Controlling the taxonomic composition of biological information storage in 16S ribosomal RNA

Kiara Reyes Gamas[1], Travis R. Seamons[1], Matthew J. Dysart[1], Lin Fang[1], James Chappell[2,3], Lauren B. Stadler[4,*], and Jonathan J. Silberg[2,3,5,*]

Affiliations:

[1]Systems, Synthetic, and Physical Biology Graduate Program, Rice University, 6100 Main Street, MS-180, Houston, Texas 77005, United States of America.

[2]Department of BioSciences, Rice University, MS-140, 6100 Main Street, Houston, TX, 77005, United States of America.

[3]Department of Bioengineering, Rice University, MS-142, 6100 Main Street, Houston, TX, 77005, United States of America.

[4]Department of Civil and Environmental Engineering, Rice University, MS-352, 6100 Main Street, Houston, TX, 77005, United States of America.

[5]Department of Chemical and Biomolecular Engineering, Rice University, MS-362, 6100 Main Street, Houston, TX, 77005, United States of America.

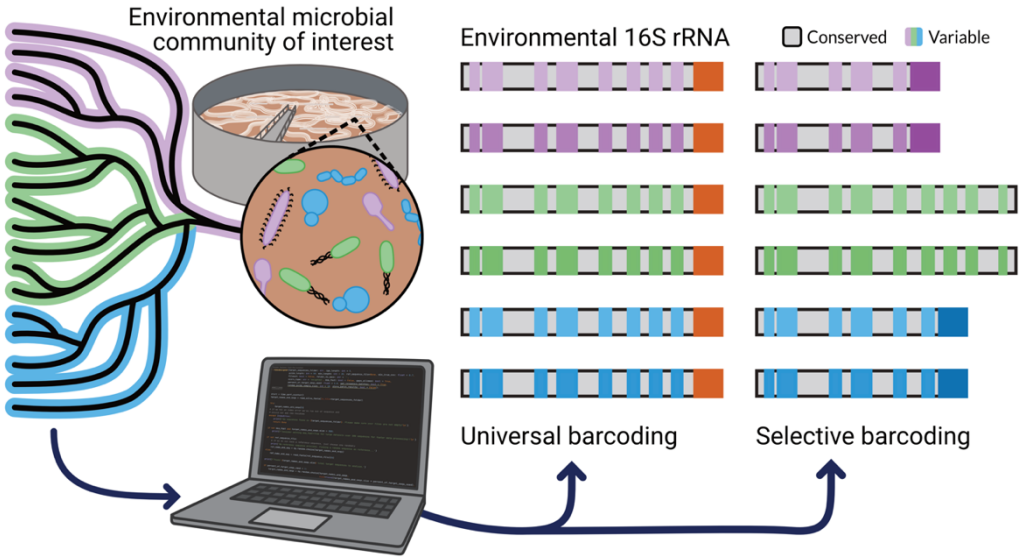
* Address Correspondence to: joff@rice.edu and lauren.stadler@rice.edu

Keywords: barcoding, conjugation, horizontal gene transfer, information storage, ribozyme, RNA, and synthetic biology

ABSTRACT

Microbes can be programmed to record participation in gene transfer by coding biological-recording devices into mobile DNA. Upon DNA uptake, these devices transcribe a catalytic RNA (cat-RNA) that binds to conserved sequences within ribosomal RNA (rRNA) and perform a trans-splicing reaction that adds a barcode to the rRNA. Existing cat-RNA designs were manually generated to be broad-host range, providing no control over the organisms that were barcoded. To achieve control over the organisms barcoded by cat-RNA, we created a program called Ribodesigner that uses input sets of rRNA sequences to create designs with varying specificities. We show how this algorithm can be used to identify designs that enable kingdom-wide barcoding, or selective barcoding of specific taxonomic groups within a kingdom. We use Ribodesigner to create cat-RNA designs that target Pseudomonadales while avoiding Enterobacterales, and we compare the performance of one design to a cat-RNA that was previously found to be broad host range. When conjugated into a mixture of *Escherichia coli* and *Pseudomonas putida*, the new design presents increased selectivity compared to a broad host range cat-RNA. Ribodesigner is expected to aid in developing cat-RNA that store information within user-defined sets of microbes in environmental communities for gene transfer studies.

GRAPHICAL ABSTRACT



INTRODUCTION

Biosensors have the potential to transform our understanding of environmental processes,[1,2] since they can couple the detection of specific environmental conditions to the production of easy-to-detect output signals.[3–5] Broadly, there are two classes of microbial sensors. First, real-time biosensors provide continuous information on environmental chemicals, generating visual or chemical outputs continuously when they encounter the inputs.[6–9] Second, memory-based biosensors record and store information about sensing using mechanisms that can be read out at a later point in time.[10–13] These latter biosensors have advantages with monitoring transient processes within environmental communities, such as gut,[14,15] soil,[16,17] and wastewater[18] communities. Further, memory biosensors can record information about complex biological processes like gene transfer,[19] which underlies the spread of antibiotic resistance[20,21] and is critical to understand when engineering microbes for environmental applications.[22]

Memory biosensors can record information in nucleic acids using either permanent sequence modifications,[23–25] or using mechanisms that enable the system to be reset.[26,27] A variety of genetic memory systems have been developed to support studies of gene transfer. Notably, CRISPR recording tapes being have been created that sense gene transfer between a gut microbial community and an engineered *Escherichia coli*.[14] More recently, mobile genetic elements built using the *mariner* transposon have been used to record gene transfer in a microbial community.[19] With this latter approach for studying gene transfer, mobile DNA host range is obtained through metagenomic sequencing, like high-throughput chromosome conformation capture.[28] While these approaches can store

information across diverse species within microbial communities, all genetic material in an environmental sample must be sequenced to read out stored information.

Recently, a catalytic RNA (cat-RNA) was described that can be coded into mobile DNA and used to record information about gene transfer host range.[29] As shown in Figure 1a, this cat-RNA, which is built using the *Tetrahymena* ribozyme,[30,31] splices an RNA barcode onto 16S ribosomal RNA (rRNA) adjacent to variable regions used for taxonomic identification. Targeted sequencing of barcoded 16S rRNA can then be used to establish which microbes in a community participated in gene transfer. Proof-of-concept measurements with this RNA addressable modification (RAM) device revealed that cat-RNA can record community-level information about participation in gene transfer within diverse gram-negative bacteria. While the prototype cat-RNA achieved broad host range activity in a wastewater community, it was designed using an alignment of five 16S rRNA. As such, there is a need for a more rigorous design approach for cat-RNA.

Herein, we describe a computational program called Ribodesigner that identifies cat-RNA designs that either maximize information storage across a set of rRNA sequences or selectively target specific groups within the set. Ribodesigner leverages design principles that have been successful in primer design software.[32–34] We use Ribodesigner to identify universal cat-RNA that are expected to maximize information storage across Archaea, Bacteria, or Eukaryotes. We also show how Ribodesigner can be used to design cat-RNA for selectively storing information in one order while avoiding a second order. Finally, we compare a pair of designs that are predicted to vary in their barcoding of Pseudomonadales and Enterobacterales and observe the predicted variation in barcoding selectivity.

RESULTS

Designing universal cat-RNA. We hypothesized that three design parameters would be important for programming cat-RNA to barcode rRNA across diverse taxonomic groups (Figure 1b). First, we hypothesized that the target 16S rRNA must contain an uracil at the targeted splicing site to support the formation of a G•U wobble and catalysis.[35,36] Second, to maximize catalytic efficiency, we posited that the pentanucleotide sequence upstream of the catalytic U must be complementary to the internal guide sequence (IGS) in the synthetic ribozyme across all five base pairs. Prior studies of trans-splicing ribozymes have revealed that splicing activity depends upon formation of a base-paired helix of P1 between the IGS and sequences adjacent to the splice site.[37,38] Third, we assumed that the ribozyme must have a guide sequence that directs annealing to the 16S rRNA sequence downstream of the catalytic U to support cat-RNA and 16S rRNA association at the desired location. The guide was included as a requirement since a recent study showed that split ribozymes can be spliced together when they are designed to anneal to a template RNA.[39] All guide designs were 50 base pairs long, because this length presented the highest efficiency in a recent trans-splicing study.[29]

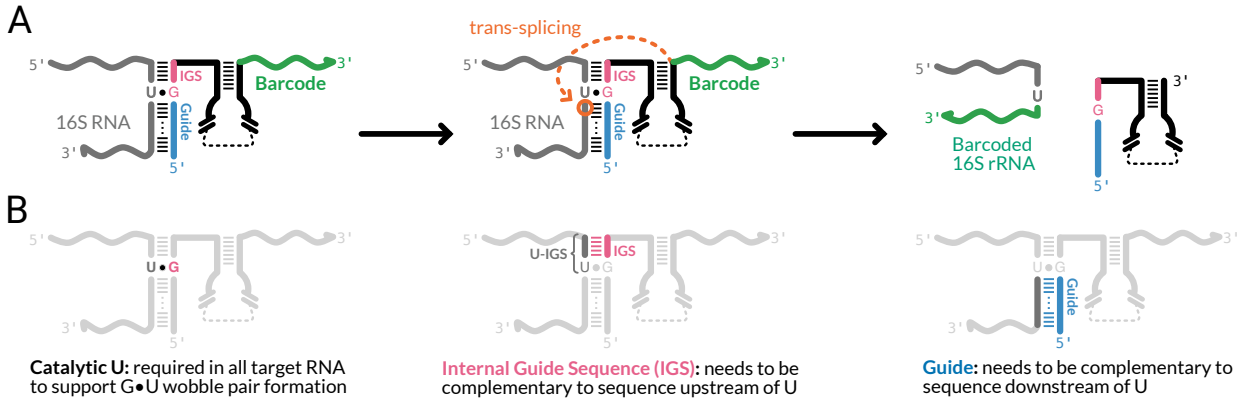


Figure 1. Cat-RNA design considerations for 16S rRNA recording. (**A**) Cat-RNA have a G that is flanked by a guide (blue) and IGS (pink), as well as a barcode (green) that is

ultimately amended to 16S rRNA. When the guide and IGS anneal to 16S rRNA, the G forms a G·U wobble pair with the U targeted for splicing, and the barcode gets spliced downstream of this site. (**B**) Ribodesigner makes three assumptions: (i) a uracil is critical for the catalytic reaction, (ii) the splicing efficiency depends upon the IGS being complementary across all five base pairs to the sequence upstream of the targeted U, and (iii) the guide must anneal to the 16S rRNA downstream of the U targeted for splicing.

To design cat-RNAs for recording information in communities, Ribodesigner identifies barcoding sites in user-defined sets of rRNA through four steps. *First*, the user provides a set of rRNA sequences for creating designs, defined as the 'design set', and a reference rRNA sequence for indexing all sequences during the analysis. The user selects the guide length and a minimum level of complementarity desired in the cat-RNA designs. *Second*, the algorithm aligns each unique rRNA sequence in the design set to the reference sequence and scores U frequencies at every position. *Third*, the algorithm assesses the frequencies of different pentanucleotide sequences upstream of each U in the rRNA set, which are designated U-IGS sequences. These sequences will ultimately serve as the IGS-binding sites for the cat-RNA. *Fourth*, the algorithm creates guide sequences for every unique U-IGS and scores them based on complementarity to a second set of rRNA sequences, defined as the 'evaluation set'.

To understand the conservation of uracils in rRNA and adjacent IGS-binding pentanucleotide sequences across bacteria, we applied Ribodesigner to a set of 2410 bacterial 16S rRNA (Figure 2a). This design set, which came from 834 bacterial orders, was chosen to capture the breadth of bacterial diversity in the SILVA rRNA database[40] and minimize degeneracy of closely-related species. Across the sequence space of this bacterial 16S rRNA design set, a wide range of U frequencies were observed. Highly conserved uracils, defined as sites found in >75% of test sequences, were dispersed across the 16S rRNA sequence (Figure 2b). Within the 16S rRNA variable regions, uracils

were more frequently flanked by IGS-binding motifs that varied in sequence (Figure 2c), while uracils between the variable regions had IGS-binding motifs with the highest conservation (Figure 2d). Across the 2410 bacterial rRNA, 44 of the positions had U-IGS with >75% identity. Similar trends were observed when Ribodesigner was applied to sets of 337 Archaeal 16S rRNA (Figure S1) and 2618 Eukaryotic 18S rRNA (Figure S2). The optimal designs observed when targeting each kingdom differed. Thus, Ribodesigner can identify highly conserved U-IGS when provided different sets of rRNA sequences.

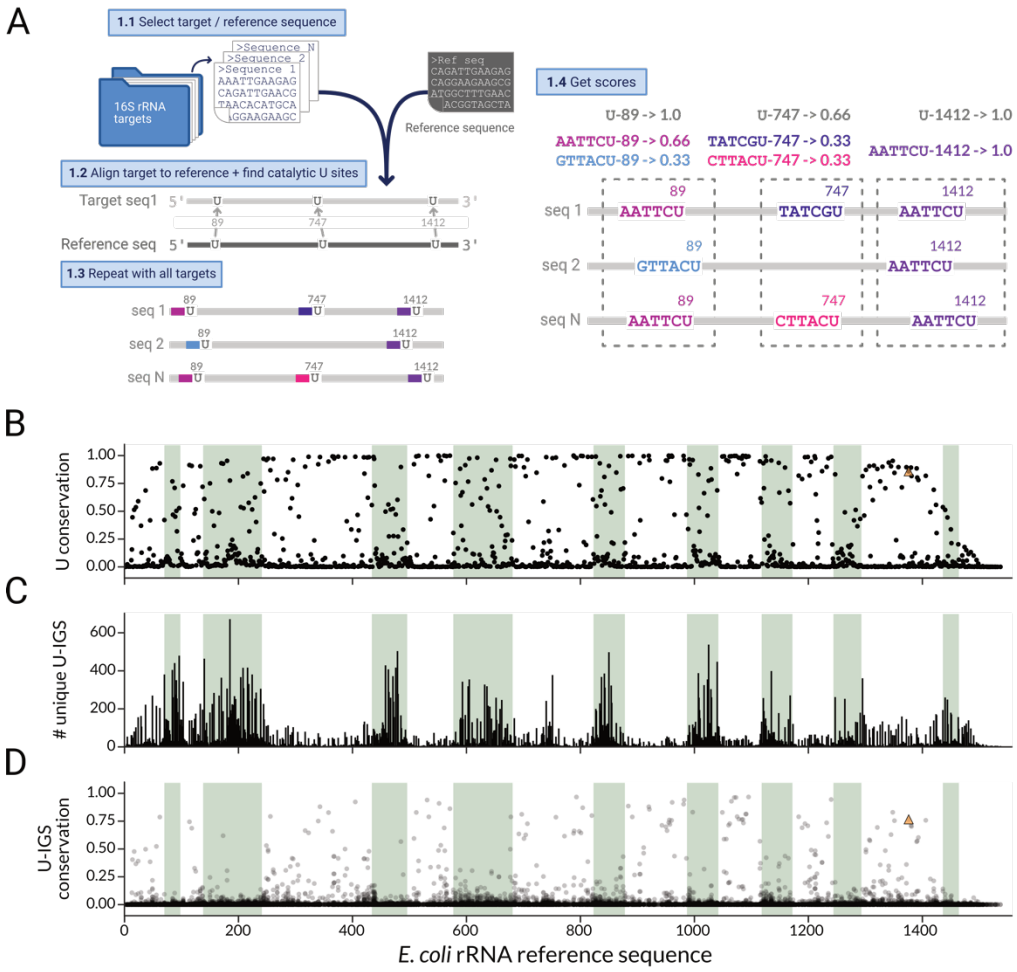


Figure 2. Uracil conservation across bacterial 16S rRNA. (**A**) The workflow that Ribodesigner uses to calculate: (**B**) the conservation of U at each 16S rRNA position in the design set, (**C**) the diversity of IGS-binding pentanucleotide sequences adjacent to each U (U-IGS), and (**D**) the conservation of each U-IGS across the design set of 16S rRNA. The bacterial 16S rRNA design set included 2410 rRNA sequences. The

hypervariable regions (green) are indexed to the reference sequence, *E. coli* MG1655. The previously characterized broad-host range cat-RNA is shown as an orange triangle.[29]

To design cat-RNA guides that maximize annealing, the 16S rRNA sequences downstream of each U-IGS are analyzed. This step is performed by generating guide sequences adjacent to each U-IGS using small batches (n = 10) of rRNA sequences from the design set and then generating a consensus sequence for each of those batches (Figure 3a). For each unique U-IGS, the consensus sequence with the lowest ambiguity (Figure S3) is identified by scoring it through alignment with 16S rRNA in the bacterial evaluation set (Figure 3b), which is distinct from the set used for design. To further improve the annealing scores of the guides, the guide sequences with the lowest ambiguities from each batch are pooled and used to find consensus sequences using the same process (Figure 3c). When the fifty designs having the highest U-IGS scores were mapped onto the 16S rRNA sequence (Figure 3d), a subset of designs were found to have high guide scores (Figure 3e), which were dispersed across the different conserved motifs. The cat-RNA previously designed manually for broad host range barcoding was among the top designs,[29] although it did not present the highest IGS and guide scores. Similar high-quality cat-RNA designs could be identified when performing this analysis with sets of 337 Archaeal rRNA (Figure S4) and 2618 Eukaryotic rRNA (Figure S5). The optimal designs targeting each kingdom were different. These analyses show how Ribodesigner can be used to identify guide designs that are predicted to support binding to diverse rRNA adjacent to highly conserved U-IGS within user-defined sets of rRNA. In addition, they suggest that different cat-RNA designs should be used to achieve broad-host range barcoding of rRNA in organisms from different kingdoms.

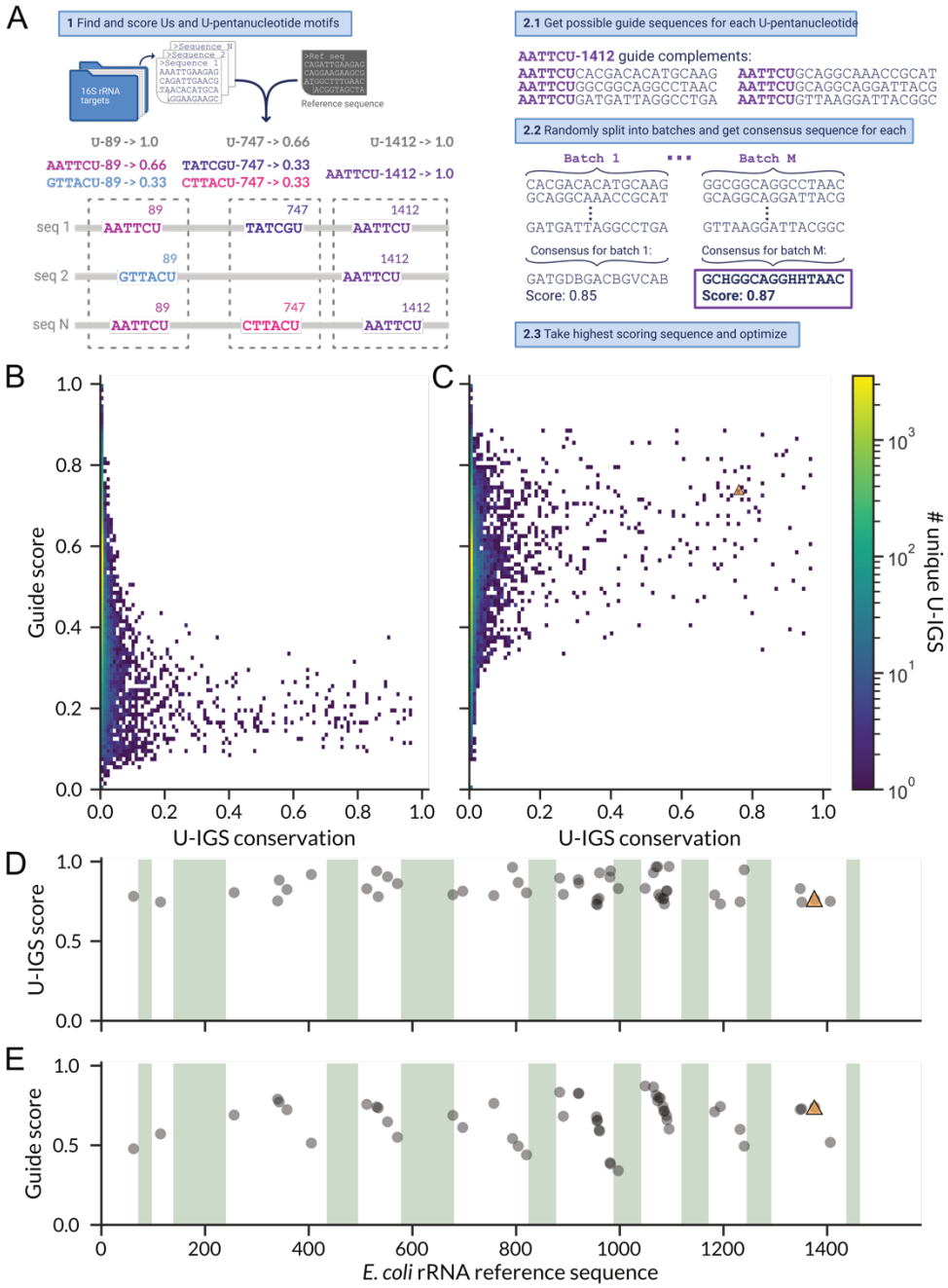


Figure 3. Designing cat-RNA guide sequences for bacteria. (**A**) For every unique U-IGS, guides are created by generating consensus sequences using small batches of 16S rRNA sequences upstream of each targeted U (n = 10). Using a design set of 2410 16S rRNA, guides were designed against all 65155 unique U-IGS motifs. (**B**) The U-IGS conservation and guide scores for all 65155 cat-RNA generated using the evaluation set of 16S rRNA. (**C**) Guide score improvements achieved by using the winning guides from each batch to create optimized guides. (**D**) For the 50 designs having the highest U-IGS scores, the location of U targeted for barcoding is mapped onto *E. coli* 16S rRNA. (**E**) For those same designs, the guide scores are also mapped onto *E. coli* 16S rRNA. The cat-RNA previously applied in a wastewater is shown as an orange triangle.[29]

Creating selective cat-RNA. The previously characterized cat-RNA exhibits broad host range in communities.[29] However, there may be cases where it is desirable to program cat-RNA to more selectively record information about gene transfer in a subset of community members that take up mobile DNA. To design cat-RNA with narrower taxonomic selectivity, we created an option in Ribodesigner to generate a selectivity score (Figure 4a). For this analysis, three sets of rRNA must be designated, including: (i) a design set of rRNA targeted for barcoding, (ii) an off-target rRNA evaluation set that are to be avoided, and (iii) an on-target evaluation set that differs from the design set. First, using the design set, cat-RNA are generated and scored. Second, the designs from this analysis are scored independently using the on- and off-target evaluation sets. Finally, a selectivity score is calculated for each design by subtracting the off-target score from the on-target score. With this approach, scores having a value of one indicate high selectivity for the target rRNA set, while lower scores indicate poorer selectivity.

To evaluate the extent to which microbes from different taxonomic groups can be differentiated using the selectivity score, we designed taxonomically-selective cat-RNA using microbes commonly found in wastewater microbiomes. For these calculations, we constructed an *in silico* community made up of microbes from three phyla, including gram-negative (*Proteobacteria*) and gram-positive (*Actinobacteriota* and *Firmicutes*) bacteria that are often abundant in wastewater.[29] This *in silico* community was assembled by creating a set of 16S rRNA sequences from the SILVA database.[40] To determine the extent to which cat-RNA can be selectively designed for each taxonomic group in the community, we calculated: (i) the conservation of every U and U-IGS for different target groups in this community, (ii) the conservation of U and U-IGS at the same sites in all

other organisms in the community, which represent non-target groups, and (iii) the differences between these scores for uracils (ΔU score) and U-IGS (ΔU-IGS score). A wide range of selectivity scores were observed with this analysis (Figure 4b). Among these, scores indicative of high selectivity (>0.75) were observed when evaluating designs targeting a subset of organisms for barcoding at the genus, family, and order levels. These findings suggest that it may be possible to achieve barcoding selectivity among different organisms at the genus, family, and order levels by targeting U for splicing that present high ΔU and ΔU-IGS scores.

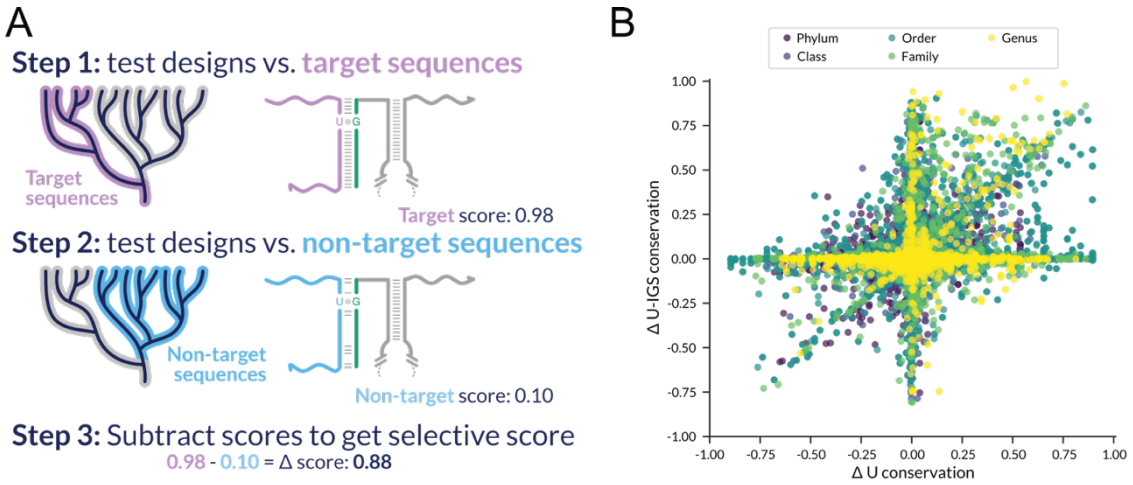


Figure 4. Evaluating the selectivity of cat-RNA designs at different taxonomic levels. (**A**) To estimate cat-RNA selectivity, Ribodesigner separately scores the cat-RNA components against a subset of the community (target sequences) and all other members of the community (non-target sequences). The program then subtracts the scores obtained for each, such that the difference indicates the selectivity for the target organisms. (**B**) For an *in silico* community, a specificity score for each U (ΔU conservation) and each U-IGS (ΔU-IGS conservation) was calculated for each taxonomic group at the phyla, class, order, family, and genera level.

The selectivity calculations with Ribodesigner suggested that cat-RNA designs can be created that barcode 16S rRNA in some organisms while avoiding barcoding in other organisms. To test this idea, we used Ribodesigner to identify cat-RNA designs that are predicted to selectively barcode Pseudomonadales while minimizing barcoding of

Enterobacterales. These calculations were performed using 3000 16S rRNA sequences from each order, which were compiled using the SILVA database.[40] The Pseudomonadales design with the highest calculated selectivity (Pcat-RNA), which is predicted to perform a trans-splicing reaction after variable region six within 16S rRNA, was compared with a broad-host range, universal cat-RNA design previously characterized in a wastewater community. When evaluating both designs with the same set of Enterobacterales used for design, the U and U-IGS scores were larger for the universal cat-RNA compared to Pcat-RNA (Figures. 5a-b). In contrast, when evaluating these designs with against the set of Pseudomonadales used for design, similar high U and U-IGS scores were obtained for both the universal cat-RNA and Pcat-RNA. Both the universal cat-RNA and Pcat-RNA presented high guide scores with both target and non-target organisms (Figure 5c). These findings identify a Pcat-RNA design that is predicted to present distinct barcoding selectivity compared to the broad-host range cat-RNA, selectively barcoding 16S rRNA in Pseudomonadales but not in Enterobacterales.

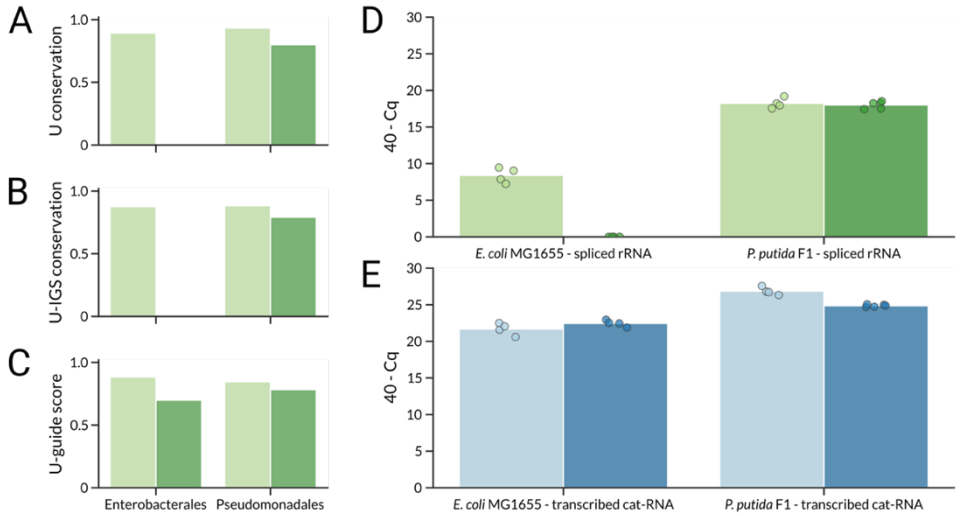


Figure 5. Comparing the selectivity of different cat-RNA designs. *In silico* comparison of the (**A**) U, (**B**) U-IGS, and (**C**) guide scores for Pcat-RNA (dark green) and universal cat-RNA (light green). Values were calculated using evaluation sets of 3000 Enterobacterales and 3000 Pseudomonadales. (**D**) The splicing efficiency of Pcat-RNA

was compared with the universal cat-RNA using qPCR in model organisms from the targeted order (*P. putida* F1) and non-targeted order (*E. coli* MG1655). With *E. coli* MG1655, the qPCR signal was below the limit of detection. (**E**) Total transcribed cat-RNA was quantified using qPCR in cells transcribing the universal cat-RNA (light blue) and Pcat-RNA (dark blue). For each experiment, four biological replicates were performed.

We next created a plasmid that transcribes the Pcat-RNA by altering the IGS and guide sequence in the previously described broad host range cat-RNA.[29] To evaluate if Pcat-RNA selectively targets *Pseudomonadales*, the plasmid was transformed into *E. coli* MG1655 and *Pseudomonas putida* F1, respectively. Strains were also transformed with a previously described broad host range cat-RNA ("universal cat-RNA") as a control.[29] Cells were grown overnight, total RNA was isolated, and total cat-RNA and barcoded 16S rRNA were quantified using reverse-transcription quantitative PCR (RT-qPCR). With this analysis, the universal cat-RNA yielded a signal for barcoded-rRNA in both *E. coli* MG1655 and *P. putida* F1 (Figure 5d). In contrast, Pcat-RNA only presented a barcoded-rRNA signal in *P. putida* F1. No signal was observed in *E. coli* MG1655. As a control, we evaluated the level of transcribed ribozymes in *E. coli* MG1655 and *P. putida* F1. Figure 5e shows that cat-RNA and Pcat-RNA were transcribed at similar levels in each organism. Further, similar levels of total 16S rRNA were observed in each of the samples (Figure S6). Together these findings demonstrate evidence of selectivity, as Pcat-RNA yielded a barcoding signal in Pseudomonadales but not in Enterobacterales, despite being transcribed to a similar extent as the universal cat-RNA.

To investigate if next-generation sequencing (NGS) can detect selective barcoding in a gene transfer experiment, we transformed *E. coli* MFDpir[41] with conjugative plasmids that transcribe Pcat-RNA and the universal cat-RNA and added each strain to a community containing *E. coli* MG1655 and *P. putida* KT2440 at similar titers (Figures 6a-b). Following conjugation for 24 hours, total RNA was isolated from each sample and

converted into complementary DNA (cDNA), and amplicon sequencing was performed (Figures S7-S8). With the universal cat-RNA, a comparison of the ratios of barcoded rRNA to native 16S rRNA read counts (Figure 6c) revealed that the signal in *E. coli* MG1655 was 9-fold higher than in *P. putida* KT2440. In contrast, Pcat-RNA presented a signal in *P. putida* KT2440 that was 26-fold higher than the ratio of barcoded rRNA to native 16S rRNA in *E. coli* MG1655. Only the signals with the Pcat-RNA were significantly different across the strains. These findings demonstrate that Ribodesigner conjugation of selective cat-RNA can store information in a targeted subset of community members.

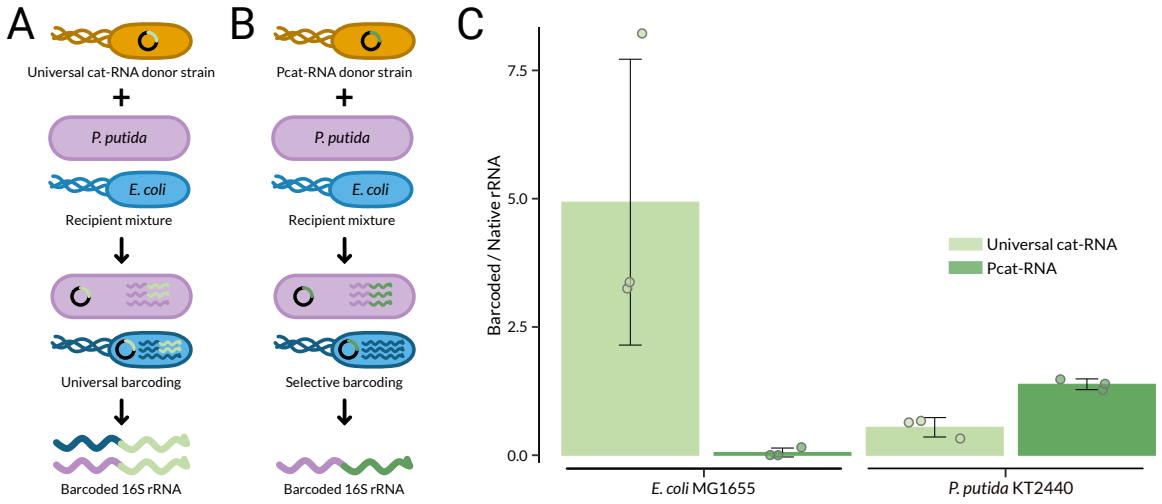


Figure 6. Cat-RNA selectivity achieved in a synthetic community. *E. coli* MFDpir was used to conjugate either (**A**) a broad host range cat-RNA (universal) or (**B**) a cat-RNA designed to be Pseudomonadales selective (Pcat-RNA) into a two-member community containing *E. coli* MG1655 and *P. putida* KT2440. Following conjugation, both native and barcoded 16S rRNA were sequenced. (**C**) Barcoded-rRNA counts normalized to 16S rRNA for the universal (light green) and Pcat-RNA (dark green). Each bar represents the ratio between rarefied counts, while error bars represent ±1 standard deviation from three biological replicates. The signal for Pcat-RNA is significantly lower in *E. coli* MG1655 compared to *P. putida* KT2440 (Welch's two tailed, unpaired t-test, $p = 0.006$), while the universal cat-RNA was not significantly different across the strains ($p = 0.1$).

Implications. Ribodesigner represents a new computational tool for designing cat-RNA with user-defined barcoding host ranges. By calculating a selectivity score, we show how

Ribodesigner can identify cat-RNA designs that are predicted to selectively barcode microbes in the order Pseudomonadales while avoiding a signal in Enterobacterales. We validate one selective cat-RNA design by showing that it can be used for conjugation in a synthetic community to increase the barcoded-rRNA signal in a targeted microbe compared with a previously described broad-host range cat-RNA. In future studies, it will be interesting to investigate how this selective cat-RNA performs in communities containing a greater diversity of Enterobacterales and Pseudomonadales, such as wastewater samples.[42,43] Selective cat-RNA designs are expected to improve RAM's ability to report on gene transfer within low-abundance microbes in communities, which will be important for studies of engineered bacteria for environmental release.[44] By enabling the selective barcoding of rRNA in a subset of organisms that participate in gene transfer, a greater fraction of the barcoded-rRNA sequencing reads is expected to represent the targeted organisms, similar to increases in sensitivity that are achieved with hybridization capture enrichment approaches.[45] Ribodesigner will also be useful for designing broad-host range cat-RNA for monitoring gene transfer within the different kingdoms of life. The archaeal designs should be useful for studying the host range of liposome-mediated transformation methods commonly used with isolates by applying them in communities rich in archaea,[46,47] while the eukaryote designs can be used to study the host range of Agrobacterium-mediated transformation methods and horizontal gene transfer in fungi.[48,49]

METHODS

Calculations. Ribodesigner, which can be found on Github (kiararey/Ribodesigner), takes in user-defined sets of small subunit (SSU) rRNA sequences and finds every possible U splicing site in each sequence and adjacent pentanucleotide sequences upstream. When creating broad-host range designs, the algorithm performs five different calculations. First, the algorithm aligns each target rRNA sequence in the 'design set' to a reference sequence, which is defined by the user, and indexes each potential splice site. Second, all potential splice sites in the design set of rRNA are scored by calculating the fraction of sequences in the set with a conserved U at the same location. Third, the algorithm calculates the conservation of the five base pairs adjacent to each U, which will ultimately bind to the IGS in the cat-RNA. In some cases, a given site in the rRNA has multiple U-IGS, which differ in their frequencies. Third, using the subset of rRNA sequences having conserved U-IGS sequences at a given site, Ribodesigner extracts every possible guide sequence that would match the 50 bases downstream of the conserved U. It uses these sequences to calculate a consensus sequence that maximizes the base pairing conservation across all target sequences. Fourth, the program scores each of these guide sequences based on their ambiguity. Fifth, the algorithm scores each cat-RNA design against an evaluation set of rRNA sequences, defined by the user which is distinct from the design set. To score quality, three parameters are calculated for every location in the indexed rRNAs, including: (i) the fraction of rRNA sequences containing a U, (ii) the fraction of rRNA sequences containing each observed U-IGS sequence, and (iii) the average correct base pairing of the guide to the test sequences. To create taxa-specific designs, Ribodesigner uses the aforementioned protocol with a design set of

rRNA sequences. Then, it scores the designs generated against an 'on-target evaluation set' of rRNA where barcoding is desired and an 'off-target evaluation set' where barcoding is to be avoided. Finally, for each cat-RNA design, it subtracts the score from the off-target evaluation from the on-target evaluation to estimate the selectivity.

Step 1: Alignment of target sequences to a reference sequence. The input to the ribodesigner.py program is a FASTA file containing a design set of target SSU sequences. These sequences are aligned to a reference sequence, defined by the user, using a global alignment with Bio.Align.PairwiseAligner in Biopython. The user also defines the reference sequences and SSU rRNA variable regions. For this study, the SSU rRNA sequences for bacteria (*Escherichia coli* MG1655), eukaryotes (*Saccharomyces cerevisiae*), and archaea (*Methanobrevibacter smithii*) differed. Variable regions were labeled for *E. coli* and *S. cerevisiae* based on literature,[50,51] while the *M. smithii* variable regions were determined by finding conserved regions with a pairwise alignment with the *E. coli* 16S rRNA sequence using the function adjust_var_regs in Ribodesigner.

Step 2: Identification and evaluation of U-IGS sequences. The program finds all possible catalytic sites, defined as all uracils in each sequence that are at least six nucleotides downstream from the 5' end and at least 50 nucleotides away from the 3' end. The latter constraint is longer because there must be sufficient sequence for cat-RNA guide annealing. Each U site is given an index corresponding to the position of that base pair in the reference sequence. The program then finds the corresponding IGS and guide sequences that a cat-RNA would require to target each U, defined as the reverse complement of the five bases upstream of the catalytic U site and the reverse complement of the 50 bases downstream of the catalytic site, respectively. This yields a variety of cat-

RNA designs, each consisting of a guide sequence at the 5' end fused to a G and an IGS. Ribodesigner aligns conserved SSU rRNA sequences, so the computational time generating adequate putative sequences can be drastically reduced by first aligning each target sequence to a shared reference sequence and finding the equivalent reference position (reference index) for each catalytic site. Ribodesigner thus does a pairwise alignment between a given reference sequence and each target sequence using a global alignment with Bio.Align.PairwiseAligner. Ribodesigner scores each design by calculating the U-IGS true coverage, defined as the fraction of all sequences in the design set with that exact U-IGS sequence at the same reference index position.

Step 3: Guide sequence generation. Once all possible U-IGS have been identified and scored in the design set of rRNA, the program extracts and pools together the corresponding complementary guide sequences, which will ultimately be used to build cat-RNA. These guides are randomly grouped into batches of 10. The guide sequences in each batch are then aligned with MUSCLE5 multiple sequence alignment.[52] A consensus sequence is generated for each batch. To generate the consensus sequence, we modified the consensus sequence generation function in RSAT to include ambiguity nucleotide codes.[53] Briefly, a list of sequences are fed into the "words2countmatrix" function, as well as a priori containing the base frequency of each base. In this study, the priori was kept as 0.25 for either A, T, G, C. This function counts the base of each sequence and adds up the frequency of each base at each position, returning a matrix with base frequency counts at each position for the input set of sequences. Then, the function "consensus" counts the number of bases at each position and returns the most likely base. If there are equal probabilities of more than one base at one position, the

corresponding ambiguity code is returned. For example, if both A and G have equal likelihood in one position, the code “R” is returned.

Step 4: Guide sequence optimization. Each batch of guides results in one consensus guide sequence. Each consensus guide sequence is scored using the weighted scoring method described in the appendix, and the highest-scoring sequence is selected as the template design for the guide. If the template contains ambiguous bases, an additional optimization protocol is performed to reduce ambiguity. Ambiguity is reduced by performing pairwise alignments between the template and each of the other batch consensus guide sequences, in order of highest to lowest score. For each pairwise alignment, at each location in the template with an ambiguous base, the corresponding base at the same location in the other sequence is compared. If the corresponding base can reduce the ambiguity (e.g., contains a shared base), then the base in the template design is replaced by the shared base. For example, if a base in the template design is ‘H’, corresponding to bases ‘A’, ‘C’, or ‘T’, and the corresponding base in the other sequence is a ‘T’, the base in the template design is replaced by the shared base ‘T’. This process repeats until either the guide score increases to 1, includes no ambiguous bases, or until the consensus sequence of each batch has been compared. For a guide sequence *x* with *n* total bases and length of *m*, the weighted score ($S_{weighted}$) is defined as:

$$S_{weighted}(x) = \frac{\sum_{i=0}^{n-1} a(x_i)}{m}$$

where $a(x_i)$ is the degenerate score for the base *x* at location *I* such that:

$$a(x_i) = \begin{cases} 1, & x_i = [A, T, U, G, C] \\ 0.5, & x_i = [R, Y, M, K, S, W] \\ \frac{1}{3}, & x_i = [H, B, D, V] \\ 0, & x_i = [N, -] \end{cases}$$

Step 5. Design evaluation using test sequences. The quality of broad host range cat-RNA designs is scored using an evaluation set of rRNA. With this evaluation, three aspects of the cat-RNA are scored, including: (i) the conservation of the U targeted in the rRNA sequences in the evaluation set, (ii) the conservation of the U-IGS sequence at the same location in the evaluation set, and (iii) the average guide conservation (guide score). The U conservation of each design is defined as the fraction of test sequences containing a U at the same index targeted by the design. The U-IGS conservation is defined as the true coverage of the design's U-IGS sequence in the test sequences with a catalytic U at the same index targeted by the design. Finally, the guide score is defined as the mean of the weighted guide score of the consensus sequence arising from the alignment between the design's guide sequence and each test sequence. The quality of selective cat-RNA designs is scored using a pair of evaluation sets of rRNA, including on- and off-target sets. Each of these scores is calculated as described for the broad host range designs. Then the specificity score is generated by subtracting the off-target score from the on-target score.

RNA sequences used for calculations. Table S1 provides a summary of the datasets used for Ribodesigner calculations, which were generated from the SILVA SSU Ref NR99 database. These data sets can be found on Github (github.com/kiararey/Ribodesigner/tree/main/Common_sequences). These data sets fall into three categories, including: (i) three design sets of target rRNA sequences from each kingdom used for creating broad-host range designs, (ii) a set of microbes whose diversity mirrors that in a wastewater community for selectivity calculations, and (iii) sets of

Enterobacterales and Pseudomonadales used for designing and testing the Pcat-RNA. The protocols for selecting these data sets are described below.

Broad-host range designs. To test whether the program could make universal designs, we needed a dataset that would be representative of a large variety of organisms while avoiding bias arising from members overrepresented in the database. To do so, we downloaded the SILVA Ref Seq SSU NR 99 dataset and chose a subset.[40] For universal design, we randomly selected 10 organisms per genus from each kingdom that did not belong to the same species. One of these datasets was used to make cat-RNA (design set) while the other was used to test designs (evaluation set). If there were fewer than 10 unique genera in an order, the design and evaluation sequences contained overlapping sequences. This resulted in 2 reduced SSU datasets for each kingdom.

Wastewater designs. To evaluate how well selectivity can be achieved, we made datasets corresponding to different taxonomic classification levels of the top three orders observed in wastewater from a previous RAM study,[29] which included Aeromonadales, Enterobacterales, and Pseudomonadales, as well as gram-positive organisms. In the SILVA database, the orders Aeromonadales and Enterobacterales are both classified as Enterobacterales. These datasets included 10 sequences per genus. We made datasets including the targeted organisms (included datasets) as well as datasets including all orders, we saw in our initial study but excluding the targeted organisms (excluded datasets). The orders we saw in our initial study were: Aeromonadales, Enterobacterales, Pasteurellales, Alteromonadales, Vibrionales, Pseudomonadales, Cellvibrionales, Xanthomonadales, Steroidobacterales, Bacillales, Pseudonocardiales,

Corynebacteriales, Streptosporangiales, Verrucomicrobiales, Blastocatellales, SJA-28, Nitrospirales, Rubrobacterales, and Burkholderiales.

Pseudomonadales-specific designs. Three unique data sets were used to design and evaluate the specificity of designs. Initially, cat-RNA were generated using a design set of 3000 Pseudomonadales, which represents the order targeted. Then, two evaluation sets were used, including an on-target evaluation set of 3000 Pseudomonadales that differed from the initial design set and an off-target evaluation set of 3000 Enterobacterales. The sequences within the design and evaluation sets for Pseudomonadales are of unique species, such that one species is not found more than once in each dataset. However, the design and evaluation sets for Pseudomonadales can have overlapping sequences between each other.

Strains. The genotypes of the strains used are provided in Table S2. NEB Turbo competent *E. coli* (New England Biolabs) was used for cloning, *E. coli* MG1655 and *P. putida* F1 were used for evaluating cat-RNA barcoding in pure culture, *E. coli* MG1655 and *P. putida* KT2440 were used as recipient strains for conjugation, and *E. coli* MFDpir was used as a donor for conjugation.[41] To evaluate cat-RNA barcoding in individual strains, *E. coli* MG1655 were transformed using heat shock, while *P. putida* were transformed using electroporation.[54] For all cloning, cells were grown in LB30 medium, which contained 5 mM magnesium and was prepared by mixing 10 g tryptone, 30 g yeast extract, 10 g NaCl, and 1 g $MgCl_2 \cdot 6\ H_2O$ in 1 L H_2O. All other experiments used LB medium. Unless noted, *E. coli* was grown at 37°C while *P. putida* was grown at 30°C.

Plasmids. Plasmids are described in Table S3. A plasmid (pRAM24) that transcribes a broad-host range cat-RNA was used as a universal barcoding system,[29] which is

kanamycin resistant and has a pBBR1 origin. To create a plasmid that selectively targets Pseudomonadales (pPribo-B0), the guide and IGS in pRAM24 were modified using Golden Gate cloning.[55] The barcode of pRAM24 was modified to yield a plasmid (pU1376-B2) that transcribes the universal barcoding system but uses a barcode that differs from the Pcat-RNA in pRibo-B0. Finally, a plasmid (pRAM25) that represses cat-RNA transcription by expressing CymR was used in the donor strain for conjugation experiments to repress barcoding;[29] this plasmid confers chloramphenicol resistance. All plasmids were sequence verified and deposited in Addgene.

RNA Extraction. Cultures of *E. coli* MG1655 or *P. putida* F1 were transformed with plasmids expressing a universal cat-RNA (pU1376-B2) and Pcat-RNA (pPribo-B0). Transformations were spread on LB-agar plates containing chloramphenicol (34 µg/mL) and kanamycin (50 µg/mL). After growing overnight, individual colonies were used to inoculate LB cultures containing chloramphenicol (34 µg/mL) and kanamycin (50 µg/mL), which were grown for 16 hours at 30°C with shaking (250 rpm). RNA was extracted using the Maxwell® Purefood GMO and Authentication Kit (Promega). Pelleted cells were resuspended in CTAB buffer (1 mL) and heated at 90°C for 5 minutes to lyse cells. Proteinase K (40 µL, 20 mg/mL) was added, mixed by vortexing, and incubated at 40°C for 10 minutes. Cell lysate (300 µL) was processed using the Maxwell® RSC 48 instrument using the Maxwell RSC PureFood GMO and Authentication Kit (Promega AS1600) and eluted using 50 µL of provided elution buffer, following the manufacturer's protocol. The eluted samples were treated with 1 µL of TURBO DNase from the DNA-free™ DNA Removal Kit (Invitrogen, AM1906) and cleaned up using the inactivation

reagent beads provided with the kit according to the manufacturer's routine DNase treatment protocol.

Quantitative PCR. Total transcribed cat-RNA and barcoded rRNA were measured via reverse transcriptase (RT) quantitative polymerase chain reaction (qPCR). Purified RNA was first diluted 125-fold to eliminate residual DNase. The abundance of transcribed cat-RNA was measured using primers and probes that bind within the barcode region of the cat-RNA. Primers and probes were synthesized by IDT (Table S4). All probes used a FAM fluorophore and ZEN quencher. The spliced RNA was measured using primers and probes that span the splice site of the barcoded 16S rRNA molecule. Reactions were prepared in 10 µL volumes using 4 µL of diluted RNA (diluted 125x) and the Clara Probe 1-Step Lo-ROX Mix (PCR Biosystems, PB25.81-01) was used for the reactions. Assays to measure native 16S rRNA were performed in 10 µL reactions containing 4 µL of diluted RNA (125X) and the qPCRBIO Sygreen 1-Step Go Lo-ROX mix (PCR Biosystems, PB25.31-01). Forward and reverse primers had a final concentration of 400 nM, while the probe was 200 nM. All reactions were prepared in MicroAmp™ Fast Optical 96-Well Reaction Plates, 0.1 mL (Applied Biosystems, 4346907). Reactions were carried out using a QuantStudio 5 thermocycler (Applied Biosystems). Reverse transcription was performed at 52°C for 10 minutes, followed by polymerase activation at 95°C for 3 minutes. Denaturation at 95°C for 15 seconds and annealing/extension at 62°C for 30 seconds was repeated for 50 cycles. A melt analysis was included for native 16S rRNA assays. Cq values were determined using Design and Analysis Software (Applied Biosystems, v 2.8.0).

Conjugation. As donor strains, *E. coli* MFDpir harboring pRAM25, which encodes the CymR repressor and represses cat-RNA expression in the donor, was transformed with pU1376-B2 or pPribo-B0. Individual colonies of these donor strains were grown overnight in LB medium (3 mL) with kanamycin, chloramphenicol, and 2,6-diaminopimelic acid (DAP, 0.3 mM). These cultures were used to inoculate cultures (3 mL) containing LB medium, DAP, and antibiotics. Each culture was washed with PBS three times to remove antibiotics and resuspended at an OD of 1. *E. coli* MG1655 and *P. putida* F1 cultures prepared in a similar manner using LB medium lacking antibiotics and DAP were mixed at a 1:1 (v/v) ratio to constitute the two-member recipient community mixture. Donor cells and the recipient community mixture were then mixed at a 1:1 (v/v) ratio. Each donor-recipient mixture (50 µL) was spotted onto a nitrocellulose filter that had been placed on LB–agar medium containing DAP (0.3 mM). After incubating at 30°C for 24 hours, filters were placed in a sterile 1.5 mL tube and resuspended in 1 mL of PBS by vortexing for 10 seconds. Cells were pelleted by centrifugation (3 minutes, 6800 x g), and filters were removed. Cell pellets were stored at -80°C until RNA extraction.

Amplicon Sequencing. Four combinations of primers were used for next-generation sequencing. For amplifying universal cat-RNA barcoded rRNA, a forward primer (oRM20) and a custom reverse primer (oRM21) that anneals to the barcode were used. For amplifying Pcat-RNA barcoded rRNA, a different forward primer (oMJD147) was selected to ensure the formation of ~500 base pair amplicon products due to differential splicing sites between the two cat-RNAs. To amplify native 16S rRNA, reverse primers were designed for both forward primers (oRM20 and oRM18, oMJD147 and oMJD150) for amplifying a ~500 base pair region between the V6 and V8 regions of the 16S rRNA.

LunaScript® Multiplex One-Step RT-PCR Kit (NEB) was used for one-step reverse-transcription PCR reactions. Reactions (25 μL) were prepared following the manufacturer's protocol with the supplementation of 10% DMSO. The following PCR setting were used: (i) 60 °C for 1 minute, 55 °C for 10 minutes, 98°C of 1 minute; (ii) followed by either 25 cycles (native 16S rRNA) or 40 cycles (barcoded rRNA) of 98°C for 10 seconds, 60°C for 20 seconds, and 72°C for 12 seconds; and (iii) 72°C for 5 minutes and hold at 4°C. PCR products were separated on a 2% agarose gel. Products with the expected size were then gel-purified, purified using the Monarch DNA Gel Extraction Kit Protocol (NEB #T1020), and sequenced using Amplicon-EZ (Azenta).

NGS Analysis Pipeline. Data analysis of sequenced amplicons was performed as described with modifications in trimming.[29] All raw reads were imported to Qiime2 (version 2024.2). The Cutadapt plugin was used for trimming raw reads of primer sequences and cat-RNA barcodes, using the default allowed error rate of 10%. For Universal cat-RNA treated samples' native rRNA amplicons, forward primer (GCAACGCGAAGAACCTTACC) and reverse primer (TGACGGGCGGTGWGTRCA) were removed. For universal cat-RNA barcoded rRNA amplicons, the forward primer (GCAACGCGAAGAACCTTACC), reverse primer (AAGTCATGCCGTTTCATGTGATC), and barcode (CGGATAACCACTCGGGAAAAGCATTGAACACCAT) were removed. For Pcat-RNA treated samples' native rRNA amplicons, the forward primer (GGATTAGATACCCTGGTAGTC) and reverse primer (CATTGTAGCACGTGTGTAGC) were removed. For Pcat-RNA barcoded rRNA amplicons, the forward primer (GGATTAGATACCCTGGTAGTC), reverse primer (AAGTCATGCCGTTTCATGTGATC), and barcode (CGGATAACGGGAAAAGCATTGAACACCAT) were removed. The DADA2

plugin was used to denoise and dereplicate reads. The resulting feature tables were filtered to retain ASVs with a representative sequence of at least 250 base pairs. Additionally, only ASVs present in at least two out of the three biological replicates and with at least three reads were kept for downstream analysis. Taxonomy was assigned using feature-classifier classify-sklearn using the pre-formatted SILVA 138 SSURef NR99 full-length classifier. All ASVs classified under Enterobacterales were considered as *E. coli*, and all ASVs classified under Pseudomonadales were considered as *P. putida*. All other ASVs (on average 1.23% of reads per sample) were discarded.

Statistical Analysis. Taxonomy-assigned read counts were rarefied to 100,000 reads prior to any statistical analysis. Pair and unpaired two-tailed Welch's T-tests were used to evaluate differences between rarefied ASV read counts. All statistical analyses were performed in RStudio (version 2023.12.1).

AUTHOR INFORMATION

Corresponding Authors

Lauren B. Stadler – Department of Civil and Environmental Engineering, Rice University, Houston, Texas, 77005, United States; http://orcid.org/0000-0001-7469-1981; ls58@rice.edu

Jonathan J. Silberg – Department of BioSciences, Rice University, Houston, Texas, 77005, United States; http://orcid.org/0000-0001-5612-0667; joff@rice.edu

Authors

Kiara Reyes Gamas – Systems, Synthetic, and Physical Biology program, Rice University, Houston, Texas, 77005, United States; http://orcid.org/0000-0002-2226-6403

Travis R. Seamons – Systems, Synthetic, and Physical Biology program, Rice University, Houston, Texas, 77005, United States; http://orcid.org/0000-0003-2099-3700

Matthew J. Dysart – Systems, Synthetic, and Physical Biology program, Rice University, Houston, Texas, 77005, United States; http://orcid.org/0000-0003-4267-3257

Lin Fang – Systems, Synthetic, and Physical Biology program, Rice University, Houston, Texas, 77005, United States; http://orcid.org/0009-0001-3932-2184

James Chappell – BioSciences Department, Rice University, Houston, Texas, 77005, United States; http://orcid.org/0000-0001-7367-1524

Author contributions

K.R.G, L.B.S, and J.J.S: Conceptualization; K.R.G, T.S., L.F., and M.J.D: Investigation; K.R.G, T.S., L.F., and M.J.D: Analysis; K.R.G, T.S., L.F., M.J.D, L.B.S, and J.J.S: Drafting manuscript; K.R.G, T.S., L.F., M.J.D, J.C., L.B.S, and J.J.S: Manuscript review and editing; K.R.G, T.S., L.F., M.J.D, L.B.S, and J.J.S: Visualization; J.C., L.B.S, and J.J.S: Funding; L.B.S and J.J.S: Supervision.

Conflict Of Interest

The authors declare no competing financial interest.

Funding

This research was supported by the United States Department of Agriculture Biotechnology Risk Assessment grant 2021-33522-35356 (to JC, JJS, LBS), and National Science Foundation grants 2227526 (to JC, JJS), 2237052 (to LBS) and 2237512 (to JC). Research was sponsored by the Army Research Office and was accomplished under Cooperative Agreement Number W911NF-24-2-0073. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government.

SUPPORTING INFORMATION

Supplementary figures and tables referenced throughout the main text, providing additional analysis of archaeal and eukaryote designs, more details on next-generation sequencing from conjugation measurements, and primers and probes used for analysis.

REFERENCES

(1) Del Valle, I.; Fulk, E. M.; Kalvapalle, P.; Silberg, J. J.; Masiello, C. A.; Stadler, L. B. Translating New Synthetic Biology Advances for Biosensing Into the Earth and Environmental Sciences. *Front. Microbiol.* **2021**, *11*, 618373. https://doi.org/10.3389/fmicb.2020.618373.

(2) Moraskie, M.; Roshid, M. H. O.; O'Connor, G.; Dikici, E.; Zingg, J.-M.; Deo, S.; Daunert, S. Microbial Whole-Cell Biosensors: Current Applications, Challenges, and Future Perspectives. *Biosensors and Bioelectronics* **2021**, *191*, 113359. https://doi.org/10.1016/j.bios.2021.113359.

(3) Ding, N.; Zhou, S.; Deng, Y. Transcription-Factor-Based Biosensor Engineering for Applications in Synthetic Biology. *ACS Synth. Biol.* **2021**, *10* (5), 911–922. https://doi.org/10.1021/acssynbio.0c00252.

(4) Bereza-Malcolm, L. T.; Mann, G.; Franks, A. E. Environmental Sensing of Heavy Metals Through Whole Cell Microbial Biosensors: A Synthetic Biology Approach. *ACS Synth. Biol.* **2015**, *4* (5), 535–546. https://doi.org/10.1021/sb500286r.

(5) Sridhar, S.; Ajo-Franklin, C. M.; Masiello, C. A. A Framework for the Systematic Selection of Biosensor Chassis for Environmental Synthetic Biology. *ACS Synth. Biol.* **2022**, *11* (9), 2909–2916. https://doi.org/10.1021/acssynbio.2c00079.

(6) Lazar, J. T.; Tabor, J. J. Bacterial Two-Component Systems as Sensors for Synthetic Biology Applications. *Current Opinion in Systems Biology* **2021**, *28*, 100398. https://doi.org/10.1016/j.coisb.2021.100398.

(7) Butcher, R. J.; Tabor, J. J. Real-Time Detection of Response Regulator Phosphorylation Dynamics in Live Bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (35), e2201204119. https://doi.org/10.1073/pnas.2201204119.

(8) Meyer, A. J.; Segall-Shapiro, T. H.; Glassey, E.; Zhang, J.; Voigt, C. A. Escherichia Coli "Marionette" Strains with 12 Highly Optimized Small-Molecule Sensors. *Nat Chem Biol* **2019**, *15* (2), 196–204. https://doi.org/10.1038/s41589-018-0168-3.

(9) Cheng, H.-Y.; Masiello, C. A.; Del Valle, I.; Gao, X.; Bennett, G. N.; Silberg, J. J. Ratiometric Gas Reporting: A Nondisruptive Approach To Monitor Gene Expression in Soils. *ACS Synth. Biol.* **2018**, *7* (3), 903–911. https://doi.org/10.1021/acssynbio.7b00405.

(10) Sheth, R. U.; Wang, H. H. DNA-Based Memory Devices for Recording Cellular Events. *Nat Rev Genet* **2018**, *19* (11), 718–732. https://doi.org/10.1038/s41576-018-0052-8.

(11) Short, A. E.; Kim, D.; Milner, P. T.; Wilson, C. J. Next Generation Synthetic Memory via Intercepting Recombinase Function. *Nat Commun* **2023**, *14* (1), 5255. https://doi.org/10.1038/s41467-023-41043-w.

(12) Yang, L.; Nielsen, A. A. K.; Fernandez-Rodriguez, J.; McClune, C. J.; Laub, M. T.; Lu, T. K.; Voigt, C. A. Permanent Genetic Memory with >1-Byte Capacity. *Nat Methods* **2014**, *11* (12), 1261–1266. https://doi.org/10.1038/nmeth.3147.

(13) Loveless, T. B.; Carlson, C. K.; Dentzel Helmy, C. A.; Hu, V. J.; Ross, S. K.; Demelo, M. C.; Murtaza, A.; Liang, G.; Ficht, M.; Singhai, A.; Pajoh-Casco, M. J.; Liu, C. C. Open-Ended Molecular Recording of Sequential Cellular Events into DNA. *Nat Chem Biol* **2025**, *21* (4), 512–521. https://doi.org/10.1038/s41589-024-01764-5.

(14) 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* **2020**, *11* (1), 95. https://doi.org/10.1038/s41467-019-14012-5.

(15) Schmidt, F.; Zimmermann, J.; Tanna, T.; Farouni, R.; Conway, T.; Macpherson, A. J.; Platt, R. J. Noninvasive Assessment of Gut Function Using Transcriptional Recording Sentinel Cells. *Science* **2022**, *376* (6594), eabm6038. https://doi.org/10.1126/science.abm6038.

(16) Casavant, N. C.; Beattie, G. A.; Phillips, G. J.; Halverson, L. J. Site-Specific Recombination-Based Genetic System for Reporting Transient or Low-Level Gene Expression. *Appl Environ Microbiol* **2002**, *68* (7), 3588–3596. https://doi.org/10.1128/AEM.68.7.3588-3596.2002.

(17) Essington, E. A.; Vezeau, G. E.; Cetnar, D. P.; Grandinette, E.; Bell, T. H.; Salis, H. M. An Autonomous Microbial Sensor Enables Long-Term Detection of TNT Explosive in Natural Soil. *Nat Commun* **2024**, *15* (1), 10471. https://doi.org/10.1038/s41467-024-54866-y.

(18) Kalvapalle, P. B.; Sridhar, S.; Silberg, J. J.; Stadler, L. B. Long-Duration Environmental Biosensing by Recording Analyte Detection in DNA Using Recombinase Memory. *Appl Environ Microbiol* **2024**, *90* (4), e02363-23. https://doi.org/10.1128/aem.02363-23.

(19) Rubin, B. E.; Diamond, S.; Cress, B. F.; Crits-Christoph, A.; Lou, Y. C.; Borges, A. L.; Shivram, H.; He, C.; Xu, M.; Zhou, Z.; Smith, S. J.; Rovinsky, R.; Smock, D. C. J.; Tang, K.; Owens, T. K.; Krishnappa, N.; Sachdeva, R.; Barrangou, R.; Deutschbauer, A. M.; Banfield, J. F.; Doudna, J. A. Species- and Site-Specific Genome Editing in Complex Bacterial Communities. *Nat Microbiol* **2021**, *7* (1), 34–47. https://doi.org/10.1038/s41564-021-01014-7.

(20) Von Wintersdorff, C. J. H.; Penders, J.; Van Niekerk, J. M.; Mills, N. D.; Majumder, S.; Van Alphen, L. B.; Savelkoul, P. H. M.; Wolffs, P. F. G. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Front. Microbiol.* **2016**, *7*. https://doi.org/10.3389/fmicb.2016.00173.

(21) He, Y.; Yuan, Q.; Mathieu, J.; Stadler, L.; Senehi, N.; Sun, R.; Alvarez, P. J. J. Antibiotic Resistance Genes from Livestock Waste: Occurrence, Dissemination, and Treatment. *npj Clean Water* **2020**, *3* (1), 4. https://doi.org/10.1038/s41545-020-0051-0.

(22) Gillett, D. L.; Selinidis, M.; Seamons, T.; George, D.; Igwe, A. N.; Del Valle, I.; Egbert, R. G.; Hofmockel, K. S.; Johnson, A. L.; Matthews, K. R. W.; Masiello, C. A.; Stadler, L. B.; Chappell, J.; Silberg, J. J. A Roadmap to Understanding and

Anticipating Microbial Gene Transfer in Soil Communities. *Microbiol Mol Biol Rev* **2025**, e00225-24. https://doi.org/10.1128/mmbr.00225-24.

(23) McKenna, A.; Findlay, G. M.; Gagnon, J. A.; Horwitz, M. S.; Schier, A. F.; Shendure, J. Whole-Organism Lineage Tracing by Combinatorial and Cumulative Genome Editing. *Science* **2016**, *353* (6298), aaf7907. https://doi.org/10.1126/science.aaf7907.

(24) Schmidt, S. T.; Zimmerman, S. M.; Wang, J.; Kim, S. K.; Quake, S. R. Quantitative Analysis of Synthetic Cell Lineage Tracing Using Nuclease Barcoding. *ACS Synth. Biol.* **2017**, *6* (6), 936–942. https://doi.org/10.1021/acssynbio.6b00309.

(25) Frieda, K. L.; Linton, J. M.; Hormoz, S.; Choi, J.; Chow, K.-H. K.; Singer, Z. S.; Budde, M. W.; Elowitz, M. B.; Cai, L. Synthetic Recording and in Situ Readout of Lineage Information in Single Cells. *Nature* **2017**, *541* (7635), 107–111. https://doi.org/10.1038/nature20777.

(26) Bonnet, J.; Subsoontorn, P.; Endy, D. Rewritable Digital Data Storage in Live Cells via Engineered Control of Recombination Directionality. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (23), 8884–8889. https://doi.org/10.1073/pnas.1202344109.

(27) Fernandez-Rodriguez, J.; Yang, L.; Gorochowski, T. E.; Gordon, D. B.; Voigt, C. A. Memory and Combinatorial Logic Based on DNA Inversions: Dynamics and Evolutionary Stability. *ACS Synth. Biol.* **2015**, *4* (12), 1361–1372. https://doi.org/10.1021/acssynbio.5b00170.

(28) Wu, R.; Davison, M. R.; Nelson, W. C.; Smith, M. L.; Lipton, M. S.; Jansson, J. K.; McClure, R. S.; McDermott, J. E.; Hofmockel, K. S. Hi-C Metagenome Sequencing Reveals Soil Phage–Host Interactions. *Nat Commun* **2023**, *14* (1), 7666. https://doi.org/10.1038/s41467-023-42967-z.

(29) Kalvapalle, P. B.; Staubus, A.; Dysart, M. J.; Gambill, L.; Reyes Gamas, K.; Lu, L. C.; Silberg, J. J.; Stadler, L. B.; Chappell, J. Information Storage across a Microbial Community Using Universal RNA Barcoding. *Nat Biotechnol* **2025**. https://doi.org/10.1038/s41587-025-02593-0.

(30) Waring, R. B.; Towner, P.; Minter, S. J.; Davies, R. W. Splice-Site Selection by a Self-Splicing RNA of Tetrahymena. *Nature* **1986**, *321* (6066), 133–139. https://doi.org/10.1038/321133a0.

(31) Cech, T. R. The Chemistry of Self-Splicing RNA and RNA Enzymes. *Science* **1987**, *236* (4808), 1532–1539. https://doi.org/10.1126/science.2438771.

(32) Jarman, S. N. Amplicon: Software for Designing PCR Primers on Aligned DNA Sequences. *Bioinformatics* **2004**, *20* (10), 1644–1645. https://doi.org/10.1093/bioinformatics/bth121.

(33) Pessoa, A. M.; Pereira, S.; Teixeira, J. PrimerIdent: A Web Based Tool for Conserved Primer Design. *Bioinformation* **2010**, *5* (2), 52–54. https://doi.org/10.6026/97320630005052.

(34) Vijaya Satya, R.; Kumar, K.; Zavaljevski, N.; Reifman, J. A High-Throughput Pipeline for the Design of Real-Time PCR Signatures. *BMC Bioinformatics* **2010**, *11* (1), 340. https://doi.org/10.1186/1471-2105-11-340.

(35) Strobel, S. A.; Cech, T. R. Minor Groove Recognition of the Conserved G·U Pair at the *Tetrahymena Ribozyme* Reaction Site. *Science* **1995**, *267* (5198), 675–679. https://doi.org/10.1126/science.7839142.

(36) Michel, F.; Hanna, M.; Green, R.; Bartel, D. P.; Szostak, J. W. The Guanosine Binding Site of the Tetrahymena Ribozyme. *Nature* **1989**, *342* (6248), 391–395. https://doi.org/10.1038/342391a0.

(37) 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* **1986**, *47* (2), 207–216. https://doi.org/10.1016/0092-8674(86)90443-5.

(38) Köhler, U.; Ayre, B. G.; Goodman, H. M.; Haseloff, J. Trans -Splicing Ribozymes for Targeted Gene Delivery 1 1Edited by J. Karn. *Journal of Molecular Biology* **1999**, *285* (5), 1935–1950. https://doi.org/10.1006/jmbi.1998.2447.

(39) 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* **2023**, *14* (1), 543. https://doi.org/10.1038/s41467-023-36073-3.

(40) Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F. O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Research* **2012**, *41* (D1), D590–D596. https://doi.org/10.1093/nar/gks1219.

(41) Ferrières, L.; Hémery, G.; Nham, T.; Guérout, A.-M.; Mazel, D.; Beloin, C.; Ghigo, J.-M. 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* **2010**, *192* (24), 6418–6427. https://doi.org/10.1128/JB.00621-10.

(42) Rodríguez-Chueca, J.; Morales, M.; Mosteo, R.; Ormad, M. P.; Ovelleiro, J. L. Inactivation of Enterococcus Faecalis, Pseudomonas Aeruginosa and Escherichia Coli Present in Treated Urban Wastewater by Coagulation—Flocculation and Photo-Fenton Processes. *Photochem Photobiol Sci* **2013**, *12* (5), 864–871. https://doi.org/10.1039/c3pp25352j.

(43) Filipič, J.; Kraigher, B.; Tepuš, B.; Kokol, V.; Mandic-Mulec, I. Effects of Low-Density Static Magnetic Fields on the Growth and Activities of Wastewater Bacteria Escherichia Coli and Pseudomonas Putida. *Bioresource Technology* **2012**, *120*, 225–232. https://doi.org/10.1016/j.biortech.2012.06.023.

(44) Chemla, Y.; Sweeney, C. J.; Wozniak, C. A.; Voigt, C. A. Design and Regulation of Engineered Bacteria for Environmental Release. *Nat Microbiol* **2025**, *10* (2), 281–300. https://doi.org/10.1038/s41564-024-01918-0.

(45) Beaudry, M. S.; Wang, J.; Kieran, T. J.; Thomas, J.; Bayona-Vásquez, N. J.; Gao, B.; Devault, A.; Brunelle, B.; Lu, K.; Wang, J.-S.; Rhodes, O. E.; Glenn, T. C. Improved Microbial Community Characterization of 16S rRNA via Metagenome

Hybridization Capture Enrichment. *Front. Microbiol.* **2021**, *12*, 644662. https://doi.org/10.3389/fmicb.2021.644662.

(46) Metcalf, W. W.; Zhang, J. K.; Apolinario, E.; Sowers, K. R.; Wolfe, R. S. A Genetic System for Archaea of the Genus *Methanosarcina* : Liposome-Mediated Transformation and Construction of Shuttle Vectors. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (6), 2626–2631. https://doi.org/10.1073/pnas.94.6.2626.

(47) Zhang, W.; Alessi, A. M.; Heaven, S.; Chong, J. P. J.; Banks, C. J. Dynamic Changes in Anaerobic Digester Metabolic Pathways and Microbial Populations during Acclimatisation to Increasing Ammonium Concentrations. *Waste Management* **2021**, *135*, 409–419. https://doi.org/10.1016/j.wasman.2021.09.017.

(48) Michielse, C. B.; Hooykaas, P. J. J.; Van Den Hondel, C. A. M. J. J.; Ram, A. F. J. Agrobacterium-Mediated Transformation as a Tool for Functional Genomics in Fungi. *Curr Genet* **2005**, *48* (1), 1–17. https://doi.org/10.1007/s00294-005-0578-0.

(49) Gonçalves, P.; Gonçalves, C. Horizontal Gene Transfer in Yeasts. *Current Opinion in Genetics & Development* **2022**, *76*, 101950. https://doi.org/10.1016/j.gde.2022.101950.

(50) Reich, M.; Labes, A. How to Boost Marine Fungal Research: A First Step towards a Multidisciplinary Approach by Combining Molecular Fungal Ecology and Natural Products Chemistry. *Marine Genomics* **2017**, *36*, 57–75. https://doi.org/10.1016/j.margen.2017.09.007.

(51) 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* **2007**, *69* (2), 330–339. https://doi.org/10.1016/j.mimet.2007.02.005.

(52) Edgar, R. C. Muscle5: High-Accuracy Alignment Ensembles Enable Unbiased Assessments of Sequence Homology and Phylogeny. *Nat Commun* **2022**, *13* (1), 6968. https://doi.org/10.1038/s41467-022-34630-w.

(53) Santana-Garcia, W.; Castro-Mondragon, J. A.; Padilla-Gálvez, M.; Nguyen, N. T. T.; Elizondo-Salas, A.; Ksouri, N.; Gerbes, F.; Thieffry, D.; Vincens, P.; Contreras-Moreira, B.; van Helden, J.; Thomas-Chollier, M.; Medina-Rivera, A. RSAT 2022: Regulatory Sequence Analysis Tools. *Nucleic Acids Research* **2022**, *50* (W1), W670–W676. https://doi.org/10.1093/nar/gkac312.

(54) Chaves, J. E.; Wilton, R.; Gao, Y.; Munoz, N. M.; Burnet, M. C.; Schmitz, Z.; Rowan, J.; Burdick, L. H.; Elmore, J.; Guss, A.; Close, D.; Magnuson, J. K.; Burnum-Johnson, K. E.; Michener, J. K. Evaluation of Chromosomal Insertion Loci in the Pseudomonas Putida KT2440 Genome for Predictable Biosystems Design. *Metabolic Engineering Communications* **2020**, *11*, e00139. https://doi.org/10.1016/j.mec.2020.e00139.

(55) Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE* **2009**, *4* (5), e5553. https://doi.org/10.1371/journal.pone.0005553.