Title: Public good exploitation in natural bacterioplankton communities

Authors: Shaul Pollak[1†], Matti Gralka[1†], Yuya Sato[2], Julia Schwartzman[1], Lu Lu[1], Otto X. Cordero[1*].

Affiliations:

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, 02139, Massachusetts, USA.

[2]Management Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 16-1 Onogawa, Tsukuba, Ibaraki, 305-8569, Japan.

*Correspondence to: ottox@mit.edu.

†Equal contribution.

Abstract:

Public good exploitation has been studied extensively from an evolutionary lens, but little is known about the occurrence and impact of public good exploiters in natural communities. Here, we develop a reverse ecology approach to systematically identify bacteria that can exploit public goods produced during the degradation of polysaccharides. Focusing on chitin – the second most abundant biopolymer on the planet, we show that public good exploiters hinder the growth of degraders and invade marine microbial communities during early stages of colonization. Unlike cheaters in social evolution, exploiters and polysaccharide degraders (cooperators) come together by a process of community assembly, belong to distant lineages, and can stably coexist. Thus, our approach opens novel avenues to interpret the wealth of genomic data through an ecological lens.

Main Text:

One of the main challenges in modern microbiology is to interpret the enormous diversity of organisms that make up natural microbial communities in terms of ecological roles. Plant and animal ecologists have typically assigned taxa to roles such as primary producers, grazers, etc., based on experience and direct observation. This simplification has allowed ecologists to build simple models of ecosystems, bypassing the need to account for the enormous diversity of species (*1*). For microbes, however, such a functional categorization must rely on experiments or genomic data (*2*). In this paper, we address the challenge of identifying organisms that can act as public good exploiters in communities that degrade complex forms of organic matter. The degradation of complex organic matter by microbes 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, akin to cheaters in social evolution, which consume the breakdown byproducts but do not contribute to the pool of enzymes (*3*). However, community assembly experiments on complex carbon sources have revealed that, although secondary consumers (non-degraders) are numerically dominant, they do not grow on oligosaccharides and instead prefer to consume metabolic waste products, such as organic acids and biomass precursors (*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 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, without relying on annotation, which organisms can act as chitooligosaccharide exploiters and which are obligate metabolic waste scavengers. 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. Finally,

we showed that exploiters sampled from natural communities can also stably coexist in the lab 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.

Genomic signatures of public good exploitation

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, 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 byproducts 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 co-located in the chromosome (*7*, *8*).

We developed a computational approach to detect genes that coevolve or are colocalized with chitinases across 8752 non-redundant complete bacterial genomes spanning the tree of bacterial life (Table S1, Data S1,S2). 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 evolution (Methods), we found 2097 gene families that significantly coevolve with chitinases in bacteria (Fig. 1B). In addition, we identified 1479 genes that co-localize together with chitinases in the genome, for a total of 3576 candidate predictor genes (Data S3). These genes had diverse KEGG annotations such as chemotaxis and motility, attachment and biofilm formation as well as 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 (Fig. 1C,S1 Data S4)(*9*). 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.

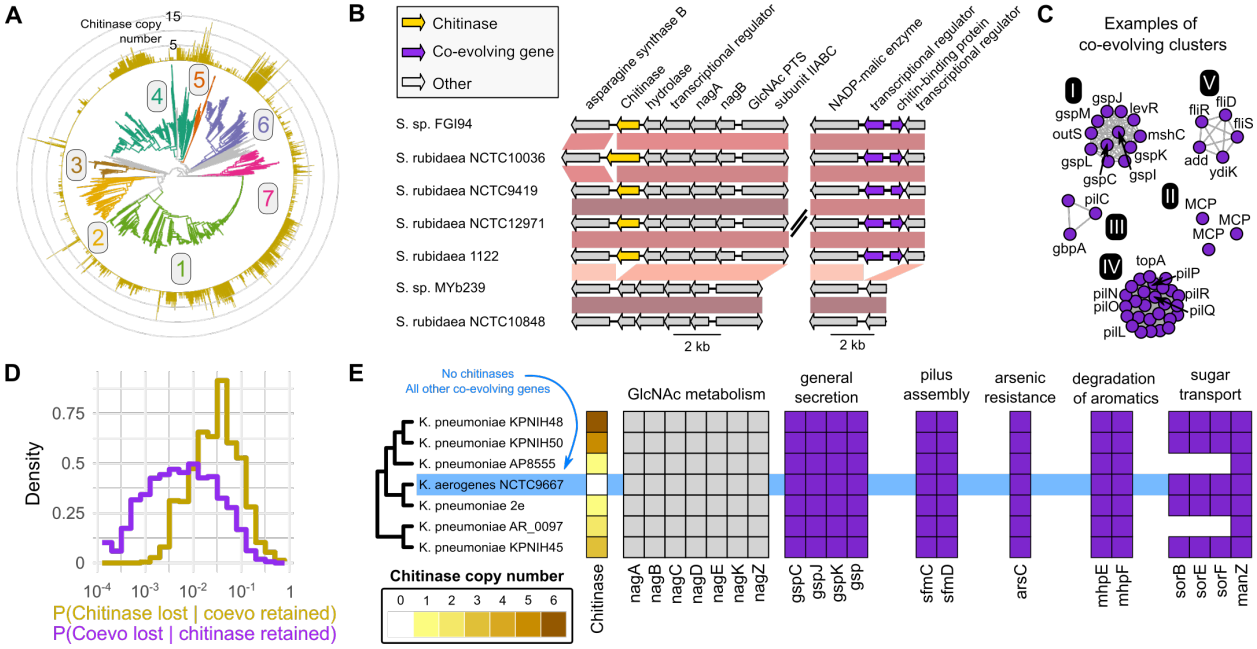

Fig. 1 Large-scale detection of genes that co-evolve with chitinases reveals genomic features of diverse chitin degraders and putative exploiters. (A) Genetically distinct (ANI < 99.9%) closed genomes used in this study, with chitinase copy-number indicated on each tip (gold bars). The different phyla are indicated by edge colors, with the identity of each phylum indicated by the phylum number. See Table S1 for full names. (B) Example of the genomic signatures detected by our co-evolution approach in a group of closely related Shewanella genomes. Specific excision events at distal loci drive the observed patterns, and not mutations that result in reduced homology. (C) examples of co-evolving clusters: I - general secretion pathway proteins. II – Chemotaxis receptor proteins. III – Pilus genes that mediate attachment to surfaces. IV – Pilus genes, and genes of unknown function. V - motility genes encoding for flagella proteins. (D) Conditional probability density of losing a chitinase while retaining a coevolving gene and vice-versa. The conditional probabilities are calculated based on the reconstructed gene gain and loss events on each branch of the tree of life. The result shows that chitinases have a higher tendency to be lost while co-evolving genes are retained, as compared to the opposite scenario. (E) The consequence of rapid loss of chitinases is the evolution of oligosaccharide exploiters. Shown is a group of closely related *Klebsiella pneumoniae* strains, where strain NCTC9667 lost the chitinase while retaining all other co-evolving genes, hinting at a possible natural exploiter.

Although by definition the histories of gain and losses of chitinases and accessory clusters were correlated, chitinases had a higher loss rate than their coevolving genes. This means that through evolution there were multiple events in which chitinases were lost but accessory clusters retained, likely leading to the evolution of an oligosaccharide (public good) exploiter. The opposite event - loss of an accessory gene and chitinase retention - was less likely to occur (mean odds-ratio = 0.755, t-test p-value < 5e-5) (Figure 1D, Supplementary Figure S2). An example is shown in Figure 1E, 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-HPP catabolism (*mhpEF*)(*13*), and several sugar transport proteins (*sorBEF, manZ*)(*14*, *15*), all of which coevolve significantly 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 that we hypothesize is needed to compete for chitin oligosaccharides but that does not produce hydrolytic enzymes.

Predicting exploitation potential from genomes

Motivated by these observations, we devised an approach to systematically differentiate degraders, exploiters, and waste product consumers, further referred to as 'scavengers', solely on the basis of genomic data. Our approach is as follows: using an elastic-net regression model (Methods), we predict the expected number of chitinases (E_{chi}) a genome should have based on its repertoire of accessory clusters (on average 190.2 gene clusters are needed to make these predictions, min 37, max 460. Average cross-validated $R^2 = 0.825$, Table S2). These predictors were significantly better than those based on random subsets of genes in most phyla, further supporting the idea that the accessory clusters contain significant ecological information (Fig. S3). Genomes with $E_{chi} < 1$ are predicted to be scavengers since they lack the genetic machinery required to compete for oligosaccharides. Genomes with $E_{chi} > 1$ are predicted to be either exploiters or degraders. To distinguish these two classes, we use the observed number of chitinases (O_{chi}). An agreement between the observed and expected number of chitinases ($O_{chi} > 1$) predicts a degrader, whereas a disagreement ($O_{chi} < 1$, $E_{chi} > 1$) predicts an exploiter (Fig. 2A). Since distantly related genomes have little overlap in their gene content, we trained the

regression model separately on well-sampled phyla, such as Bacteroidetes, Proteobacteria, or Firmicutes (Methods, Table S2).

With our predictions in hand, our next task was to validate them in the context of a naturally assembled chitin-degrading community. To this end, we leveraged a collection of 57 isolates of copiotrophic marine bacteria, co-isolated 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 = 0.604). We predicted 11 degraders, 17 exploiters, and 29 scavengers in this collection. Notably, these functional groups were phylogenetically diverse, belonging to such diverse orders as Flavobacteriales, Alteromonadales, Rhodobacterales, and Vibrionales (Fig. S4). We interpreted our functional classification (degrader, exploiter, and scavenger) in terms of growth phenotypes that could be tested in-vitro: 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. 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 (*16*).

Overall, our model performed better than the annotation-based predictions, making ~40% fewer errors (14 vs 23 errors), confirming the value of our approach (Fig. 2B). Moreover, 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. 2C,S5, Data S5). 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. S6). This is consistent with 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 *E. coli*, *B. subtilis*, or *V. cholerae*. In contrast, our nearly annotation-free 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.

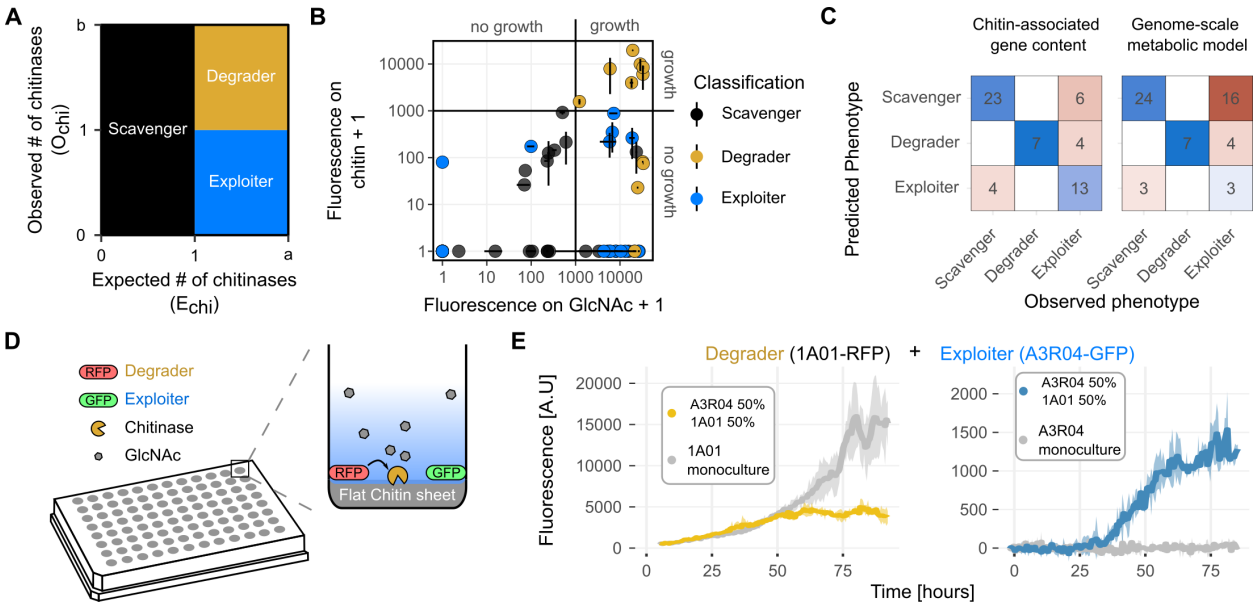

Fig. 2 Evolution-guided classification predicts isolate phenotypes in mono- and co-culture. (A) Schematic illustration of the classification scheme. The expected number of chitinases is predicted using an elastic-net regression that uses accessory-cluster content as features. Classification is carried out by comparing the observed to the expected number of chitinases. (B) Final yield of 57 marine isolates on Chitin and its monomer N-Acetylglucosamine. Yield was measured as fluorescence emitted from Sybr-gold incorporated DNA dye (methods). Each point represents the mean of at least 3 biological repeats, and bars represent standard deviations. (C) accessory-cluster gene content-based classification predicts phenotype better than genome-scale metabolic models. Confusion matrices comparing the predicted phenotypes of each model to the observed phenotypes. (D) Illustration of the flat-chitin experimental setup used to measure co-culture dynamics. (E) Degrader-Exploiter co-culture reveals the consequence of exploitation among co-inhabiting environmental isolates. The Y-axis displays the fluorescence intensity emitted from a plasmid-encoded fluorescent protein, as depicted in panel D. Grey lines represent mono-culture readings, while colored lines represent readings from co-cultures.

To further validate the prediction that exploiters should grow on chitin in co-culture with a degrader, but not in monoculture, we built an experimental system to study a degrader-exploiter co-culture on chitin. Briefly, we coated the bottom of a plastic microwell with chitin obtained from crab shells to create a flat, ~20 µm thick, transparent layer through which we could measure optical density and fluorescence (Methods). Using a new plasmid design, we labeled a degrader (vibrio1A01) and an exploiter (alteroA3R04) with fluorescent reporters to track population dynamics in co-culture (Fig. 2D). As expected, the results show that the exploiter grew only when co-cultured with the degrader. Moreover, the exploiter acted as a 'parasite' and reduced the yield of the chitinase producer, likely due to competition for the public good (GlcNAc or other oligosaccharides) (Fig. 2E,S7,S8).

Ecological dynamics of exploiters

It is still unclear whether 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 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). This allowed us to analyze the colonization dynamics of each mapped isolate across the different functional roles (Fig. 3B). We found that the 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 (12-16 h for exploiters, 8 h for degraders), while scavengers peaked at late successional stages (>80 h) (Fig. S9). This observation implies that the accessory gene clusters, containing the genetic machinery to chemotax towards GlcNAc, adhere to chitin surfaces, etc., are the main determinant of early colonization, and not chitinases. In other words, our response variable, E_{chi}, should be a better predictor of colonization dynamics in a wild community than chitinase copy number (O_{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 (Methods, Fig. 3C). As expected, we found that E_{chi} was a better predictor of colonization speed than the observed number of chitinases, O_{chi}. Genomes predicted to contain chitinases had a 19-fold higher colonization rate on average compared to those predicted to lack chitinases (p-value = 0.02). In contrast, genomes encoding chitinases did not display a significant difference in colonization rate compared to those that lacked chitinases (p-value = 0.37, average fold-difference = 2.1).

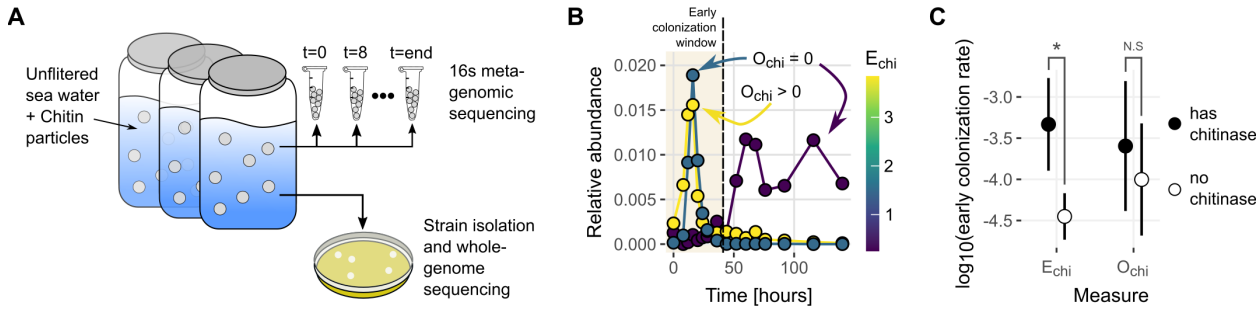

Fig. 3 Exploiter dynamics mimic those of degraders in complex natural communities. (A) Schematic depiction of the experimental setup. Natural sea-water communities were incubated with chitin as the sole carbon source, and samples were taken periodically for DNA extraction and isolate collection. (B) Colonization dynamics of 3 16S sequences, representing 3 strains. The early colonization window was defined as in (*5*). By mapping isolate genomes' 16S sequences back to those of the community we were able to determine the observed (O_{chi}) and expected (E_{chi}) number of chitinases of each one of these community members. Lines are colored according to E_{chi}. (C) The expected number of chitinases (E_{chi}) is a better predictor of early colonization rate than the observed number of chitinases (O_{chi}). The distribution of early colonization rates of isolates with or without chitinases was compared based on E_{chi} or O_{chi}. asterisk marks p-value < 0.05. N.S = p-value > 0.3).

In the natural chitin-degrading communities the dynamics of colonization and growth are successional, meaning stable coexistence cannot be attained on a single particle. Instead, degraders, 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. To assess coexistence, we serially passaged the co-culture over 11 dilution-growth cycles (Fig. 4A,B, Data S6) and repeated these experiments with three different dilution rates (which determine the minimum growth rate required to survive serial passages) to assess the robustness of coexistence to different growth conditions (Methods).

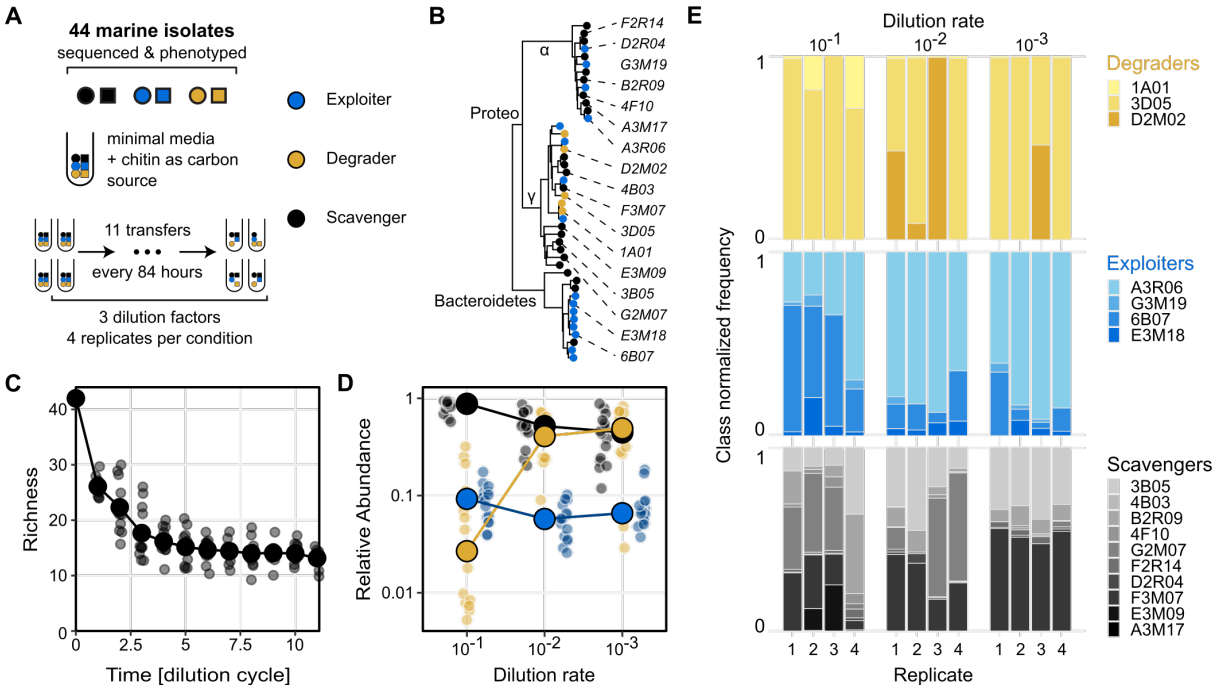

Fig. 4 Synthetic communities display stable coexistence of ecological roles, with stochastic strain replacement. (A) Illustration of the experimental setup. (B) phylogenetic tree of the isolates in this experiment. The isolates that survived to the last time point are indicated by name (C) The number of species present in a community as a function of time. Communities reach a steady state in terms of the number of coexisting species after approximately 6 transfers. (D) Steady-state relative abundance (transfer ≥ 8) of each ecological class in the synthetic communities. Each point represents the total class abundance from a given sample at a given time point. Large points represent dilution rate medians. Colors represent ecological classes as in panel A. (E) The niche space of each functional role is filled by different strains in different repeats/dilution factors. Class normalized relative abundances are shown from the final timepoint. Colors as in panel A.

Following a short transient where species are rapidly purged, community richness stabilized on 10-15 members (Fig. 4C), with coexistence of all three roles across all dilution rates (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 functional group (20-80%), suggesting a widespread export of metabolic waste by exploiters and degraders. 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.

Closer inspection of the community compositions across dilution factors and replicates revealed different patterns of diversity within functional roles. In most communities, degraders were dominated by a single strain, *Pseudoalteromonas 3D05*, and two exploiters (*Maribacter 6B07* and *Tritonibacter A3R06*), while four to five scavengers coexisted in the same community. However, there were clear cases in which other strains became dominant: *3D05* coexisted with other degraders at appreciable abundance in at least four out of the twelve communities; was altogether replaced by *Colwellia D2M02* in one replicate, while one of the two exploiters seemed to reach higher relative abundances at higher dilution rates. 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 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.

Discussion

In this paper, we uncovered a previously cryptic class of secondary consumers: exploiters, which share many genomic features with degraders but do not encode the hydrolytic enzymes. Although consistent with the notion of a social 'cheater', there is an important difference: cheaters are loss-of-function mutants that rise in frequency due to a short-term fitness advantage over a wild-type cooperator phenotype (*17*). Exploiters, by contrast, invade communities by a process of community assembly, implying that exploitative interactions are likely to take place between organisms with very distinct genetic backgrounds. Indeed, exploiters in naturally assembled chitin-degrading communities are distantly related to degraders and have diverse metabolic potentials. There are therefore many possible mechanisms by which exploiters and degraders could stably coexist, since these two classes of organisms can differ along multiple phenotypic dimensions, including, for example, their preferred oligosaccharide chain lengths (*18*). Another important consideration is the fact that the public good, in this case chitooligosaccharides, is not the only carbon source made available by degraders. Metabolic waste products are also released to the environment, supporting the growth of scavengers. Therefore, there is a potential for exploiters (but not scavengers) to utilize both

chitooligosaccharides and other metabolic waste products depending on the availability of substrates in the community.

One of our main contributions is the development of a computational approach that allowed us to differentiate exploiters from scavengers. A key feature of our method is that it is independent of 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 (*19*). 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.

The patterns observed at the level of functional roles and species are consistent with the emergent view that functional groups, encompassing species with cohesive phenotypic properties, represent the 'right' variables to describe microbiomes (*2*). Our approach provides a path to predict these functional groups from genomic data, in particular in the context of processes mediated by public goods, such as polysaccharide degradation, nitrogen fixation, iron utilization, phosphate solubilization, etc. Further work should explore these applications as well as alternative strategies to leverage the wealth of microbial genomic data to infer resource preferences and inter-specific interactions in the environment.

References and Notes:

1. J. A. Estes, J. Terborgh, J. S. Brashares, M. E. Power, J. Berger, W. J. Bond, S. R. Carpenter, T. E. Essington, R. D. Holt, J. B. C. Jackson, R. J. Marquis, L. Oksanen, T. Oksanen, R. T. Paine, E. K. Pikitch, W. J. Ripple, S. A. Sandin, M. Scheffer, T. W. Schoener, J. B. Shurin, A. R. E. Sinclair, M. E. Soulé, R. Virtanen, D. A. Wardle, Trophic Downgrading of Planet Earth. *Science*. **333**, 301–306 (2011).

2. S. Louca, M. F. Polz, F. Mazel, M. B. N. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann, A. S. Hahn, D. S. Srivastava, S. A. Crowe, M. Doebeli, L. W. Parfrey, Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* **2**, 936–943 (2018).

3. R. Axelrod, W. D. Hamilton, The evolution of cooperation. *Science*. **211**, 1390–1396 (1981).

4. 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).

5. 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).

6. S. Beier, S. Bertilsson, Bacterial chitin degradation—mechanisms and ecophysiological strategies. *Front. Microbiol.* **4** (2013), doi:10.3389/fmicb.2013.00149.

7. 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** (2018), doi:10.1126/science.aar4120.

8. O. X. Cordero, B. Snel, P. Hogeweg, Coevolution of gene families in prokaryotes. *Genome Res.* **18**, 462–468 (2008).

9. 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).

10. A. P. Vogler, J. W. Lengeler, Analysis of the nag regulon from Escberichia coli K12 and Klebsiella pneumoniae and of its regulation. *Mol. Gen. Genet. MGG*. **219**, 97–105 (1989).

11. C.-G. Korea, J.-M. Ghigo, C. Beloin, The sweet connection: Solving the riddle of multiple sugar-binding fimbrial adhesins in Escherichia coli. *BioEssays*. **33**, 300–311 (2011).

12. 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.* **89**, 9474–9478 (1992).

13. 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).

14. U. F. Wehmeier, J. W. Lengeler, Sequence of the sor-operon for l-sorbose utilization from Klebsiella pneumoniae KAY2026. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **1208**, 348–351 (1994).

15. 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).

16. 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).

17. G. J. Velicer, L. Kroos, R. E. Lenski, Developmental cheating in the social bacterium Myxococcus xanthus. *Nature*. **404**, 598–601 (2000).

18. J.-H. Hehemann, 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).

19. 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).

20. 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).

21. 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.

22. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, 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**, 1–8 (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. *InterJournal, Complex Systems*, 1695. https://igraph.org (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).

Data and materials availability: All data is available in the main text or the supplementary materials. Code available at https://github.com/sigmap666/NaturalExploiters