

## GENETICS

# Public good exploitation in natural bacterioplankton communities

Shaul Pollak<sup>1†</sup>, Matti Gralka<sup>1†</sup>, Yuya Sato<sup>1,2</sup>, Julia Schwartzman<sup>1</sup>, Lu Lu<sup>1</sup>, Otto X. Cordero<sup>1\*</sup>

Bacteria often interact with their environment through extracellular molecules that increase access to limiting resources. These secretions can act as public goods, creating incentives for exploiters to invade and “steal” public goods away from producers. This phenomenon has been studied extensively *in vitro*, but little is known about the occurrence and impact of public good exploiters in the environment. Here, we develop a genomic approach to systematically identify bacteria that can exploit public goods produced during the degradation of polysaccharides. Focusing on chitin, a highly abundant marine biopolymer, we show that public good exploiters are active in natural chitin degrading microbial communities, invading early during colonization, and potentially hindering degradation. In contrast to *in vitro* studies, we find that exploiters and degraders belong to distant lineages, facilitating their coexistence. Our approach opens novel avenues to use the wealth of genomic data available to infer ecological roles and interactions among microbes.

## INTRODUCTION

To access insoluble nutrients in their environment, microorganisms frequently need to invest in costly secretions such as iron-scavenging molecules and hydrolytic enzymes, to name a few (1, 2). This form of extracellular metabolism opens a niche for public good exploiters, or “cheaters,” that do not invest in those secretions but reap their benefits. Although many studies have investigated the *in vitro* dynamics of otherwise clonal producers and exploiters, the environmental relevance of public good exploitation in a complex environment such as the ocean remains poorly constrained. This is largely due to the difficulty of predicting microbial interactions in natural communities. Here, we developed a novel approach to predict such interactions in communities that degrade complex forms of organic matter. The microbial degradation of complex organic matter is mediated by extracellular enzymes that break down biopolymers, releasing its fragments to the local environment. Polymer fragments (e.g., oligosaccharides) act as public goods, available as primary carbon and nitrogen sources for all members of the community and not just the enzyme producer. The prevalence of this process leads to the prediction that such communities should be riddled with public good exploiters, or cheaters, which compete directly with degraders by consuming the breakdown by-products but do not contribute to the pool of enzymes (3). However, community assembly experiments on complex carbon sources have revealed that, although secondary consumers (nondegraders) are numerically dominant, they do not necessarily grow on oligosaccharides and instead prefer to consume metabolic waste products, such as organic acids (4, 5). Therefore, it is unclear whether public good exploitation is relevant in natural biopolymer-degrading communities and, if so, how it would contribute to community dynamics and function.

Here, we address this problem in the context of the degradation of chitin, the second most abundant biopolymer on the planet. By detecting genes that are evolutionarily and/or physically linked to

chitinases (enzymes that break chitin polymers), we found that exploitation has evolved multiple times across different phyla, suggesting that exploitative lifestyles should be common in natural chitin-degrading communities. The study of the genomic signatures associated with chitinases helped us predict which organisms can act as chito-oligosaccharide exploiters, without relying on prior protein functional annotation. After validating our predictions, we showed that public good exploiters are present in naturally assembled communities in the ocean and that their dynamics mirror those of degraders. Last, we showed that exploiters sampled from natural communities can also stably coexist in the laboratory with degraders and waste product scavengers, showing that even in these simplified conditions, exploiters do not cause the “tragedy of the commons,” nor are they fully suppressed.

## RESULTS

### Predicting exploitation potential from genomes

In theory, detecting a public good exploiter from its genome should be straightforward: An exploiter should lack the genes for public good production (extracellular chitinases, in our case) but have the genetic machinery that allows it to compete for uptake and utilization of the public good against producers. However, as we will show below, predicting which organisms can grow on chitin oligosaccharides in a community is not a simple task. The reason is that competition for oligosaccharide uptake in a community is a complex trait involving multiple processes encoded by different genes (6). These genes need not be in the canonical pathway for chitin utilization, as they can mediate surface attachment, biofilm formation, and chemotactic behavior, among many other phenotypes. Moreover, given the limitations of gene functional annotations, especially in poorly studied taxa, many of the ecologically relevant genes could lack predicted functions. With this in mind, we first aimed to identify a set of genes that could be used as predictors of the ability of an organism to compete for chitin degradation by-products in a manner that is independent of functional annotations. We reasoned that if a gene increases the capacity of an organism to grow on chitin oligosaccharides, it should be “linked” to chitinases, either by sharing a similar evolutionary history or by being collocated in the chromosome (7, 8).

Copyright © 2021  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>2</sup>Environmental Management Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 16-1 Onogawa, Tsukuba, Ibaraki 305-8569, Japan.

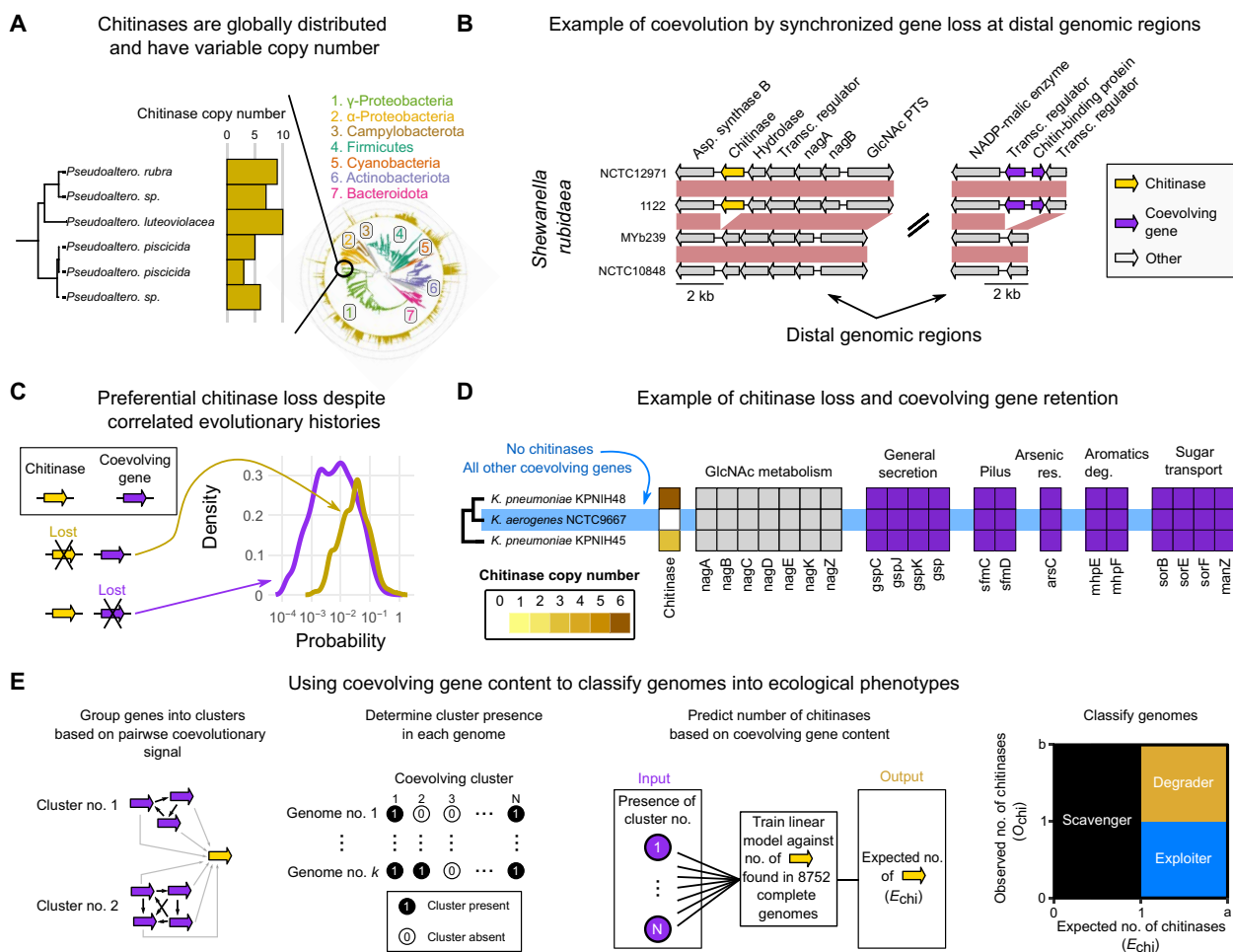
\*Corresponding author. Email: ottox@mit.edu

†These authors contributed equally to this work.

We developed a computational approach to detect genes that coevolve or are colocalized with chitinases across 8752 nonredundant complete bacterial genomes spanning the tree of bacterial life (table S1 and data S1 and S2). The genomes belong to both marine- and non-marine-derived isolates. To infer coevolution, we took advantage of the fact that the number of chitinases in a genome can change drastically even among close relatives, ranging from 0 to 15 chitinases (Fig. 1A), implying frequent gene gain and loss events. By reconstructing the history of gene gain and loss for 3,237,392 gene families across the tree of bacterial life and contrasting these histories against a null model of gene content coevolution (Materials and Methods), we found 2097 gene families that coevolve with chitinases in bacteria (Fig. 1B). In addition, we identified 1479 genes that colocalize together with chitinases in the genome, for a total of 3576 candidate predictor genes (data S3). In addition to canonical pathway genes, these genes

had diverse Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations, such as chemotaxis and motility, attachment and biofilm formation, and the synthesis of antibiotics, but were most enriched for genes of unknown function and mobile element-related functions, highlighting the need for an annotation-agnostic approach (figs. S1 and S2 and data S4) (9). Chitinase-linked genes with nearly identical presence-absence distributions across our species set (e.g., two proteins in the same operon) were further clustered into 1905 “accessory clusters,” containing either individual gene families or small gene family clusters.

Although, by definition, the histories of gain and loss of chitinases and accessory clusters were correlated, chitinases had a higher loss rate than their coevolving genes (means odds ratio = 1.4,  $t$  test  $P < 1 \times 10^6$ ,  $t = 4.93$ ,  $df = 5085.7$ ; Fig. 1C and fig. S3). This means that, through evolution, there were multiple events in which chitinases were lost but accessory clusters were retained, likely leading to the evolution



**Fig. 1. Large-scale detection of genes that coevolve with chitinases allows gene content-based genome classification.** (A) Chitinase copy number in a group of *Pseudoalteromonas* strains. Inset shows the GTDB-tk-derived phylogenetic tree of 8752 genetically distinct (ANI < 99.9%) closed genomes used in this study, with chitinase copy number indicated on each tip (gold bars). Major phylogenetic groups are highlighted on the tree. (B) An example of coupled gain/loss of a chitinase and two distally located coevolving genes in a group of closely related *Shewanella rubidaea* strains. (C) Conditional probability density of losing a chitinase while retaining a coevolving gene (gold) and vice versa (purple). The probabilities are calculated for each coevolving gene, and the distribution is over all coevolving genes. (D) A group of *Klebsiella pneumoniae* strains, where strain NCTC9667 lost all chitinases while retaining coevolving genes. Nag genes do not coevolve with chitinases in this group but are shown to highlight pathway completeness. (E) Genome classification from coevolving gene content. Coevolving genes are grouped into clusters based on their patterns of coevolution with each other. Each genome is represented by a bit-string that marks the presence and absence of each cluster. A linear model is trained on all fully sequenced genomes to predict chitinase copy number, given their coevolving cluster content. New genomes are classified into ecological phenotypes based on the (un) coupling between observed ( $O_{chi}$ ) and expected ( $E_{chi}$ ) number of chitinases.

of an oligosaccharide (public good) exploiter. An example is shown in Fig. 1D, where a *Klebsiella pneumoniae* strain lost its chitinases, while retaining the genes required for amino-sugar transport and metabolism (*nagABCDEKZ*) (10), along with attachment and secretion pili (*sfmCD*, *gsp/gspCJK*) (11), as well as other genes involved in arsenic resistance (*arsC*) (12), 3-(3-hydroxyphenyl)propionate (3-HPP) catabolism (*mhpEF*) (13), and several sugar transport proteins (*sorBEF*, *manZ*) (14, 15), all of which coevolve with chitinases. The retention of these genes may be explained by the fact that they are involved in multiple functions, not just chitin degradation. However, independent of the selective pressure that keeps them in the genome, the result of chitinase losses and accessory cluster retention is the evolution of an organism that carries all the genomic machinery needed to compete for chitin oligosaccharides but that does not produce hydrolytic enzymes.

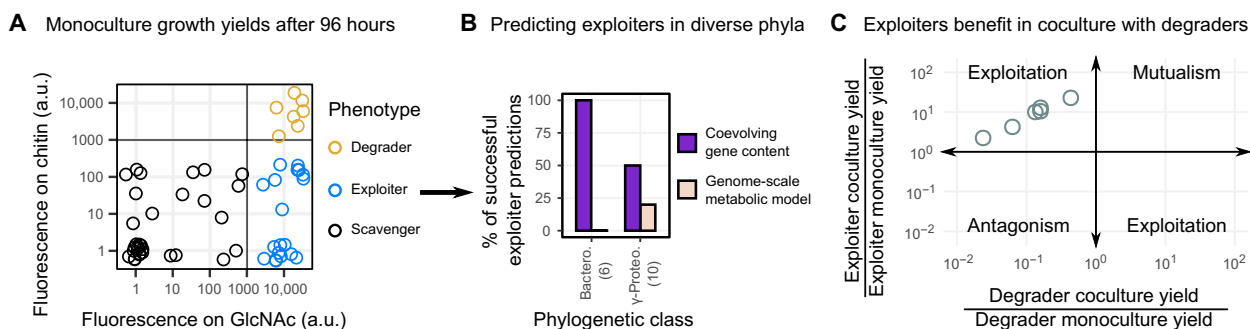
Motivated by these observations, we devised an approach to systematically differentiate degraders, exploiters, and waste product consumers (“scavengers”), solely on the basis of genomic data. The large list of accessory clusters found to coevolve with chitinases provided us with a set of possible variables to predict an organism’s phenotype, in particular its ability to grow on chitooligosaccharides. Assuming that the number of hydrolases in a genome should correlate with the ability of an organism to grow on the corresponding soluble sugars (16), we trained an elastic-net regression model (Materials and Methods) to predict the number of chitinases in a genome based on its accessory cluster content. The model was trained separately for each phylum (Materials and Methods and table S2) and out of the 1905 coevolving clusters; on average, ~190 clusters were sufficient to predict the number of chitinases with high accuracy (average fivefold cross-validated  $R^2 = 0.84$ ; figs. S4 and S5). Interpreting the expected number of chitinases predicted by the model ( $E_{\text{chi}}$ ) as a measure of the potential of the organism to compete for chitooligosaccharides, we classified genomes as degraders, exploiters, or scavengers based on the contrast between  $E_{\text{chi}}$  and the actual number of chitinases observed in the genome ( $O_{\text{chi}}$ ). If the observed number of chitinases was zero and the expected number was greater than or equal to one ( $O_{\text{chi}} = 0$  and  $E_{\text{chi}} \geq 1$ ), we predicted the organism to be an exploiter, because it had the genetic machinery that typically accompanies

chitinases in genomes of degraders but lacked the hydrolytic enzyme. If  $O_{\text{chi}} \geq 1$  and  $E_{\text{chi}} \geq 1$ , we expected the organism to act as a degrader, and if  $E_{\text{chi}} < 1$ , we predicted that the organism was unable to grow on chitooligosaccharides and therefore must act as a waste scavenger if present in a chitin-degrading community (Fig. 1E).

### Coevolving gene content successfully predicts mono- and coculture phenotypes

Having trained our model, our next task was to validate its predictions in the context of a naturally assembled chitin-degrading community. To this end, we leveraged a collection of 57 isolates of copiotrophic marine bacteria, coisolated from the coastal ocean using paramagnetic model marine particles (4, 5). Although our model was not trained on the genomes of this set, we were still able to accurately predict their chitinase content ( $R^2 = 0.65$ ,  $P < 1 \times 10^{15}$ ,  $df = 63$ ). We predicted 16 degraders, 16 exploiters, and 33 scavengers in this collection. Notably, these functional groups were phylogenetically diverse, belonging to such diverse orders as *Flavobacteriales*, *Alteromonadales*, *Rhodobacterales*, and *Vibrionales* (fig. S6). We interpreted our ecological classification (degrader, exploiter, and scavenger) in terms of growth phenotypes that could first be tested using *in vitro* monocultures: Degraders should grow on chitin and its monomer GlcNAc, exploiters should grow on GlcNAc but not on chitin, and scavengers should grow on neither of the two substrates (Fig. 2A and fig. S7). We assessed our predictions against those made by genome-scale metabolic models generated using CarveMe, a state-of-the-art tool that derives organism-specific models based on a universal, manually curated set of reactions (17).

Overall, our model performed better than the annotation-based predictions, making ~30% fewer errors (16 versus 22 errors, Matthews correlation = 0.55 versus 0.34), confirming the value of our approach (fig. S8). Critically, virtually all of the improvement came from the ability of our model to correctly predict exploiters, which the genome-scale metabolic model misclassified as scavengers (Fig. 2B and table S3). In particular, we find that the gap-filled genome-scale metabolic model underperforms when predicting the phenotypes of two relatively understudied families, Flavobacteriaceae and Alteromonadaceae (Fig. 2B and fig. S9). This is consistent with



**Fig. 2. Evolution-guided classification predicts isolate phenotypes in mono- and coculture.** (A) Final yield of 57 marine isolates after 96 hours of growth on chitin and its monomer *N*-acetylglucosamine. Yield was measured as fluorescence emitted from incorporated SYBR Gold DNA dye (Materials and Methods). Each point represents the mean of at least three biological repeats. Points are colored according to their observed phenotype. a.u., arbitrary units. (B) Agreement between predictions made using different genome-informed methods and actual observed phenotypes. Phenotypically, exploiters are isolates that grow on GlcNAc (>1000 a.u.) but not on chitin (<1000 a.u.). (C) Cocultures between two degraders and three flavobacteria exploiters. Exploiters were misclassified by the genome-scale metabolic model but correctly classified by their coevolving gene content. Growth was conducted for 72 hours in minimal media with chitin as the sole carbon source. The x axis depicts the fold change in degrader yield during coculture compared to monoculture, and the y axis shows the same quantity for the exploiter. Each point represents a different coculture and represents the mean fold change from two independent replicates. Raw data for the cocultures are presented in table S4.

the notion that models that rely on functional annotations should perform poorly on organisms that are distant from the few that have been experimentally characterized, such as *Escherichia coli*, *Bacillus subtilis*, or *Vibrio cholerae*. In contrast, our approach only requires sampling enough genomic diversity and a single annotation (in this case, chitinases) and can therefore be applied to any uncultured and poorly annotated organism.

Ultimately, our predictions make a statement about interactions between cocultured community members. Plasmid-encoded fluorescent reporters allowed us to study the dynamics of a *vibrio1A01* (*Vibrio*) and *alteroA3R04* (*Alteromonas*) coculture in detail, when these two strains were grown with chitin as the sole carbon source. This experiment showed that the exploiter (*alteroA3R04*) was only able to grow in the presence of the degrader, at the cost of the degrader's final yield but not its maximal growth rate (figs. S10 to S12). This result prompted us to examine more pairs to confirm our ability to predict exploitative interactions based on the final yields of the cocultured members. To this end, we cocultured several degrader-exploiter pairs in minimal media containing chitin as the sole carbon source for 72 hours. We used two different degraders and three isolates that were predicted by our method to be exploiters but were classified as scavengers by the genome-scale metabolic model. We found that, as predicted, in all cases, the interaction was exploitative: Putative exploiters produced, on average, ~10-fold more colony-forming units (CFUs) compared to growth in monoculture. Degraders, on the other hand, suffered from the presence of exploiters, decreasing their yield by a factor of ~6 on average (Fig. 2C). We found that in those cases where degraders were suppressed by cheating, the exploiter did not reach a high yield and vice versa. This trade-off between exploitation and degrader yield can be recapitulated with simple consumer-resource models that include public good exploitation (fig. S13), lending further support to the notion that our method is able to successfully predict GlcNAc consumers (i.e., exploiters) where annotation-based methods fail.

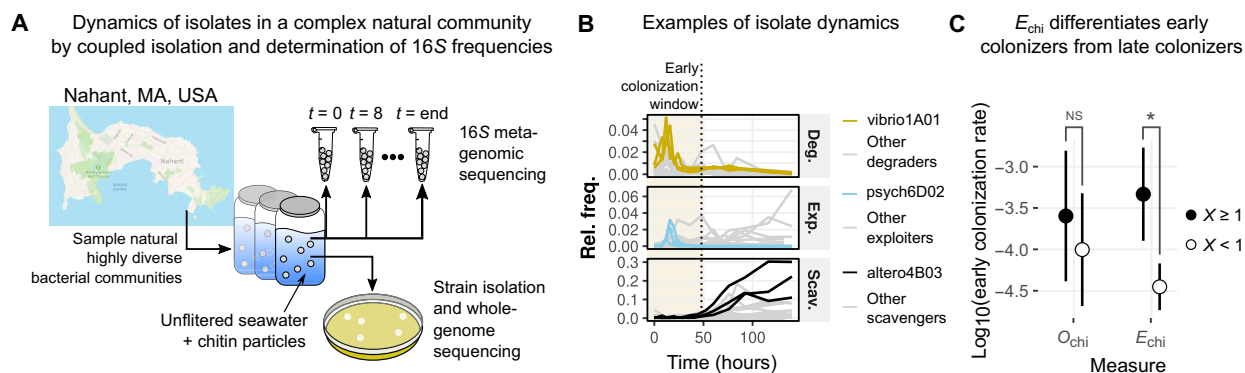
### Dynamics of exploiters in the environment

Our in vitro results do not guarantee that our putative exploiters can invade natural chitin-degrading communities in a manner consistent

with their purported role. To address this question, we studied the population dynamics of our isolates in their native seawater community. We mapped the isolates' 16S ribosomal RNA (rRNA) sequences to the time-series data of the original enrichment experiment, in which the community assembled from seawater onto chitin particles for a period of 244 hours (Fig. 3A) (4, 5). This allowed us to analyze the colonization dynamics of each mapped isolate across the different ecological roles (Fig. 3B). We found that the early colonization dynamics of exploiters were more similar to that of degraders than to that of scavengers, consistent with their role as parasites of degraders. Exploiters reached maximum frequency in the succession shortly after degraders (12 to 16 hours for exploiters and 8 hours for degraders), while scavengers peaked at late successional stages (>80 hours) (fig. S14). We find that some exploiters are able to increase in frequency at both early and late stages, implying that they may have a dual ecological role (fig. S14). Together, these observations imply that the accessory gene clusters, containing the genetic machinery to chemotax toward GlcNAc, adhere to chitin surfaces, etc. are the main determinant of early colonization and not chitinases themselves. In other words, our response variable,  $E_{\text{chi}}$ , should be a better predictor of early colonization dynamics in a wild community than chitinase copy number ( $O_{\text{chi}}$ ). To test this assertion, we calculated the maximal speed of colonization of each isolate during the early phases of assembly, before the community became dominated by scavengers (Materials and Methods and Fig. 3C). As expected, we found that  $E_{\text{chi}}$  was a better predictor of colonization speed than the observed number of chitinases,  $O_{\text{chi}}$ . Genomes predicted to contain chitinases had a 19-fold higher colonization rate on average compared to those predicted to lack chitinases ( $t$  test  $P = 0.02$ ,  $t = 3$ ,  $df = 8.03$ ). In contrast, genomes encoding chitinases did not display a significant difference in colonization rate compared to those that lacked chitinases ( $t$  test  $P = 0.37$ ,  $t = 0.94$ ,  $df = 7.8$ ).

### Exploiters, degraders, and waste scavengers coexist in synthetic communities

In the natural chitin-degrading communities, the dynamics of colonization and growth are successional, meaning stable coexistence cannot be attained on a single chitin particle. Instead, degraders,



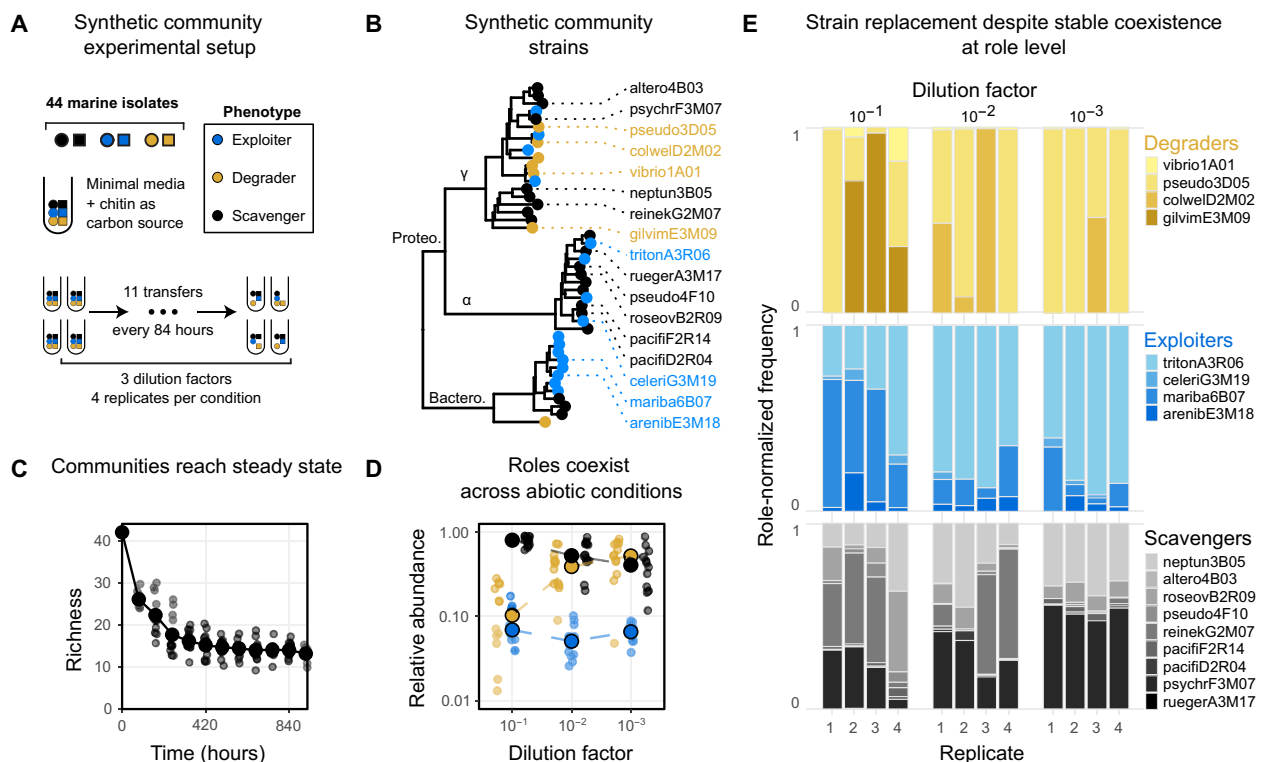
**Fig. 3. Exploiter dynamics mimic those of degraders in complex natural communities.** (A) Natural, highly diverse bacterial communities were sampled as seawater samples taken from Nahant, MA, USA, and incubated with chitin as the sole carbon source. Samples from the incubation were taken periodically for DNA extraction and isolate collection. (B) Colonization dynamics of 16S sequences that perfectly matched the 16S sequences of isolated strains. Isolates were assigned an ecological role according to their chitinase-linked gene content. Each panel shows the trajectories of a different ecological role, and each line represents the trajectory of a single 16S sequence in a single repeat. One representative strain from each role is highlighted as indicated by the legend. The early colonization window was defined as in (4). For a heatmap of each strain through time, see fig. S14. (C) Grouping strains by the expected number of chitinases ( $E_{\text{chi}}$ ), and not by their observed number ( $O_{\text{chi}}$ ), results in a clean separation between early colonizers (degraders and exploiters) and late colonizers (scavengers). \* $P < 0.05$ ; not significant (NS) =  $P > 0.3$ .

exploiters, and scavengers could stably coexist at the larger scale of many particles (the metapopulation). To test whether the degraders and exploiters could coexist in a closed system with chitin particles as the sole carbon source or whether their overlap in resource preference would lead to an eventual community collapse, we assembled synthetic communities with 44 strains sampled from our isolate collection and from the three different roles. The strains were chosen based on 16S sequence distance [ $>5$  single nucleotide polymorphisms (SNPs) between any pair] and robust preculture growth. To assess coexistence, we serially passaged the coculture over 11 dilution-growth cycles (Fig. 4, A and B, and data S5) and repeated these experiments with three different dilution factors (which determine the minimum growth rate required to survive serial passages) to assess the robustness of coexistence to different growth conditions (Materials and Methods).

Following a short transient where species are rapidly purged, community richness stabilized on 10 to 15 members (Fig. 4C), with coexistence of all three roles across all dilution factors (Fig. 4D). Over the final five transfers, after community richness had equilibrated, exploiters were stably maintained but at low abundance, recruiting roughly 10% of all reads. Scavengers, by contrast, were the most abundant ecological group (20 to 80%), supporting previous findings of widespread export of metabolic waste by exploiters and degraders

(4, 5). Scavenger and degrader abundances were controlled by the dilution factor, with degraders increasing in relative abundance with increasing dilution strength (Fig. 4D). This trend can be interpreted as a consequence of the lower time-averaged growth rates of scavengers, which have to wait for the accumulation of secreted metabolites to grow.

It is evident that replicates can be variable in terms of the frequencies of the different ecological roles (Fig. 4D). Closer inspection of the community compositions across dilution factors and replicates revealed different patterns of diversity within ecological roles. In most communities, degraders were dominated by a single strain, *pseudo3D05* (*Pseudoalteromonas*), and two exploiters, *mariba6B07* (*Maribacter*) and *tritonA3R06* (*Tritonibacter*), while four to five scavengers coexisted in the same community. However, there were clear cases in which other strains became dominant: *pseudo3D05* coexisted with other degraders at appreciable abundance in at least 4 of the 12 communities and was altogether replaced by *colwelD2M02* (*Colwellia*) in one replicate and *gilvimE3M09* (*Gilvmarinus*) in another, while one of the two exploiters seemed to reach higher relative abundances at higher dilution factors. The abundance of different scavengers varied across replicates in a seemingly stochastic manner, especially at low dilution factors (Fig. 4E). These compositional changes within the different roles seemed to be independent of each other; i.e., the different strains were substituted with no apparent



**Fig. 4. Synthetic communities display stable coexistence of ecological roles, with stochastic strain replacement.** (A) Synthetic communities were assembled from 44 isolates that represent different phylogenetic groups and ecological types (data S5). The initial composition was identical for all communities, as they were sampled from the same initial mix. Communities were allowed to grow for 84 hours and were transferred 11 times. We performed three different dilution factors ( $10^{-1}$ ,  $10^{-2}$ , and  $10^{-3}$ ), with four repeats per condition. (B) GTDB-tk generated phylogenetic tree of the isolates used in this experiment. Isolate phenotypes are indicated by colored circles at the tip of the tree. Isolates that survived to the end of the experiment in at least one condition are indicated by name. (C) The number of species present in a community (richness) as a function of time (measured in number of growth-dilution cycles). (D) Steady-state relative abundance (transfer  $>8$ ) of each ecological role in the synthetic communities. Each point represents the total role abundance from a given sample at a specific time point. Large circles represent the medians for each role as a function of the dilution factor. Colors represent ecological roles as in (A). (E) Each bar represents the role-normalized abundance of all isolates belonging to a certain ecological role at the final time point. After normalization, abundances of strains with the same role add up to 1.

consequence for the rest of the community. The fact that the roles stably coexisted despite variation in community composition suggests that degraders, exploiters, and scavengers represent three stable metabolic niches in polysaccharide-degrading communities.

## DISCUSSION

Here, we developed a novel computational approach that allowed us to predict phenotype from genome, in particular the ability of an organism to act as a chitoligosaccharide consumer. A key feature of our method is that it is agnostic to functional annotations and instead builds on evolutionary patterns that can be inferred directly from genomic data. The fact that genome repositories continue to grow at an accelerated pace, whereas functional annotations remain limited to what can be learned from a few model organisms, underscores the relevance and timeliness of our approach. Ancestral genome reconstructions have been used for many years to infer gene-gene coevolution, with the goal to expand and guide the discovery of protein-protein interactions in microorganisms (18). However, the work presented in this paper is the first to leverage these methods to infer and validate the growth phenotypes of microbes and their ecology in complex communities.

Using this approach, we showed that oligosaccharide exploiters, which share many genomic features with degraders but do not encode the relevant hydrolytic enzymes, are abundant in natural communities and have the potential to hinder degradation. The frequent losses of chitinases during the evolution of bacteria are likely the result of short-term evolutionary incentives created by the release of public goods during extracellular hydrolysis. For degraders, in turn, the invasion of exploiters can create incentives to privatize the public good, via spatial aggregation or tethering of enzymes to the membrane (19). Recent studies have also shown that some *Flavobacteria* are capable of a “selfish” mode of uptake, whereby long oligomers are directly incorporated inside the cell, bypassing the need for extracellular hydrolysis (20). The relative advantage of one strategy over the other and the conditions where each may be more relevant remain unknown. However, what is clear is that the constant tension between degraders and exploiters seems to have shaped the evolution of polysaccharide degradation strategies among bacteria, highlighting the fact that microbial interactions and the evolutionary dynamics they generate can affect key ecosystem processes, such as the degradation of organic matter.

Although consistent with the notion of a social cheater, there is an important difference between the exploiters we identified and cheaters, as typically conceived in the social evolution literature: Cheaters are loss-of-function mutants that rise in frequency due to a short-term fitness advantage over a wild-type cooperator phenotype (21). Natural exploiters of polysaccharide degraders, by contrast, invade communities by a process of community assembly. Consequently, exploiters in naturally assembled chitin-degrading communities are distantly related to degraders and have diverse metabolic potentials. This increases the possibilities for coexistence, since organisms belonging to these two roles can differ along multiple phenotypic dimensions, but it also makes it harder to predict interspecific interactions. Our in vitro studies indicate that, under simple batch culture conditions, public good exploiters hinder the growth of degraders, potentially slowing down the turnover of organic matter. However, further work is needed to elucidate the impact of this common ecological interaction in more

realistic conditions, such as under fluid flow and multiple nutrient limitations.

## MATERIALS AND METHODS

### Genome databases and functional annotation (Fig. 1 and fig. S1)

Genome bundles were downloaded from the National Center for Biotechnology Information (NCBI) Refseq FTP website on 13 July 2019 (22). Only the latest versions of complete genomes were downloaded, resulting in 13,737 genomes. We then removed redundant genomes from our dataset by clustering genomes that had >99.9% average nucleotide identity (ANI) over 90% of their genome (23), keeping the longest genome from each cluster, resulting in 8753 non-redundant genomes (24, 25). The phylogenetic tree, as well as the classification of each genome, was obtained using the Genome Taxonomy Database Toolkit (GTDB-tk) (26). For all downstream analyses, a concatenated file was created where only coding sequences and their corresponding translated proteins were retained, with pseudogenes (as annotated by NCBI) discarded.

Chitinases (EC.3.2.1.14) were identified using a slightly modified dbCAN annotation workflow (27). The modified workflow consisted of identifying proteins that significantly aligned to the dbCAN hidden Markov models ( $e$  value  $<10^{-10}$ ) and contained conserved  $k$ -mers, as identified by the dbCAN hotpep program [with hits-cutoff set to 1 and freq cutoff set to 0.2; these parameters were found to be the best performing on bacterial genomes (27)]. Last, hits were filtered to only contain GH18 and GH19, as these are the most prevalent and specific chitinase domains found in the CAZy database.

eggno-mapper v2 was used to functionally annotate proteins, with parameters `--go_evidence non-electronic --target_orthologs all --seed_ortholog_eval 0.001 --seed_ortholog_score 60` (28). For KEGG annotation of coevolving genes (fig. S1), unknown function genes were genes that either had no KEGG orthology hit and no description or were not found in the eggno output file (had no hits to seed orthologs). Mobile genes were genes that had one of the following terms in their description: “phage,” “tail,” “capsid,” “transposon,” “transposase,” “conjugation,” or “insertion element.” Other functions were extracted from the KEGG pathway annotations of genes. The final distribution of functions of coevolving genes was calculated as follows: the probability of finding an annotation in a given gene was calculated, and then the sum of each annotation across all genes is the final number reported in data S3.

### Protein-family determination and ancestral-state reconstruction (Fig. 1)

All 49,638,395 genes from the 13,373 complete genomes were clustered into gene families based on homology of their protein products using `mmseqs2`, with command-line options `--cov-mode 0 --min-seq-id 0.5 -c 0.8 -s 7 --max-seqs 300 --cluster-mode 1` (29). This means that to be clustered, genes must be at least 50% identical over 80% of the length of the longer protein product. After clustering, which resulted in 3,237,392 families, we determined the copy number of each gene family in each genome. To determine gain/loss events, we used the “`asr_squared_change_parsimony`” function in the R `Castor` package to reconstruct the ancestral copy numbers of each gene family using a maximum-likelihood approach, given the extent of the genome’s copy number and the GTDB-tk-generated species tree (30). Changes were then calculated for each edge in the

tree and every gene family as the difference between the ancestral copy number and the descending node's copy number (rounded to the nearest integer).

The conditional probability distributions shown in Fig. 1D were computed as follows: For each of the 3576 linked genes, we found the most recent common ancestor (MRCA) of all genomes containing that gene. The MRCA was used to extract a subtree descending from that MRCA node, over which the analysis for each gene was carried out. The probability of losing chitinases while retaining the linked gene was the ratio between (i) the number of edges over which the chitinase copy number went from some positive number to zero, while the copy number of the linked gene remained positive (e.g., the parent node had one chitinase and one copy of the coevolving gene, while the child node had zero chitinases and one copy of the coevolving gene), and (ii) the number of edges over which the copy number of the linked gene remained positive. The second conditional distribution was computed similarly, with the categories switched.

### Detection of chitinase-linked genes (Fig. 1)

Coevolution was defined as similarities in gene gain/loss events (8, 31). Similarities in evolutionary gain/loss events [after correcting for changes in total genome sizes, as done in (8)] were computed using the cosine similarity metric. As high similarity could be due to random chance if the observed number of changes is too small, we only considered correlations between chitinase and genes that had at least three changes on the phylogeny and were found in at least five genomes (31). This reduced the number of genes analyzed to 270,461 (1,889,229 gene families were only found in a single genome). We used a randomization test to assess the significance of the observed similarities with chitinase. For each gene, we identified the subtree in which it exists and verified that chitinase also has >2 changes in that subtree. We then generated  $10^4$  random vectors from the entries of the change matrix (edge  $\times$  gene) in that specific subtree, by randomly permuting the entries of each edge across different genes, while only sampling from genes that had >2 changes in that subtree. This process was performed independently for each edge in the subtree. The  $P$  value of the correlation was the fraction of random similarities that showed a higher similarity than the one observed for the chitinase-gene pair. In cases where the distribution of a gene was not monophyletic, the common ancestor of all genomes carrying the gene was used to define the subtree in which it exists.

To detect genes that colocalize frequently with chitinases, we first created a list of all protein sequences that are proximal to chitinases ( $\leq 5$  genes away on the chromosome). If a genome contained more than 1 chitinase, this procedure was performed for each chitinase in the genome. Each one of these proteins belonged to a gene family, as described above. This allowed us to determine all gene families that appear next to chitinases at least once. For each gene family, we calculated the probability,  $P$ , of being proximal to a chitinase as the number of genomes in which it is proximal to a chitinase, divided by the total number of genomes it is found in. Following (32), we set a relaxed threshold of  $P \leq 0.1$  to consider a gene as colocalized with chitinases.

Clusters of coevolving genes were first determined using the weighted minhash, using a low threshold (0.3) for similarity (33). Because of the low threshold and probabilistic nature of the minhash distance and resulting clustering, we then broke up clusters identified using the weighted minhash, as these clusters may contain illegitimately clustered vectors. This was done as described above

for detecting evolutionary histories that are similar to those of chitinases, while skipping the significance testing step. This resulted in a network of genes, where nodes in the network are genes and edges connect genes with similar histories. To detect communities in the graph, we used the Louvain clustering method as implemented in the R igraph package.

### Determination of $E_{\text{chi}}$ and isolate ecological role classification (Figs. 1 to 4)

An elastic-net linear model was trained on all nonredundant publicly available complete genomes, with chitin-linked gene cluster content as features, and actual chitinase counts as the response. The chitin-linked gene content of each genome was determined on the basis of the protein clustering performed in the previous segment. Each genome was represented by a binary vector, where each entry represented a cluster of coevolving genes. Clusters were marked as present if the genome contained more than 60% of genes belonging to that cluster. For a full sensitivity analysis for this threshold, see fig. S5. These vectors, together with a vector of associated chitinase copy numbers in each genome, were fed to an elastic-net regression, as implemented in the `cv.glmnet` function from the R `glmnet` package. The parameters used were `nfold = 5`, `alpha = 0.2`, `nlambda = 1 \times 10^3`. The model was trained on each phylum separately (except for proteobacteria, which were split into alpha- and gamma-proteobacteria) and on the entire combined dataset with fivefold cross-validation. The statistics of each model in terms of mean square error and cross-validated  $R^2$  are found in table S2.

The values of  $E_{\text{chi}}$  were rounded to the nearest integer and genomes were classified according to the following scheme: Genomes with  $O_{\text{chi}} \geq 1$  and  $E_{\text{chi}} \geq 1$  were classified as degraders. Genomes with  $O_{\text{chi}} = 0$  and  $E_{\text{chi}} \geq 1$  were classified as exploiters. Genomes with  $O_{\text{chi}} = 0$  and  $E_{\text{chi}} = 0$  were classified as scavengers. In addition, we reasoned that organisms that have chitinases but are not retaining their chitinase-associated genes were not ecological primary degraders of chitin, and so genomes with  $O_{\text{chi}} \geq 1$  and  $E_{\text{chi}} = 0$  were classified as scavengers. On the basis of the relative rate of chitinase loss and chitinase-associated gene retention, compared to the opposite scenario, we postulate that such genomes evolved from chitinase-containing ancestors, which lost their chitinases but retained many of their chitinase-associated genes.

### Genome-scale metabolic modeling (Fig. 2)

Genome-scale metabolic model construction was performed with CarveMe (17). Genes were predicted using prodigal (34), and functional annotation was performed in CarveMe with gap-filling on MBL media with pyruvate as the sole carbon source, as most isolates grew in that condition. Flux-balance analysis was performed using CobraPy with default parameters (35). Growth on GlcNAc was defined as a nonzero simulated biomass yield on MBL + 20 mM GlcNAc. Degraders were defined as genomes predicted to grow on GlcNAc and that have a chitinase, exploiters were predicted to grow on GlcNAc but have no chitinases, and scavengers were genomes predicted to not grow on GlcNAc.

### Growth media (Figs. 2 and 4)

Marine broth (MB) was purchased from Millipore-Sigma (catalog no. 76448). The minimal media used throughout (MBL) was prepared by mixing  $10^{-2}$  M  $\text{NH}_4\text{Cl}$ ,  $10^{-3}$  M  $\text{Na}_2\text{HPO}_4$ ,  $10^{-3}$  M  $\text{Na}_2\text{SO}_4$ , 0.05 M Hepes (pH 8.2), NaCl (20 g/liter),  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  (3 g/liter),

CaCl<sub>2</sub>\*2H<sub>2</sub>O (0.15 g/liter), and KCl (0.5 g/liter). The following trace metals and vitamin solutions were added at a final concentration, which is 1/1000 the indicated concentrations. Trace metals: FeSO<sub>4</sub>\*7H<sub>2</sub>O (2100 mg/liter), H<sub>3</sub>BO<sub>3</sub> (30 mg/liter), MnCl<sub>2</sub>\*4H<sub>2</sub>O (100 mg/liter), CoCl<sub>2</sub>\*6H<sub>2</sub>O (190 mg/liter), NiCl<sub>2</sub>\*6H<sub>2</sub>O (24 mg/liter), CuCl<sub>2</sub>\*2H<sub>2</sub>O (2 mg/liter), ZnSO<sub>4</sub>\*7H<sub>2</sub>O (144 mg/liter), Na<sub>2</sub>MoO<sub>4</sub>\*2H<sub>2</sub>O (36 mg/liter), NaVO<sub>3</sub> (25 mg/liter), NaWO<sub>4</sub>\*2H<sub>2</sub>O (25 mg/liter), and Na<sub>2</sub>SeO<sub>3</sub>\*5H<sub>2</sub>O (6 mg/liter). Vitamins [dissolved in 10 mM Mops (pH 7.2)]: riboflavin (100 mg/liter), D-biotin (30 mg/liter), thiamine hydrochloride (100 mg/liter), L-ascorbic acid (100 mg/liter), Ca-D-pantothenate (100 mg/liter), folate (100 mg/liter), nicotinate (100 mg/liter), 4-aminobenzoic acid (100 mg/liter), pyridoxine HCl (100 mg/liter), lipoic acid (100 mg/liter), NAD (100 mg/liter), thiamine pyrophosphate (100 mg/liter), and cyanocobalamin (10 mg/liter).

### Isolate phenotyping (Fig. 2A)

Strains were streaked from glycerol frozen stocks onto MB agar plates and allowed to grow for 48 hours at 25°C. Single colonies were inoculated into 14-ml polypropylene growth tubes (VWR no. 60818) containing 1 ml of MBL media supplemented with either 20 mM pyruvate or 20 mM glucose and grown for a maximal duration of 48 hours at 25°C with shaking. Whenever possible, cultures grown for 24 hours in MBL + 20 mM pyruvate were used. If a strain did not show visible growth, as measured by optical density, after 24 hours in MBL + 20 mM pyruvate, the MBL + 20 mM glucose was assayed for optical density. If the glucose culture also showed no visible growth, this strain's culture was allowed to grow for an additional 24 hours, and the cultures were checked in the same order again. After pregrowth, cells were washed three times in artificial seawater (Millipore-Sigma no. S9883) by centrifugation at 5000 rpm and diluted 1/20 into the final growth media. Assays were performed in (biological) triplicate in deep 96-well plates containing 1 ml of MBL media supplemented with either 20 mM GlcNAc or colloidal chitin (2 g/liter). Growth was determined using SYBR Gold staining. Samples (100 µl) were taken from cultures and allowed to incubate for 15 min in the dark with SYBR Gold, after which they were measured with a plate reader (Tecan Spark) using a SYBR Green filter set.

### Degrader-exploiter co-cultures (Fig. 2C)

Strains were pregrown as described for monocultures. Degrader-exploiter pairs were then mixed at 1:1 ratios, as determined by CFU counts. Cocultures were allowed to grow in MBL + colloidal chitin (2 g/liter) for 72 hours. After growth, CFU counts of each cocultured strain were determined by plating, using the fact that the exploiters used in these experiments form orange/yellow colonies, while degraders form white colonies.

### Consumer-resource model (Fig. 2C)

Cocultures were modeled using a simple consumer resource model with the following equations

$$\frac{d\text{Chitin}}{dt} = -\alpha * \text{Degrader} * \text{Chitin}$$

$$\frac{d\text{GlcNAc}}{dt} = \alpha * \text{Degrader} * \text{Chitin} - \text{GlcNAc} * (\gamma_{\text{deg}} * \text{Degrader} + \gamma_{\text{exp}} * \text{Exploiter})$$

$$\frac{d\text{Degrader}}{dt} = \text{Degrader} * (\gamma_{\text{deg}} * \text{GlcNAc} - \lambda)$$

$$\frac{d\text{Exploiter}}{dt} = \text{Exploiter} * (\gamma_{\text{exp}} * \text{GlcNAc} - \lambda)$$

with  $\alpha$  representing the hydrolysis rate,  $\gamma$  representing the uptake rate (assuming 100% efficiency of biomass conversion for simplicity), and  $\lambda$  representing a mortality rate. The model was solved numerically using the Julia package "DifferentialEquations" with default parameters. The model parameters were set to  $\alpha = 2 \times 10^{-4}$ ,  $\gamma_{\text{degrader}} = 10^{-2}$ , and  $\lambda = 10^{-2}$ . Initial conditions were  $\text{Chitin}_0 = 10^3$ ,  $\text{GlcNAc}_0 = 0$ ,  $\text{degrader}_0 = 10$ , and  $\text{exploiter}_0 = 10$ . The model was simulated for 150 time integration steps. Different exploiter advantage ratios ( $\gamma_{\text{exploiter}}/\gamma_{\text{degrader}}$ ) were tested and are plotted in fig. S12.

### Metagenomic 16S rRNA dynamics and early colonization rate calculation (Fig. 3)

The 16S rRNA sequence of each genome was determined by first amplifying it using the 27F (5-AGAGTTTGATCMTGGCTCAG-3) + 1492R (5-GGTTACCTTGTTACGACTT-3) universal bacterial primers (4, 5), followed by Sanger sequencing. Our strain collection was isolated from two experiments previously performed in the laboratory, which studied community succession on model particulate matter using natural seawater communities (4, 5). We mapped the 16S sequences of our isolates back to the metagenomic 16S trajectories (PRJNA319196 and PRJNA478695) using blast, only keeping perfect matches (36). Each sequence was only mapped to the experiment from which it was isolated.

To calculate the early colonization dynamics, read counts were transformed into relative frequencies and only entries with a relative abundance  $>10^{-3}$  were retained for downstream analysis. Each experiment was originally performed in triplicate, and we filtered the data for each isolate by only considering repeats that were coherent with each other, with a threshold of a Pearson correlation of 0.3 between repeats. The early colonization window was defined as the time window between 0 and 48 hours for data derived from (4) or 0 to 60 hours for data derived from (5), as the dynamics in this experiment were slower. The early colonization rate was defined as the average rate of increase in relative frequency between the lowest and highest points in the early colonization window. This rate was calculated for each repeat, and the average rate was returned for each isolate.

### Synthetic community experiment (Fig. 4)

For the data presented in Fig. 4, all isolates were pregrown in individual wells of a 96-well plate containing MB media for 48 hours. The individual precultures were then combined in equal volumes, washed in MBL salts, and inoculated into deep-well 96-well plates containing MBL with 40 mg of colloidal chitin as the sole carbon source. In each dilution cycle, communities were allowed to grow for 84 hours and then diluted into fresh media with the appropriate dilution factor (1/10, 1/100, or 1/1000). To determine community dynamics through time, DNA was extracted using the Beckman-Coulter DNAadvance Kit and sequenced using the EMP 16S amplicon protocol [using 515F (parada)-806R(apprill) primers] at the Environmental Sample Preparation and Sequencing Facility (ESPSF), which is located in the Argonne National Laboratory.

### Flat-chitin coculture experiments (fig. S8)

A predicted exploiter candidate, *alteroA3R04*, labeled with a plasmid-encoded enhanced green fluorescent protein (eGFP) (plasmid pLL014),

and a chitin degrader, *vibrio1A01*, labeled with a plasmid-encoded DsRed (plasmid pVSV208), were cocultured with a chitin sheet as the sole carbon source. The transparent chitin film was generated at the bottom of the well of a 96-well plate by adding 32  $\mu$ l of chitin solution (10 mg/ml) (dissolved in HFIP) and allowing the solvent to dry. The cells of the two strains were precultured in MBL + 20 mM GlcNAc at 25°C and diluted using MBL salts to an OD<sub>600</sub> (optical density at 600 nm) of 0.0625. The two strains were then mixed at the ratio of 1:9, 5:5, or 9:1. Mixed cells (150  $\mu$ l) were inoculated onto the wells of the chitin sheet-coated 96-well plate and incubated at 25°C for more than 80 hours. The growth profiles of the exploiter and the degrader were evaluated by measuring fluorescence intensity and OD<sub>600</sub> at 30-min intervals with a plate reader (Tecan Spark).

### Fluorescent marker plasmids (figs. S8 to S10)

Plasmid pVSV208 was a kind gift from the Ruby laboratory (37). It contains an origin of replication from the plasmid pES213, the R6K origin of replication, and an origin of transfer, and encodes chloramphenicol resistance and the fluorescent protein DsRed. Plasmid pVSV208 was maintained in *E. coli* strain DH5-alpha lambda pir or PIR1 (Invitrogen). The strain was routinely grown in LB with chloramphenicol (25  $\mu$ g/ml). Plasmid pBTK569 was a gift from the Barrick laboratory and was acquired from Addgene (Addgene plasmid no. 110614; <http://n2t.net/addgene:110614>; RRID:Addgene\_110614). It contains the RSF1010 origin of replication and an origin of transfer, and encodes spectinomycin resistance and the fluorescent protein mCrimson. Plasmid pBTK569 and its derivative pLL014 were maintained in *E. coli* Top10 (Invitrogen). Spectinomycin (50  $\mu$ g/ml) was used to select for pBTK569, and chloramphenicol (25  $\mu$ g/ml) was used to select for pLL014. *E. coli* strain DH5-alpha was used to maintain pEVS104 and was grown in LB containing kanamycin (25  $\mu$ g/ml).

To construct pLL014, a variant of pBTK569 encoding chloramphenicol resistance and eGFP, we synthesized a segment of DNA coding for eGFP and the chloramphenicol resistance gene *cat* (GeneArt, Invitrogen). We amplified this segment with primers engineered with the restriction sites Bsr GI and Xba I, which are present at unique sites in pBTK569, using high-fidelity Q5 polymerase (NEB). We digested pBTK569 and the Cm-eGFP DNA with Bsr GI and Xba I (NEB) and treated the plasmid digest with Antarctic phosphatase to prevent religation. We ligated the digested Cm-eGFP and pBTK569 with T4 DNA ligase (NEB) and electroporated the products into Top10 *E. coli* (Invitrogen). The construct was confirmed by Sanger sequencing.

Plasmids pLL014 and pVSV208 were introduced into strains of interest by conjugation. The helper plasmid pEVS104 (38), which carries in-trans copies of the conjugative transfer genes, was used to mobilize the oriT-containing plasmids. Equal parts of overnight cultures of *E. coli* strains carrying pVSV208 or pLL014 (donor), pEVS104, or the recipient strain were washed, resuspended in antibiotic-free MB, combined, and pelleted. The pellet was resuspended in a small amount of MB and plated on MB 1.5% agar. After overnight incubation at 25°C, the conjugation mixture was plated onto MB agar containing chloramphenicol (12.5  $\mu$ g/ml). Colonies were plated on two rounds of MB agar (12.5  $\mu$ g/ml) to isolate plasmid-carrying recipient strains from any residual *E. coli*. The transconjugants were confirmed by sequencing the 16S rRNA gene using universal primers 8F (5-AGAGTTTGATCCTGGCTCAG-3) and 1522R (5-AAGGAGGTGATCCANCCRCA-3).

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/31/eabi4717/DC1>

### REFERENCES AND NOTES

- O. X. Cordero, L.-A. Ventouras, E. F. DeLong, M. F. Polz, Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20059–20064 (2012).
- E. Rosenberg, K. H. Keller, M. Dworkin, Cell density-dependent growth of *Myxococcus xanthus* on casein. *J. Bacteriol.* **129**, 770–777 (1977).
- R. Axelrod, W. D. Hamilton, The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
- M. S. Datta, E. Sliwerska, J. Gore, M. F. Polz, O. X. Cordero, Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat. Commun.* **7**, 11965 (2016).
- T. N. Enke, M. S. Datta, J. Schwartzman, N. Cermak, D. Schmitz, J. Barrere, A. Pascual-García, O. X. Cordero, Modular assembly of polysaccharide-degrading marine microbial communities. *Curr. Biol.* **29**, 1528–1535.e6 (2019).
- S. Beier, S. Bertilsson, Bacterial chitin degradation—Mechanisms and ecophysiological strategies. *Front. Microbiol.* **4**, 149 (2013).
- S. Doron, S. Melamed, G. Ofir, A. Leavitt, A. Lopatina, M. Keren, G. Amitai, R. Sorek, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
- O. X. Cordero, B. Snel, P. Hogeweg, Coevolution of gene families in prokaryotes. *Genome Res.* **18**, 462–468 (2008).
- M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
- A. P. Vogler, J. W. Lengeler, Analysis of the *nag* regulon from *Escherichia coli* K12 and *Klebsiella pneumoniae* and of its regulation. *Mol. Gen. Genet.* **219**, 97–105 (1989).
- C.-G. Korea, J.-M. Ghigo, C. Beloin, The sweet connection: Solving the riddle of multiple sugar-binding fimbrial adhesins in *Escherichia coli*: Multiple *E. coli* fimbriae form a versatile arsenal of sugar-binding lectins potentially involved in surface-colonisation and tissue tropism. *Bioessays* **33**, 300–311 (2011).
- G. Ji, S. Silver, Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9474–9478 (1992).
- S.-J. Lee, J. Ko, H.-Y. Kang, Y. Lee, Coupled expression of MhpE aldolase and MhpF dehydrogenase in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **346**, 1009–1015 (2006).
- U. F. Wehmeier, J. W. Lengeler, Sequence of the sor-operon for L-sorbose utilization from *Klebsiella pneumoniae* KAY2026. *Biochim. Biophys. Acta* **1208**, 348–351 (1994).
- C. L. Bouma, S. Roseman, Sugar transport by the marine chitinolytic bacterium *Vibrio furnissii*: Molecular cloning and analysis of the mannose/glucose permease. *J. Biol. Chem.* **271**, 33468–33475 (1996).
- J.-H. Heheman, P. Arevalo, M. S. Datta, X. Yu, C. H. Corzett, A. Henschel, S. P. Preheim, S. Timberlake, E. J. Alm, M. F. Polz, Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nat. Commun.* **7**, 12860 (2016).
- D. Machado, S. Andrejev, M. Tramontano, K. R. Patil, Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
- D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. von Mering, STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
- A. Ebrahimi, J. Schwartzman, O. X. Cordero, Cooperation and spatial self-organization determine rate and efficiency of particulate organic matter degradation in marine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23309–23316 (2019).
- G. Reintjes, C. Arnosti, B. M. Fuchs, R. Amann, An alternative polysaccharide uptake mechanism of marine bacteria. *ISME J.* **11**, 1640–1650 (2017).
- G. J. Velicer, L. Kroos, R. E. Lenski, Developmental cheating in the social bacterium *Myxococcus xanthus*. *Nature* **404**, 598–601 (2000).
- N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciuffo, D. Haddad, R. McVeigh, B. Rajput, B. Robertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

23. C. Jain, L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, S. Aluru, High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
24. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
25. G. Csardi, T. Nepusz, The igraph software package for complex network research (2006).
26. P.-A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
27. H. Zhang, T. Yohe, L. Huang, S. Entwistle, P. Wu, Z. Yang, P. K. Busk, Y. Xu, Y. Yin, dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
28. J. Huerta-Cepas, K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen, C. von Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
29. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
30. S. Louca, M. Doebeli, Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
31. O. Cohen, H. Ashkenazy, D. Burstein, T. Pupko, Uncovering the co-evolutionary network among prokaryotic genes. *Bioinformatics* **28**, i389–i394 (2012).
32. L. Gao, H. Altae-Tran, F. Böhning, K. S. Makarova, M. Segel, J. L. Schmid-Burgk, J. Koob, Y. I. Wolf, E. V. Koonin, F. Zhang, Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
33. D. Moi, L. Kilchoer, P. S. Aguilar, C. Dessimoz, Scalable phylogenetic profiling using MinHash uncovers likely eukaryotic sexual reproduction genes. *PLOS Comput. Biol.* **16**, e1007553 (2020).
34. D. Hyatt, G.-L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
35. A. Ebrahim, J. A. Lerman, B. O. Palsson, D. R. Hyduke, COBRAPy: COnstraints-based reconstruction and analysis for python. *BMC Syst. Biol.* **7**, 74 (2013).
36. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
37. A. K. Dunn, D. S. Millikan, D. M. Adin, J. L. Bose, E. V. Stabb, New *rfp*- and pES213-derived tools for analyzing symbiotic *Vibrio fischeri* reveal patterns of infection and *lux* expression in situ. *Appl. Environ. Microbiol.* **72**, 802–810 (2006).
38. E. V. Stabb, E. G. Ruby, in *Methods in Enzymology* (Academic Press, 2002); <http://www.sciencedirect.com/science/article/pii/S0076687902581064>, vol. 358 of *Bacterial Pathogenesis Part C: Identification, Regulation, and Function of Virulence Factors*, pp. 413–426.

**Acknowledgments:** We thank M. Polz, S. Kuehn, A. Eldar, A. Goyal, and members of the Cordero laboratory for helpful comments. **Funding:** This work was supported by the Simons Collaboration: Principles of Microbial Ecosystems (PrIME) award no. 542395. S.P. was supported by the EMBO ALTF grant no. 800-2017. M.G. was supported by Simons Foundation Postdoctoral Fellowship Award 599207. **Author contributions:** S.P.: conceptualization, investigation, methodology, visualization, and writing. M.G.: investigation, visualization, and writing. Y.S.: methodology, J.S.: methodology. O.X.C.: conceptualization, project administration, funding, supervision, and writing. L.L.: methodology. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Code available at <https://github.com/sigmap666/NaturalExploiters>.

Submitted 11 March 2021

Accepted 10 June 2021

Published 28 July 2021

10.1126/sciadv.abi4717

**Citation:** S. Pollak, M. Gralka, Y. Sato, J. Schwartzman, L. Lu, O. X. Cordero, Public good exploitation in natural bacterioplankton communities. *Sci. Adv.* **7**, eabi4717 (2021).

## Public good exploitation in natural bacterioplankton communities

Shaul Pollak, Matti Gralka, Yuya Sato, Julia Schwartzman, Lu Lu and Otto X. Cordero

*Sci Adv* 7 (31), eabi4717.  
DOI: 10.1126/sciadv.abi4717

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/7/31/eabi4717>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2021/07/26/7.31.eabi4717.DC1>

### REFERENCES

This article cites 36 articles, 10 of which you can access for free  
<http://advances.sciencemag.org/content/7/31/eabi4717#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).