networks. *Ecology Letters* **25**, 1410–1420 (2022).
25. Aziz, F. A. A. *et al.* Interspecies interactions are an integral determinant of microbial community dynamics. *Front. Microbiol.* **6**, (2015).
26. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. (2018).
27. Pacheco, A. R., Osborne, M. L. & Segrè, D. Non-additive microbial community responses to environmental complexity. *Nat Commun* **12**, 2365 (2021).
28. Bayer, B. *et al.* Metabolite release by nitrifiers facilitates metabolic interactions in the ocean. *The ISME Journal* wrae172 (2024) doi:10.1093/ismejo/wrae172.

29. Grant, M. A. A., Kazamia, E., Cicuta, P. & Smith, A. G. Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal–bacterial cocultures. *The ISME Journal* **8**, 1418–1427 (2014).
30. Hillesland, K. L. & Stahl, D. A. Rapid evolution of stability and productivity at the origin of a microbial mutualism. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2124–2129 (2010).
31. Chang, C.-Y., Bajić, D., Vila, J. C. C., Estrela, S. & Sanchez, A. Emergent coexistence in multispecies microbial communities. *Science* **381**, 343–348 (2023).
32. Yu, X., Polz, M. F. & Alm, E. J. Interactions in self-assembled microbial communities saturate with diversity. *The ISME Journal* **13**, 1602–1617 (2019).
33. Hammarlund, S. P., Gedeon, T., Carlson, R. P. & Harcombe, W. R. Limitation by a shared mutualist promotes coexistence of multiple competing partners. *Nat Commun* **12**, 619 (2021).
34. Pande, S. *et al.* Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J* **8**, 953–962 (2014).
35. Garcia, S. L. *et al.* Model Communities Hint at Promiscuous Metabolic Linkages between Ubiquitous Free-Living Freshwater Bacteria. *mSphere* **3**, e00202-18 (2018).
36. Ratzke, C., Barrere, J. & Gore, J. Strength of species interactions determines biodiversity and stability in microbial communities. *Nat Ecol Evol* **4**, 376–383 (2020).
37. Steen, A. D. *et al.* High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J* **13**, 3126–3130 (2019).
38. Garcia, S. Mixed cultures as model communities: hunting for ubiquitous microorganisms, their partners, and interactions. *Aquat. Microb. Ecol.* **77**, 79–85 (2016).
39. Alneberg, J. *et al.* Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun Biol* **3**, (2020).
40. Alneberg, J. *et al.* BARM and BalticMicrobeDB, a reference metagenome and interface to meta-omic data for the Baltic Sea. *Sci Data* **5**, (2018).
41. Larsson, J. *et al.* Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *The ISME Journal* **8**, 1892–1903 (2014).
42. Pacheco-Valenciana, A., Garcia, S. L., Dharamshi, J. E., Delgado-Zambrano, L. F. & Andersson, A. F. The BalticMAG catalog. https://doi.org/10.17044/scilifelab.28746086.v1 (2025).
43. Rodríguez-Gijón, A. *et al.* A Genomic Perspective Across Earth's Microbiomes Reveals That Genome Size in Archaea and Bacteria Is Linked to Ecosystem Type and Trophic Strategy. *Front. Microbiol.* **12**, 761869 (2022).

44. Herlemann, D. P. *et al.* Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* **5**, 1571–1579 (2011).
45. Rodríguez-Gijón, A. *et al.* Linking prokaryotic genome size variation to metabolic potential and environment. *ISME COMMUN.* **3**, 25 (2023).
46. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
47. Giovannoni, S. J. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annu. Rev. Mar. Sci.* **9**, 231–255 (2017).
48. Herlemann, D. P. R., Woelk, J., Labrenz, M. & Jürgens, K. Diversity and abundance of "Pelagibacterales" (SAR11) in the Baltic Sea salinity gradient. *Systematic and Applied Microbiology* **37**, 601–604 (2014).
49. Gich, F., Schubert, K., Bruns, A., Hoffelner, H. & Overmann, J. Specific Detection, Isolation, and Characterization of Selected, Previously Uncultured Members of the Freshwater Bacterioplankton Community. *Appl Environ Microbiol* **71**, 5908–5919 (2005).
50. Janssen, P. H., Yates, P. S., Grinton, B. E., Taylor, P. M. & Sait, M. Improved Culturability of Soil Bacteria and Isolation in Pure Culture of Novel Members of the Divisions *Acidobacteria* , *Actinobacteria* , *Proteobacteria* , and *Verrucomicrobia*. *Appl Environ Microbiol* **68**, 2391–2396 (2002).
51. Kaeberlein, T., Lewis, K. & Epstein, S. S. Isolating 'Uncultivable' Microorganisms in Pure Culture in a Simulated Natural Environment. *Science* **296**, 1127–1129 (2002).
52. Könneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).
53. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633 (2002).
54. Schut, F. *et al.* Isolation of Typical Marine Bacteria by Dilution Culture: Growth, Maintenance, and Characteristics of Isolates under Laboratory Conditions. *Appl Environ Microbiol* **59**, 2150–2160 (1993).
55. Mella-Flores, D. *et al.* Prochlorococcus and Synechococcus have Evolved Different Adaptive Mechanisms to Cope with Light and UV Stress. *Front. Microbio.* **3**, (2012).
56. Orr, P. T. & Jones, G. J. Relationship between microcystin production and cell division rates in nitrogen-limited Microcystis aeruginosa cultures. *Limnology & Oceanography* **43**, 1604–1614 (1998).

57. Sohm, J. A., Edwards, B. R., Wilson, B. G. & Webb, E. A. Constitutive Extracellular Polysaccharide (EPS) Production by Specific Isolates of Crocosphaera watsonii. *Front. Microbio.* **2**, (2011).
58. Fischer, F., Zillig, W., Stetter, K. O. & Schreiber, G. Chemolithoautotrophic metabolism of anaerobic extremely thermophilic archaebacteria. *Nature* **301**, 511–513 (1983).
59. Leschine, S. B., Holwell, K. & Canale-Parola, E. Nitrogen Fixation by Anaerobic Cellulolytic Bacteria. *Science* **242**, 1157–1159 (1988).
60. Zehnder, A. J. & Brock, T. D. Methane formation and methane oxidation by methanogenic bacteria. *J Bacteriol* **137**, 420–432 (1979).
61. Zhang, T., Shi, X.-C., Ding, R., Xu, K. & Tremblay, P.-L. The hidden chemolithoautotrophic metabolism of *Geobacter sulfurreducens* uncovered by adaptation to formate. *The ISME Journal* **14**, 2078–2089 (2020).
62. Belzer, C. *et al.* Microbial Metabolic Networks at the Mucus Layer Lead to Diet-Independent Butyrate and Vitamin B_{12} Production by Intestinal Symbionts. *mBio* **8**, e00770-17 (2017).
63. Kazamia, E. *et al.* Mutualistic interactions between vitamin B_{12} -dependent algae and heterotrophic bacteria exhibit regulation. *Environmental Microbiology* **14**, 1466–1476 (2012).
64. Chodkowski, J. L. & Shade, A. Bioactive exometabolites drive maintenance competition in simple bacterial communities. *mSystems* e00064-24 (2024) doi:10.1128/msystems.00064-24.
65. Harrison, F., Paul, J., Massey, R. C. & Buckling, A. Interspecific competition and siderophore-mediated cooperation in *Pseudomonas aeruginosa*. *The ISME Journal* **2**, 49–55 (2008).
66. Rao, D., Webb, J. S. & Kjelleberg, S. Competitive Interactions in Mixed-Species Biofilms Containing the Marine Bacterium *Pseudoalteromonas tunicata*. *Appl Environ Microbiol* **71**, 1729–1736 (2005).
67. Han, K. *et al.* Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. *Sci Rep* **3**, 2101 (2013).
68. Bentkowski, P., Van Oosterhout, C. & Mock, T. A Model of Genome Size Evolution for Prokaryotes in Stable and Fluctuating Environments. *Genome Biol Evol* **7**, 2344–2351 (2015).

69. Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J. & Crosby, L. Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *mSystems* **3**, e00055-18 (2018).
70. Kim, S., Kang, I., Seo, J.-H. & Cho, J.-C. Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical 'helper' catalase. *ISME J* **13**, 2252–2263 (2019).
71. Bialevich, V., Zachleder, V. & Bišová, K. The Effect of Variable Light Source and Light Intensity on the Growth of Three Algal Species. *Cells* **11**, 1293 (2022).
72. Jiang, L. & Morin, P. J. Temperature fluctuation facilitates coexistence of competing species in experimental microbial communities. *Journal of Animal Ecology* **76**, 660–668 (2007).
73. Johnson, W. M. *et al.* Auxotrophic interactions: a stabilizing attribute of aquatic microbial communities? *FEMS Microbiology Ecology* **96**, fiaa115 (2020).
74. Morris, J. J., Lenski, R. E. & Zinser, E. R. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. *mBio* **3**, e00036-12 (2012).
75. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J* **8**, 1553–1565 (2014).
76. Sokolovskaya, O. M., Shelton, A. N. & Taga, M. E. Sharing vitamins: Cobamides unveil microbial interactions. *Science* **369**, (2020).
77. Kost, C., Patil, K. R., Friedman, J., Garcia, S. L. & Ralser, M. Metabolic exchanges are ubiquitous in natural microbial communities. *Nat Microbiol* **8**, 2244–2252 (2023).
78. Zelezniak, A. *et al.* Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6449–6454 (2015).
79. Zengler, K. & Zaramela, L. S. The social network of microorganisms — how auxotrophies shape complex communities. *Nat Rev Microbiol* **16**, 383–390 (2018).
80. Jurdzinski, K. T. *et al.* Large-scale phylogenomics of aquatic bacteria reveal molecular mechanisms for adaptation to salinity. *Sci. Adv.* **9**, eadg2059 (2023).
81. Hugerth, L. W. *et al.* Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol* **16**, 279 (2015).
82. Garcia, S. L., McMahon, K. D., Grossart, H.-P. & Warnecke, F. Successful enrichment of the ubiquitous freshwater acI Actinobacteria. *Environmental Microbiology Reports* **6**, 21–27 (2014).
83. Blanco, L. *et al.* Highly Efficient DNA Synthesis by the Phage ϕ 29 DNA Polymerase. *Journal of Biological Chemistry* **264**, 8935–8940 (1989).

802 84. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
803 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
804 85. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for
805 genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
806 86. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-
807 node solution for large and complex metagenomics assembly via succinct *de Bruijn*
808 graph. *Bioinformatics* **31**, 1674–1676 (2015).
809 87. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications
810 to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).
811 88. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for
812 genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
813 89. Mattock, J. & Watson, M. A comparison of single-coverage and multi-coverage
814 metagenomic binning reveals extensive hidden contamination. *Nat Methods* **20**, 1170–
815 1173 (2023).
816 90. Sjöqvist, C., Delgado, L. F., Alneberg, J. & Andersson, A. F. Ecologically coherent
817 population structure of uncultivated bacterioplankton. *The ISME Journal* **15**, 3034–3049
818 (2021).
819 91. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
820 *Methods* **9**, 357–359 (2012).
821 92. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
822 2078–2079 (2009).
823 93. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient
824 genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
825 94. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning
826 algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**,
827 605–607 (2016).
828 95. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat*
829 *Methods* **11**, 1144–1146 (2014).
830 96. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
831 assessing the quality of microbial genomes recovered from isolates, single cells, and
832 metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
833 97. Buck, M., Mehrshad, M. & Bertilsson, S. mOTUpan: a robust Bayesian approach to
834 leverage metagenome-assembled genomes for core-genome estimation. *NAR Genomics*
835 *and Bioinformatics* **4**, lqac060 (2022).

98. Sahlin, K. Strobealign: flexible seed size enables ultra-fast and accurate read alignment. *Genome Biol* **23**, 260 (2022).
99. Eren, A. M. *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
100. Eren, A. M. *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* **6**, 3–6 (2020).
101. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
102. Veseli, I. *et al.* Microbes with higher metabolic independence are enriched in human gut microbiomes under stress. Preprint at https://doi.org/10.7554/eLife.89862.2 (2024).
103. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2024).
104. RStudio Team. RStudio: Integrated Development for R. (2024).
105. Royston, J. P. An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. *Applied Statistics* **31**, 115 (1982).
106. Bauer, D. F. Constructing Confidence Sets Using Rank Statistics. *Journal of the American Statistical Association* **67**, 687–690 (1972).
107. Kruskal, W. H. & Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* **47**, 583–621 (1952).
108. Dunn, O. J. Multiple Comparisons Using Rank Sums. *Technometrics* **6**, 241–252 (1964).