Title: Information storage across a microbial community using universal RNA barcoding

Authors:

Prashant B. Kalvapalle[1,2,3]†, August Staubus[2,4]†, Matthew J. Dysart[1,2,3]†, Lauren Gambill[1,2]†, Kiara Reyes Gamas[1,2,3], Li Chieh Lu[2,4], Jonathan J. Silberg[2,5,6,*], Lauren B. Stadler[3,*], and James Chappell[2,6,*]

Affiliations:

[1]Systems, Synthetic, and Physical Biology Graduate Program, Rice University; Houston, Texas 77005, United States of America.
[2]Department of BioSciences, Rice University; Houston, Texas 77005, United States of America.
[3]Department of Civil and Environmental Engineering, Rice University; Houston, Texas 77005, United States of America.
[4]Biochemistry and Cell Biology Graduate Program, Rice University; Houston, Texas 77005, United States of America.
[5]Department of Chemical and Biomolecular Engineering, Rice University; Houston, Texas 77005, United States of America.
[6]Department of Bioengineering, Rice University; Houston, Texas 77005, United States of America.
*Corresponding author. Emails: jc125@rice.edu, lauren.stadler@rice.edu, and joff@rice.edu
† These authors contributed equally to this work

Abstract: Biosensors can code information in nucleic acids, but they remain challenging to apply in complex microbial communities. To provide information about gene transfer host range among microbiome members, a synthetic catalytic RNA was used to barcode a highly conserved segment of ribosomal RNA (rRNA). By writing information into rRNA using a ribozyme and reading out native and modified rRNA using amplicon sequencing, we find that microbial community members from twenty taxonomic orders participate in plasmid conjugation with an *Escherichia coli* donor strain and observe differences in 16S rRNA barcode signal across amplicon sequence variants. Multiplexed rRNA barcoding using conjugative plasmids with pBBR1 or ColE1 origins of replication revealed differences in host range. This autonomous

RNA-addressable modification (RAM), which provides information about gene transfer without requiring translation, will enable microbiome engineering across diverse ecological settings and studies of environmental controls on gene transfer and cellular uptake of extracellular materials.

Main Text:

Nucleic acids acquired through gene transfer can alter cell-cell interactions [1,2], facilitate niche expansion [3], drive bacterial resistance to phage [4], and control gene stability [5]. Gene transfer is also critical to cellular domestication for synthetic biology [6–8] and a challenge for the safe application of such technologies [9]. Because gene transfer can occur across species from different genera, phyla, and even kingdoms [10,11], there is a need to understand how it varies across communities. Of particular importance is understanding the host ranges of conjugative plasmids [12], bacteriophages [13], and environmental DNA uptake [10], as well as the effects of biotic and abiotic parameters on gene transfer rates [14–19]. Such knowledge is crucial to prevent the unwanted spread of harmful antibiotic resistance genes [20] and DNA from genetically-engineered biotechnologies [21].

Genetically-encoded reporters are commonly used to study gene transfer [22,23]. Fluorescent proteins and antibiotic-resistance genes can be coded into mobilizable genetic elements and used to identify cells that participate in gene transfer [24]. These approaches led to the discovery of novel DNA exchange mechanisms [25,26] and provided mechanistic insight into gene transfer processes [27]. However, since these methods require microbial growth and propagation, as well as specialized instrumentation like cell sorting [24], they require significant processing before sequencing to assess mobile DNA host range.

Culture-independent strategies can autonomously record information about gene transfer in

a community without the need for imaging. For instance, mobile DNA containing a transposase and transposon can be used to randomly integrate a transposon into the genomes of cells that take up genetic material, and DNA uptake can be read out using environmental transformation sequencing (ET-seq) [28]. However, this approach requires metagenomic sequencing to detect the transposon and identify transconjugants, limiting sensitivity, it cannot measure plasmid dynamics as the insertion is irreversible, and it depends upon transposase transcription, translation, and folding [29], which varies across microbes. Chemical and enzymatic strategies for genetic barcoding overcome some of these limitations [30,31], but they cannot be performed autonomously within living cells and can require arduous chemical manipulations.

To monitor gene transfer across microbiomes, we designed catalytic RNAs (cat-RNAs) that autonomously barcode host RNA upon gene transfer without the need for translation (Fig. 1A). These small, genetically-encoded cat-RNA are composed of: (i) a designable *RNA guide* that localizes the system to a target RNA through base pairing interactions, (ii) a *catalytic core* derived from a splicing ribozyme that serves as a writer to catalyze the barcoding reaction [32,33], and (iii) a non-coding *RNA barcode*. Upon interaction with the target RNA, cat-RNAs are designed to amend the RNA barcode onto the target RNA, forming transient genetically-encoded information (Fig. S1), which we designate RNA-addressable modification (RAM).

To determine cat-RNA efficiencies across different microbes, we developed a split *gfp* gene that functions as a visual reporter for RNA splicing (Fig. 1B). The first *gfp* fragment, which is translated, contains the coding sequence for GFP residues 1 to 65 (gfp-F1) fused to a uracil that is targeted for splicing and followed by a non-coding RNA (ncRNA) sequence. The second *gfp* fragment, which is fused to the end of a cat-RNA writer, has a guide that targets the last six nucleotides (nt) of gfp-F1 and 50 nt of the ncRNA sequence. The latter cat-RNA writer amends

an RNA barcode composed of the coding sequence for GFP residues 66-238 (gfp-F2) to the end of gfp-F1 to create a native GFP transcript. When these RNA were transcribed in *Escherichia coli* from a pBBR1 plasmid, whole cell fluorescence exceeded that of cells transformed with empty plasmid (cells) and was ~30% of that in cells that constitutively express full-length GFP (Fig. 1C, Fig. S2). The growth rates of cells expressing the reporter were not significantly different from cells lacking it (Fig. S3). These results show that cat-RNA efficiently generates chimeric transcripts through RNA-mediated barcoding with no detectable effects on cell fitness under these growth conditions.

We next investigated how RNA guide length affects barcoding. Cat-RNA writers that generate a GFP output were designed with RNA guides ranging in length from 0 to 600 nt. A plasmid expressing each cat-RNA and target RNA was transformed into *E. coli*, and cellular fluorescence was measured (Fig. 1D). A wide range of guide lengths (25 to 375 nt) presented a signal, with the highest occurring with the 50 nt guide. All subsequent experiments used cat-RNA guides that were 50 nt long. To determine if the cat-RNA writers function in different microbes, we introduced a broad host plasmid encoding the cat-RNA reporter into microbes from diverse environments, including gut (*E. coli*), ocean (*Vibrio natriegens*), soil (*Pseudomonas putida*), and freshwater (*Shewanella oneidensis* MR1). In all cases, a fluorescence signal was observed for the cat-RNA reporter (Fig. 1E). Thus, a single cat-RNA writer functions across microbes from different orders without optimization of the cat-RNA sequence or the transcription cassette.

Given the centrality of 16S rRNA in taxonomic analysis, we hypothesized that a universal cat-RNA writer could be created by modifying the RNA guide so that it is complementary to conserved 16S rRNA regions (Fig. 2A). We first identified variable regions of

the 16S rRNA gene for five microbes, including four Gram negative and one Gram positive species, by aligning their 16S rRNA sequences and calculating Shannon entropy. To design universal cat-RNAs, we then performed a 16S rRNA multiple sequence alignment using the same microbes to identify conserved regions (6 nt) that end in a uracil (Fig. 2B). The 5 nt adjacent to the uracil will ultimately bind to the internal guide sequence (IGS) to form a paired domain called P1 [34–36]. We then designed 50 nt RNA guide sequences that are complementary to the rRNA sequences adjacent to each conserved IGS-U (Fig. S4). We scored each variant based on the conservation of the guide-binding region and identified 54 designs across the targeted species with >80% identity (Fig. S5). This finding shows a broad-host range RNA guide can be designed that targets cat-RNA to diverse 16S rRNA sequences for information storage.

Four cat-RNA writers were built that target conserved rRNA sequences (Fig. S6). These cat-RNAs were expressed from plasmids in *E. coli* using a constitutive promoter (Fig. 2C), and the native and barcoded 16S rRNA was quantified by reverse transcription-quantitative polymerase chain reaction (RT-qPCR). All four of the designs produced barcoded 16S rRNA. A comparison of the barcoded 16S rRNA to total rRNA abundance revealed that these cat-RNAs presented similar barcoding efficiencies, with 5 to 28 barcoded rRNA per million total rRNA molecules. Assuming that exponentially growing cells contain ~72,000 copies of rRNA [37], this finding shows that 0.4 to 2 copies of rRNA per cell are barcoded.

To establish the sequence of the barcoded rRNA, we performed RT-PCR and amplicon sequencing of the spliced product arising from the cat-RNA that targets U1376 (Fig. 2D). The major product arose from ligation of the barcode after U1376, although low frequency off-pathway splicing occurred at sites with uracils and related P1 sequences (Fig. S7). We next investigated the stability and the reversibility of the barcoded rRNA (Fig. S8). This analysis

revealed stability that is similar to that of a typical bacterial mRNA (3 to 8 minute half-lives) [38]. Taken together, these findings show that cat-RNA can barcode a conserved rRNA sequence, creating a form of information storage that is expected to be less stable than information stored within DNA[39,40]. The cat-RNA that targets U1376 was used for all subsequent analyses.

To determine whether cat-RNA efficiently barcodes rRNA in different microbes, we cloned it onto a broad host plasmid (pBBR1) and transformed this plasmid into *E. coli*, *P. putida*, *Pseudomonas stutzeri*, *S. oneidensis*, and *V. natriegens*. RT-qPCR detected barcoded rRNA in all five species although species-to-species variation in the fraction of barcoded 16S rRNA was observed (Fig. 2E). We also saw variation in cat-RNA transcription in each microbe with *E. coli* having the highest signal (Fig. S9). These findings demonstrate that a universal cat-RNA coded into a single broad-host-range plasmid can record information within a conserved rRNA sequence in diverse microbes, and they show that the gene transfer signal within each organism fluctuates due to variation in the activity of the promoter used to transcribe the cat-RNA.

The finding that cat-RNA can barcode rRNA in a range of microbes suggested it could be applied in a microbiome from the environment to record information about who participates in gene transfer. Because 16S rRNA sequences are used for taxonomic classification, we hypothesized that cat-RNA would generate an ensemble of chimeric rRNA that could be sequenced and analyzed to identify the organisms that participated in gene transfer. Since conjugation rates can vary widely [41], we programmed *E. coli* to minimize the donor rRNA barcoding signal by repressing cat-RNA transcription ~90 fold using repression with the transcription factor CymR (Fig. S10). We used this strain as a donor for conjugation into a wastewater sludge microbial community sampled from the West University Wastewater Treatment Plant in Houston, Texas (Fig. 3A). Filter mating assays were performed aerobically,

147 total RNA was isolated and converted to cDNA, and amplicon sequencing was performed.
148 Analysis of barcoded and native rRNA using high-throughput sequencing revealed 279 amplicon
149 sequence variants (ASVs), with microbes from twenty orders (Fig. 3B). Among all ASVs
150 detected, 140 presented 16S rRNA-barcode signals. Among these ASVs, twenty were only
151 detected when sequencing the barcoded rRNA (Fig. S11). Thus, amplicon sequencing can be
152 used with cat-RNA to read out information about who is present in a community and who
153 participates in gene transfer.

154 To understand how ASV relatedness correlates with information storage, we quantified
155 the fraction of ASVs barcoded across each order (Fig. 3C). A majority of the Proteobacteria
156 (60%), which represented almost 90% of the ASVs in the community, contained barcoded 16S
157 rRNA. The order *Aeromonadales* presented the largest fraction of barcoded ASVs (~70%). The
158 barcoding signal decreased with ASV taxonomic distance from this order, as did ASV abundance
159 in each order. This data shows that diverse microbes in the wastewater community are capable of
160 information storage about DNA uptake, and it shows that *Aeromonadales*, *Enterobacterales*, and
161 *Pseudomonadales* are active participants in gene transfer, consistent with prior studies [3,42,43].

162 To evaluate whether ASV abundance affects barcoding, we analyzed the relationship
163 between total and barcoded 16S rRNA for the ASVs that presented detectable levels of both (Fig.
164 3D). This analysis yielded a positive correlation, suggesting that microbial abundance in a
165 community contributes to the barcode signal. However, the trend could not be fully explained by
166 a simple linear model, suggesting that additional parameters contribute to the signals, such as
167 conjugation efficiency, cat-RNA transcription, cat-RNA activity, and rRNA-barcode stability.

168 To investigate if we could detect variation in barcoding efficiency across ASV pairs
169 within the community, we compared the ratios of the barcoded to total 16S rRNA amplicon

relative abundances. When this analysis was performed across only those ASVs that had detectable barcoded and total rRNA in all six replicates (Fig. 3E), we found that the ratio varied by up to ~43,000 fold. A pairwise analysis of this data revealed a range of differences in their ratios across ASVs within the same order and between different orders (Fig. 3F). Bootstrap analysis of the data revealed that many pairwise comparisons presented differences that were outside of the 95% confidence intervals (Fig. S12). These findings show that closely related organisms present variation in barcoding, which is thought to arise from differences in ASV conjugation rates, promoter activities, and the stabilities of cat-RNA and the mobile element coding for cat-RNA.

To test if trans-splicing activity is required for 16S rRNA barcoding, we evaluated the signal arising from cat-RNA variants expected to abolish activity, including a cat-RNA having a G264A mutation and a variant lacking a guide (Fig. S13). When transformed into *E. coli* and conjugated into a community, these cat-RNA did not present barcoding signals. To provide further validation that the RAM signal arises from conjugation, we benchmarked the cat-RNA signal against a fluorescent reporter of gene transfer. Specifically, we created a plasmid that expresses cat-RNA and mScarlet under control of CymR (Fig. S14) and established the sensitivities of their signals (Fig. S15). This plasmid was then conjugated into a wastewater community, and samples were sequenced before and after sorting for mScarlet positive cells (Fig. S16A). In total, 905 ASVs were detected (Fig. S16B), including 845 in the total 16S rRNA sequencing, 548 in the mScarlet positive cell sequencing, and 306 in the 16S rRNA-barcode sequencing. Among the ASVs presenting a cat-RNA signal, 90% also exhibited a signal in the mScarlet positive cells. Of the 23 barcoded ASVs which were not observed in the total 16S rRNA sequencing, 70% had a mScarlet signal. Finally, the ASVs that presented signals in all

three sample types represented those ASVs with the highest abundances (Fig. S16C). Thus, a majority of the ASVs that presented a cat-RNA signal also presented visual reporter signals, including many with low abundances.

To establish whether cat-RNA can store information about the uptake of different conjugative plasmids in parallel, we created orthogonal cat-RNAs that amend distinct barcodes at U1376 on 16S rRNA. When transcribed in *E. coli*, both cat-RNA designs barcoded 16S rRNA (Fig. S17). The cat-RNA initially used for community barcoding was on a plasmid with a pBBR1 origin, while the orthogonal cat-RNA design was encoded in the same plasmid backbone modified to contain a ColE1 origin of replication. To evaluate the relative barcoding efficiencies of the cat-RNA plasmids with these different origins, referred to as pBBR1-cat and pColE1-cat, we transformed the two plasmids into separate *E. coli* cells and mixed them at different ratios. We then sequenced the barcoded rRNA of the mixtures and determined the relative counts of the orthogonal barcodes representing each plasmid (Fig. S18). This analysis revealed that pColE1-cat produced a greater barcoding signal on average compared to pBBR1-cat, which we speculate was due to the differences in origins of replication that can affect copy number[44]. These findings show that orthogonal cat-RNA can amend different barcodes to the same site within 16S rRNA, and the information stored can be read out from mixtures of cells.

We next investigated whether information storage could be multiplexed across a single community to record the host range of multiple conjugative plasmids in parallel. We transformed pBBR1-cat and pColE1-cat into *E. coli* MDFpir and used a mixture of these strains as donors for conjugation into a wastewater sludge community (Fig. 4A). Following filter mating, analysis of community 16S rRNA revealed ASVs from twenty orders (Fig. 4B). Analysis of the barcoded rRNA revealed that 188 ASVs had the barcode from pBBR1-cat, while 187 ASVs contained the

barcode from pColE1-cat (Fig. 4C). Among all ASVs, 145 had barcodes from both plasmids, and 85 had barcodes from a single plasmid. These results show that the host range of conjugative plasmids that differ only in their origin of replication can be monitored in parallel within an environmental community.

To investigate if the origin of replication in the plasmid affects conjugative host range, we evaluated the relative abundances of ASVs having one or both barcodes representing pBBR1-cat and pColE1-cat, respectively. *Enterobacterales* was the most prevalent order among the ASVs only barcoded by pBBR1-cat (Fig. 4B), while *Pseudomonadales* was the most prevalent among organisms only barcoded by pColE1-cat. We also observed that pBBR1-cat presents a distinct barcoding pattern from pColE1-cat (Fig. S19). This can be seen in the barcoding of the four most prevalent orders. *Aeromonadaceae*, *Moraxellaceae*, and *Enterobacteriaceae* all present similar barcoding ratios for the two plasmids, as observed in the *E. coli* donor. However, *Psuedomonadaceae* presents stronger signals for the ASVs containing pBBR1-cat. Consistent with this observation, prior gene transfer measurements have shown that pBBR1 and ColE1 origins can support plasmid replication in both *Enterobacteriales* and *Pseudomonadales* [24,45], while bioinformatics has found that ColE1 plasmids are rare in *Pseudomonadales* [45]. Our results show how parallel analysis of plasmid conjugation across a microbiome can be applied to compare the relative host range of plasmids with different origins of replication.

To determine if there are significant differences in the barcoding host range of pBBR1-cat and pColE1-cat, principal component analysis (PCA) was used to compare the microbial community that was barcoded by each plasmid and the native community (Fig. 4D). This analysis revealed significant differences between the communities with the pBBR1-cat barcode, those that had the pColE1-cat barcode, and the native microbial community (Fig. S20). In

addition, we identified ASVs and taxonomic groups with barcode abundances that were significantly different between the pColE1 and pBBR1 (Fig. 4E). Taken together, these findings illustrate how differences in the host range of two conjugative plasmids can be monitored in parallel within a single microbial community using orthogonal cat-RNA.

The facile information storage enabled by cat-RNA has several advantages compared with existing nucleic acid barcoding systems [46–52]. First, cat-RNA records information within universally-expressed biomolecules at highly conserved sequences adjacent to variable sequences that can be used to distinguish taxa [53]. As such, well-established rRNA sequencing pipelines used for environmental microbiology can be used to read out community information [54]. Second, cat-RNA uses a small biomolecular writer (<500 nt) that only requires a single promoter to synthesize, while existing autonomous barcoding systems require transcription and translation [28]. Thus, RAM can easily store information in microbes across a complex microbiome by coupling cat-RNA synthesis to broad-host-range promoters [55]. Third, the use of RNA for writing and storing information is expected to minimize resource burdens, since the cost of protein synthesis required by existing memory systems is much greater than RNA synthesis [56]. While no growth defect was observed with cells expressing a cat-RNA that barcodes an essential transcript, fitness could be negatively affected under alternative growth conditions. Fourth, after cells are cured of a mobile genetic element expressing cat-RNA, the recorded information will not be vertically inherited, making it compatible with transient information storage. This characteristic can be contrasted with DNA memory devices, such as lineage tracers [57], whose power lies in the vertical inheritance of information.

To diversify this proof-of-concept cat-RNA for information storage, alternative promoters can be used to regulate cat-RNA synthesis, and donor strains that generate different

DNA modifications can be applied to minimize mobile DNA degradation by restriction-modification systems [58]. Alternative RNA guides can be created to selectively store information in a subset of consortia members by targeting RNA sequences that are only conserved in those species. By diversifying RNA barcodes, cat-RNA can be created that store information about the activity of different promoters across a range of microbes. Small, cat-RNA that store information about cellular uptake of molecules across microbial communities situated in their natural environment will revolutionize our ability to program cells using synthetic biology by allowing researchers to perform the design-build-test-learn cycle in environmental samples, and it will advance our ability to study and understand the environmental controls on the mobility of natural and synthetic DNA that is exchanged via conjugation, transduction, and transformation. Future work will be required to extend cat-RNA to other mobile genetic elements and establish its sensitivity in real-world samples and more representative environments. 16S rRNA barcoding should transform our ability to study how environmental conditions affect gene transfer host range and efficiencies across environmental communities.

Methods

Plasmid construction

Table S1 lists all plasmids. Plasmids were constructed through either PCR cloning, Gibson assembly, or Golden Gate assembly [59,60]. Sanger or whole plasmid sequencing was used to verify all constructs. The synthetic plasmids used for community-scale information storage were identical in all respects, except they differed by 5 bp in their barcode amended and their origins of replication. One plasmid used a pBBR1 origin, while the other used a ColE1 origin derived from

pMB1 (Supplementary Table 1). The sequences of these origins of replication are provided as supporting information.

Strains, growth, and transformations

Table S2 lists all strains. For all manipulations, single colonies were used to inoculate liquid cultures. *Escherichia coli:* cells were made competent by growing overnight in Lysogeny Broth (LB) medium at 37°C and 225 rpm, diluting 1:100 in fresh LB, and growing to an optical density (OD) of ~0.5. Cells were harvested by centrifugation, 3,000 g for ≥10 minutes (min), resuspended in TSS (1% w/v PEG3500, 0.5% V/V DMSO, 20 mM MgCl2 in LB) medium using a volume that concentrates 40 fold, and stored at -80°C in aliquots (200 µL). To transform, a cell aliquot was mixed with plasmid DNA (50 ng), incubated for 20 min on ice, heat shocked at 42°C for 45 seconds (s), grown for 1 hour (hr) in 2x YT medium (0.5 mL) at 37°C and 225 rpm, and plated on LB-agar medium containing antibiotics. *Shewanella oneidensis MR-1:* cells were made competent by growing in LB medium for 17 hrs at 30°C and 250 rpm, were washed 3x by pelleting using centrifugation (7,900 g, 2 min) and resuspending in 10% glycerol (V/V), and ultimately resuspended in a volume of 10% glycerol that concentrates 15 fold. Unmethylated plasmid DNA (100 ng) purified from *dam*$^-$*/dcm*$^-$ *E. coli* (New England Biolabs) was mixed with freshly prepared cells (50 µL), incubated on ice for 1 min, and electroporated using 1.2 kV for ~5 milliseconds (ms) with a 0.1 cm cuvette. After recovery in LB medium (1 mL) for 2 hrs at 30°C and 250 rpm, cells were plated on LB-agar containing antibiotics. *Vibrio natriegens:* cells were made competent by growing in Brain Heart Infusion (BHI) medium with V2 salts (204 mM NaCl, 4.2 mM KCl, 23.1 mM $MgCl_2$) for 2 hrs 37°C and 250 rpm [61]. When cells reached an OD of 0.5, they were pelleted, washed 3x times using electroporation buffer (680 mM sucrose and 7 mM K_2HPO_4, pH = 7),

resuspended in a volume that concentrates 100 fold, and frozen in aliquots (50 µL) at -80°C. To transform, a cell aliquot was mixed with plasmid DNA (100 ng), incubated on ice for 1 min, and electroporated using a 0.9 kV pulse for ~5 ms using a 0.1 cm cuvette. After recovering in BHI medium (0.5 mL) supplemented with 680 mM sucrose and V2 salts for 1 hr at 37°C and 250 rpm, cells were plated on LB-agar containing V2 salts and antibiotics. *Pseudomonas putida:* *E. coli* MFDpir, a diaminopimelic acid (DAP) auxotroph [62], containing plasmid DNA was grown in LB supplemented with 0.3 mM DAP for 10 hrs at 37°C and 250 rpm, while *P. putida* was grown in LB for 10 hrs at 37°C and 250 rpm. After mixing these cultures at a 1:1 ratio, 50 µL was added to a nitrocellulose filter on top of LB-agar containing 0.3 mM DAP, and this plate was grown overnight at 25°C. The filter was placed in a 1.5 mL tube, and phosphate-buffered saline (PBS) was added (1 mL) and vortexed for 10 s. Cells were pelleted (6,800 g for 3 min), washed with PBS (1 mL), and plated on LB-agar containing antibiotics. *Pseudomonas stutzeri:* cells were made competent by growing for 16 hrs at 30°C and 250 rpm in LB. After harvesting cells by centrifugation (16,000 g, 2 min), they were washed 4 times by resuspending in 300 mM sucrose solution. Fresh cells (70 µL) that had been concentrated 86 fold were mixed with plasmid DNA (≥200 ng), incubated at 23°C for 1 min, and electroporated using a 2.5 kV pulse for ~5 ms with a 0.1 cm width cuvette. After allowing cells to recover in LB (1 mL) for 2 hrs at 30°C and 250 rpm, cells were plated on LB-agar containing antibiotics.

Fluorescence measurements

To assess ribozyme activity across diverse microbes, our visual reporter (pRAM18) was tested in *E. coli, S. oneidensis*, *P. putida,* and *V. natriegens*. Single colonies containing pRAM18, 4 for each strain, were used to inoculate cultures containing antibiotics in a 2 mL 96-well block and grown

for ~17 hrs at 1,000 rpm in a VorTemp 56 (Labnet) shaker. LB medium was used for all cultures, except *V. natriegens* which used LB containing v2 salts. After inoculating fresh LB (1:50 dilution), cultures were grown for 3 to 8 hrs at 30 or 37°C and 1,000 rpm. Whole-cell fluorescence was performed using cells (50 µL) mixed with PBS (50 µL) in a 96-well plate. Optical density (OD) at 600 nm and sfGFP fluorescence (FL) were measured using a Tecan Spark (excitation: 485 nm; emission: 510 nm). OD and FL values were corrected by subtracting the corresponding mean values of the media blank. The FL to OD ratio (FL/OD) was calculated for each well, which was derived from a single colony. Normalized FL/OD was calculated for each colony by dividing the mean of a specified condition (described in each figure legend) and each data point presented. For analysis of conjugative vector repression, we transformed a vector that expresses mScarlet under control of the CymR promoter (pRAM24) into *E. coli* MG1655 containing or lacking a CymR expression plasmid (pRAM25). Single colonies were used to inoculate LB cultures (1 mL) containing antibiotics. After growing for 18 hrs at 37°C and 650 rpm, OD and FL were measured (excitation: 569 nm; emission: 610 nm). For flow cytometry analysis, overnight cultures of cells harboring pRAM18 and appropriate controls (pRAM16, pRAM17) were diluted 1:100 and analyzed using a Sony Biotechnology SH800 Cell Sorter. In total, 50,000 events were collected per sample (excitation: 487 nm; emission: 525 nm). Flowjo (BD Biosciences) was used to process flow cytometry data. An elliptical gate was used to isolate the densest population on the plot containing SSC-H vs FSC-H channels, containing ~80% of the population. A second gate was used on the FSC-H vs FSC-A for doublet discrimination.

Growth measurements

Single colonies (n = 12 each) of *E. coli* MG1655 harboring plasmids that express the visual reporter (pRAM18), a sfGFP positive control (pRAM17), and an empty vector (pRAM16) were used to inoculate LB cultures (1 mL) containing antibiotics. After growing cultures overnight at 30°C and 650 rpm, 1 µL was used to inoculate LB medium (99 µL) containing antibiotics in Nunc Edge 2.0 96-well plates (Thermo Fisher). Cells were grown for 24 hrs at 30°C with double orbital rotation (3 mm amplitude) in a Tecan M1000 while measuring the OD every 10 min. Maximum specific growth rates were obtained using Growth Curve Modeler [63].

RNA extraction

E. coli, S. oneidensis, *P. putida, P. stutzeri,* and *V. natriegens* transformed with plasmids for expressing different cat-RNA were grown in LB as described for sfGFP fluorescence. Stationary phase cultures were diluted 1:100 into LB medium (1 mL) containing antibiotic, except *V. natriegens* which was diluted into LB medium containing V2 salts. After growing for ~4 hrs, cells were pelleted (4,000 g, 5 min). The Maxwell® Purefood GMO and Authentication Kit (Promega) was used to extract total DNA and RNA using an automated magnetic bead purification with a Maxwell® RSC 48 instrument. In brief, cell pellets were resuspended by vortexing in CTAB buffer (1 mL), heated for 5 min at 95°C, incubated at 23°C for 2 min, and vortexed. Proteinase K (40 µL, 20 mg/mL) was added, vortexed, and incubated at 70 °C for 10 min. Lysed samples (300 µL) were added to the receiver well of the extraction cartridges along with lysis buffer (300 µL), and elution buffer (50 µL) was used to obtain total RNA and DNA. DNA was then removed by adding 2 µL of Turbo DNAse (Thermofisher) per 10 µg RNA in a 50 µL reaction volume. The sample was cleaned up using a Turbo DNA free kit (Invitrogen) by adding the chelating bead suspension (5 µL), incubating for 10 min while agitating every 2 min to resuspend the beads, and centrifuging at

10,000 g for 3 min. To obtain the RNA concentrations, the resulting supernatant (47 µL) was analyzed using a Qubit™ Broad Range (BR) Assay Kit. To ensure that all inhibitors from the kit were diluted, all samples were adjusted to an RNA concentration that was 10 fold lower than the sample having the lowest concentration following purification.

Reverse transcription quantitative PCR (RT-qPCR)

Primers (Table S3) and probes used for RT-qPCR were designed using Primer3plus with user-defined thermodynamic parameters [64]. RT-qPCR reactions quantifying barcoding by different cat-RNA designs were performed using a dye-based qPCRbio Sygreen 1-step kit (PCR Biosystems, PB25.11) in a Quantstudio-3 qPCR thermocycler (ThermoFisher). RNA template (4 µL) was used in a 10 µL total reaction volume with 0.4 µM of each primer and the provided RTase Go reverse transcriptase (0.5 µL). To determine an initial concentration, RT-qPCR amplification curves were analyzed using LinRegPCR [65], which background subtracts, log-transforms, and finds the linear phase of the amplification curve. To compare the concentration of native and barcoded 16S rRNA across microbes and measure the stability of barcoded rRNA, RT-qPCR reactions were performed with probes using PCRbio Probe 1-step Detect master mix (PCR Biosystems, PB25.41). RNA template (4 µL) was added into a 10 µL total reaction volume with 0.4 µM of each primer and 0.2 µM of each probe along with RTase Go reverse transcriptase (0.1 µL). For the same samples, the 16S rRNA quantification was performed in a 10 µL reaction using a dye-based qPCRbio Sygreen blue mix (PCR Biosystems, PB20.15) with RTase Go reverse transcriptase spiked in (0.5 µL). A custom R script was used to convert the Cq data exported from Thermo Fisher Quantstudio software into absolute quantities in copies/µL template using standard curves. In all cases, sample concentrations were converted to copies/µL template using standard curves made with known

concentrations of commercially-synthesized DNA (IDT) or purified PCR products. Scripts are available on GitHub (github.com/ppreshant/qPCR-analysis).

RNA stability measurements

To measure RNA stability, *E. coli* harboring vectors that express GFP (pRAM17) or the cat-RNA that targets U1376 (pRAM23) were grown to stationary phase in LB medium (3 mL) containing antibiotics, diluted 1 to 100 into fresh LB (3 mL), and grown for 3 hrs at 37°C. Cells were diluted to an OD of 0.5 into fresh LB medium (5 mL) containing or lacking the transcription inhibitor rifampicin (150 ng/µL). Samples (750 uL) harvested from these cultures after 0, 10, 30, 60, and 180 min incubations were frozen prior to RNA extraction and RT-qPCR analysis.

Universal cat-RNA guide design

To identify possible cat-RNA designs, a 16S rRNA sequence alignment was created using four *Gammaproteobacteria* strains and one *Bacillus* strain, including *E. coli* K-12 MG1655 (NCBI: NC_000913.3, bases 223771-225312), *P. stutzeri* NCTC10475, (NCBI: LR134482.1, bases 939610-941134), *S. oneidensis* MR-1, (NCBI: AE014299.2, bases 46116-47649), *V. natriegens* NBRC 15636, (NCBI: CP009977.1, bases 55-1617), and *Bacillus subtilis subsp. subtilis* 168, NCBI: CP053102.1, bases 30279-31832). First, we identified the uracils that were conserved across all of the 16S rRNA from each species. Those uracils within 35 nucleotides (nt) of the 3' end of the 16S rRNA were discarded because we set a minimum sequence length of 35 nt for complementary sequence to the 50 nt guide RNA. All uracils within 5 nt of the 5' end of the 16S rRNA were also discarded, because they could not be sequenced if barcoded. Second, we used Clustal Omega multiple sequence alignment to establish the conservation in the 5 nt upstream to

421 the conserved uracils [66], where the IGS in the cat-RNA will be designed to anneal, which is
422 designated the P1 stem. To maximize IGS binding in synthetic cat-RNA designs, only those uracils
423 and adjacent P1 stems with 100% sequence conservation were used for further analysis. Third, we
424 created a consensus sequence corresponding to the 50 nt downstream of every completely
425 conserved P1-uracil, which corresponds to the region that is complementary to the synthetic guide
426 sequences in cat-RNA. To do this, we used BioPython's "dumb_consensus" with a threshold of
427 0.70, and denoted any base pairs as ambiguous *N* if below this threshold [67]. The consensus score
428 *S(x)*, which ranged from 0 to 1, was calculated using the consensus sequence as follows:

429 $$S(x) = \frac{\sum_{i=1}^{i=n} \alpha(x_i)}{n}$$

430 where x_i is the identity of the base at location i of the consensus sequence having length n, and
431 $a(x_i)$ is 1 if x_i = A, U, C, or G and $a(x_i)$ is 0 if x_i is N. As such, the consensus score provides a
432 simple metric of sequence conservation. Finally, we designed complementary cat-RNA guide
433 sequences (50 nt) based on 4 sequences with the high consensus scores, which were used to guide
434 the construction of plasmids that express four different cat-RNA. These are named based on the
435 uracil targeted for the trans splicing reaction in *E. coli* 16S rRNA, which included U17, U891,
436 U1376, and U1490 (Table S4). Community conjugation studies all used cat-RNA U1376. The
437 variable regions in *E. coli* 16S rRNA used for comparisons were previously reported [68].
438
439 Amplicon sequencing of barcoded rRNA in *E. coli*
440 To convert extracted RNA to DNA for amplicon sequencing, purified RNA (~500 ng) was
441 combined with primers (0.5 µL each, 2 µM), dNTPs (0.5 µL, 10 mM), and water to a final volume
442 of 6.5 µL. RNA was denatured at 65°C (5 min), cooled on ice for 5 min, and the following reagents
443 (Thermofisher) were added: dithiothreitol (0.5 µL, 100 mM), 5x first strand buffer (2 µL),

RNAseOUT ribonuclease inhibitor (0.5 µL, 40 U/µl), Superscript III reverse transcriptase (0.25 µL, 200 U/µL), and water (2.5 µL). Reverse transcription was carried out at 55°C for 1 hr, heat inactivation was done at 75°C for 15 min, and the resulting cDNA was stored at -80°C. PCR amplification was performed in a 25 µL reaction by combining cDNA (0.5 to 1 µL), forward and reverse primers (0.125 µL each, 100 mM), dNTPs (0.5 µL, 10 mM), 5x Q5 reaction buffer (5 µL), and Q5 DNA polymerase (0.25 µL, 2 U/µL). The following PCR protocol was used: 1 cycle of 98°C for 1 min, 25 cycles of 98 °C for 10 s, 60°C for 25 s, and 72°C for 20 s, and 1 cycle of 72°C for 2 min. An aliquot of the product (3 µL) was analyzed on a 1% TAE-agarose gel to visualize the amplicon. Gel analysis confirmed the presence of the expected barcoded product (552 bp) along with an additional side product. All potential PCR products having a range of sizes were purified using a silica spin column (Epoch) and sequenced by Genewiz using their Amplicon-EZ protocol.

Barcoding specificity analysis

Deep sequencing of barcoded rRNA in *E. coli* was returned as paired-end reads, with the forward and reverse reads in separate files. Even though all PCR products are expected to contain a splice site, since they were generated using one primer that anneals to the 16S rRNA and a second primer that anneals to the barcode, the forward and reverse sequencing reads will only detect the splice site in cases where it is proximal to the sequencing adapters, since Amplicon-EZ sequencing yields short reads. The reverse reads are all expected to begin with the barcode sequence, which is proximal to the splice site junction between 16S rRNA and the barcode (<79 bp), so all of these reads are predicted to contain splicing information. The forward reads begin with 16S rRNA sequence information, so only a subset of these reads are expected to yield splicing information,

given their wide range of potential sizes. For example, forward reads of small PCR products (<250 bp) are expected to contain a splice site, while forward reads of longer PCR products (>250 bp) are expected to yield sequences that are completely 16S rRNA sequence as the reads do not extend to the splicing site. Due to these constraints, only the reverse sequencing reads were used for the identification of the splice site within 16S rRNA. Cutadapt was used to first find reads with the reverse primer sequence (AACCTTCGGGCATGG) used for PCR, and then it was used to remove the primer sequences from these reads using the default parameters [69], with the exception of setting the minimum overlap to 5 and minimum length to 1. The forward PCR primer site (AACGCGAAGAACCTTAC) was also removed for small PCR products where the reverse read covered the entire barcode, 16S rRNA, and forward PCR primer sequence. If no reverse primer sequence was identified in a read, the read was discarded. Using Cutadapt, the barcode sequence was identified and removed in each read to yield each 16S rRNA sequence ending in the corresponding splice site. If no barcode sequence was identified in a read, the read was discarded. Using a custom python script, identical reads were then dereplicated and the reverse complement of each unique read was generated. The splice site was determined by aligning each unique read to the *E. coli* 16S rRNA sequence. The point where the alignment ended was designated the splice site. The P1 stem was defined as the nucleotides immediately upstream of the splice site in the 16S rRNA sequence.

Wastewater sampling

Mixed liquor (50 mL) was collected on March 8th, 2022 (Fig. 3) and January 28th, 2024 (Fig. 4) from the aeration basin of the West University Wastewater Treatment plant in Houston, Texas. The sample was immediately placed on ice and transported to the laboratory where conjugation

was performed within 4 hrs of sampling. Prior to conjugation, the wastewater sample was homogenized by vortexing it. All experiments that involved wastewater samples were performed at room temperature (~25 ^{o}C).

Wastewater conjugations

E. coli MFDpir was used as a donor strain for conjugation [62]. Donor cells transformed with vectors that express CymR (pRAM25) and the cat-RNA designed to target U1376 (pRAM24 or pRAM28) were grown to late log phase (10 hrs) in LB (3 mL) containing 0.3 mM DAP and antibiotics. These cells (OD = 1) were washed 3x times with PBS (1 mL) to remove antibiotics and mixed with the wastewater microbial community at a 1:1 (V/V) ratio. An aliquot of each mixture (50 μL) was added onto a nitrocellulose filter that had been placed on LB-agar medium containing DAP (0.3 mM). After incubating at 25°C for 24 hrs, filters were removed, placed in a 1.5 mL tube, and washed with PBS (1 mL). Cells were recovered by vortexing (10 s) and centrifuging (3 min, 6800 g). The filter paper was removed, and the cells were washed 2x using PBS (1 mL). At this point, for the samples extracted on March 8 (Fig. 3), three of the six samples were spiked with ~10^5 *Vibrio natriegens*. This spike in served as a control for signals arising from splicing that occurs after conjugation during our extraction and amplification protocols. For RNA extraction, which was performed immediately following conjugation, a modified protocol of the Maxwell® RSC PureFood GMO and Authentication Kit (Promega) was used to extract nucleic acids from the cell pellet. Alterations included excluding RNAse, incubating the proteinase K at 37°C, and using 50 uL of elution buffer per sample. Samples were DNAse treated prior to reverse transcription of RNA to cDNA, as described above in “RNA Extraction.”

513 Amplicon sequencing of the wastewater microbial community native and barcoded 16S rRNA

514 As a forward primer for NGS sequencing, we used the universal 968F primer with a partial
515 Illumina adapter sequence added (described as AOS20B in Table S3). This primer anneals to
516 nucleotides 968–984 of the *E. coli* 16S rRNA. Two different reverse primers were used with this
517 forward primer. To amplify 16S rRNA, AOS57B was used. This primer is a universal primer
518 (1391R) with a partial Illumina adapter sequence added, which anneals to nucleotides 1397-1401
519 of the *E. coli* 16S rRNA. To amplify the barcoded rRNA, a custom primer (AOS62B) was used
520 that anneals to the barcode. These primers amplify variable regions V6-V8 of the 16S rRNA. Total
521 RNA was converted to cDNA as described for *E. coli*. These cDNA libraries were PCR amplified
522 in 100 µL reactions by combining cDNA (10 µL), forward and reverse primers (0.5 µL each, 100
523 µM), dNTPs (2 µL, 10 mM), 5x Q5 reaction buffer (20 µL), and Q5 DNA polymerase (1 µL, 2
524 units/µL). After dividing samples into four aliquots, touchdown PCR was used: (step 1) 98°C for
525 2 min followed by 12 cycles of 98°C for 30s, 72°C for 45 s with a 1°C/cycle decrease in
526 temperature, and 72°C for 1 min; (step 2) 2 cycles of 98°C for 30s, 60°C for 45 s, and 72°C for 1
527 min; (step 3) 19 cycles of 98°C for 30s and 72°C for 1 min 45 s, and (step 4) 72°C for 2 min and
528 4°C for 15 min. PCR products were separated on a 1% agarose gel, the expected products were
529 gel purified (516 bp for barcoded rRNA and 505 bp for native rRNA), and these products were
530 sequenced using the Amplicon-EZ service (Genewiz). When performing agarose gel analysis of
531 the PCR product, unexpected bands were observed, which were removed by gel purification.

532

533 Data analysis pipeline

534 The workflow for data analysis is shown in Fig. S21. Cutadapt was used to demultiplex the
535 barcoded reads and trim the raw sequencing reads so that both native and barcoded rRNA

sequences comprised the same homologous regions [69]. To do this, the forward primer sequences (AACGCGAAGAACCTTAC) for both barcoded and native rRNA amplicons were removed, as well as the reverse primer sequences, which differed for the native (TGACGGGCGGTGWGTRCA) and barcoded (GTTTCATGTGATCCGGATAAC) barcoded amplicons. The barcoded rRNA amplicon reads were then demultiplexed based on a 5 nt difference between the pBBR1-cat and pColE1-cat barcodes (GGGAA vs CACTC, respectively). For the barcoded rRNA amplicon reads, the barcode sequence (ATGGTGTTCAATGCTTTTCCC for pBBR1-cat and ATGGTGTTCAATGCTT for pColE1-cat) was removed from each read. Because the native 16S rRNA amplicon contained a short sequence (ACGTTCCCGGGCCT) after the splice site, this sequence was also trimmed to make the two amplicons completely homologous. Reads were required to have a minimum amount of the trimmed sequence align to the end of the read (5 nt for the primers and the sequence after the splice site, 10 nt for the barcode) and could not have mismatches that exceeded a threshold (10% for primers and barcoded, 14% for the sequence after the splice site). Any reads that lacked any of these trimmed sequences or exceeded the error tolerance were discarded. After trimming, NGS reads were imported into QIIME 2 version 2021.11 [70]. Sequence quality assessment using `demux summarize` revealed that sequences were high quality and required no additional end trimming. The DADA2 plugin was used to quality filter, denoise, demultiplex, and dereplicate the reads. Chimeras were also removed using DADA2 [71]. This processing resulted in a feature table showing a representative sequence of each unique amplicon sequencing variant (ASV), their relative abundances, and the total number of ASV counts observed in each sample. The feature table was filtered to retain only ASVs with a representative sequence $\geq$ 250 nt, that appeared in at least two samples, and had at least three total reads. Additionally, a few ASVs were derived from cat-RNA sequences, and these reads were

559 removed. A phylogenetic tree of the filtered ASVs was generated with a fragment-insertion
560 approach using the plugin `fragment-insertion sepp` [72]. Finally, taxonomy was assigned
561 to each ASV using `feature-classifier classify-sklearn` and using a Naive Bayes
562 classifier trained on the amplified 16S rRNA region [73–75]. A weighted unifrac distance matrix, and
563 corresponding principal coordinate analysis plot, was calculated using the `diversity core-`
564 `metrics-phylogenetic` plugin. Finally, differentially abundant taxa were determined by
565 analysis of the composition of microbiomes with bias correction (ANCOM-BC) by collapsing the
566 ASV table to the specified level (*i.e.*, family, genus, or none), then using the `composition`
567 `ancombc` plugin.
568
569 Comparing native and barcoded rRNA levels across ASVs
570 Sequence read abundances of native and barcoded rRNA from 6 replicate samples were vectorized
571 and linearly transformed with centered log-ratio (CLR) transformations using ALDEx2 software
572 [76]. The log transformation is base 2, reporting fold changes. To understand the variance within our
573 consortium conjugation data, samples were split into two pools, and ~10^5 *V. natriegens* cells were
574 added to three of the six replicates directly prior to RNA extraction, as described above. *V.*
575 *natriegens* was selected because it is a marine microbe that is not expected to be found in
576 wastewater. The spiked and unspiked replicates were compared using the ALDEx2 built in
577 function difference between (diff.btw). Custom software was used to further analyze the CLR
578 transformed vectors. This analysis showed that the spike in of *Vibrio* did not affect sequence
579 diversity (Fig. S22). The spike in experiment also served as a control for RAM signals arising from
580 PCR artifacts following conjugation. With this experiment, 16S rRNA sequencing of biological
581 replicates (n = 3) yielded 6.3 x 10^5 reads for microbes from the conjugation reaction (replicate

reads were 180,907, 274,921, and 180,332) and 1,184 reads of *V. natriegens* (replicate reads were 570, 252, and 362). While sequencing of barcoded 16S rRNA yielded a similar number of reads from the conjugation reaction (3.0×10^5 reads; replicate reads were 110,410, 90,797, and 99,308), we did not detect any barcodes on *V. natriegens* 16S rRNA in this reaction. This analysis illustrates the specificity of the signal for conjugation.

Determining confidence intervals for barcoding efficiencies

To quantify the relative barcoding efficiency for each ASV (*i.e.*, the barcoded to native rRNA CLR ratios), only those ASVs which yielded detectable barcoded and native rRNA in all six replicates were further analyzed. This represented ASVs from three of the orders observed in the wastewater microbial community. To compare barcoding between ASVs, a bootstrap approach (n = 1,000) was used to compute the difference between means for each pairwise comparison, *i.e.*, 28 total ASVs, and the magnitude of each difference is reported as a heatmap. To assess the certainty that a given ASV pair has distinct barcoding levels (Fig. S12), the 95% confidence interval was calculated using the bootstrap analysis. In cases where this 95% confidence level does not include a value of 0 as a potential mean difference, the pairwise difference of the ASVs was reported as certain within the heatmap.

Statistical analysis of wastewater microbial community data

A Welch t test was used to compare samples containing or lacking a *V. natriegens* spike in using the ALDEx2 software [76], with a p value < 0.1 being used as a threshold for significance. All other p-values were determined by performing two-tailed, independent t tests with using $p < 0.05$ as a threshold for significance.

605

606 Calculation of free energies for cat-RNA and target RNA binding

607 A locally-installed version of the Nucleic Acids Package (NUPACK) version 4.0.0.27 was used to

608 predict free energy terms using default parameters [77].

609

610 Shannon entropy calculations

611 To calculate the Shannon entropy $H(x_i)$, we first aligned m 16S rRNA sequences to *E. coli* 16S

612 rRNA. At each position i in each 16S rRNA sequence j, we calculated $P_c(x_i)$ as the sum of the

613 normalized probabilities $p_c(x_{i,j})$ of nucleotide c (*i.e.*, A, C, G, U, or gap) in each sequence j, where

614 there are $n = 5$ possible states considered. $P_c(x_i)$ was calculated using m sequences as:

$$P_c(x_i) = \frac{\sum_{j=0}^{m-1} p_c(x_{i,j})}{n}$$

615

616 For sequences with a single canonical bp (A, C, G, U, or gap) at position i, the probability $p_c(x_{i,j})$

617 of finding residue c in a given sequence was designated 1. However, if positions presented

618 ambiguity, *e.g.*, a purine (R), then we designated each base pair (A, G) as having an equivalent

619 probability of 0.5. $H(x_i)$ was calculated at each position i as:

$$H(x_i) = \sum_{c \in [A,U,G,C,-]} -P_c(x_i) * log_2(P_c(x_i))$$

620

621

Barcoding using non-splicing cat-RNA variants

To evaluate whether cat-RNA requires catalytic activity and an external guide sequence targeting 16S rRNA for barcoding, we created two different cat-RNA variants using the design that targets U1376. The first variant lacked the guide sequence that targets 16S rRNA, while the second variant had the guanine at position 264 mutated to an adenine; this guanine is thought to be critical for maximum activity as it participates in a critical G·U wobble pair with the target rRNA [33]. *E. coli* was transformed with constructs transcribing the original cat-RNA, cat-RNA lacking the guide, and the G264A mutant. Total RNA was extracted from cells harboring these constructs and RT-qPCR analysis was performed to quantify 16S rRNA and barcoded rRNA. These three constructs were also conjugated into a wastewater community using *E. coli* MFDpir as a donor strain, as well as an empty vector. After conjugation, total RNA was then purified, and 16S rRNA and barcoded rRNA were quantified using RT-qPCR.

Comparison of Cat-RNA and mScarlet reporters

To enable a comparison of cat-RNA with a visual reporter of conjugation, we created a plasmid having a pBBR1 origin where the production of cat-RNA and mScarlet are both under control of the same pCymR promoter (pRAM29). To evaluate the relative signals from these reporters, this plasmid was transformed into *E. coli* MG1655 containing or lacking a plasmid that expresses the CymR repressor (pRAM25). The mScarlet fluorescence for each strain was measured using a Sony SA3800 spectral flow cytometer. In brief, cultures grown in LB (500 uL) were centrifuged, resuspended in PBS, and diluted 1:100 in PBS for flow cytometry. For each sample, at least 30,000 cell-like scattering events were acquired and processed using Python based FlowCal, before plotting using the R based openCyto and ggcyto packages. The cat-RNA signal was

evaluated by RNA extraction and RT-qPCR as described for the construct that transcribes cat-RNA U1376. To compare the limit of detection for mScarlet and cat-RNA, *E. coli* MG1655 was transformed with pRAM29 and mixed at different dilutions (1:1, 1:10 to $1:10^6$) with *E. coli* transformed with a plasmid that constitutively expresses GFP for ease of flow cytometry. The cell mixtures were then analyzed using flow cytometry and RT-qPCR.

Fluorescence activated cell sorting

To compare cat-RNA and mScarlet as conjugation reporters, wastewater sludge was mixed with *E. coli* MFDpir transformed with pRAM29. This mixture was conjugated as described for pRAM24. After conjugation, the sample (1 mL in PBS) was split to allow for parallel analysis of 16S rRNA barcoding and cell sorting. Total RNA was immediately extracted from a portion of the sample (90%) and processed for sequencing of barcoded 16S rRNA in the community. The remainder of the sample (10%) was incubated at 4°C for 20 hours to allow for fluorescent protein maturation before FACS sorting following a previously reported protocol [78]. To generate a solution that yields optimal sorting density of ~10,000 events per second, cell suspensions (6.5 µL) were diluted into PBS (1 mL) containing 0.1% Tween 20 to minimize cell aggregation. A Sony SH800S sorter with a 70 µm nozzle sorting chip was used to isolate transconjugants using a two-way sort into 15 mL tubes. Sorting was performed in two stages. First, enrichment sorting was performed in the normal sorting mode, which enhanced the transconjugant fraction from ~2% to ~70%. This step was performed at a scale that yielded ~100,000 events. Second, the sorted cells were centrifuged to collect the droplets, transferred into 1.5 mL tubes, and re-sorted using the ultra-pure sorting mode. This step was performed at a scale that generated >30,000 events per replicate.

For all sorting experiments, cells were thresholded at 0.05% back scatter (BSC), with gain values of 30% for BSC, 4% for forward scatter (FSC), and 45% for fluorescence. Gates were set to identify sub-populations of cells using pure samples of wastewater sludge, conjugative donor samples expressing mGreenlantern fluorescence, and model recipient cells (*E. coli* MG1655). Cell-like particles were gated in the highest density region of the forward-side scatter (FSC-SSC) plot, and a subset of these events were marked as singlets along the diagonal of the FSC-Width vs FSC-Area graph. Events with green fluorescence higher than baseline determined using *E. coli* controls were marked as donors. Events with red fluorescence higher than baseline were marked as potential transconjugants. True transconjugants for sorting were selected using a boolean gate of [potential transconjugants] & [singlet cells] NOT [donor cells].

To compare the ASVs found in the wastewater community, those ASVs presenting a red fluorescence signal, and those ASVs presenting a 16S rRNA signal, RNA was extracted from the unsorted conjugation and the sorted transconjugants in the wastewater conjugation samples. Extracted RNA was reverse transcribed, amplified, and sequenced. Both barcoded and native rRNA was sequenced for the unsorted conjugation, whereas only native rRNA was sequenced for sorted transconjugants. NGS reads were analyzed as described using the data analysis pipeline. To allow for comparison across the different control samples, each sample was rarefied to 22,500 reads prior to phylogenetic tree generation.

References

1. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).

2. Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The Evolution and Ecology of Bacterial Warfare. *Current Biology* **29**, R521–R537 (2019).
3. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nat Rev Genet* **16**, 472–482 (2015).
4. Hussain, F. A. *et al.* Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science* **374**, 488–492 (2021).
5. Wang, T. *et al.* Horizontal gene transfer enables programmable gene stability in synthetic microbiota. *Nat Chem Biol* **18**, 1245–1252 (2022).
6. Kreitz, J. *et al.* Programmable protein delivery with a bacterial contractile injection system. *Nature* (2023) doi:10.1038/s41586-023-05870-7.
7. Nikel, P. I., Martínez-García, E. & de Lorenzo, V. Biotechnological domestication of pseudomonads using synthetic biology. *Nat Rev Microbiol* **12**, 368–379 (2014).
8. Brophy, J. A. N. *et al.* Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria. *Nat Microbiol* **3**, 1043–1053 (2018).
9. Blazejewski, T., Ho, H.-I. & Wang, H. H. Synthetic sequence entanglement augments stability and containment of genetic information in cells. *Science* **365**, 595–598 (2019).
10. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat Rev Microbiol* **3**, 711–721 (2005).
11. Heinemann, J. A. & Sprague, G. F. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205–209 (1989).
12. Lopatkin, A. J. *et al.* Antibiotics as a selective driver for conjugation dynamics. *Nat Microbiol* **1**, 16044 (2016).
13. Lu, T. K. & Collins, J. J. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4629–4634 (2009).
14. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241–244 (2011).
15. Zhou, H., Beltrán, J. F. & Brito, I. L. Functions predict horizontal gene transfer and the emergence of antibiotic resistance. *Sci. Adv.* **7**, eabj5056 (2021).

16. Nazarian, P., Tran, F. & Boedicker, J. Q. Modeling Multispecies Gene Flow Dynamics Reveals the Unique Roles of Different Horizontal Gene Transfer Mechanisms. *Front. Microbiol.* **9**, 2978 (2018).
17. Cheng, H.-Y., Masiello, C. A., Bennett, G. N. & Silberg, J. J. Volatile Gas Production by Methyl Halide Transferase: An In Situ Reporter Of Microbial Gene Expression In Soil. *Environ. Sci. Technol.* **50**, 8750–8759 (2016).
18. Bethke, J. H. *et al.* Environmental and genetic determinants of plasmid mobility in pathogenic *Escherichia coli*. *Sci. Adv.* **6**, eaax3173 (2020).
19. Davison, J. Genetic Exchange between Bacteria in the Environment. *Plasmid* **42**, 73–91 (1999).
20. Hughes, V. M. & Datta, N. Conjugative plasmids in bacteria of the 'pre-antibiotic' era. *Nature* **302**, 725–726 (1983).
21. Schmidt, M. & de Lorenzo, V. Synthetic constructs in/for the environment: Managing the interplay between natural and engineered Biology. *FEBS Letters* **586**, 2199–2206 (2012).
22. Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. & Wuertz, S. Studying plasmid horizontal transfer in situ: a critical review. *Nat Rev Microbiol* **3**, 700–710 (2005).
23. Brito, I. L. Examining horizontal gene transfer in microbial communities. *Nat Rev Microbiol* **19**, 442–453 (2021).
24. Ronda, C., Chen, S. P., Cabral, V., Yaung, S. J. & Wang, H. H. Metagenomic engineering of the mammalian gut microbiome in situ. *Nat Methods* **16**, 167–170 (2019).
25. Stegemann, S. & Bock, R. Exchange of Genetic Material Between Cells in Plant Tissue Grafts. *Science* **324**, 649–651 (2009).
26. Hertle, A. P., Haberl, B. & Bock, R. Horizontal genome transfer by cell-to-cell travel of whole organelles. *Sci. Adv.* **7**, eabd8215 (2021).
27. Babić, A., Lindner, A. B., Vulić, M., Stewart, E. J. & Radman, M. Direct Visualization of Horizontal Gene Transfer. *Science* **319**, 1533–1536 (2008).
28. Rubin, B. E. *et al.* Species- and site-specific genome editing in complex bacterial communities. *Nat Microbiol* **7**, 34–47 (2022).
29. Morris, E. R., Grey, H., McKenzie, G., Jones, A. C. & Richardson, J. M. A bend, flip and trap mechanism for transposon integration. *eLife* **5**, e15537 (2016).

30. Spencer, S. J. *et al.* Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J* **10**, 427–436 (2016).
31. Yaffe, E. & Relman, D. A. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nat Microbiol* **5**, 343–353 (2019).
32. Cech, T. R. The Chemistry of Self-Splicing RNA and RNA Enzymes. *Science* **236**, 1532–1539 (1987).
33. Gambill, L., Staubus, A., Mo, K. W., Ameruoso, A. & Chappell, J. A split ribozyme that links detection of a native RNA to orthogonal protein outputs. *Nat Commun* **14**, 543 (2023).
34. Waring, R. B., Towner, P., Minter, S. J. & Davies, R. W. Splice-site selection by a self-splicing RNA of Tetrahymena. *Nature* **321**, 133–139 (1986).
35. Sullenger, B. A. & Cech, T. R. Ribozyme-mediated repair of defective mRNA by targeted trans-splicing. *Nature* **371**, 619–622 (1994).
36. Been, M. D. & Cech, T. R. One binding site determines sequence specificity of Tetrahymena pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity. *Cell* **47**, 207–216 (1986).
37. Bremer, H. & Dennis, P. P. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* **3**, (2008).
38. Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S. & Cohen, S. N. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9697–9702 (2002).
39. Bhattarai-Kline, S. *et al.* Recording gene expression order in DNA by CRISPR addition of retron barcodes. *Nature* **608**, 217–225 (2022).
40. Loveless, T. B. *et al.* Open-ended molecular recording of sequential cellular events into DNA. Preprint at https://doi.org/10.1101/2021.11.05.467507 (2021).
41. Neil, K., Allard, N., Grenier, F., Burrus, V. & Rodrigue, S. Highly efficient gene transfer in the mouse gut microbiota is enabled by the Incl2 conjugative plasmid TP114. *Commun Biol* **3**, 523 (2020).
42. Alderliesten, J. B. *et al.* Effect of donor-recipient relatedness on the plasmid conjugation frequency: a meta-analysis. *BMC Microbiol* **20**, 135 (2020).
43. Li, L. *et al.* Estimating the Transfer Range of Plasmids Encoding Antimicrobial Resistance in a Wastewater Treatment Plant Microbial Community. *Environ. Sci. Technol. Lett.* **5**, 260–265 (2018).

44. Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H. & Müller, S. Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR. *Microb Cell Fact* **15**, 211 (2016).
45. Ares-Arroyo, M., Rocha, E. P. C. & Gonzalez-Zorn, B. Evolution of ColE1-like plasmids across γ-Proteobacteria: From bacteriocin production to antimicrobial resistance. *PLoS Genet* **17**, e1009919 (2021).
46. Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat Rev Genet* **19**, 718–732 (2018).
47. Munck, C., Sheth, R. U., Freedberg, D. E. & Wang, H. H. Recording mobile DNA in the gut microbiota using an Escherichia coli CRISPR-Cas spacer acquisition platform. *Nat Commun* **11**, 95 (2020).
48. Schmidt, F. *et al.* Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science* **376**, eabm6038 (2022).
49. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nat Methods* **11**, 1261–1266 (2014).
50. Sheth, R. U., Yim, S. S., Wu, F. L. & Wang, H. H. Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017).
51. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
52. Kempton, H. R., Love, K. S., Guo, L. Y. & Qi, L. S. Scalable biological signal recording in mammalian cells using Cas12a base editors. *Nat Chem Biol* **18**, 742–750 (2022).
53. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596 (2012).
54. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621–1624 (2012).
55. Johns, N. I. *et al.* Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nat Methods* **15**, 323–329 (2018).
56. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat Methods* **12**, 415–418 (2015).
57. Loveless, T. B. *et al.* Lineage tracing and analog recording in mammalian cells by single-site DNA writing. *Nat Chem Biol* **17**, 739–747 (2021).

58. Jeltsch, A. & Pingoud, A. Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J Mol Evol* **42**, 91–96 (1996).
59. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).
60. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
61. Weinstock, M. T., Hesek, E. D., Wilson, C. M. & Gibson, D. G. Vibrio natriegens as a fast-growing host for molecular biology. *Nat Methods* **13**, 849–851 (2016).
62. Ferrières, L. *et al.* Silent mischief: bacteriophage Mu insertions contaminate products of Escherichia coli random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J Bacteriol* **192**, 6418–6427 (2010).
63. Hoeflinger, J. L., Hoeflinger, D. E. & Miller, M. J. A dynamic regression analysis tool for quantitative assessment of bacterial growth written in Python. *J Microbiol Methods* **132**, 83–85 (2017).
64. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71-74 (2007).
65. Untergasser, A., Ruijter, J. M., Benes, V. & van den Hoff, M. J. B. Web-based LinRegPCR: application for the visualization and analysis of (RT)-qPCR amplification and melting data. *BMC Bioinformatics* **22**, 398 (2021).
66. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* **50**, W276–W279 (2022).
67. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
68. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods* **69**, 330–339 (2007).
69. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).
70. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857 (2019).

71. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).
72. Janssen, S. *et al.* Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems* **3**, e00021-18 (2018).
73. Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).
74. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825–2830 (2011).
75. Robeson, M. S. *et al.* RESCRIPt: Reproducible sequence taxonomy reference database management. *PLoS Comput Biol* **17**, e1009581 (2021).
76. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
77. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
78. Klümper, U. *et al.* Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J* **9**, 934–945 (2015).

Acknowledgments: We thank Dr. Kreso Josic (University of Houston) for discussions about bootstrap analysis. We thank the West University Wastewater Treatment Plant operators for their assistance with collecting samples.

Funding:

US Department of Agriculture (USDA) Biotechnology Risk Assessment Grants (BRAG) 2021-33522-35356 (JC, JJS, LBS)

National Science Foundation grant 1805901 (LBS, JJS)

National Science Foundation grant 1828869 (JJS)

National Science Foundation grant 2227526 (JC, JJS)

National Science Foundation CAREER grant 2237052 (LBS)

National Science Foundation CAREER grant 2237512 (JC)

Subcontract from the US Department of Energy under FWP 78814 at PNNL (JJS)

Army Research Office W911NF-24-2-0073 (JJS, JC, LBS)

Author contributions:

Contributions are noted in alphabetical order.

Conceptualization: JC, JJS, LBS

Methodology: AS, JC, JJS, KRG, LBS, LCL, LG, MJD, and PBK

Investigation: AS, JC, JJS, KRG, LBS, LCL, LG, MJD, and PBK

Visualization: AS, JC, JJS, KRG, LBS, LCL, LG, MJD, and PBK

Funding acquisition: JC, JJS, and LBS

Project administration: JC, JJS, and LBS

Supervision: JC, JJS, and LBS

Writing – original draft: JC, JJS, and LBS

Writing – review & editing: AS, JC, JJS, KRG, LBS, LCL, LG, MJD, and PBK

Competing interests: JC and JJS submitted a provisional patent 63/487,372 on a related technology.

Data and materials availability: Genetic constructs will be made available in Addgene and custom computer scripts uploaded to GitHub. Sequencing data will be made available through NCBI Sequence Read Archive (PRJNA954059). Other data are available in the main text or the supplementary materials.

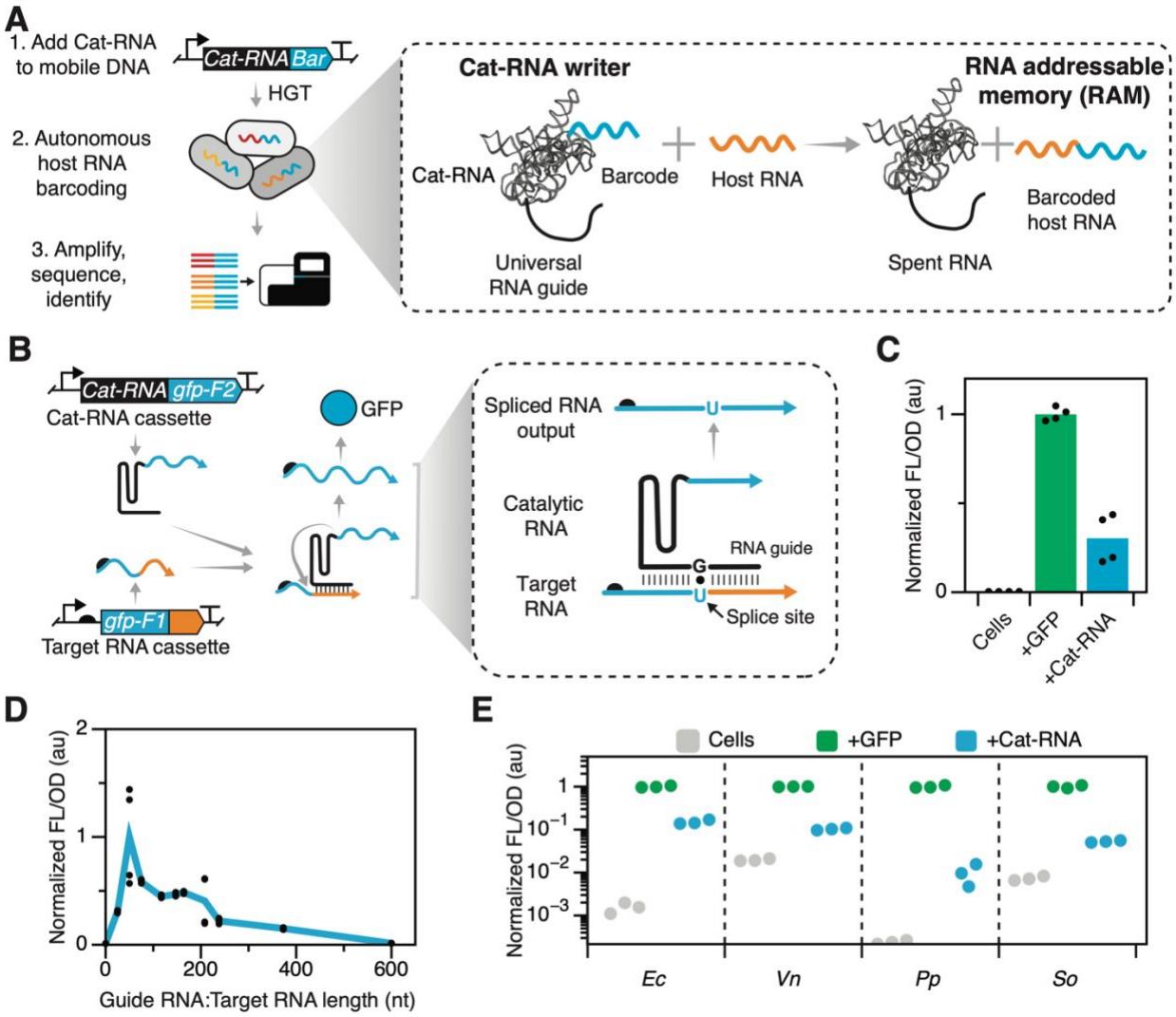

Fig. 1. RNA-Addressable Memory (RAM) is an efficient barcoding technology that functions across diverse microbes. (**A**) RAM uses a cat-RNA to splice synthetic RNA barcodes onto host RNA to record mobile DNA uptake by horizontal gene transfer (HGT). (**B**) A fluorescence assay was created that uses cat-RNA barcoding of a target RNA to produce a mRNA product that is competent for translating GFP. (**C**) Fluorescence of *E. coli* expressing the cat-RNA reporter (+Cat-RNA) or a constitutively-expressed GFP positive control (+GFP) is compared with cells harboring an empty vector (Cells). The signal from cat-RNA is significantly higher than the vector control (two-tailed, unpaired t test, $p<0.05$). (**D**) RNA guide length affects the reporter signal. RNA guides and corresponding target RNA, ranging from 25 to 373 nt, present signals that are significantly higher than a cat-RNA design lacking a guide (two-tailed, unpaired t test, $p<0.05$). Replicates are shown as points, while the line connects average values. (**E**) Fluorescence characterization of the visual reporter with a 50 nt guide in *E. coli* (*Ec*), *V. natriegens* (*Vn*), *P. putida* (*Pp*), and *S. oneidensis* (*So*). All cells present fluorescence (+Cat-RNA) that is significantly higher than the negative control (Cells), like the positive control (+GFP) (two-tailed, unpaired t test, $p<0.05$). Three or more biological replicates were acquired for each experiment.

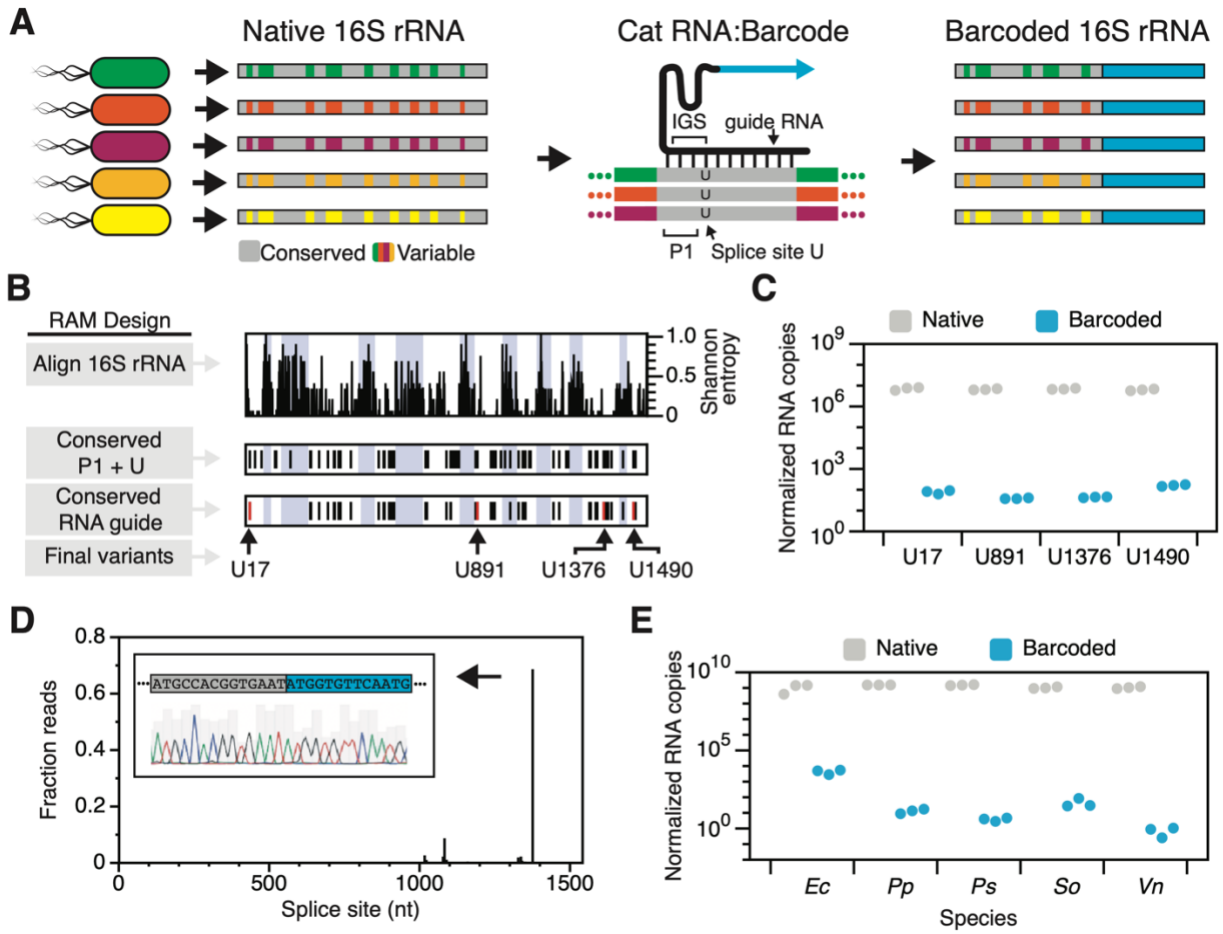

Fig. 2. Universal cat-RNA can record information in conserved 16S rRNA sequences. (**A**) Cat-RNAs that target conserved regions of 16S rRNA (gray) add barcodes (blue) such that amplicon sequencing can identify the cells that participate in gene transfer (variable). In 16S rRNA, P1 sequences adjacent to the targeted uracils (U) form duplex interactions with the internal guide sequence (IGS) in cat-RNAs. (**B**) Bioinformatic analysis of 16S rRNAs reveal the Shannon entropy, potential P1 and U splice sites conserved across microbes, and 50 nt guide RNA targets having high annealing strengths. The designs tested are in red. (**C**) Quantification of native and barcoded 16S rRNA in *E. coli* expressing each cat-RNA using RT-qPCR. (**D**) The cat-RNA designed to target U1376 generates the expected barcoded product, revealed by amplicon and Sanger sequencing (inset), with minimal non-specific barcoding. The sequence shown is from a single amplicon. (**E**) Quantification by RT-qPCR of native and barcoded 16S rRNA in *E. coli* (*Ec*), *P. putida* (*Pp*), *P. stutzeri* (*Ps*), *S. oneidensis* (*So*), and *V. natriegens* (*Vn*) transformed with plasmids encoding the cat-RNA that targets U1376. Data represent three biological replicates.

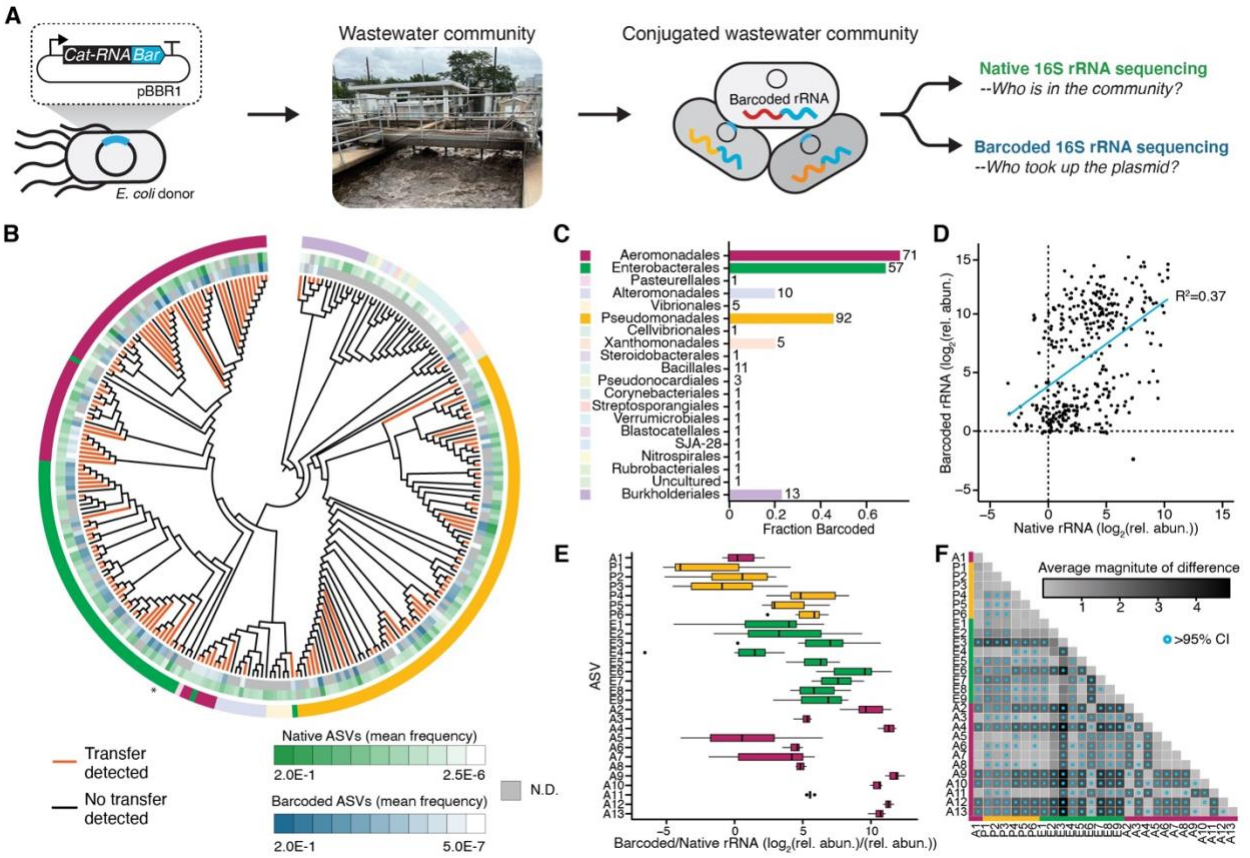

Fig. 3. Information storage in a Houston wastewater microbial community. (**A**) RAM was used to record information about conjugation within a Houston wastewater microbial community, and amplicon sequencing of the native and barcoded 16S rRNA showed which microbes were present and who participated in conjugation. (**B**) Conjugation host range (orange) is mapped onto an evolutionary tree including all amplicon sequence variants (ASVs) observed in the community. The relative abundances of native (green) and barcoded (blue) ASVs are shown, with the outer leaves showing taxonomic order. The asterisk indicates the position of the *E. coli* donor on the tree. (**C**) For each order, the fraction of total ASVs that participated in conjugation is shown, as well as the number of ASVs. (**D**) A comparison of native and barcoded 16S rRNA frequencies from individual biological replicates reveals a weak linear correlation ($y = 3.5 + 0.9x$; $R^2 = 0.37$). (**E**) For those ASVs that yielded detectable native and barcoded 16S rRNA in all six biological replicates, the median ratio of barcoded/total 16S rRNA is shown using whisker plots. (**F**) Pairwise differences in the barcoded/native 16S rRNA ratios. Bootstrapping was used to determine which differences lie outside of the 95% confidence intervals (CI) (blue points).

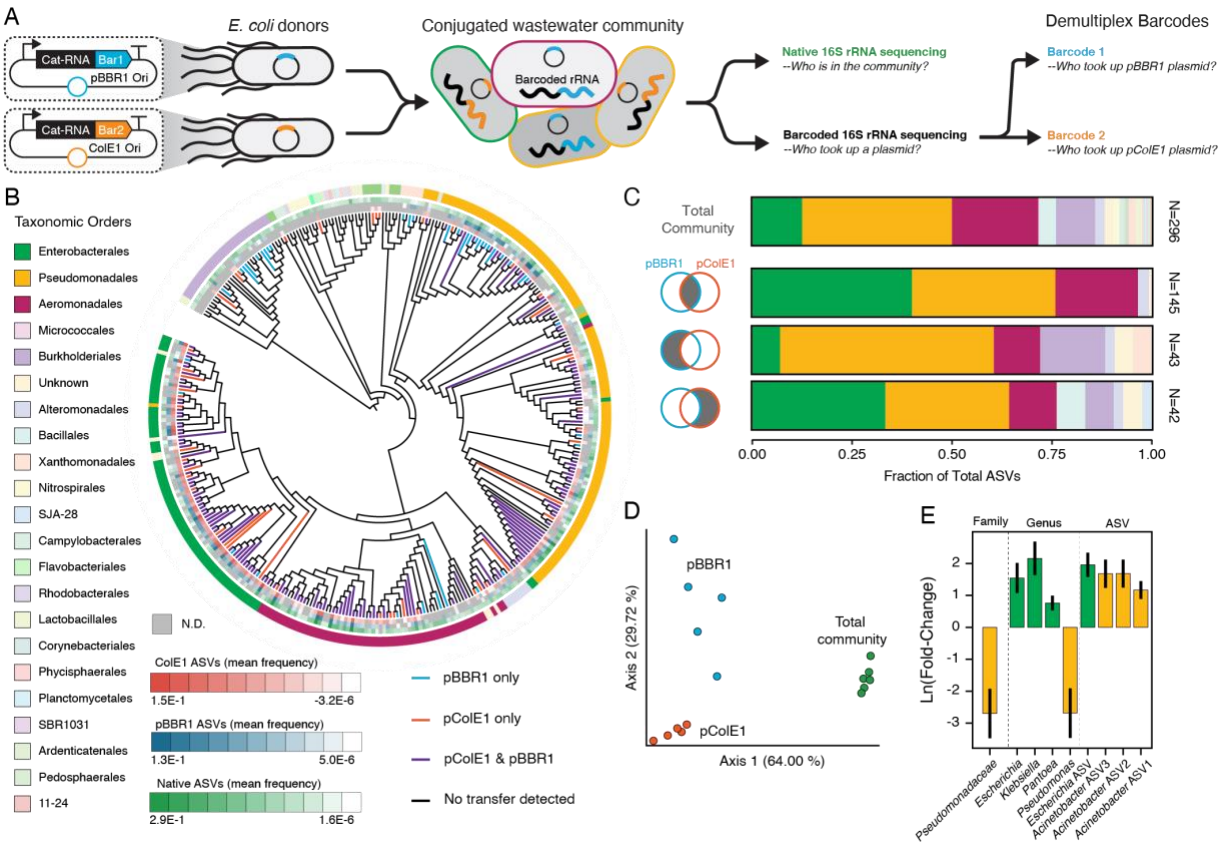

Fig. 4. Multiplexed information storage reveals differences in plasmid host range. (**A**) *E. coli* transformed with pBBR1-cat or pColE1-cat were simultaneously added to a wastewater community, and amplicon sequencing was used to evaluate conjugation. (**B**) The host range for each plasmid is mapped onto an evolutionary tree (black) to show which ASVs presented signals from the pBBR1-cat (blue), pColE1-cat (red), or both plasmids (purple). The outer leaflets of the tree show the relative abundances of native ASVs (green) and ASVs presenting signals from pBBR1-cat (blue) and pColE1-cat (red). (**C**) The relative fraction of total ASVs observed in 22 different orders colored as in B. Each bar graph was generated using data from 6 biological replicates. ASVs above the limit of detection were counted once, regardless of the total reads. (**D**) Principle coordinate analysis (PCoA) reveals distinct clustering of the pBBR1-cat (blue) and pColE1-cat (red) barcoded populations, and this clustering is distinct from the community (green). PCoA was performed on the weighted unifrac distance matrix between populations. The significance of these differences ($q \leq 0.01$) was evaluated using permutational multivariate analysis of variance (Fig. S20). (**E**) The relative abundance of taxonomic groups that presented significant differences ($q < 0.05$) in pBBR1-cat and pColE1-cat signals at the family, genera, and ASV levels. A positive value indicates a greater abundance in the pColE1-barcoded samples, while a negative value shows a greater abundance in the pBBR1-barcoded samples. Bars are colored based on taxonomic order, as in B. Error bars indicate standard error.