Northwestern University

Design and Optimization of a Genome-Engineering Platform

for Systems-Level Optimization of Synthetic Translation Systems

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemical and Biological Engineering

By

Samuel Luke Walker Gowland

Evanston, IL

December 2022

Abstract

Cellular translation is responsible for the synthesis of proteins, a highly diverse class of macromolecules that form the basis of biological function. In *Escherichia coli*, harnessing and engineering of the biomolecular components of translation, such as ribosomes, transfer RNAs (tRNAs) and aminoacyl-tRNA synthetases, has led to both biotechnology products (i.e., amylases, insulin) and an expanded genetic code. However, the engineering potential of molecular translation is hampered by the limited capabilities for rapidly sampling the large genomic space necessary to evolve well-coordinated synthetic translation networks inside cells. To address this limitation, I developed a genome engineering method inspired by the action of mobile genetic elements termed mobilization. Mobilization utilizes the stochastic action of the recombinase flippase (FLP) to generate up to ~400 million genomic insertions, deletions, or rearrangements at short flippase recognition target (FRT) sites per mL culture per OD in living *E. coli* cells. As a model, I applied this approach to evolve faster-growing *E. coli* strains living exclusively off genomically expressed tethered ribosomes. In an iterative "pulse-passaging scheme," I generated genomic libraries of cells via induction of FLP recombinase (pulse) followed by passaging the population without induction of FLP to enrich the resulting population for cells with higher fitness. I observed large structural genomic diversity across these cells, with the fastest growing strains exhibiting a 71% increase in growth rate compared to the ancestral strain. I anticipate both these strains, and the mobilization method will be useful tools for synthetic biology efforts to engineer translation systems.

Acknowledgments

This journey would not have been possible without the support and inspiration of many people. Firstly, I would like to thank my family for their constant love and support, without which I would not be the man I am today. I would like to thank all of my teachers and mentors from grade school through grad school, particularly Dr. Tsafrir Mor and Dr. Wade Van Horn from Arizona State University for continuing to provide me with opportunities even through many struggles. I would like to thank the members of my graduate school cohort at Northwestern University for being wonderful friends and great colleagues, especially Blaise Kimmel, Grant Marsden, Sam Leach, and Blake Rasor. I thank my advisor, Mike Jewett, for the opportunity to work on such interesting and rewarding projects and the oft-needed space to figure things out for myself. Thank you to the current and former members of the Jewett Lab, including my mentors Do Soon Kim and Anne d'Aquino, contributors to the ribosome subgroup including Mike Hammerling, Antje Krüger, Joon Lee, Dani Yoesep, Camila Kofman, Kosuke Seki, and Jess Willi, and others including Grant Rybnicky, Katie Warfel, Ariel Thames, Adam Silverman, Jazzy Hershewe, Andrew Hunt, Charlotte Abrahamson, Ashty Karim, Lauren Clark, Maria Cabezas, Markus Jeschek, and others who are all incredibly talented and hard-working colleagues that have made me a better and more well-rounded scientist.

Preface

At some point in the last decade or two, I unconsciously committed myself to exploring the deepest, most complex problems that bar us (humans) from living up to our ideals as members of a global species, committed to principles of compassion, justice, community, and flourishment. I had hoped to find the correct levers to pull or the right boundaries to push to tip the scales toward the basic goodness found in all people. While I have certainly lost much of my starry-eyed idealism about the way human society works since then, this set of ideals still seems to be the place that hope always brings me back to.

I began my journey seeking to better understand the human condition better by studying life itself at its most fundamental scales. My college studies in molecular biology and biochemistry led to my continuing, motivating awe at the power and complexity of biological systems as they build from the molecular scale. But I soon realized that I was jealous. The capability of the technology inside of every living cell in existence still far exceeds that of any technology manufactured by human hands. If we could just repurpose that biological power – the power of proteins and DNA and molecular biological systems and emergent function – toward our own human technologies, we could go so far in making the world a more sustainable, equitable, and joyful place.

This motivation ultimately led me to pursue my PhD in Chemical and Biological Engineering at Northwestern, and it led to my fascination with the ribosome and translation engineering project in the Jewett lab. The ambition of this project felt a bit like what it must have felt like to consider launching a rocket at the moon with people in it sixty years ago – somewhat

thoroughly insane, with no one having a clear idea of what to do or how exactly to get there, but knowing that we'll certainly learn a hell of a lot just by the act of trying.

I soon came into charge of a full-time science project for the first time in my life: we wanted to get Ribo-T onto the genome of *E. coli*. In Mike's signature advising style, the "how" and even the "why" of getting this accomplished were mostly left up to me. But there were so many questions we didn't have answers to, and that I felt we *couldn't* get answers to, in the time allotted for my PhD. Where among the 4.6 million base pairs of the *E. coli* genome would be the best place for Ribo-T? How many times would we want to put it on there? And, given that we will inevitably want to include more and more synthetic translation parts integrated into cells like this in the future, won't this challenge (with compound interest) just repeat it itself again and again?

With these ideas and ideals bouncing around my head along with a notable lack of cautious temperance that a more seasoned scientist might have, I did my absolute best to take this project off the deep end, trying to break everything at once to see how we could better build back up. In that, my personal "moonshot" mission certainly succeeded.

And, true to my own stubborn and idealistic nature, I did it by developing a toolbox that I hope can be used by others to create a ripple effect of advancements in translation engineering and synthetic biology.

List of Abbreviations

aaRS – aminoacyl-tRNA synthetase

FLP – flippase or Flp recombinase

FRT – FLP recognition target sequence

mRNA – messenger RNA

n- – native or cell-supporting translation component

o- – orthogonal translation component

OTS – orthogonal translation system

PCR (cPCR) – polymerase chain reaction (colony polymerase chain reaction)

rRNA – ribosomal RNA

RTvX – Ribo-T (tethered ribosome), version X (in chronological order of development and publication)

SQ171 – Squires 171 strain[1]

tRNA – transfer RNA

Dedication

To my mom, Kathe.

Table of Contents

List of Tables and Figures

1. Figures

2. Tables

1. Introduction

1. Introduction to translation

Translation is the cellular process responsible for the synthesis of proteins, a highly diverse class of macromolecules that form the basis of biological function. Translation is so fundamental to cell growth and maintenance that this system exists – mostly unchanged – in every known organism on Earth. The process itself is coordinated by ribosomes, large complexes comprised of rRNA and proteins organized into two major subunits. Ribosomes, with the help of amino-acid-delivering tRNAs, translate mRNAs transcribed from the genome into proteins.

The proteins made by translation then go on to perform many functional tasks for the greater organism, including catalyzing metabolic reactions, providing structural support, signaling, and regulation of homeostatic conditions, among many others. The incredible breadth of tasks achievable by these molecules is made possible by the incredibly large design space the translation system gives access to. That is, for a typical protein 300 amino acids in length, there are $300^{20} = 3.5*10^{49}$ different possibilities for its sequence.

As a biologist, I look at this system and appreciate how it enables life as we know it. As an engineer, I want to understand how we could modify or emulate translation to advance our technologies and give us the capabilities to improve our world. While we already commonly harness translation for the manufacture of products such as insulin and amylase, what could we make if we expanded the very chemistry available to proteins? How might we go about engineering translation to find practical pathways toward useful products?

2. The core challenge of engineering translation

Explicit efforts to engineer translation have been ongoing for the past 20+ years. These efforts include work done both *in vivo*, primarily through orthogonal translation systems

(OTSs),[2-9] and *in vitro*, utilizing the open reaction environment to construct and test ribosomes in a cell-free environment.[10-14] These approaches were both developed to circumvent the central challenge of ribosome and translation engineering: given that the translation system is so crucial to cell health and growth, any modifications made to a cell that disrupt its translation system even slightly have very strong negative effects on cell health, making typical mutational studies of the translation system much more difficult. OTSs circumvent this problem by working independently, with minimized interruption of or crosstalk with the native translation system, while cell-free approaches are inherently (mostly) unconstrained by cell health considerations. That is, modifications that are too toxic to incorporate into a chassis strain for cell lysate production can typically be made via creative purification, mixing, and expression in the open reaction environment of cell-free systems. These relationships are summarized in **Fig. 1-1**.
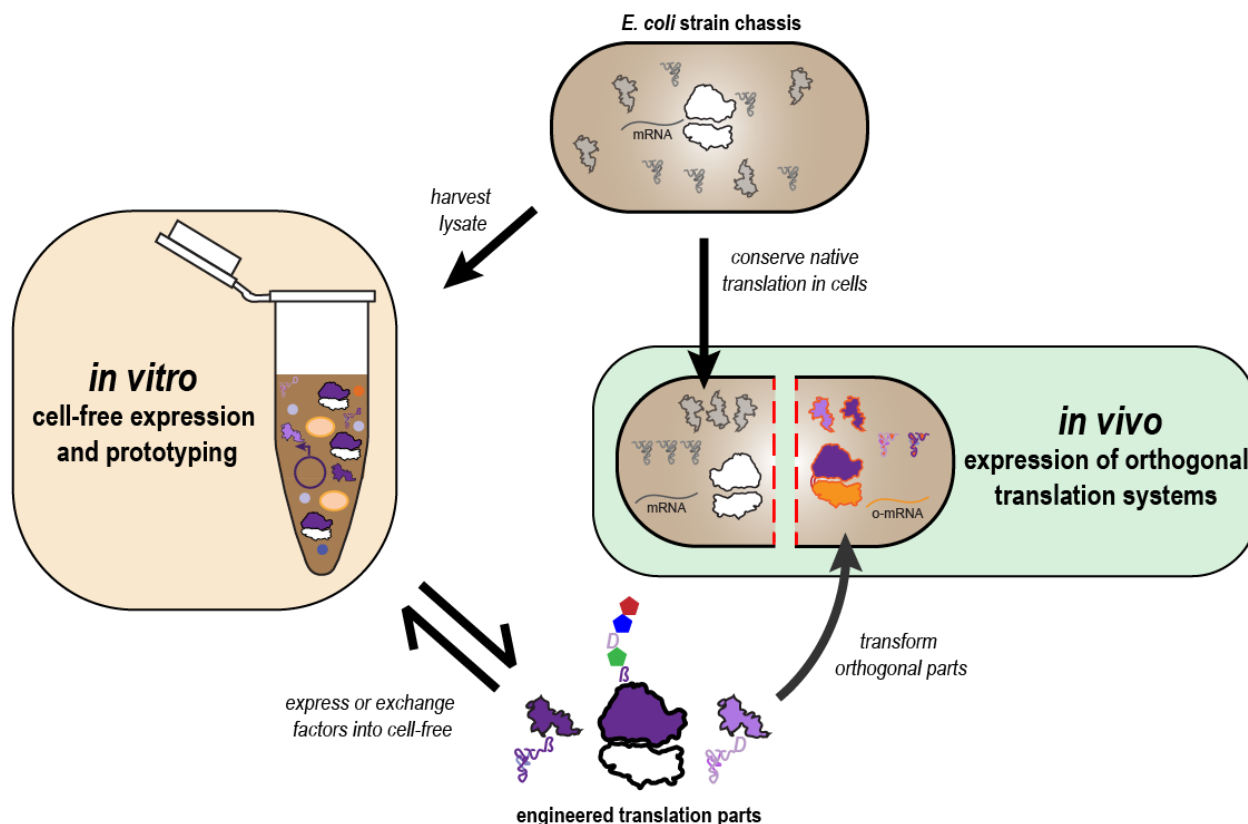
Figure 1-1. Relationships between in vitro and in vivo translation engineering platforms. Progenitor strain chassis can be harvested into cell lysates for cell-free prototyping platforms which allow highly parallelized expression and/or exchange of translation part libraries. The same or different strain chassis may be used to support orthogonal translation systems by incorporating engineered translation machinery into orthogonal circuits. These engineered cells may later become strain chassis themselves, forming a baseline for future research and production endeavors.

3. Strengths and weaknesses of in vivo and in vitro platforms for translation engineering

Both *in vivo* and *in vitro* approaches for translation engineering have significant strengths and weaknesses, and a balanced strategy that plays to the strengths of both major platforms is crucial for realizing the potential of translation engineering. Specifically, in the ribosome-engineering space, the Jewett lab's in-house-developed platform named integrated synthesis, assembly, and translation (iSAT)[10] allows the expression and assembly of ribosomes entirely outside of living cells. This allows iSAT to be an ideal prototyping platform for ribosomal variants. Other more complex platforms can also be built starting from the base iSAT platform;

for example, ribosome synthesis and evolution (RISE) utilizes the iSAT platform to express and evolve ribosome libraries.[11] Additionally, the open reaction environment of iSAT allows free exchange of factors such as synthetically-charged tRNAs into the reaction environment, providing valuable means to study questions that involve more factors than just modified ribosomes.[13] However, the iSAT platform typically produces protein yields much lower than other platforms, restricting its use for endpoint production of synthetic products. This weakness appears to occur due to the complex ribosome assembly process occurring less efficiently and with less fidelity than in cells.

By comparison, *in vivo* platforms are more restricted by cell growth and cell barrier constraints than *in vitro* platforms, but they also possess some key strengths. For one, expression and assembly of all factors occurs in a native environment honed toward efficient expression, assembly, and coordination of each individual component. When compared to cell-free expression platforms, cells show remarkable resilience and flexibility over a relatively wide range of environmental conditions such as temperature, ion concentration, and metabolic inputs. This removes the need for time-consuming optimizations necessary for cell-free expression systems. When considering a system as complex as the translation system, these self-optimization capabilities become especially useful – as optimizations in a typical design of experiments scale multiplicatively with the number of factors being optimized – and integration of efficient synthetic translation systems into chassis strains can provide new benchmarks for further *in vivo* and *in vitro* research efforts (**Fig. 1-2**).

Goal:
Production of genetically-encoded polymers at scale

Integrate parts into cells for systems-level function

Engineer parts for orthogonal function *in vivo*

Goal:
Dissemination of synthetic translation capabilities to the broader community

Engineer parts for function *in vitro*
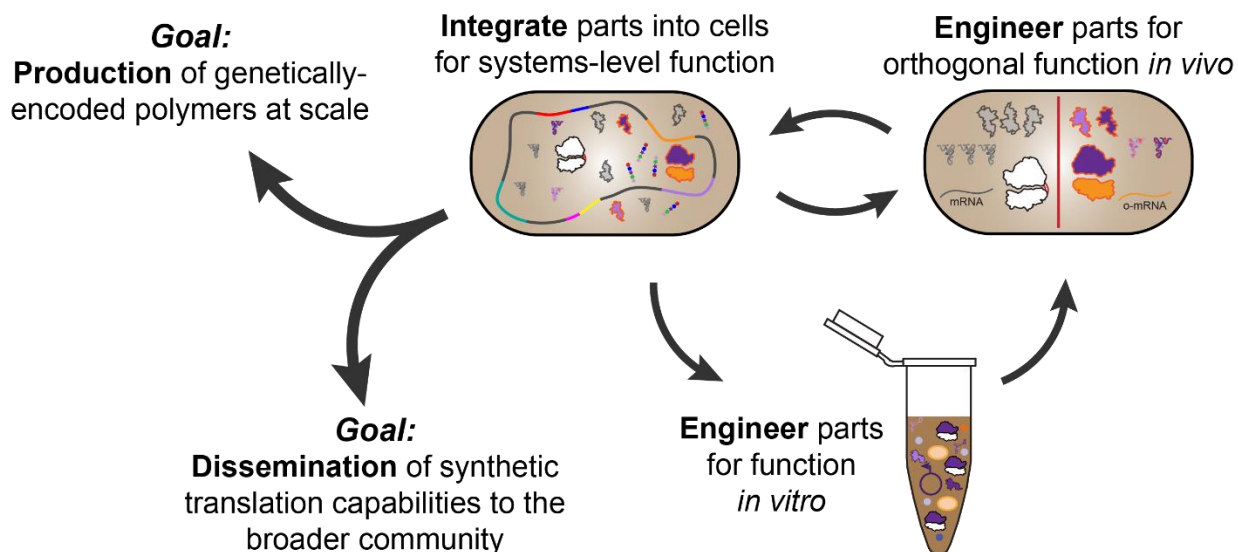
Figure 1-2: Overall workflow for efficient construction of synthetic translation system capabilities for downstream use. The cycle at right describes an iterative process that takes advantage of the unique capabilities of both *in vivo* and *in vitro* systems for engineering the translation system. The high-throughput testing capacity and flexibility of *in vitro* platforms makes them ideal for accelerating the pace of part engineering, while *in vivo* platforms are integral to building chassis strains that set the benchmark for future research and production efforts.

Additionally, although library sizes of engineered components transformed into cells are bottlenecked by transformation efficiency when compared to *in vitro* methods, downstream selections on such libraries tend to be much simpler *in vivo* than *in vitro*. For instance, a selection operated upon successful translation of an antibiotic resistance marker is extremely simple to carry out once library construction is complete: the population containing the library is simply plated on the correct antibiotic, and putative selection "winners" can be isolated from any resulting colonies. In contrast, while *in vitro* selections are possible, they tend to require a much greater amount of work for optimization of conditions and workflow as well as data collection and analysis.

Finally, *in vivo* translation engineering efforts have access to a large breadth of tools developed for cell engineering by the synthetic biology community that provide different engineering access points than *in vitro* tools, including fluorescence-activated cell sorting

(FACS) and life-death selections. These methods are generally representative of a top-down engineering approach in which selective pressure is applied to cells with a specific engineering goal in mind, but with specific mechanisms or mutations left unspecified. For example, cells forced under selective pressure to express and optimize an orthogonal tRNA/aaRS pair may make mutations in untargeted regions that improve functionality of the orthogonal tRNA/aaRS pair within the context of the cell, or they may even make mutations in seemingly unrelated genes altogether that improve the efficiency of the cell to accomplish the selected engineering goal. In contrast, *in vitro* systems for translation engineering tend to embody a bottom-up approach, in which individual components of the engineered system are defined and specific interactions between components can be parsed more easily. Ultimately, the strengths of both of these methods complement one another toward accomplishing the goal of efficient engineered translation systems.

4. Previous work toward in vivo translation engineering platforms

The advent of genomically recoded *Escherichia coli* organisms[7, 15-19] as well as engineered tRNA/aaRS pairs[20, 21] and tethered ribosomes[2-4, 8, 9, 22, 23] represent progress toward engineering individual elements of translation systems for manufacturing proteins with noncanonical amino acids (ncAAs). Specifically, recoded organisms take advantage of redundancy in the natural codon table to replace natural codons and their associated decoding machinery with orthogonal translation parts, this allowing incorporation of ncAAs into proteins *in vivo*. This freed-up coding space is typically used by engineered tRNA/aaRS pairs to encode noncanonical amino acids into proteins sequences.[20] Additionally, orthogonalization of Shine-Dalgarno interactions allowed orthogonalization of the 16S rRNA,[6] followed by inclusion of the 23S rRNA sequence (which includes the catalytic core of the ribosome) via tethering of the 16S and 23S rRNA sequences.[2, 3, 5, 8, 12] These advances all share a common goal: elaboration of complex, orthogonal translation systems for the production of new classes of genetically-encoded polymers at scale (**Fig. 1-2**).

5. Central challenges addressed by this work

The storyline of my PhD begins shortly after the innovation of recoded *Escherichia coli* organisms[7, 15-19] as well as tethered ribosomes[2-4, 8, 9, 22, 23]. With the tethered ribosome, we now had a tool to access engineering the 23S rRNA in orthogonal translation systems. For the first time, the catalytic core of the ribosome could be evolved toward new functions. However, in considering the direction of my PhD work during my first years of grad school, I began to realize that our approach to the problem wasn't taking in the full picture – that is, while the ribosome is the central coordinating and catalytic complex of translation, translation requires the precise coordination and involvement of many more factors than just the ribosome; translation is a complex process that involves numerous independently expressed factors, including ribosomal RNA (rRNA) and proteins, transfer RNAs (tRNAs), aminoacyl-tRNA synthetases, and initiation, elongation, and release factors. Indeed, while the ribosome is the key coordinating complex of translation, it relies on tens to hundreds of additional elements (depending on how you count them) in order to fully carry out its translation functions. While coordinately tuning and optimizing all of these factors into efficient and stable translation systems is fundamental to all known life forms, mimicking this network efficiency in synthetic translation systems – here including both restructured native translation systems (i.e., synthetic parts and/or novel expression architectures incorporated into cell-supporting translation networks) and orthogonal translation systems – remains a challenge (**Fig. 1-3**).[24]
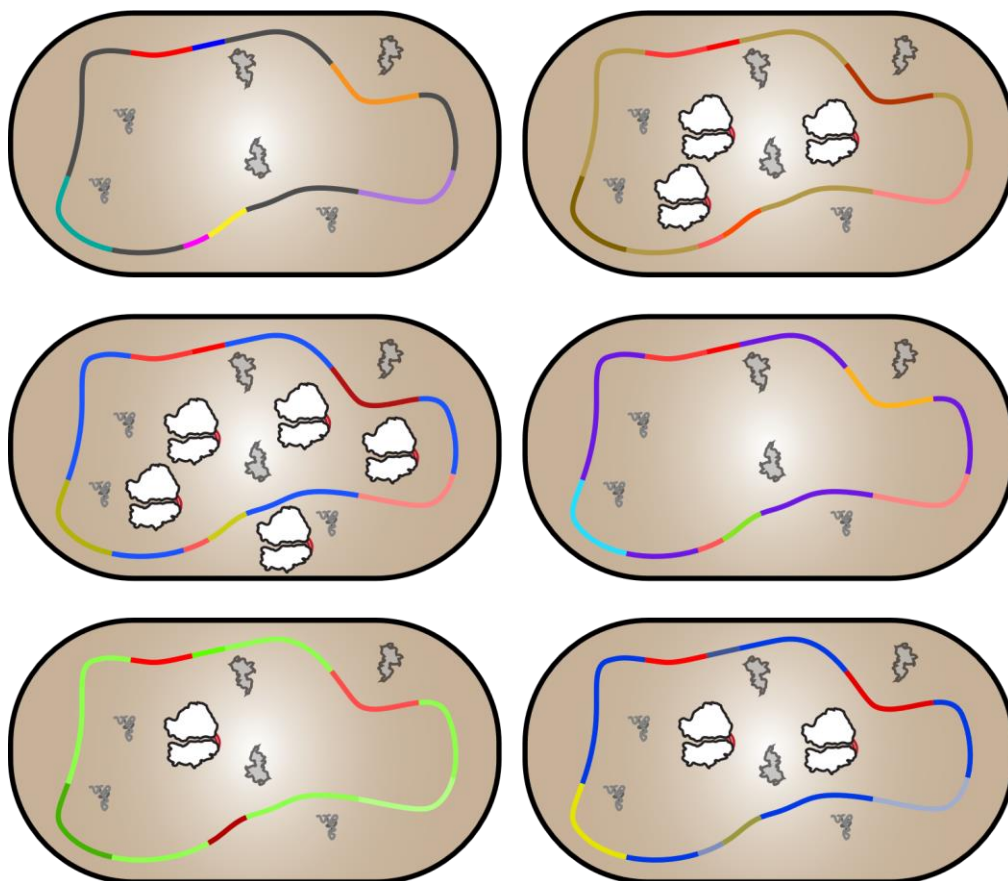
Figure 1-3: Many possible designs for synthetic translation systems exist. Even with designs of engineered translation components in-hand, many questions remain about how to build optimal genetic architectures that incorporate the engineered part into a desired chassis strain. Here, multicolored lines in each cell figuratively represent some of the many genomic permutations possible for expressing tethered ribosomes from the genome. These genomic permutations could include differences in copy number, location, direction, and genetic context of gene cassettes on the genome, among other factors.

One of my first major goals in my PhD was to put tethered ribosomes specifically, Ribo-T-v2 (RTv2),[2] onto the genome of *E. coli* as the cell-supporting ribosome. Ribo-T-v2 (RTv2)[2] is a ribosome with covalently tethered subunits where core 16S and 23S rRNAs form a single chimeric molecule. This would enable a "flipped orthogonal system" in which a dissociable ribosome could be used as the orthogonal ribosome in the cell. We hypothesized that this inverted system would have two major engineering benefits: (1) enabling larger reconstructions of an orthogonal ribosome without the structural constraint of the ribosomal tether, and (2) easier interconversion between ribosomes engineered via iSAT *in vitro* and orthogonal

ribosomes *in vivo*, given that iSAT struggles to reliably express tethered ribosomes. While previous efforts toward such a flipped orthogonal system had been made, these relied on the cell-supporting, tethered ribosome and the dissociable, orthogonal ribosome both being expressed from plasmids in SQ171.[9] We hypothesized that integrating ribosomes onto the genome would enable faster-growing cells with freed-up episomal space when compared to these efforts, enabling further selective power to be leveraged on future efforts to engineer OTSs *in vivo*.

However, conceptualizing designs for such a strain was not straightforward. For instance, with tethered ribosome sequences in hand, what does the cell that makes best use of a tethered ribosome look like? Does it use a tethered ribosome as the cell-supporting ribosome or the orthogonal ribosome? What genetic architectures are used to express it – are they genome-bound, plasmid-bound, and where are they and in what copy number? Do these answers change when more factors, such as tRNA/aaRS pairs, are incorporated into the orthogonal network? How does a cell that hosts a complex orthogonal translation system balance its necessary growth functions with our engineering goals? And… how do we build it?

2. Design and construction of the mobilization genome-engineering platform

1. Introduction

To solve the challenge of integrating RTv2 onto the genome, a variety of genome engineering tools are available. For example, the Datsenko-Wanner method and CRISPR-Cas mediated approaches[25, 26] can deliver whole gene cassettes onto the genome of *E. coli*.[27] Unfortunately, these approaches typically require a minimum of 1-2 days per edit made in series, and are not used for parallel and continuous directed evolution of genomes. Given the many open-ended questions raised previously about how best to build efficient translation networks incorporating tethered ribosomes, the rate at which these genome engineering approaches explore genomic space is unsatisfying at best. However, in yeast, the SCrAmbLE system has demonstrated more rapid exploration of genomic space than the Datsenko-Wanner and CRISPR-Cas mediated approaches.[28-31] The SCrAmbLE system allows rapid generation of millions of structural yeast genome variants in one pot by simple induction of Cre recombinase. Unfortunately, equivalent recombinase-based techniques to generate massive genomic diversity were underdeveloped for synthetic translation systems in *E. coli*.

In thinking about these challenges, I came to a realization: evolutionary dynamics *should* encourage a strain such as SQ171 toward integrating their ribosomal DNA onto stable genomic sites (and indeed, unpublished accounts of such events happening through homologous recombination pathways are common), but *the genome engineering tools we use do not provide practical or controlled pathways toward these desirable evolutionary states*. If I were able to allow the RTv2 cassette to mimic the action of mobile genetic elements, which are much more plastic in their copy numbers and expression profiles over evolutionary time than typical genomic elements, SQ171 should be able to evolve toward a more fit state incorporating copies of RTv2 onto the genome. When I realized that SQ171 still harbored seven FRT sites on its

genome at former ribosomal operon locations, a plan began to form, and after researching the mechanism of FLP/FRT recombination, I began to put together designs for how such a system as I was imagining might work. FLP/FRT recombination aligns two FRT sites contained in dsDNA and swaps the partnered upstream and downstream regions (**Fig. 2-1a**). In a well-established use, excision of FRT-flanked cassettes allows removal of an antibiotic resistance cassette from the genome after antibiotic selection for genomic integration of a larger cassette that includes the antibiotic resistance marker (**Fig. 2-1b**). From this mechanism, I hypothesized that a plasmid bisected by two parallel FRT sites should allow the integration of the plasmid onto the genome at available FRT sites and subsequent unpairing of the two cassettes via subsequent recombination events (**Fig. 2-1c**).
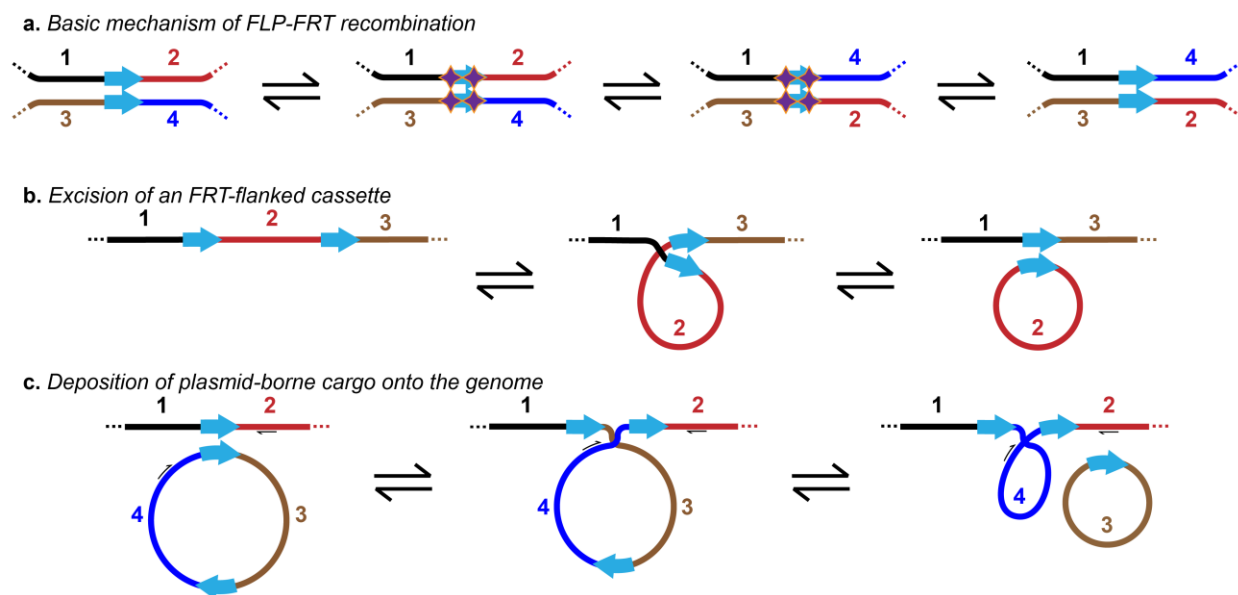
Figure 2-1: Mechanism of FLP/FRT recombination. (a) FLP/FRT recombination occurs when flippase (FLP) forms a complete tetrameric complex in conjunction with two FRT sites. This complex catalyzes the swapping of upstream-of-FRT (black, brown) and downstream-of-FRT (red, blue) sequences relative to one another. (b) An FRT-flanked cassette within a large DNA molecule can be circularized and excised via FLP/FRT recombination occurring between its flanking FRT sites. (c) An FRT-bisected plasmid can deposit cargo onto the genome in a two-recombination-step process. Small black harpoon arrows represent PCR primer sites that can be used to screen for such integration events.

This line of thinking, with dedicated work, evolved into a platform we have termed *mobilization* that allows synthetic biologists to rapidly sample large sets of genome permutations around a small set of specifically targeted gene cassettes. Inspired by the action of mobile genetic elements, which are much more plastic in their copy numbers and expression profiles over evolutionary time than typical genomic elements, I show here that we "mobilize" targeted elements of synthetic translation systems by utilizing flippase (FLP)/flippase recognition target (FRT)-mediated recombination to stochastically recombine FRT-flanked target elements into FRT genomic sites within each cell in a population. We hypothesized that mobilization would allow for rapid exploration and evolution of coordinated expression dynamics of these synthetic translation elements in the context of the native host translation system.

2. Mobilization allows targeted genomic integration of FRT-flanked cassettes

We set out to build a platform for genomic mobilization of synthetic translation systems using FLP/FRT-mediated recombination. First, we asked whether we could transfer cellular dependence in SQ171 from episomally expressed tethered ribosomes (RTv2) to genomically expressed tethered ribosomes (**Fig. 2-1c and Fig. 2-2a**). SQ171 has each of its seven genomic ribosomal operons removed and replaced with an FRT site, thus requiring two plasmids: one expressing the cells' ribosomes and one expressing essential tRNAs from the deleted ribosomal operons. These two plasmids were exchanged with plasmids that would maintain their essential function while enabling mobilization: the ribosomal plasmid was replaced with a plasmid expressing an RTv2 cassette flanked by two FRT sites on the plasmid vector (pSLG022), and the tRNA plasmid was enlarged to include an arabinose-inducible FLP recombinase cassette (pSLG033) (**Appendix B, Table 1**). Using this mobilization-capable strain, we induced FLP recombinase with 4mM arabinose and incubated cultures at 37 ºC for 8 hours. With primers

flanking each of the seven genomic FRT sites (one unique to RTv2 and one unique to the genomic

locus), we amplified the recombined genomic DNA from SQ171(pSLG022, pSLG033) (**Fig. 2-2b**).

Further study of a range of arabinose-induction conditions found robust induction of FLP activity

in all conditions tested between 1mM and 8mM arabinose induction (**Fig. 2-3**). We found that,

across induced populations of this strain, DNA encoding RTv2 was integrated into each of the

seven loci flanked by FRT sites on the genome, confirmed by Sanger sequencing. We did not

observe recombination without arabinose induction.
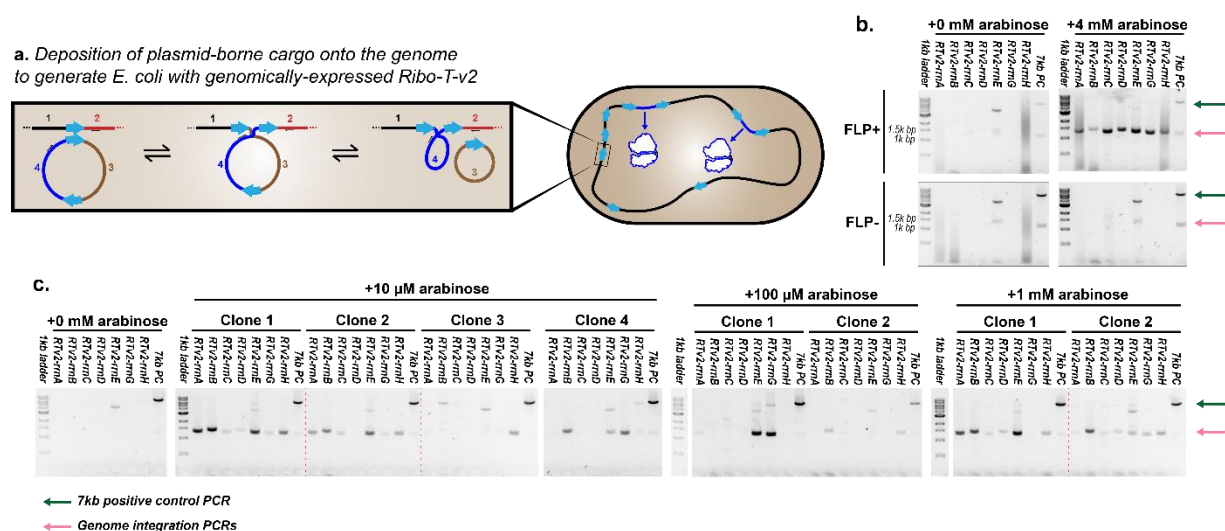
Figure 2-2: FLP-dependent integration of pSLG022 into genomic FRT sites. a. A circular plasmid (example vector shown in brown; example cargo shown in blue) with two parallel FRT sites may be integrated into the genome, then the vector can be excised in a two-recombination-step process to allow the integration of desired cargo at FRT sites in the genome. A schematic of an example product strain produced from mobilization of RTv2 onto the genome of SQ171. Mobilization has introduced two RTv2 cassettes onto the genome to translate the proteome. Primer sites for the PCR reactions shown in *b.* and *c.* are contained with the blue and red regions in this illustration, with successful amplification of a PCR product denoting presence of recombined products. b. PCR reactions on SQ171(pSLG022, pSLG033) [FLP+] and SQ171(pSLG022, pSLG028) [FLP-] cultures with and without induction by 4mM arabinose. Approximate expected band sizes for each genomic site assay PCR are shown by the pink arrow, and for the 7kb PCR positive control by the green arrow. Similar results are seen for integration of the plasmid vector, as well as induction by as little as 1mM arabinose (Fig. S1). c. Characterization of sets of FLP+ clones isolated from the same mobilized populations, with each population grouped by its associated concentration of arabinose induction. As in *b.*, pink and green arrows signify expected band sizes for the genomic site assay PCR and the 7kb PCR positive control, respectively. Differing band patterns across the seven genomic FRT sites assayed signifies independent recombinase activity between clones isolated from the same population.

We then isolated clonal populations of SQ171(pSLG022, pSLG033) induced with a range of arabinose concentrations to evaluate variability of integration of the RTv2 cassette across the seven loci. We performed colony PCR with primers flanking the seven genomic FRT sites and found that each clonal isolate has a characteristically different pattern of integration of the pSLG022 plasmid across the seven genomic FRT sites when compared to others isolated from the same population (**Fig. 2-2c**). Thus, mobilization allows the independent generation of a unique set of structural genomic edits in each of the millions of *E. coli* contained within a single culture tube. Among the clones assayed, higher levels of arabinose induction did not appear to correlate strongly with higher amounts of recombination into target sites.
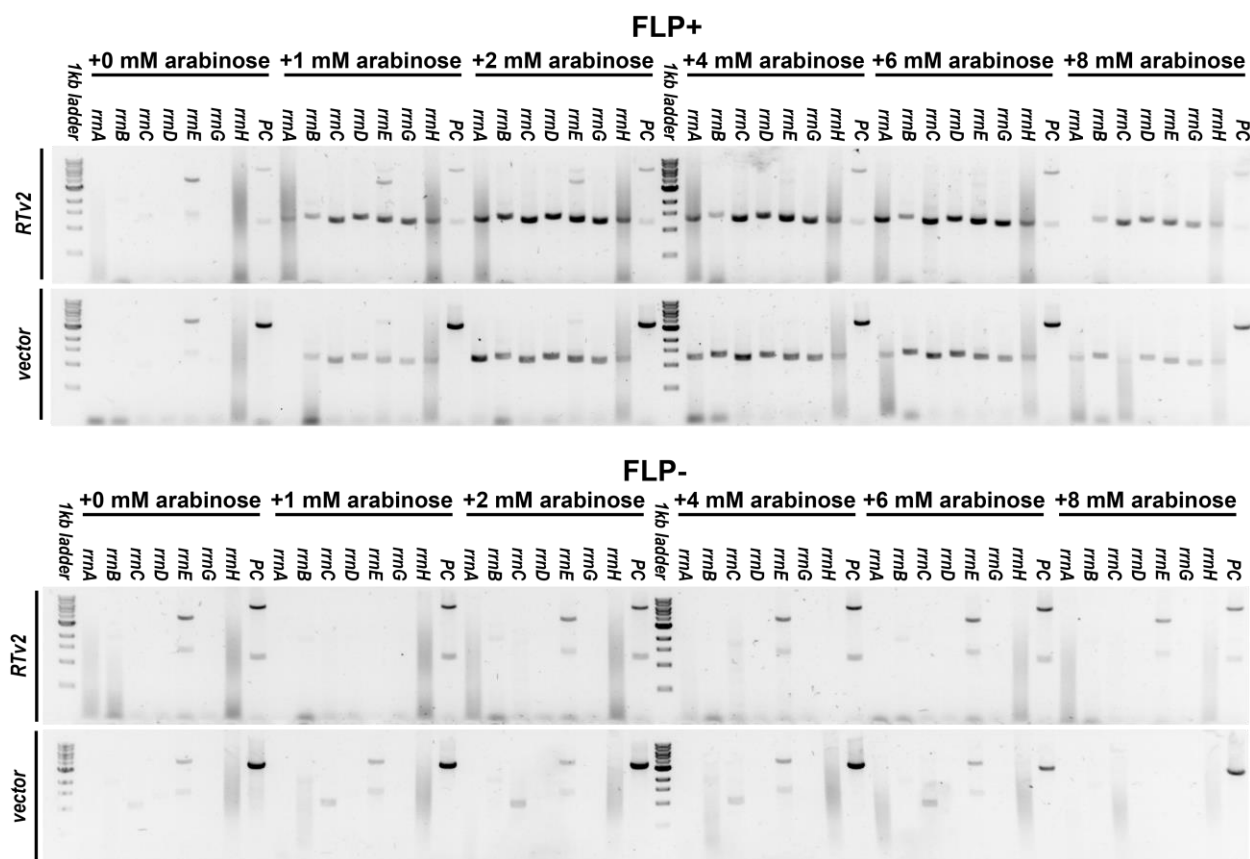
Figure 2-3: Colony PCR screening of arabinose induction conditions with (FLP+) and without (FLP-) the arabinose-inducible-FLP-containing plasmid pSLG033. FLP- cultures contain PSLG028. For each induction condition, 16 colony PCR reactions were performed: 14 assays for integration of RTv2 or plasmid vector at each of the possible *rrn* sites, and two positive control reactions (PC). The positive control reactions shown in RTv2 rows are an amplification of an unrelated 6977 bp fragment of the

SQ171 genome, and the positive controls shown in vector rows are an amplification of 16S that generates a 3617 bp product when amplified from RTv2 or a 694 bp product when amplified from wild-type ribosomes.

3. Building an NGS analysis pipeline for the construction of FRT-junction maps

While cPCR reactions were useful as an initial screen for genomic integrations with a fast turnaround, these data lacked the resolution and throughput to answer many more interesting questions. What are the approximate rates of genomic FLP/FRT recombinations happening in a population during mobilization, and how do they respond to induction strength and timing? What are the relative frequencies of different classifications of FLP/FRT recombinations that occur? To answer these questions, I turned to next-generation sequencing, a much more powerful platform that sequences millions of DNA molecules within a sample at one time. The resulting set of reads can be used to understand the DNA makeup in sample in powerful detail.

However, no preexisting computational pipeline existed for parsing the meaning of my mobilization experiments from NGS data. To analyze the resulting dataset, I built a computational pipeline in the iPython environment (Jupyter notebooks), using the Biopython package to assist alignment functions.[32] Briefly, for each sample, paired-end-read pairs were analyzed by pairwise local alignment to the 34-bp FRT sequence ("forward" direction: GAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC, "reverse" direction: GAAGTTCCTATACTTTCTAGAGAATAGGAACTTC). Additionally, each read was independently aligned to known upstream and downstream sequences of FRT sites to determine whether a putative, non-sequenced FRT site may exist on the fragment between sequenced read pairs. In either case, those read pairs with an identified putative internal FRT site were selected for further analysis.

The location of the putative FRT site was used to determine putative upstream-of-FRT and downstream-of-FRT regions within the read pair. Given FRT sites are directional, upstream-

of-FRT and downstream-of-FRT sequence sets are expected to remain constant without switching independent of the FRT-mediated recombination events that have occurred in a strain (e.g., a site that is immediately upstream of an FRT site is expected to remain upstream of an FRT site, although its downstream partner may change). Therefore, the upstream-of-FRT and downstream-of-FRT sequences from each read pair with a putative internal FRT site were aligned to the sets of known upstream-of-FRT and downstream-of-FRT sequences, respectively, to determine putative identities for the upstream-of-FRT and downstream-of-FRT sequences. Sequences returning an alignment score that is both ≥ 50 (match score: 1; mismatch score: -1.25; open gap score: -5; extend gap score: -1) and a ratio of 1.4 higher than the next-highest alignment score from the set of possible alignments were considered positively identified. Upstream or downstream regions unable to pass these criteria were classified as "not identified".

This alignment is done for both the upstream-of-FRT and downstream-of-FRT regions from the paired-end reads to generate an index pair that uniquely classifies the analyzed FRT junction as one of 81 possible types (plus 19 additional possible junctions where one or both junctions are not able to be identified). The set of index pairs for sequenced FRT junctions was used to construct resulting FRT-junction maps for each strain or population analyzed, with each identified index pair generated from one read pair adding one count to its box in the resulting FRT-junction map (**Fig. 2-4a**).

4. Optimization of FLP induction conditions

We next wanted to calibrate our mobilization strategy by finding condition(s) of arabinose induction and time of selection that produced a high number of genomic edits in viable cells – while too little expression of FLP might lead to a lack of significant diversity generation, too much expression is likely toxic to cells and could be lethal to many or most members of a

population. We assumed that the overall population diversity, *D*, was roughly proportional to a factor derived from the degree of induction of FLP, *I*, and negatively proportional to the lethality rate caused by induction of FLP, *L*, as such:

$$D \propto I(1 - L)$$

Additionally, within a range relevant for our diversity generation experiments, we would expect the lethality rate caused by induction of FLP, *L*, to be roughly proportional to our induction factor, *I*, taken to a power *n* (*n* > 0). Within this range, larger titers of induced FLP recombinase should continuously cause harsher stresses on cells, resulting in higher rates of cell death, as such:

$$L \propto I^n$$

Therefore,

$$D \propto L^{1/n}(1 - L)$$

Given our hypothesis of these coarse-grained relationships, we sought to test the lethality caused by various levels of FLP induction so as to learn approximate levels of acceptable FLP induction and enable maximization of *D*. To do this, we grew SQ171(pSLG022, pSLG033) cells in liquid media and induced with 0 mM, 1 mM, and 10 mM arabinose at OD_{600} = ~0.1. After an induction period of one hour, three hours, and six hours for each induction condition, fractions of cell culture from each condition were simultaneously genome-extracted for paired-end, next-generation sequencing (NGS) and dilution plated on LB-agar. To evaluate the data, we used a custom-built computational analysis pipeline that scans NGS read pairs for internal FRT sites, classifies the upstream-of-FRT and downstream-of-FRT flanking regions in each read pair

identified as having an internal FRT site, and records the resulting pair of classified FRT-flanking regions (**Fig. 2-4a**).

Figure 2-4: Dynamics of mobilized SQ171. a. A schematic of the FRT-junction map showing functional classes and examples of FRT junctions in this experiment. b. FRT-junction maps generated from next-generation sequencing data on genome extractions of SQ171(pSLG022, pSLG033) cells mobilized with 1mM arabinose and 10mM arabinose for one hour, three hours, and six hours. Each map is indexed by upstream-of-FRT and downstream-of-FRT regions corresponding to the two plasmid-based FRT-flanked cassettes ("RTv2" and "vector") and to the seven genomic sites ("rrnA" - "rrnH"). In each map, the number of reads identified with each possible pair of upstream-of-FRT and downstream-of-FRT regions are shown. FRT junctions identified without at least one genomic index (2x2 area in top left corner) have been excluded to focus analysis on genomic sites. The fraction of edited, or non-native, genomic FRT junctions compared to all FRT junctions calculated from each map is shown in Table 2-1 (here, native junctions are defined as the diagonal running from [*rrnA*, *rrnA*] to [*rrnH*, *rrnH*] and edited junctions are defined as all other sites shown). A simultaneous dilution plating experiment approximated lethality of each condition compared to an uninduced control condition (Table 2-1). The fraction of edited junctions and the lethality rate were used to calculate estimated edits/mL/OD in viable cells for each condition. Under the optimal conditions shown here (1mM arabinose for 6 hours), the plasmid-borne mobilization system in

SQ171(pSLG022, pSLG033) can generate approximately 11 million large (>5kb) structural genomic edits per mL per OD in viable cells.

From this analysis, we constructed FRT-junction maps that profile relative quantities of FLP-FRT recombination events present in each condition (**Fig. 2-4b**). Relatively few FLP-mediated structural genomic edits occur before six hours after induction in both 1mM and 10mM arabinose conditions, and lethality rates remain relatively low (**Table 2-1**). After six hours, large numbers of structural edits can be seen in both conditions, and lethality rates rise dramatically. The observed non-linear FLP activity over time may be caused by the cooperative behavior of FLP recombinase:[33] as recombination is catalyzed by a tetrameric FLP complex joining two FRT sites, the number of FLP-mediated recombination events catalyzed per time responds sigmoidally to FLP concentration. In the 10mM-arabinose condition, the shock of many recombination events appears to have caused a rapid die-off of these cells, resulting in a lethality rate of 99.9%. By comparison, in the 1 mM arabinose condition, the amount of FLP activity observed is appreciable for genomic library generation but not nearly as lethal. Under the best condition observed—six hours of induction with 1mM arabinose—we calculated that the plasmid-borne mobilization system in SQ171(pSLG022, pSLG033) can generate approximately 11 million large (>5kb) genomic insertions/deletions/rearrangements (structural edits) per mL per OD in viable cells (**Table 2-1**). These rates are calculated using the following equation:

$$\frac{edits\ in\ viable\ cells}{mL * OD} = \frac{edited\ FRT}{total\ FRT} * \frac{FRT}{cell} * \frac{expected\ cells}{mL * OD} * survival\ rate$$

, where edited FRT is defined as the total counts of genomic junctions in an FRT-junction map outside of those expected for the initialized system, and total FRT includes all

counts of genomic junctions. FRT/cell is a constant defined by the number of unique genomic

FRT sites expected in the initial system: 7 for the plasmid-borne system and 9 for the genome-

bound system. Expected cells/mL/OD was calculated at each timepoint from dilution plating of a

control strain that did not receive arabinose induction. Finally, the survival rate term is deduced

from a corresponding lethality experiment by taking the ratio of CFU/mL/OD from induced and

uninduced samples at a given timepoint.

Table 2-1: Calculated values from FRT-junction mapping and corresponding lethality experiment.
Fraction of edited junctions and lethality compared to uninduced control were used for each condition to
calculate an estimated edits/mL/OD in viable cells. All dilution plates contributing colony counts to these
data contained between 40 and 400 colonies.

	Value	1 hour	3 hours	6 hours
1mM arabinose	Fraction of edited junctions	0.014	0.009	0.292
	Lethality compared to uninduced control	0.417	0.435	0.983
	Estimated edits/mL/OD in viable cells	$2.39*10^6$	$1.19*10^6$	$1.11*10^7$
10mM arabinose	Fraction of edited junctions	0.012	0.005	0.087
	Lethality compared to uninduced control	0.493	0.103	0.999

	Estimated edits/mL/OD in viable cells	$1.64*10^{6}$	$7.79*10^{6}$	$7.73*10^{4}$

Under the 1mM arabinose/6-hour conditions, we observed evidence of widespread integration of the RTv2 cassette and the plasmid vector cassette into most available genomic sites (shown by the first row and column and second row and column of each FRT-junction map, respectively), as well as rearrangements of portions of the genome. Genomic integrations of the plasmid vector seem to initially appear at higher rates than the RTv2 cassette. Additionally, genomic rearrangements may be biased toward rearrangements of smaller regions or between nearby FRT sites: in the base strain, *rrnC*, *rrnA*, *rrnB*, and *rrnE FRT* sites are all contained within a 250-kbp genomic segment, and recombination events between these sites appear to occur relatively frequently compared to those between other genomic sites. These observations demonstrate that mobilization is capable of rapidly exploring structural genomic space about both targeted FRT-flanked cassettes as well as rearrangements of large portions of the genome.

5. Passaging and selection of mobilized cultures

With our optimized mobilization strategy, we next wanted to use a laboratory evolution approach to evolve highly fit, genomically integrated-RTv2-dependent strains of *E. coli*. To do this, we designed an induction scheme for recombinase-pulse passaging experiment (**Fig. 2-5, Fig. 2-6**). Ultimately, we passaged 60 cultures, split between 30 FLP+ and 30 FLP-, independently in a 96-well plate. These cultures were induced with a gradient of arabinose (0 mM, 0.001 mM, 0.01 mM, 0.1 mM, and 1 mM) twice over the two-week passaging period, resulting in six replicates for each condition. Cultures were passaged twice daily in a 1:100 dilution from the previous culture to maintain exponential growth as the dominant phase of the passaging cultures.
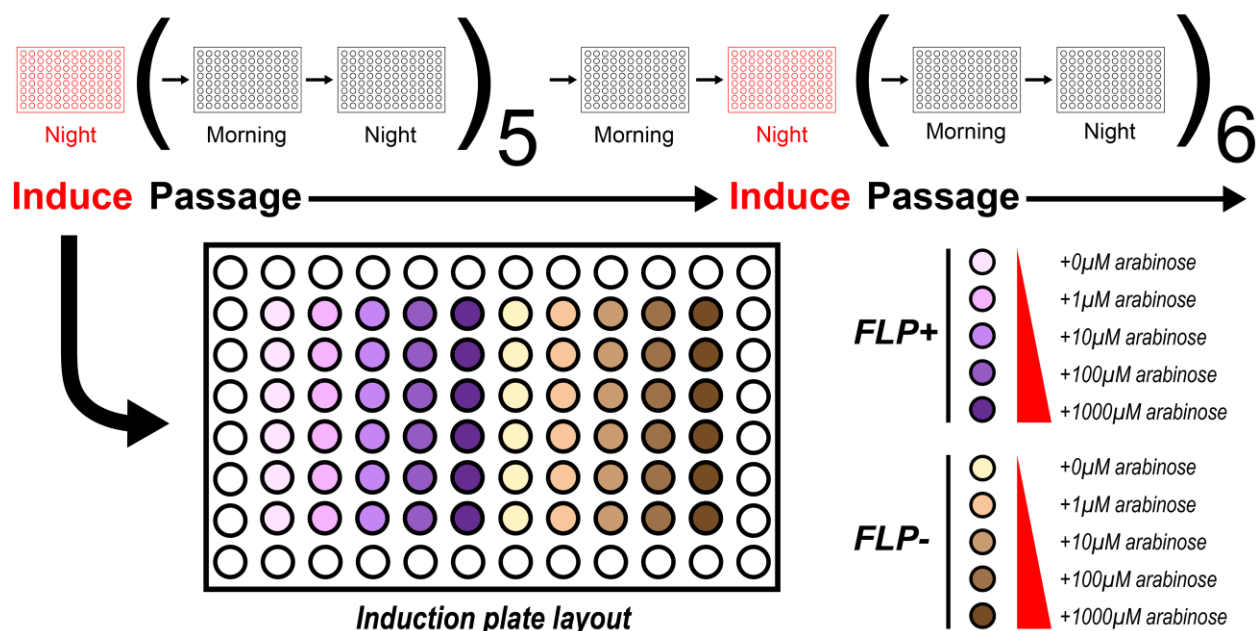
Figure 2-5: Overall pulse-passaging scheme. Induction (red plates) occurred twice over an approximately two-week passaging period in which cultures were passaged twice per day. In each induction plate, a variable amount of arabinose was added to cell culture media before induction and growth from the previous culture.

Following our passaging scheme, we extracted genomic samples from 24 of these populations (all six replicates of [FLP+, 0mM arabinose], [FLP+, 1mM arabinose], [FLP-, 0mM arabinose], and [FLP-, 1mM arabinose]) before and after passaging and submit them for next-generation sequencing to generate FRT-junction maps (**Fig. 2-6**). In the initial 1 mM arabinose condition, many FRT junction recombination events can be seen (**Fig. 2-6**). Notably, these data were generated from genome extraction of a culture outgrown from the original induced culture (necessary to generate enough material for genome extraction). As such, what appears to be a less diverse library shown in **Figure 2-6** than generated from similar conditions in **Figure 2-4** may be a result of several generations of selection having occurred between initial library generation and extraction. After the passaging period, genomic integrations of RTv2 cassette in the induced FLP+ condition are strongly enriched at each possible genomic site compared to their initial

condition. At the same time, genomic integrations of the plasmid vector in the induced FLP+ condition, while present at a similar rate as RTv2 integrations in the initial library, are significantly de-enriched after passaging. These data demonstrate that expression of RTv2 integrated onto the genome without need of its original plasmid vector can have fitness benefits for the cell and therefore can be selected for through serial passaging. Additionally, certain genome restructuring events (e.g., formation of new junctions between former *rrnC* and *rrnE* genomic sites) appear to have been enriched in these final populations.
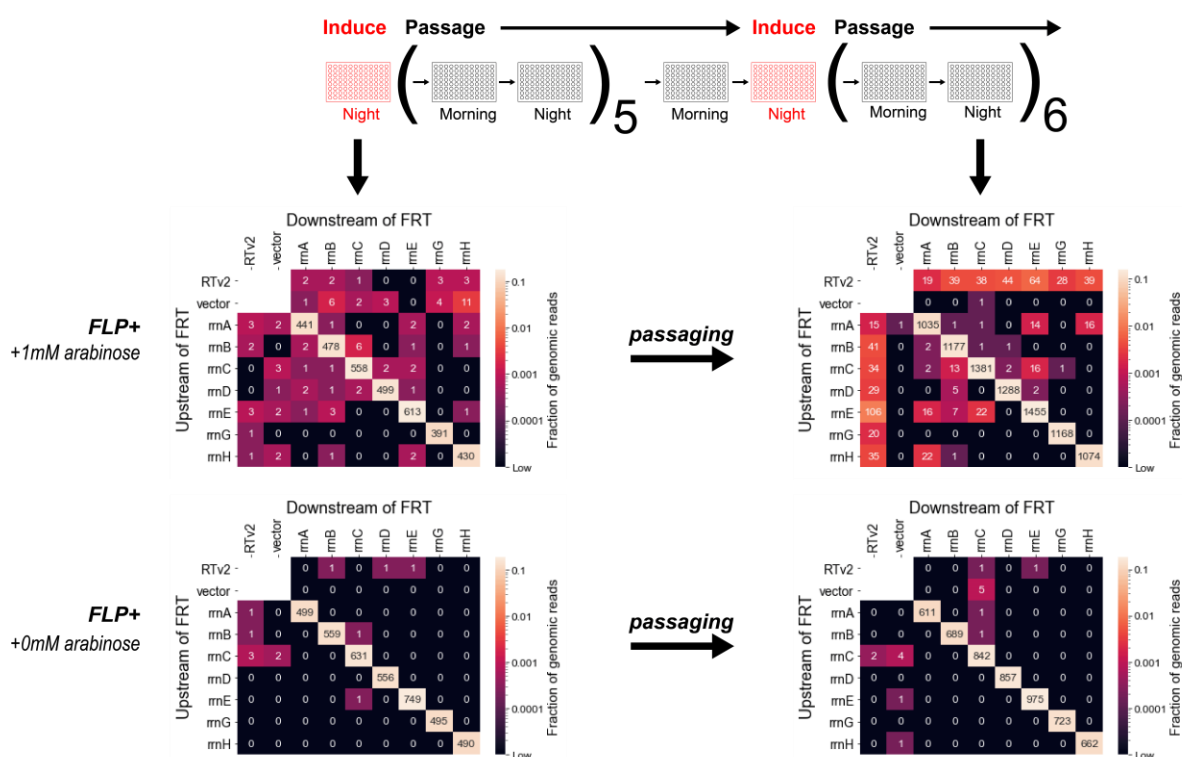
Figure 2-6: FLP-catalyzed evolution of genomic libraries toward genome-integrated-RTv2 genotypes. Passaging scheme and pooled FRT-junction maps for starting point and ending point samples induced with 1 mM arabinose or not induced with arabinose. Cultures were independently passaged twice daily in a 96-well plate in replicates of six for each condition. FLP+ denotes the presence of the arabinose-inducible FLP cassette in these cultures. Cultures were induced twice over the two-week passaging period by inoculation into culture media containing a gradient of arabinose (red plates) FRT-junction maps show starting point (after first induction, left) and ending point (after complete passaging, right) cultures for all six replicates of the (FLP+, 0mM arabinose), (FLP+, 1mM arabinose) conditions pooled together. As in *Figure 7*, FRT junctions not containing at least one native genomic site have been hidden.

Counterselection and clonal isolation from the F6 population

We observed that, of 551 total RTv2 genomic integrations found in the six [FLP+, 1 mM arabinose] populations, 509 (92.4%) occurred in the population found in well F6 ("population F6"), even though each of these populations were derived from one of six original replicates (**Fig. 2-5**). That is, each population was treated with the same passaging and arabinose induction conditions but was allowed to evolve independently. After plating F6 cultures on sucrose for counterselection of the pSLG022 vector, subsequent diagnostic PCRs show successful isolation of clones in which RTv2 is the dominant ribosomal population, but which contain no ColE1-based plasmid vector (**Fig. 2-7**). Further PCR assays on a set of 19 isolated product clones are consistent with all clones containing a highly similar integration pattern of the RTv2 cassette at the *rrnC* and *rrnG* genomic sites, and nowhere else among the assayed sites (**Fig. 2-8**). Although the counterselection appears to have bottlenecked the diversity seen in the original F6 population, the resulting clones are now fully weaned from dependence upon plasmid-expressed RTv2.
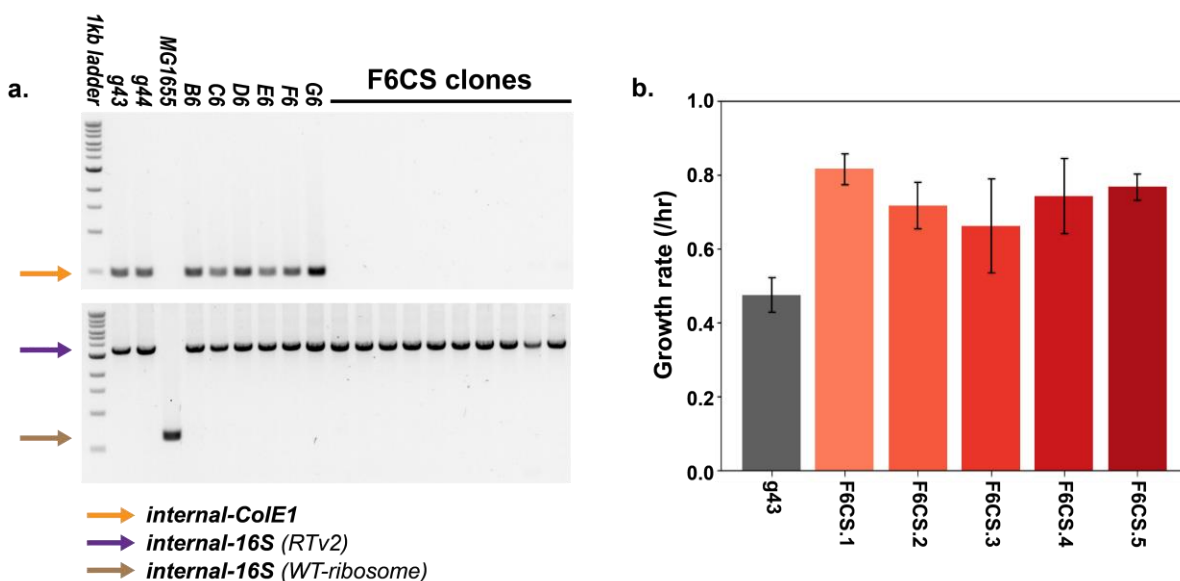
Figure 2-7: Characterization of population F6 clones after sucrose counterselection (F6CS clones).
a. Colony PCR reactions assaying for the presence of the ColE1 plasmid vector used in pSLG022 (top, orange arrow) and across the 16S rRNA (16S) of the small ribosomal subunit (bottom). g43 is the ancestral SQ171(pSLG022, pSLG033) strain; g44 is the ancestral SQ171(pSLG022, pSLG028) strain; MG1655 is a reference strain of wild-type E. coli; B6-G6 are the six 1mM-arabinose pulse-passaged replicates evolved from g43 before sucrose counterselection. The 16S rRNA PCR produces products of different lengths when amplified from WT ribosomal operons (brown arrow) compared to RTv2 (purple arrow), which is because the tethered ribosome has circularly permuted 23S rRNA inserted into the 16S rRNA. b. Comparison of growth rates of sucrose-counterselected product clones (red shades) with their evolutionary ancestor strain (gray). Data are shown for n = 4 or n = 5 independent experiments with standard deviation for error. Full kinetic data with fitted model curves are shown in **Appendix B**.
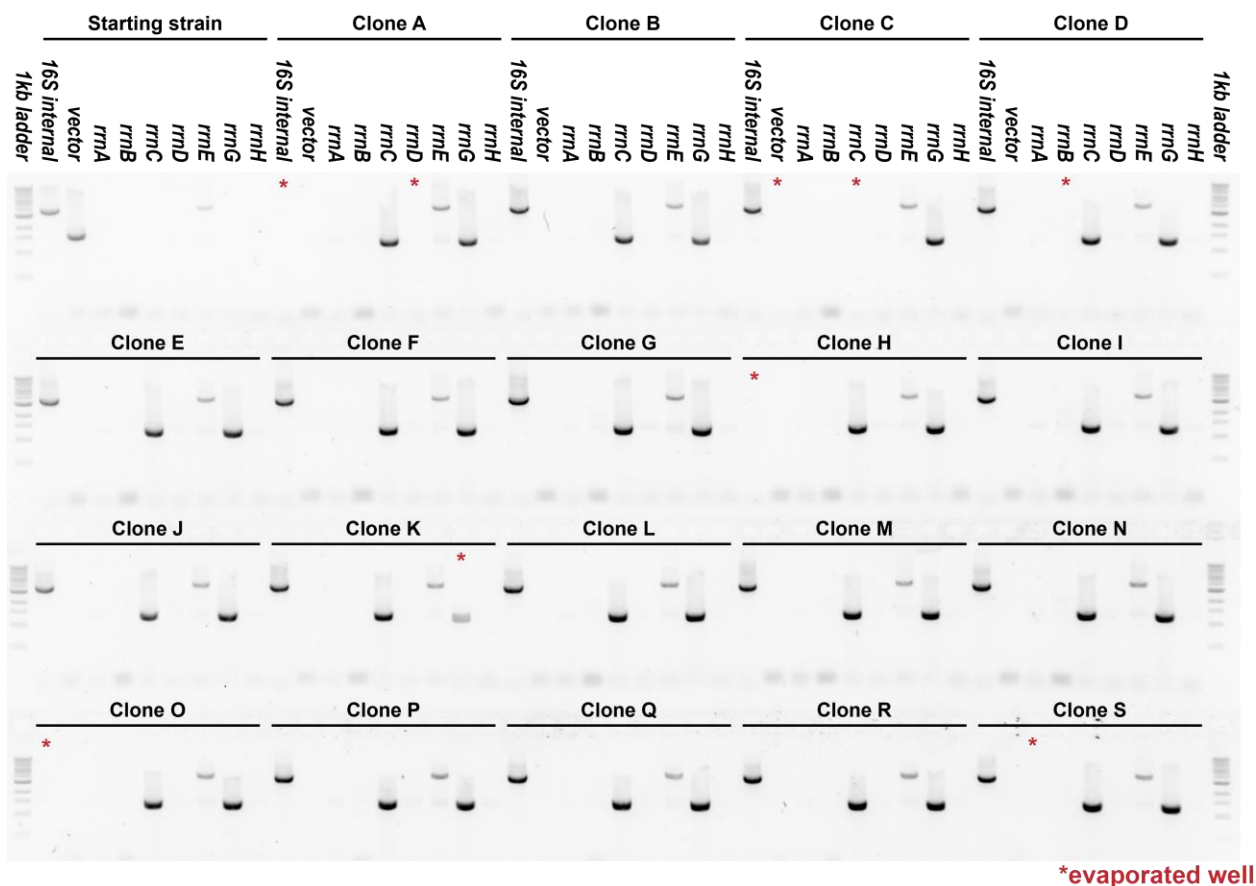
Figure 2-8: Colony PCR screening of a set of 19 F6-counterselected clones for RTv2 integrations at *rrn* genomic sites. Red asterisks indicate wells that evaporated in the PCR machine and are possible false negatives.

Further characterization of five clones from the F6 population shows significantly higher growth rates than their evolutionary ancestor strain, with strain F6CS.1 growing ~71% faster than its ancestral strain (**Fig. 2-7b**). Additionally, genome sequencing and subsequent analysis with the *breseq* package[34] confirmed loss of the plasmid vector and identified junctions between RTv2 and the reference genome at *rrnC* and *rrnG* as well as junctions with itself (as in a tandem array), with no other identified RTv2 junction sites (**Fig. 2-9, Tables 2-2 and 2-3**). Coverage of the RTv2 cassette is between 17-25 times the coverage of average genomic sites, resulting in a putative genomic RTv2 copy number of 17-25 for these strains which suggests the RTv2 cassette may be

clustered in repeats at either or both the *rrnC* and *rrnG* integration sites (**Fig. 2-8, Tables 2-2 and 2-3**).
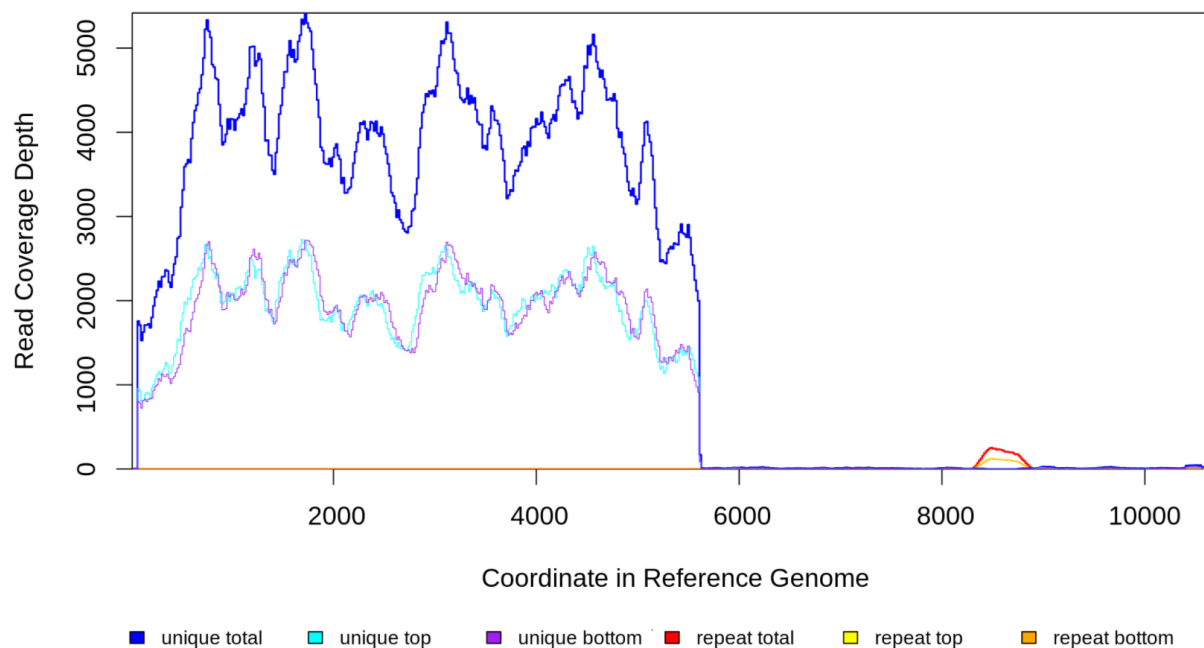
Figure 2-9: Example coverage map of pSLG022 generated from F6CS.1 using *breseq*.[34] Of the 10,741 bp in plasmid pSLG022, bases 108 - 5597 represent the RTv2 cassette and bases 5646 - …59 represent the vector cassette. Shown here is the coverage depth of sequences identified and mapped to pSLG022 from an F6CS.1 genome extraction sent for next-generation sequencing and analyzed with *breseq*. Graphed lines are labeled "unique" for reads with only one best fit from the reference plasmid and genome, and "repeat" for regions with multiple equally good matches. The reference sequence used for SQ171 genome is that of the base strain with seven genomic FRT sites and no RTv2 integrations. This map is representative of the five F6CS strains sequenced and analyzed with *breseq*.

Table 2-2: Example breseq new junction evidence, here generated from pSLG022 and SQ171 genome reference sequences for F6CS.1 genome extractions. The reference sequence used for SQ171 genome is that of the base strain with seven genomic FRT sites and no RTv2 integrations. breseq

identified integrations of RTv2 at former rrnC and rrnG genomic locations as well as tandem repeats of RTv2 (shown by base position in the reference sequence). In the SQ171 genome reference sequence, position 3,921,6xx corresponds to rrnC and position 2,717,xxx corresponds to rrnG. In the pSLG022 reference map, positions ~100 to ~5600 correspond to the RTv2 cassette. reads refers to the number of next-generation sequencing reads identified that support alignment to individual reference sequences (unmerged cells) or to a new junction between the two reference sequences (merged cell), with (cov) representing the relative coverage of the given sequence as a fraction of its expected value (as compared to overall coverage of the reference sequence(s)). skew represents the negative log10 probability of the hypothesis that the tiling of reads across a predicted junction is unusual. All junction predictions shown have a low-enough skew to not be rejected by the breseq algorithm. freq is a prediction of the junction frequency within the sample.

seq id	position	reads (cov)	reads (cov)	skew	freq
SQ171_genome	1201476 =	42 (0.210)	123 (0.680)	0.5	80.0%
SQ171_genome	1203305 =	24 (0.130)			
SQ171_genome	= 1201491	40 (0.200)	176 (0.970)	0.4	85.5%
SQ171_genome	= 1203288	24 (0.130)			
SQ171_genome	2717590 =	2 (0.010)	48 (0.050)	2.7	91.8%
pSLG022	108 =	7 (0.000)			
SQ171_genome	= 2717624	1 (0.000)	48 (0.050)	1.8	93.4%
pSLG022	= 5610	6 (0.000)			
SQ171_genome	3921636 =	2 (0.010)	77 (0.090)	0.7	94.8%
pSLG022	108 =	7 (0.000)			
SQ171_genome	= 3921670	2 (0.010)	110 (0.120)	0.5	96.7%
pSLG022	= 5610	6 (0.000)			
pSLG022	60 =	6 (0.000)	940 (0.660)	0.0	99.5%
pSLG022	= 5597	6 (0.000)			

(rrnG brace spans the rows for SQ171_genome 2717590, pSLG022 108, SQ171_genome 2717624, pSLG022 5610. rrnC brace spans the rows for SQ171_genome 3921636, pSLG022 108, SQ171_genome 3921670, pSLG022 5610.)

Table 2-3: Reference sequence information of F6CS strains generated using breseq. Fit mean and dispersion are calculated by fitting a negative binomial to a plot of number of reference positions as a function of coverage depth in reads. Fit dispersion is ratio of the variance to the mean for the negative binomial fit. Reference positions with a low coverage depth are censored from the fit to mitigate effects of deleted regions in the reference sequence on the fit. As such, the fit mean of pSLG022 for each sample corresponds roughly to the average coverage of the RTv2 region of the plasmid only, as shown in Fig. S6, whereas the fit mean of SQ171 genome represents the average coverage of the genome. Therefore, the ratio between fit means of pSLG022 and the SQ171 genome ("coverage ratio") gives a rough approximation of the genomic copy number of RTv2.

Sample	Sequence ID	length	fit mean	fit dispersion	% mapped reads	coverage ratio
F6CS.1	SQ171 genome	4,599,702	208.4	20.4	97.8%	20.1

	pSLG022	10,741	4192.7	149.5	2.2%	
F6CS.2	SQ171 genome	4,599,702	186.4	21.9	98.0%	17.7
	pSLG022	10,741	3299.8	94.8	2.0%	
F6CS.3	SQ171 genome	4,599,702	195.7	22.6	97.3%	25.0
	pSLG022	10,741	4886.3	139.9	2.7%	
F6CS.4	SQ171 genome	4,599,702	199.2	19.7	97.8%	20.4
	pSLG022	10,741	4057.4	110.8	2.2%	
F6CS.5	SQ171 genome	4,599,702	224.6	18.9	98.1%	17.8
	pSLG022	10,741	4002.6	76.5	1.9%	

6. Mobilization of a genome-bound system

We next wanted to investigate the capacity for re-mobilization of the F6 strains living solely on genomically integrated RTv2 to understand whether mobilization was possible without assistance from FRT-containing episomal elements. In contrast to the *plasmid-borne* system previously described, *genome-bound* mobilization would be defined as one in which all available FRT-flanked cassettes in a mobilization experiment originate on the genome. This could enable future mobilization efforts in which an FRT-containing helper plasmid (e.g., pSLG022) is rendered unnecessary, removing the uncertain and onerous plasmid counterselection step from the end of the mobilization strategy and allowing mobilization experiments to be carried out with much greater flexibility.

We employed a second mobilization experiment using the F6CS.3 strain which resulted in significantly less lethality while still exhibiting recombination events at similar fractions of FRT junctions when compared to the plasmid-borne mobilization system, resulting in ~400 million edits/mL/OD in viable cells (**Fig. 2-10 and Table 2-4**). Therefore, in addition to the benefits mentioned in the previous paragraph, it appears that mobilization of genome-bound systems could result in even more efficient diversity generation than plasmid-borne systems (such as the plasmid-borne system that was used to generate F6CS strains). Future efforts, therefore, should

strongly consider setting up such genome-bound mobilization systems as an alternative to plasmid-borne systems, such as the one which was used for the majority of this work.
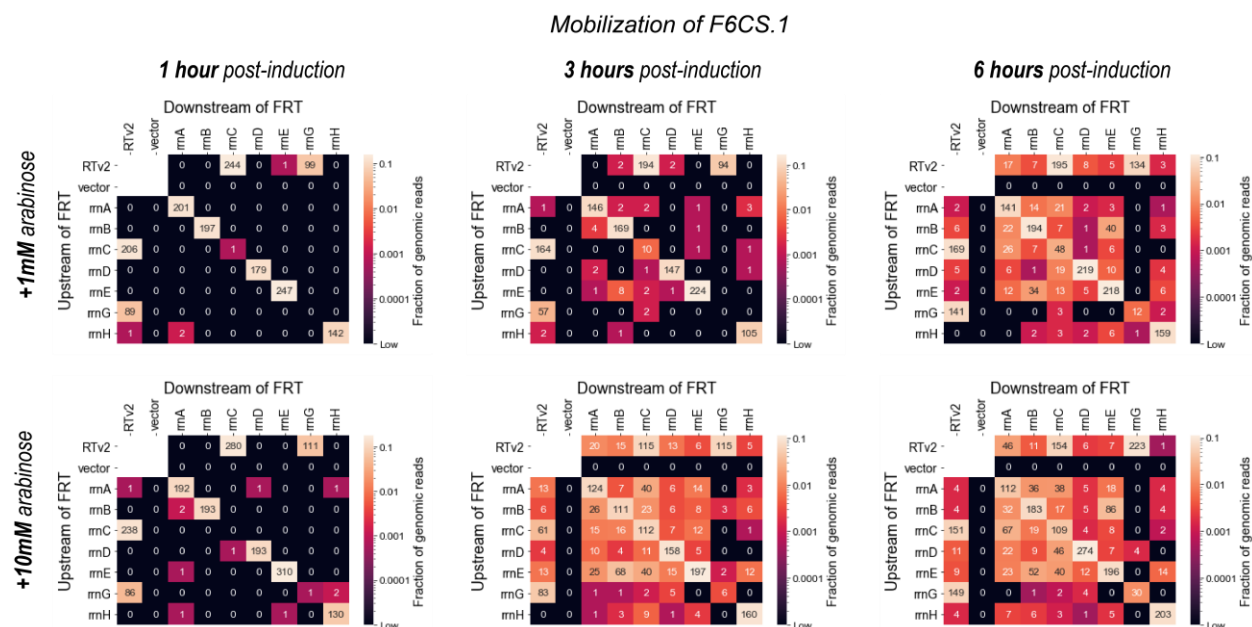
Figure 2-10: Mobilization of F6CS.3. As in Fig. 2, cultures were induced with 1mM or 10mM arabinose, then samples were taken for NGS at 1 hour, 3 hours, and 6 hours after induction.

Table 2-4: Calculated values from FRT-junction mapping and corresponding lethality experiment on F6CS mobilization.

	Value	1 hour	3 hours	6 hours
+1mM arabinose	Fraction of edited junctions	0.003	0.037	0.202
	Lethality compared to uninduced control	-0.024	0.265	0.574
	Estimated edits/mL/OD in viable cells	**4.74×10^6**	**6.74×10^7**	**4.02×10^8**
+10mM arabinose	Fraction of edited junctions	0.007	0.357	0.339
	Lethality compared to uninduced control	0.394	0.993	0.857

Estimated edits/mL/OD in viable cells	$6.99*10^6$	$6.56*10^6$	$2.28*10^8$

7. Discussion

In this work, we developed a FLP-recombinase-dependent mobilization system, which we demonstrate for the evolution of strains dependent on restructured, genomically expressed RTv2 as the cell-supporting ribosome population. By introduction of an RTv2 cassette flanked by two parallel FRT sites into *E. coli* and subsequent generation of genomic libraries via the stochastic action of FLP recombinase, we surveyed the landscape of possible solutions for a restructured, genomically expressed RTv2 strain at an approximate rate of 11 million edits per mL per OD in living cells of *E. coli* culture induced with 1mM arabinose. After two weeks of serial passaging and subsequent counterselection of the plasmid vector, we isolated clonal strains dependent on genomically expressed tethered ribosomes as their sole ribosomal population. The F6CS strains show a marked increase in growth rate compared to their ancestral strain while containing a similar number of copies of the tethered ribosome cassette, demonstrating the power of mobilization to fine-tune the expression of Ribo-T-v2 in the context of the cellular translation machinery. Furthermore, we demonstrated that F6CS strains can be re-mobilized without an FRT-containing plasmid.

Given their improved growth characteristics and their less restricted episomal space, we expect that the F6CS strains generated here will be useful chassis for further ribosome and translation engineering. That said, several improvements to the mobilization protocol might be made to improve its versatility and targetability. Use of nondirectional target recombination sites,

such as the loxPsym sites used in SCRaMbLE,[29] instead of directional FRT sites would allow sequence inversions in addition to duplications or deletions and could provide another means for a mobilized cell to fine-tune expression of the mobilized cassette. Additionally, the Cre/LoxP recombination system could offer an alternative site-specific recombination system to FLP/FRT recombination. Furthermore, multiplexed automated genome engineering (MAGE) or no-scar recombineering might be useful for inserting or deactivating/deleting targeted FRT sites over the course of a mobilization experiment, or after desired genome engineering is complete in order to produce an FRT-less strain.[35, 36] Finally, high-throughput automation and selection could further expand the power of mobilization to generate effective functional phenotypes.

One important consideration for future application of F6CS strains is genome stability. While we expect that most genome instability present during mobilization can be removed via removal/inactivation of the FLP recombinase gene, homology between FRT sites and duplicated RTv2 cassettes are another source of potential genome instability. Given our primary goal was to build more robust and faster-growing strains dependent upon RTv2 as the cell's translating ribosome, any remaining genome instability that, for instance, results in duplication, inactivation, or deletion of RTv2 cassettes would be subject to continued selection pressure for faster growth and so would likely stay in line with that goal. However, in addition to the need for fast-growing, robust chassis strains, certain experiments and production processes that use living cells rely on genomic stability to ensure controlled conditions. Characterization of the genomic stability of F6CS strains could be important to include when planning such future work.

Looking forward, we imagine the F6CS strains' most immediate application being in the evolution of orthogonal ribosomes toward new functions *in vivo*, enabled by the increased growth rates and freed-up episomal space of F6CS strains compared to their predecessor strains based

upon SQ171. For example, we expect that F6CS strains have a better capacity to support plasmid-based libraries of an orthogonal ribosome and subsequently allow a robust selection leveraged upon those orthogonal ribosome libraries via survival or fluorescence.

Additionally, we expect that mobilization could help build more complex synthetic translation systems, especially complex orthogonal translation systems with five or more orthogonal components. In this context, the combination of rich sequencing data generated from mobilization experiments and machine learning may help elucidate key cellular design principles for predicting how genomic architectures facilitate efficient synthetic translation systems. To this end, more powerful computational tools that incorporate understanding of the mechanisms of mobilization as well as experimental methods such as long-read sequencing could lead to powerful insight about evolutionary fitness and dynamics of complex genetic motifs within synthetic translation systems.

Finally, we hope that mobilization can be generalized for the construction and study of a variety of complex biological systems beyond the translation system in *E. coli*. While SQ171 has worked well for mobilization here with its seven native FRT sites, many other strains relevant for the study of various biological networks already possess one or more FRT sites as a product of historical genomic edits. If desired, additional FRT sites could easily be introduced at genomic locations of interest by a single researcher within weeks, thus generating suitable starting strains for mobilization. Ultimately, we look forward to mobilization's use as a flexible, powerful tool for studying and optimizing complex synthetic biological networks such as the translation system.

8. Summary and concluding remarks

Here, I argued the need for and demonstrated the development of a genome engineering platform capable of making at least 400 million large structural genomic edits per mL per OD in living *E. coli* cells. This mobilization platform allowed transference of cellular dependence from plasmid-expressed RTv2 to genomically incorporated RTv2 via directed evolution. I also hope that mobilization will make important contributions to the understanding of complex biomolecular networks such as synthetic translation systems through this sheer capacity to generate incredibly rich datasets and genomic libraries in high-throughput. While the work shown here has made important strides for enabling construction of more complex and powerful synthetic translation systems, there are still many remaining challenges in the field of *in vivo* translation engineering. For instance, how can we operate selections for functional behavior of an OTS *in vivo*? How can we apportion necessary coding space and metabolic requirements for a complex OTS? Such questions and others ensure that future researchers seeking to improve the effectiveness and application of synthetic translation systems will have plenty of challenges to work on.

Like all living organisms, engineered *E. coli* harboring synthetic translation systems are subject to constant evolutionary pressure during their life cycles. Given that typical lab strains of *E. coli* have a doubling time of 20-30 minutes, evolution in these organisms can occur on particularly rapid timelines. As a result, the inclusion of a synthetic translation system in *E. coli* that decreases the fitness of its host – whether due to interference / decreased efficiency of native protein expression pathways, increased metabolic loads required for the expression of additional, expensive machinery, or other factors – will result in strong selective pressure acting upon that system. If, in the evolutionary landscape of the cell, relatively simple and few-in-number changes can be made that significantly reduce the burden of the synthetic translation system (e.g., inactivation / non-expression of a gene or homologous recombination events), it is

very likely that that synthetic translation system will not be sustainable over the course of generations in a cell chassis population. In contrast, a cell that is provided with continual evolutionary reinforcement to maintain the synthetic translation system (e.g., expression of a selective marker by an OTS) and has few or no easily available pathways to "cheat" such pressure (e.g., expression of that orthogonal selective marker by the native translation system) should in theory contain a much-more evolutionarily stable system. As a result, engineering synthetic translation components for increased orthogonality not only better insulates cell-supporting translation systems from interference by the OTS, but also reduces the likelihood of cells escaping selections leveraged on the OTS for engineering and system maintenance purposes.

In my PhD work, I have frequently sought to turn this "weakness" of *in vivo* translation engineering – that evolution will often interfere and break our engineered designs – on its head. That is, how can we build synthetic translation systems that don't crack under evolutionary pressure, but instead, may even benefit from evolution making small tune-ups in terms of part design and system coordination? In order to do this, a deep understanding of the evolutionary landscape of the desired synthetic translation systems is required so that evolutionary pathways which benefit the cell's fitness are in line with our engineering goals, and pathways which do not benefit our engineering goals are suitably discouraged.

These considerations are particularly important for future efforts that aim to utilize the mobilization method and elaborate orthogonal translation systems. In its current form, mobilization relies on evolution to steer a cell population toward more fit states; the method simply provides a pathway to access these states much more quickly. In my work integrating RTv2 onto the genome, a directed evolution approach was made possible by an evolutionary fitness differential between the starting strain, SQ171, and the target strain(s) which rely on RTv2 expressed from the genome. Future efforts using mobilization to power similar directed

evolution efforts (such as for the elaboration of an OTS) should make sure to inventory possible selection pathways and how cells might escape desired selection pathways. For example, expression of an OTS is likely to always be an evolutionary burden on the cell, making the OTS a prime target for inactivation/deletion over evolutionary time. However, if the OTS expresses a selectable marker (such as an essential enzyme hidden from expression by the cell-supporting translation system), the evolutionary calculus should tip in favor of stable maintenance of a (now-essential) OTS.

3. Future directions

1. Introduction

Here, I will chart out my estimation of the best path forward for building more elaborate and powerful orthogonal translation systems. This future work builds upon the footholds we already have and often attempts to work toward two key properties of ideal orthogonal translation systems: (1) its components are stably integrated and expressed in chassis strains into well-coordinated biomolecular networks, and (2) its components are well-insulated (i.e., orthogonalized) from significant interactions with the cell-supporting translation system (**Fig. 3-1**). Additionally, the OTS would likely be responsible for constitutive expression of an essential gene or selectable marker to counterbalance evolutionary pressure to jettison the OTS due to its metabolic requirements.
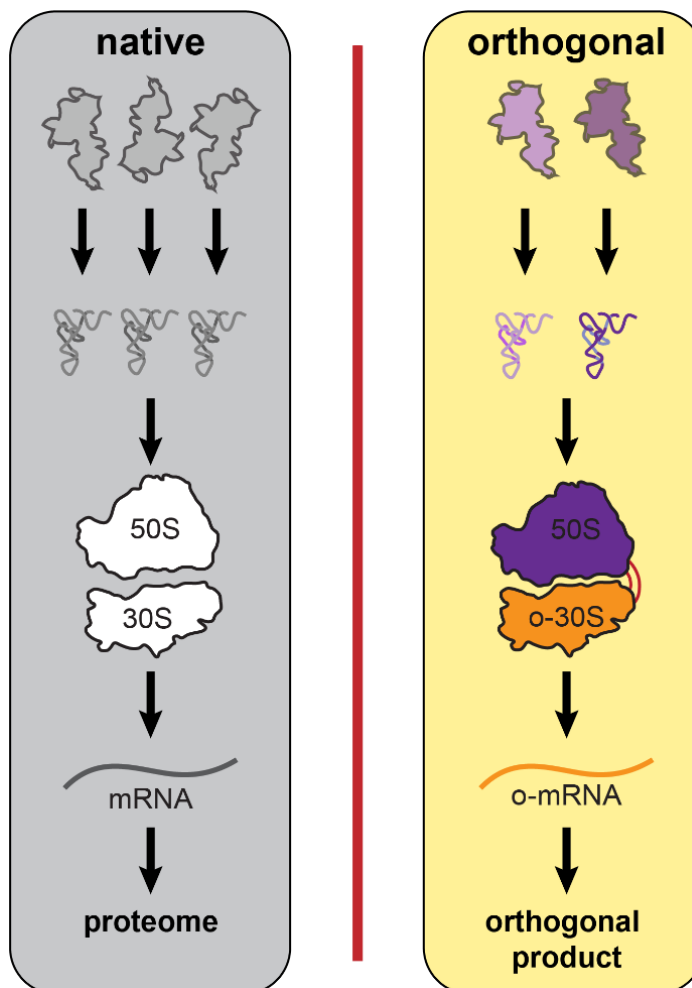
Figure 3-1: An idealized OTS is entirely self-contained, with no crosstalk between native and orthogonal components. The OTS uses its own suite of orthogonal translation components, including an orthogonal (o-) mRNA, an o-ribosome (here, a tethered ribosome) directed toward the o-mRNA, a set of o-tRNAs that specifically and selectively interact with the o-ribosome, and a suite of o-aaRS/o-tRNA pairs that are specific and selective for one another and orthogonal to all native aaRSs and tRNAs.

While mobilization is an important step toward the realization of the first goal mentioned in the previous paragraph, unfortunately, our current orthogonal translation systems continue to suffer from incomplete orthogonality within the cellular environment (**Fig. 3-1**). That is, given that native and orthogonal translation systems co-exist in the cellular environment, unwanted interactions between native and orthogonal components are unavoidable. This results in less fit, slower-growing chassis strains due to interference with the native translation system, and it

allows escape from selections leveraged on the OTS (**Fig. 3-2**). For instance, an mRNA directed toward an o-ribosome via the Shine-Dalgarno interaction remains recognizable and translatable by n-ribosomes, albeit at lower rates.[2] This leak provides an opportunity for cells to circumvent selective pressure applied through the orthogonal circuit and could ultimately render the o-ribosome and all further upstream orthogonal components functionally redundant for the cell's fitness. Similarly, n-tRNAs trafficked to the o-ribosome and used in place of desired o-tRNAs provide a pathway for selection escape in selections leveraged on o-tRNAs and/or o-aaRSs. As such, continued efforts to increase the orthogonality of interactions between native and orthogonal systems as well as efforts to provide new mechanisms for enforcement of orthogonal interactions is crucial.
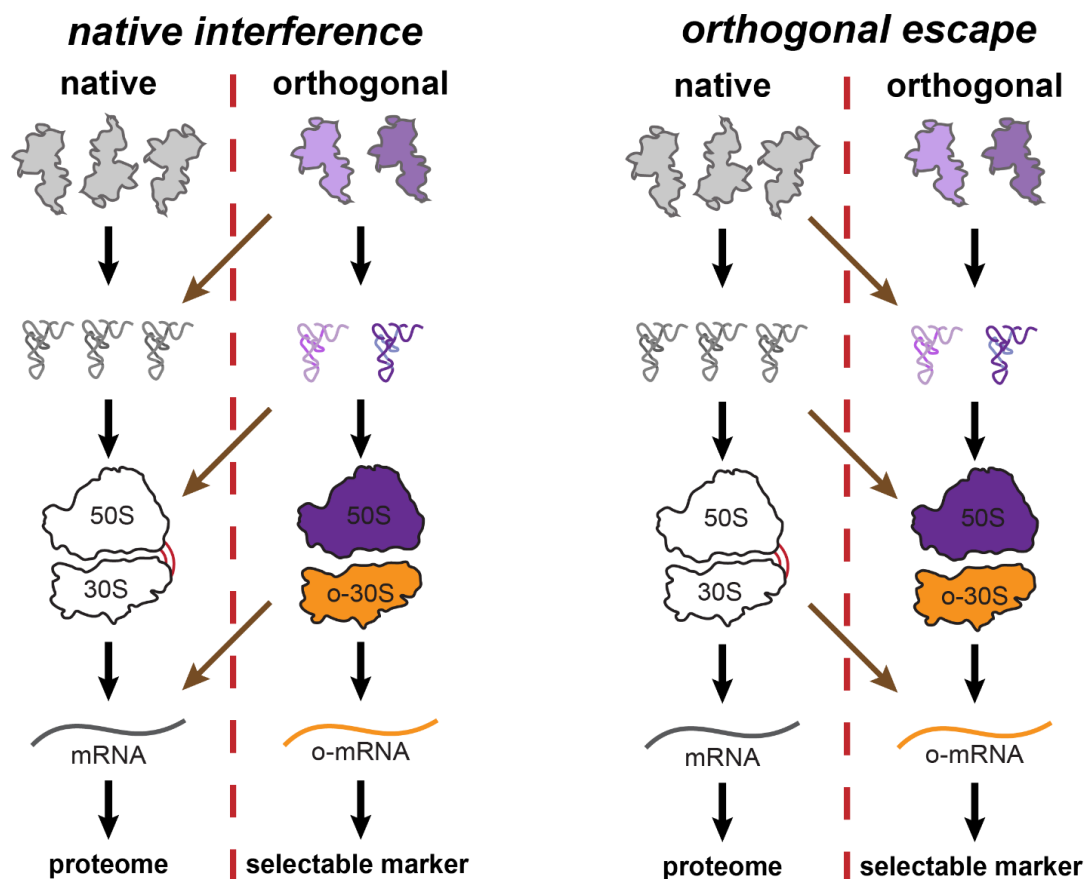
Figure 3-2: Consequences of incomplete orthogonality on downstream selections acting on a simplified OTS. At each major interaction step (1) aaRS charging of tRNAs; (2) tRNA trafficking to and interaction with the translating ribosome complex; and (3) ribosomal recognition, initiation, and complete translation of mRNAs, crosstalk between native and orthogonal systems limits the fidelity and power of selections acting on the OTS. Orthogonal components inappropriately interfering with the native translation system can cause inaccurate translation of essential protein products, reducing cell viability and increasing the cell burden of the OTS beyond its extra metabolic requirements. On the other hand, native components substituting into an orthogonal pathway are likely a primary means of escape for certain selections that would be leveraged on the OTS.

We might imagine orthogonality in this context as a chain anchored by the genetic code as written in mRNA: with the orthogonalization of successive interacting translation components beginning from the mRNA, orthogonality of the entire system strengthens and extends further outward into the orthogonal network, enabling new capabilities. For example, increasing orthogonality of mRNA-ribosome interactions ensures fewer escape routes for selections leveraged on an o-ribosome via the o-mRNA. From there, orthogonalizing interactions between

an o-ribosome and a suite of o-tRNAs could lead to an orthogonal genetic code, in which orthogonal and native translation systems decode the same genetic message differently.[37] Alternatively, orthogonalizing interactions between o-rRNA and o-ribosomal proteins could lead to o-ribosomes that are not structurally constrained by their dependence on n-ribosomal proteins. Like links in a chain, these orthogonal components connect one another back to their anchor, an o-mRNA with specific engineered or cellular function. Stronger orthogonality at each successive chain link strengthens the metaphorical chain, reducing opportunities for selection escape and improving the efficiency of synthetic translation system.

2. Building a more orthogonal o-ribosome

As previously mentioned, while ribosome-mRNA orthogonality pairing enforced through the Shine-Dalgarno interaction mediating translation initiation is effective, it remains susceptible to "leak" in which unintended ribosome-mRNA pairs can interact and initiate translation.[2] As such, developing secondary mechanisms for enforcement of orthogonal ribosome-mRNA interactions will improve future efforts that aim to leverage selective power on the orthogonal ribosome. One such mechanism that I am particularly optimistic about is engineering RF1 resistance into o-ribosomes (**Fig. 3-3**). In combination with a suitable amber-suppressor tRNA, this would allow selective expression of an amber-codon-containing o-mRNA by the o-ribosome enforced not only by orthogonalized Shine-Dalgarno interactions but also by incorporation of amber codons forcing an RF1-susceptible ribosome (as the n-ribosome should remain) to terminate translation. Simultaneously, an RF1-resistant ribosome opens coding space for a corresponding amber-suppressing tRNA for use with the OTS.
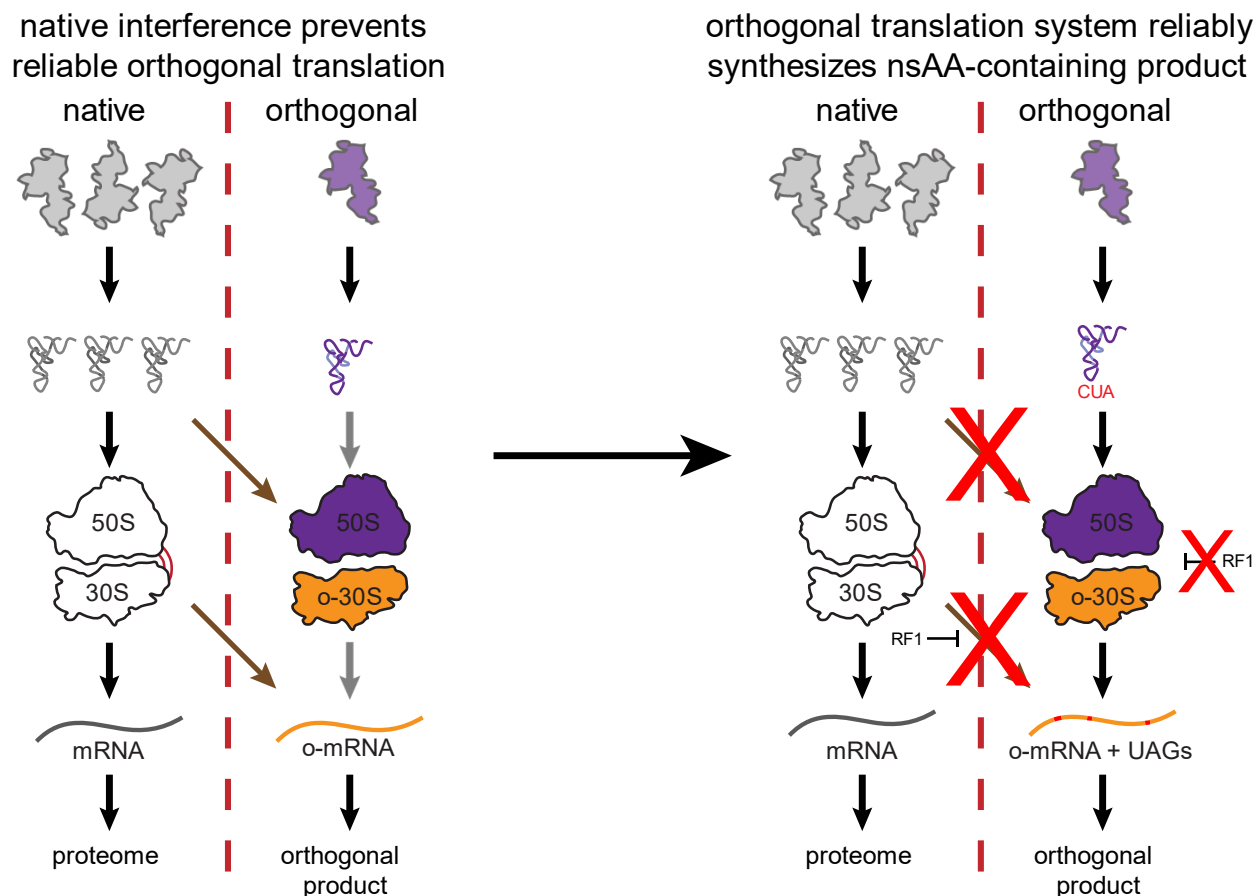
Figure 3-3: Interactions between native and orthogonal translation systems with an RF1-resistant o-ribosome. Given amber-suppressor tRNAs do not exist in native *E. coli*, the o-ribosome translating an o-mRNA with amber codons can only use the engineered amber-suppressor tRNA, preventing n-tRNA-o-ribosome crosstalk. Additionally, RF1 does not terminate o-ribosome translation at amber codons, but prevents n-ribosome-o-mRNA crosstalk by forcing n-ribosome termination at amber codons.

3. Orthogonal coding space

Another key challenge for the development of orthogonal translation systems is the

allotment of coding space for the OTS *in vivo*. Although genome-wide recoding efforts are one

potential solution to this challenge,[15, 18, 38] these efforts also have several potential design flaws

and may not be an ideal solution. For one, the allocation of coding space to orthogonal codons

can make otherwise silent or nonharmful mutations that convert a native-coding codon to a

recoded codon more threatening to the fitness of the cell. Additionally, despite more than a

decade of dedicated work, the most extensively recoded, complete strain reported to-date

contains only three recoded codons, which strictly constrains the coding space of any theoretical OTS.[18]

Another solution, enabled by the advent of robust o-ribosome systems might be the introduction of ribosome-tRNA specific interactions, such as reported by Terasaka et al. (2014).[37] Such a system, in theory, could enable free recoding of all 64 codons to orthogonal coding space without infringing upon native coding space, assuming the same orthogonality-enforcement mechanism could be introduced into other desired orthogonal tRNAs. It accomplishes this by further walling off interactions between the native and orthogonal systems, just as o-ribosomes are selectively directed toward o-mRNA transcripts.

4. Future directions – conclusions

The field of *in vivo* translation engineering is incredibly exciting, with tethered ribosomes, recoded organisms, *in-vitro*-engineered components, and mobilization all providing important footholds for future opportunities to expand OTS capabilities. Given we are dealing with living organisms, we must now find the evolutionary path that allows construction of these capabilities. Here, I have argued for the engineering of increased orthogonality into translation components, likening them to links in a chain anchored by an o-mRNA with dedicated functional purpose(s). With each new tool and platform built, we gain a stronger foothold into a future where synthetic translation systems of increasing complexity and emergent power are possible.

Appendix A – Definitions/descriptions of key strains, tools, and concepts used or referenced

SQ171

SQ171 is a strain of *E. coli* with all seven native ribosomal operons removed from the genome and replaced with two plasmids, one of which expresses the cells' ribosomes and one of which expresses essential tRNAs that were contained on the deleted ribosomal operons. This strain (and other predecessor strains in which removal/inactivation of the ribosomal operons was much less precise) has enabled a large amount of research in field of ribosome science. The ribosomal plasmid can be replaced with a plasmid expressing a ribosome of research interest via a specialized transformation procedure, enabling study of the effects of cellular dependence on variant ribosomes. However, given the disruption in regulation of core elements of its translation system, SQ171 notoriously suffers from poor growth and fitness characteristics.

Synthetic translation systems

My adoption of the term "synthetic translation system" occurred late in my PhD, in response to reviews on my published work on the topic of the majority of my PhD work. It is meant to encompass all modified or engineered translation systems – whereas we imagine orthogonal translation systems are the direct cellular circuits through which production of novel, sequenced-defined products will occur, this excludes modifications we might make to the cell-supporting translation system in order to make the cells more amenable to introduction of the orthogonal translation system. Both of these systems are crucial to expend engineering effort toward for the construction of effective chassis strains, and careful consideration to the effects that the addition of an orthogonal translation system has on the cell-supporting translation system is likely also important.

Appendix B – extraneous data, plasmid maps and sequences, primer descriptions and sequences
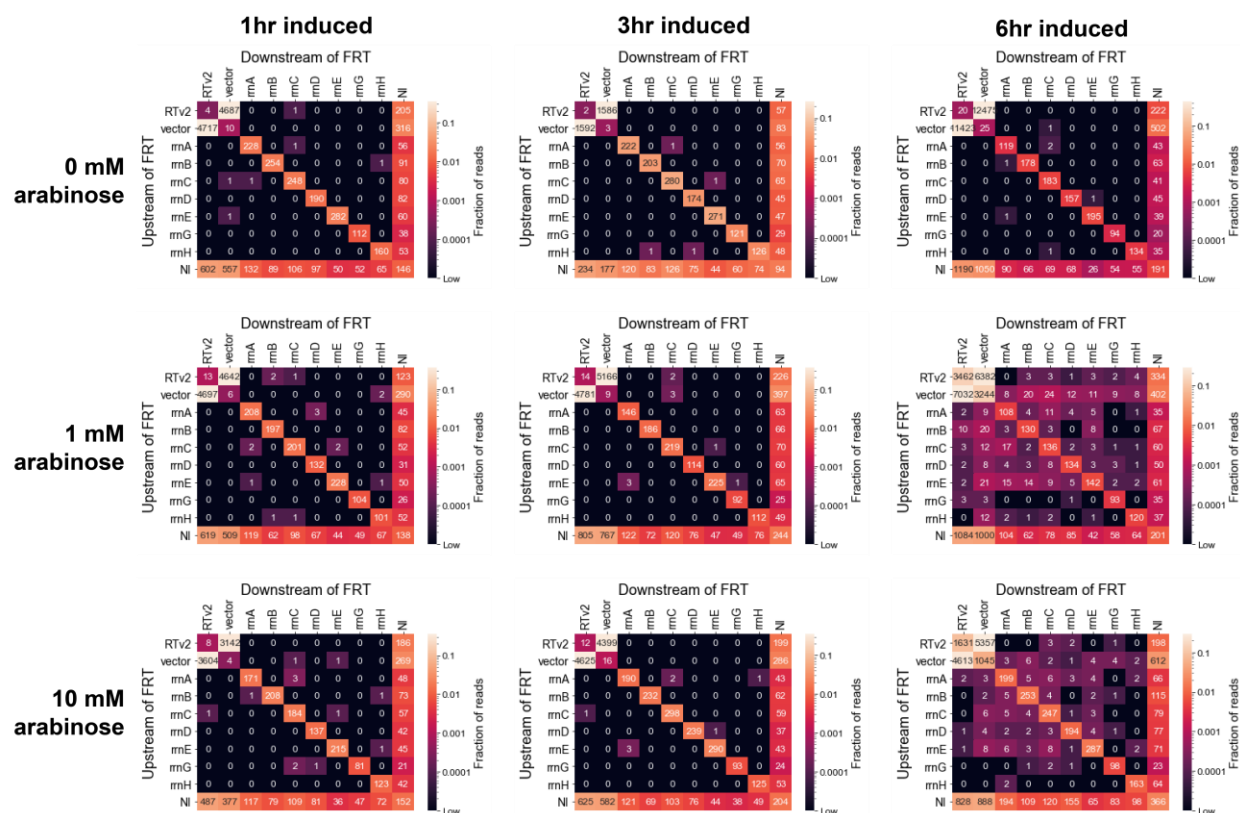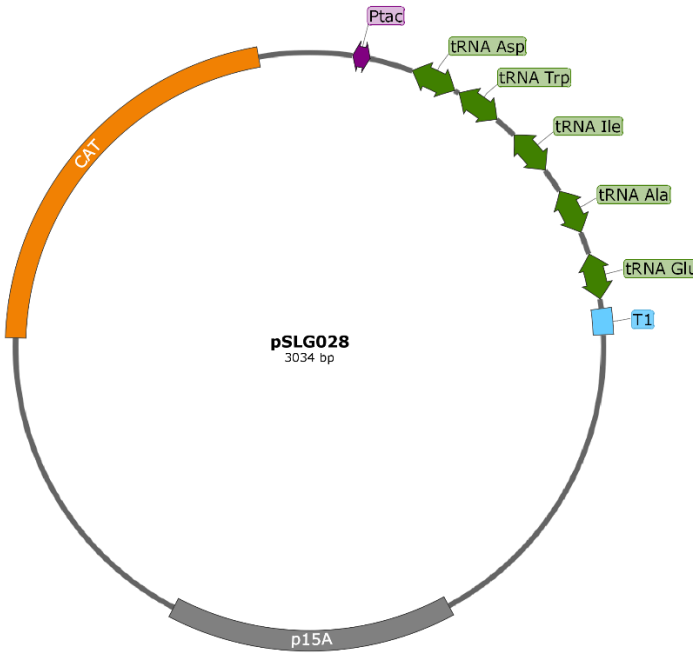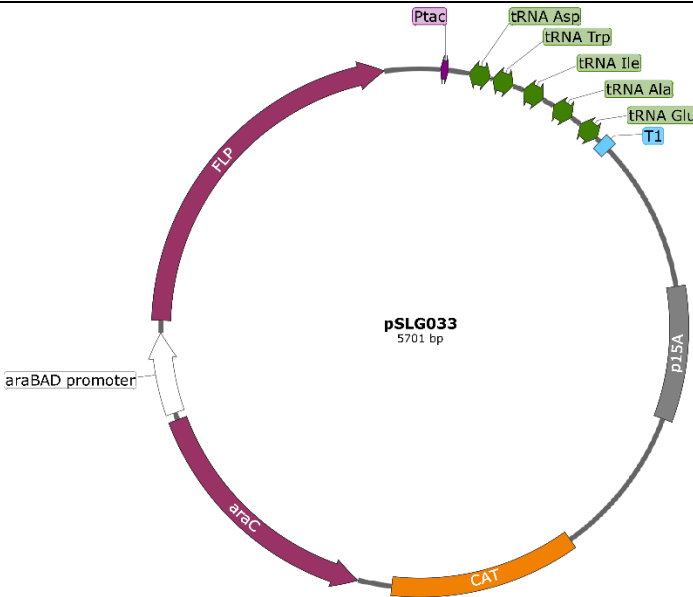
Appendix Figure 1. Unredacted results of the experiment shown in Fig. 7b. These FRT-junction maps show likely episomal read pairs (upper left corner) and read pairs in which one or both FRT-flanking regions failed to meet criteria for positive identification (NI).
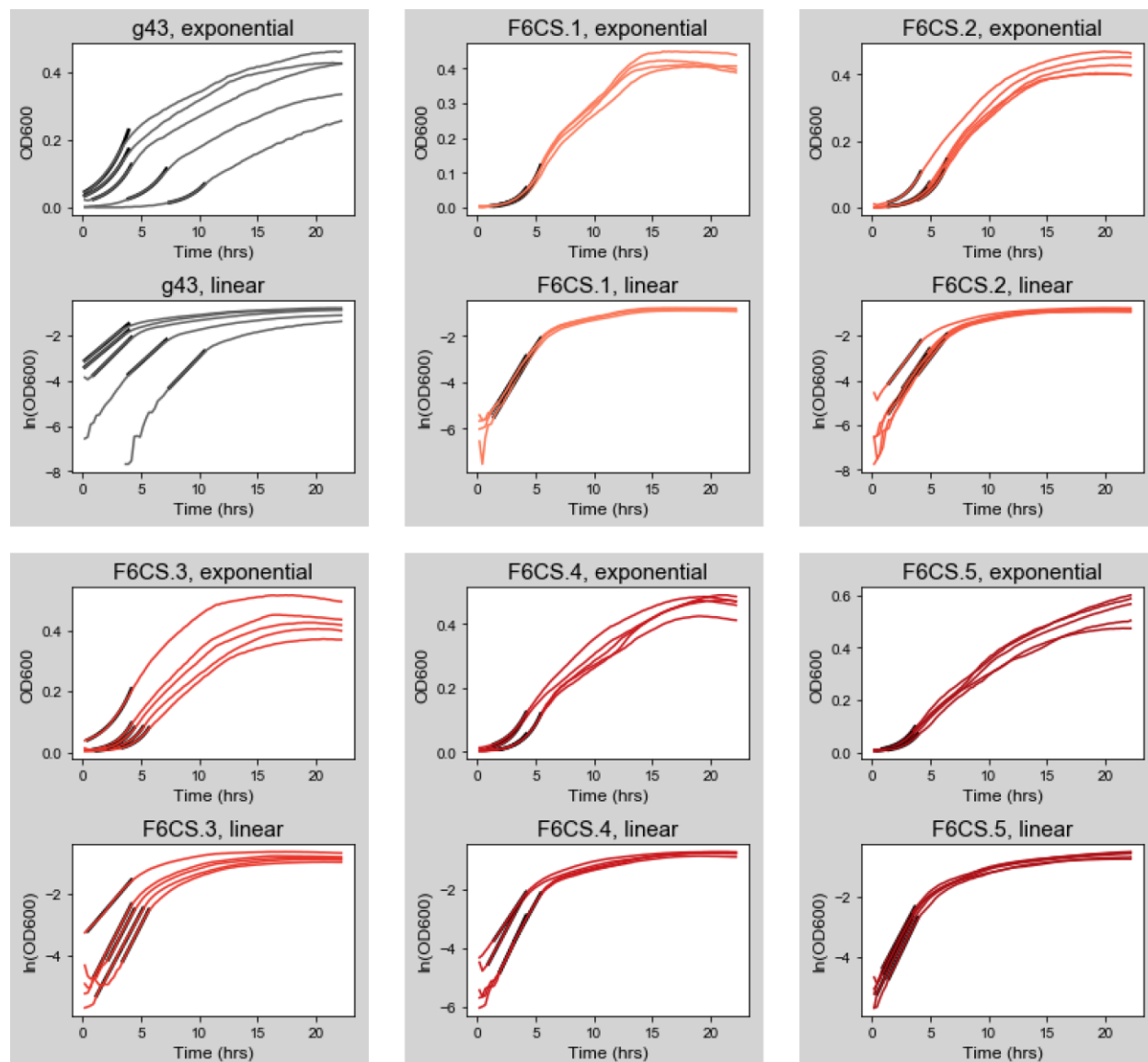
Supplementary Figure 2. Raw kinetic data and individual calculated fits for growth rate data shown in Fig. 10b. Replicates for each sample are shown plotted on the same graph. For each replicate, raw OD600 data ("exponential") and the same OD600 data, log-transformed, ("linear") are shown as a function of time in gray or shades of red. Each kinetic curve is overlaid with its calculated model function over the time window used to generate the model function in black.

Appendix Table 1. Key plasmids used in this study. "Forward" primers in the primer table are oriented clockwise from 5' to 3' in these plasmid maps, and "reverse" primers are oriented counterclockwise.

Name	Map	Key elements	Size (bp)
pSLG022		FRT: RTv2: FRT: SacB: KanR: ColE1	10,741

pSLG028		SQ171-essential tRNAs: p15A: Chloramphenicol acetyltransferase	3,034
pSLG033		SQ171-essential tRNAs: p15A: Chloramphenicol acetyltransferase: araC/araBAD/FLP	5,701

Appendix Table 2. Primers used for diagnostic colony PCR screening in this study.

Primer Name	Sequence	Binding site
SLG019	GTGGCTTCCGGAGCTAACG	16S internal, forward
EDC542	GGAGGGCGCTTACCACTTTG	3' end of 16S, forward
EDC186	TAAGGAGGTGATCCAACCGCAGG	3' end of 16S, reverse
SLG099	CTTTCTGGCGTGAATGATGGAGTAATACGTG	rrnA genomic region
SLG105	ATTGTTGAGCGAGTGGTGGCAATTG	rrnB genomic region
SLG110	CCATATCGACTTCGATCCAGCGTTCAC	rrnC genomic region
SLG115	GTTCCGGTAAATCGACAACCATTCGTTG	rrnD genomic region
SLG118	GATATCGATGTTCTCAACCGCATCTTCCAG	rrnE genomic region
SLG123	CCGGAGCCCACGTACCTATAGTTTC	rrnG genomic region
SLG126	CGATCACGGCTATGTCCATGAGATCG	rrnH genomic region
SLG155	CAAAATCCCTTAACGTGAGTTTTCGTTCCAC	5' ColE1 origin of replication, forward
SLG061	CCCTGGAAGCTCCCTCGTG	3' ColE1 origin of replication, reverse
SLG137	GGCAACTTTATGCCCATGCAACAG	Upstream of SacB cassette, forward
SLG171	CTGTTGCATGGGCATAAAGTTGC	Upstream of SacB cassette, reverse
SLG154	CGAAAACTCACGTTAAGGGATTTTGGTCATG	5' ColE1 origin of replication, reverse
SLG129	AGAAACTACCTGGCGTCAGTCCTTG	Unrelated genomic region
SLG130	AGCGAAGCTGGCATTGATGTTCATAAG	Unrelated genomic region

Appendix Table 3. Identification, primer pairs and size of diagnostic colony PCR products shown in this study.

Product ID	Primers	Source	Expected size (bp)
ColE1 ori	SLG155/SLG061	pSLG022 vector	535
RTv2-vector	EDC542/SLG171	pSLG022	1590
RTv2-rrnA	EDC542/SLG099	RTv2 genome integration at rrnA	1444
RTv2-rrnB	EDC542/SLG105	RTv2 genome integration at rrnB	1527
RTv2-rrnC	EDC542/SLG110	RTv2 genome integration at rrnC	1393
RTv2-rrnD	EDC542/SLG115	RTv2 genome integration at rrnD	1499
RTv2-rrnE	EDC542/SLG118	RTv2 genome integration at rrnE	1397
RTv2-rrnG	EDC542/SLG123	RTv2 genome integration at rrnG	1353
RTv2-rrnH	EDC542/SLG126	RTv2 genome integration at rrnH	1403
ColE1-rrnA	SLG155/SLG099	ColE1 genome integration at rrnA	1352
ColE1-rrnB	SLG155/SLG105	ColE1 genome integration at rrnB	1435
ColE1-rrnC	SLG155/SLG110	ColE1 genome integration at rrnC	1301
ColE1-rrnD	SLG155/SLG115	ColE1 genome integration at rrnD	1407
ColE1-rrnE	SLG155/SLG118	ColE1 genome integration at rrnE	1305
ColE1-rrnG	SLG155/SLG123	ColE1 genome integration at rrnG	1261
ColE1-rrnH	SLG155/SLG126	ColE1 genome integration at rrnH	1311
7kb PC	SLG129/SLG130	Unrelated genomic region	6977
Ribo-T PC	SLG019/EDC186	RTv2 / other ribosomes	3617 (RTv2); 694 (WT ribosome)

Appendix Table 4. Key strains used but not generated from this study.

ID	Use	Source
SQ171(JL1508)	Starting strain for this study before plasmid exchange	Plasmid exchange[2] from original SQ171[1]
POP2136	Cloning ribosomal constructs under the phage λ repressor	Previously established method[39]

4. Plasmid sequences

pSLG022:

gggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttggaagttcctattccgaagttcctattctctag
aaagtataggaacttcctgggcggccgcgatctctcacctaccaaacaatgcccccctgcaaaaaataaattcatataaaaaacata
cagataaccatctgcggtgataaattatctctggcggtgttgacataaataccactggcggtgatactgagcacgggtaccggccgctg
agaaaaagcgaagcggcactgctctcttaacaatttatcagacaatctgtgtgggcactcgaagatacggattcttaacgtcgcaagac
gaaaaatgaataccaagtctcaagagtgaacacgtaattcattacgaagtttaattctttgagcgtcaaacttttaaattgaagagtttgat
catggctcagattgaacgctggcggcaggcctaacacatgcaagtcgaacggtaacaggaagaagcttgcttctttgctgacgagtg
gcggacgggtgagtaatgtctgggaaactgcctgatggagggggataactactggaaacggtagctaataccgcataacgtcgcaa
gaccaaagagggggaccttcgggcctcttgccatcggatgtgcccagatgggattagctagtaggtggggtaacggctcacctaggc
gacgatccctagctggtctgagaggatgaccagccacactggaactgagacacggtccagactcctacgggaggcagcagtggg
gaatattgcacaatgggcgcaagcctgatgcagccatgccgcgtgtatgaagaaggccttcgggttgtaaagtactttcagcgggga
ggaagggagtaaagttaataccctttgctcattgacgttacccgcagaagaagcaccggctaactccgtgccagcagccgcggtaata
cggagggtgcaagcgttaatcggaattactgggcgtaaagcgcacgcaggcggtttgttaagtcagatgtgaaatccccgggctcaa
cctgggaactgcatctgatactggcaagcttgagtctcgtagagggggggtagaattccaggtgtagcggtgaaatgcgtagagatctg
gaggaataccggtggcgaaggcggccccctggacgaagactgacgctcaggtgcgaaagcgtggggagcaaacaggattagat
accctggtagtccacgccgtaaacgatgtcgacttggaggttgtgcccttgaggcgtggcttccggagctaacgcgttaagtcgaccgc
ctggggagtacggccgcaaggttaaaactcaaatgaattgacgggggcccgcacaagcggtggagcatgtggtttaattcgatgca
acgcgaagaaccttacctggtcttgacatccacggaagttttcagagatgagaatgtgccttcgggaaccgtgagacaggtgctgcat
ggctgtcgtcagctcgtgttgtgaaatgttgggttaagtcccgcaacgagcgcaacccttatcctttgttgccagcggtccggccgggaa
ctcaaaggagactgccagtgataaactggaggaaggtggggatgacgtcaagtcatcatggcccttacgaccagggctacacacgt
gctacaatggcgcatacaaagagaagcgacctcgcgagagcaagcggacctcataaagtgcgtcgtagtccggattggagtctgc
aactcgactccatgaagtcggaatcgctagtaatcgtggatcagaatgccacggtgaatacgttcccgggccttgtacacaccgcccg
tcacaccatgggagtgggttgcaaaagaagtaggtagcttaacccaatgaacaattggatgcgttgagctaaccggtactaatgaac
cgtgaggcttaaccgagaggttaagcgactaagcgtacacggtggatgccctggcagtcagaggcgatgaaggacgtgctaatctg
cgataagcgtcggtaaggtgatatgaaccgttataaccggcgatttccgaatggggaaacccagtgtgtttcgacacactatcattaac
tgaatccataggttaatgaggcgaaccggggggaactgaaacatctaagtaccccgaggaaaagaaatcaaccgagattcccccag
tagcggcgagcgaacggggagcagcccagagcctgaatcagtgtgtgtgttagtggaagcgtctggaaaggcgcgcgatacagg
gtgacagccccgtacacaaaaatgcacatgctgtgagctcgatgagtagggcgggacacgtggtatcctgtctgaatatggggggga
ccatcctccaaggctaaatactcctgactgaccgatagtgaaccagtaccgtgagggaaaggcgaaaagaaccccggcgagggg
agtgaaaaagaacctgaaaccgtgtacgtacaagcagtgggagcacgcttaggcgtgtgactgcgtaccttttgtataatgggtcagc
gacttatattctgtagcaaggttaaccgaataggggagccgaagggaaaccgagtcttaactgggcgttaagttgcagggtatagacc
cgaaacccggtgatctagccatgggcaggttgaaggttgggtaacactaactggaggaccgaaccgactaatgttgaaaaattagc
ggatgacttgtggctggggggtgaaaggccaatcaaaccgggagatagctggttctccccgaaagctatttaggtagcgcctcgtgaat
tcatctccgggggtagagcactgtttcggcaaggggggtcatcccgacttaccaacccgatgcaaactgcgaataccggagaatgttat
cacgggagacacacggcgggtgctaacgtccgtcgtgaagagggaaacaacccagaccgccagctaaggtcccaaagtcatgg
ttaagtgggaaacgatgtgggaaggcccagacagccaggatgttggcttagaagcagccatcatttaaagaaagcgtaatagctca
ctggtcgagtcggcctgcgcggaagatgtaacggggctaaaccatgcaccgaagctgcggcagcgacgcttatgcgttgttgggtag
gggagcgttctgtaagcctgcgaaggtgtgctgtgaggcatgctggaggtatcagaagtgcgaatgctgacataagtaacgataaag
cgggtgaaaagcccgctcgccggaagaccaagggttcctgtccaacgttaatcggggcagggtgagtcgacccctaaggcgagg
ccgaaaggcgtagtcgatgggaaacaggttaatattcctgtacttggtgttactgcgaaggggggacggagaaggctatgttggccg
ggcgacggttgtcccggtttaagcgtgtaggctggtttttccaggcaaatccggaaaatcaaggctgaggcgtgatgacgaggcacta
cggtgctgaagcaacaaatgccctgcttccaggaaaagcctctaagcatcaggtaacatcaaatcgtaccccaaaccgacacaggt
ggtcaggtagagaataccaaggcgcttgagagaactcgggtgaaggaactaggcaaaatggtgccgtaacttcgggagaaggca
cgctgatatgtaggtgaggtccctcgcggatggagctgaaatcagtcgaagataccagctggctgcaactgtttattaaaaacacagc

actgtgcaaacacgaaagtggacgtatacggtgtgacgcctgcccggtgccggaaggttaattgatggggttagcgcaagcgaagc
tcttgatcgaagccccggtaaacggcggccgtaactataacggtcctaaggtagcgaaattccttgtcgggtaagttccgacctgcac
gaatggcgtaatgatggccaggctgtctccacccgagactcagtgaaattgaactcgctgtgaagatgcagtgtacccgcggcaaga
cgggaagaccccgtgaacctttactatagcttgacactgaacattgagccttgatgtgtaggataggtgggaggctttgaagtgtggac
gccagtctgcatggagccgaccttgaaataccaccctttaatgtttgatgttctaacgttgacccgtaatccgggttgcggacagtgtctg
gtgggtagtttgactggggcggtctcctcctaaagagtaacggaggagcacgaaggttggctaatcctggtcggacatcaggaggtta
gtgcaatggcataagccagcttgactgcgagcgtgacggcgcgagcaggtgcgaaagcaggtcatagtgatccggtggttctgaat
ggaagggccatcgctcaacggataaaaggtactccggggataacaggctgataccgcccaagagttcatatcgacggcggtgtttg
gcacctcgatgtcggctcatcacatcctggggctgaagtaggtcccaagggtatggctgttcgccatttaaagtggtacgcgagctggg
tttagaacgtcgtgagacagttcggtccctatctgccgtgggcgctggagaactgagggggggctgctcctagtacgagaggaccgga
gtggacgcatcactggtgttcgggttgtcatgccaatggcactgcccggtagctaaatgcggaagagataagtgctgaaagcatctaa
gcacgaaacttgccccgagatgagttctccctgacccctttaagggtcctgaaggaacgttgaagacgacgacgttgataggccgggt
gtgtaagcgcagataactagtggagggcgcttaccactttgtgattcatgactggggtgaagtcgtaacaaggtaaccgtaggggaac
ctgcggttggatcacctccttaccttaaagaagcgtactttgtagtgctcacacagattgtctgatagaaagtgaaaagcaaggcgttta
cgcgttgggagtgaggctgaagagaataaggccgttcgctttctattaatgaaagctcaccctacacgaaaatatcacgcaacgcgtg
ataagcaattttcgtgtccccttcgtctagaggcccaggacaccgccctttcacggcggtaacagggggttcgaatcccctaggggacg
ccacttgctggtttgtgagtgaaagtcgccgaccttaatatctcaaaactcatcttcgggtgatgtttgagatatttgctctttaaaaatctgg
atcaagctgaaaattgaaacactgaacaacgagagttgttcgtgagtctctcaaattttcgcaacacgatgatgaatcgaaagaaaca
tcttcgggttgtgagcttaagcttacaacgccgaagctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacg
cagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaa
cgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaatgaaatcatccttagcgaaag
ctaaggatttttttgaagttcctattccgaagttcctattctctagaaagtataggaacttcttcaaaatatgtatccgctccagagaacaat
aaccctgataaatgcttcaataatattgaaaaaggaggagtatgagtattcaacatttccgtgtcgcccttattcccttttttgggggcatttt
gccttcctgttttttgctccacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaact
ggatctcaacagcggtaagatccttgagagtttttcgccccgaagaacgttttccaatgatgagcactttaaagttctgctatgtggcgcg
gtattatcccgtgttgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcaca
gaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttc
tgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgga
gctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgcaggtcgactctagaggatcgatccttttaacccatcac
atatacctgccgttcactattatttagtgaaatgagatattatgatattttctgaattgtgattaaaaaggcaactttatgcccatgcaacaga
aactataaaaaatacagagaatgaaaagaaacagatagattttttagttctttaggcccgtagtctgcaaatccttttatgattttctatcaa
acaaaagaggaaaatagaccagttgcaatccaaacgagagtctaatagaatgaggtcgaaaagtaaatcgcgcgggtttgttactg
ataaagcaggcaagacctaaaatgtgtaaagggcaaagtgtatactttggcgtcacccccttacatattttaggtctttttttattgtgcgtaa
ctaacttgccatcttcaaacaggagggctggaagaagcagaccgctaacacagtacataaaaaaggagacatgaacgatgaaca
tcaaaaagtttgcaaaacaagcaacagatatgccttactaccgcactgctggcaggaggcgcaactcaagcgtttgcgaaagaaac
gaaccaaaagccatataaggaaacatacggcatttcccatattacacgccatgatatgctgcaaatccctgaacagcaaaaaaatg
aaaaatatcaagttcctgaattcgattcgtccacaattaaaaatatctcttctgcaaaaggcctggacgtttgggacagctggccattaca
aaacgctgacggcactgtcgcaaactatcacggctaccacatcgtctttgcattagccggagatcctaaaaatgcggatgacacatcg
atttacatgttctatcaaaaagtcggcgaaacttctattgacagctggaaaaacgctggccgcgtctttaaagacagcgacaaattcga
tgcaaatgattctatcctaaaagaccaaacacaagaatggtcaggttcagccacatttacatctgacggaaaaatccgtttattctacac
tgatttctccggtaaacattacggcaaacaaacactgacaactgcacaagttaacgtatcagcatcagacagctctttgaacatcaac
ggtgtagaggattataaatcaatctttgacggtgacggaaaaacgtatcaaaatgtacagcagttcatcgatgaaggcaactacagct
caggcgacaaccatacgctgagagatcctcactacgtagaagataaaggccacaaatacttagtatttgaagcaaacactggaact
gaagatggctaccaaggcgaagaatctttatttaacaaagcatactatggcaaaagcacatcattcttccgtcaagaaagtcaaaaa
cttctgcaaagcgataaaaaacgcacggctgagttagcaaacggcgctctcggtatgattgagctaaacgatgattacacactgaaa
aaagtgatgaaaccgctgattgcatctaacacagtaacagatgaaattgaacgcgcgaacgtctttaaaatgaacggcaaatggtac

ctgttcactgactcccgcggatcaaaaatgacgattgacggcattacgtctaacgatatttacatgcttggttatgtttctaattctttaactgg
cccatacaagccgctgaacaaaactggccttgtgttaaaaatggatcttgatcctaacgatgtaacctttacttactcacacttcgctgtac
ctcaagcgaaaggaaacaatgtcgtgattacaagctatatgacaaacagaggattctacgcagacaaacaatcaacgtttgcgcca
agcttcctgctgaacatcaaaggcaagaaaacatctgttgtcaaagacagcatccttgaacaaggacaattaacagttaacaaataa
aaacgcaaaagaaatgccgatatcctattggcattttcttttatttcttatcaacataaaggtgaatcccatgaactatataaaagcag
gcaaatggctaaccgtattcctaacctttggtaatgactccaacttattgatagtgtttatgttcagataatgcccgatgactttgtcatgca
gctccaccgattttgagaacgacagcgacttccgtcccagccgtgccaggtgctgcctcagattcaggttatgccgctcaattcgctgcg
tatatcgcttgctgattacgtgcagctttcccttcaggcgggattcatacagcggccagccatccgtcatccatatcaccacgtcaaagg
gtgacagcaggctcataagacgccccagcgtcgccatagtgcgttcaccgaatacgtgcgcaacaaccgtcttccggagactgtcat
acgcgtaaaacagccagcgctggcgcgatttagccccgacatagccccactgttcgtccatttccgcgcagacgatgacgtcactgc
ccggctgtatgcgcgaggttaccgactgcggcctgagtttttaagtgacgtaaaatcgtgttgaggccaacgcccataatgcgggctgt
tgcccggcatccaacgccattcatggccatatcaatgatttctggtgcgtaccgggttgagaagcggtgtaagtgaactgcagggggg
gggggcgctgaggtctgcctcgtgaagaaggtgttgctgactcataccaggcctgaatcgccccatcatccagccagaaagtgag
ggagccacggttgatgagagctttgttgtaggtggaccagttggtgattttgaacttttgctttgccacggaacggtctgcgttgtcgggaa
gatgcgtgatctgatccttcaactcagcaaaagttcgatttattcaacaaagccgccgtcccgtcaagtcagcgtaatgctctgccagtgt
tacaaccaattaaccaattctgattagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatt
ttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccg
actcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatc
cggtgagaatggcaaaagcttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaac
caaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatg
caaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgtttttcccggg
gatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagttt
agtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatc
gatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctc
gagcaagacgtttcccgttgaatatggctcataacacccccttgtattactgtttatgtaagcagacagtttattgttcatgatgatatatttttat
cttgtgcaatgtaacatcagagattttgagacacatcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgta
gaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggttt
gtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagc
cgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcg
ataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacac
agcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaaggga
gaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccaggggggaaacgcctggta
tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggg

pSLG028:

cgcattaaaatctagcgaggatccgagatctctttcctttccatcaaaaaaatattgatgaaatgagctgttgacaattaatcatcggctc
gtataatgtgtggaattgtcacacaggaaacagaattcccggggatctggggatcatcgatggttgtaaaagaattcggtggagcgg
tagttcagtcggttagaatacctgcctgtcacgcaggggtcgcgggttcgagtcccgtccgttccgccaccctaattaggggcgtagtt
caattggtagagcaccggtctccaaaaccgggtgttgggagttcgagtctctccgcccctgccagaaatcatccttgtcgatgggagca
gtaaaacctctacaggcttgtagctcaggtggttagagcgcaccctgataagggtgaggtcggtggttcaagtccactcaggcctac
caaatttgcacggcaaatttgaagaggttttaactacatgttatggggctatagctcagctgggagagcgcctgctttgcacgcaggag
gtctgcggttcgatcccgcatagctccaccatctctgtagtgattaagagcgtgataagcaattttcgtgtccccttcgtctagaggcccag
gacaccgcccttcacggcggtaacaggggttcgaatcccctaggggacgccactctaggaaatccgccataaaacaaaaggctc
agtcggaagactgggccttttgttttatgtcgacgggcataaataggtttaattttgctacggggggcgttatttaggtttttttcttctttcgaaaa

aatctttctttatgaagttaaaagctatgtattcaatagcatattttgaatatggacatagaatagtgcttatcactattgcatatagcatcttat
ctgacacaaggaaataatacccttcgctgtttttttgttataaggtatatatatataagtgtgcagtacaggccaaataaaatattttttatgta
gtatcttaaatcccgcaagaggcccggcagtaccggcataaccaagcctatgcctacagcatccagggtgacggtgccgaggatga
cgatgagcgcattgttagatttcatacacggtgcctgactgcgttagcaatttaactgtgataaactaccgcattaaagcttatcgatgata
agctgtcaaacatgagaattacaacttatatcgtatgggggctgacttcaggtgctacatttgaagagataaaattgcactgaaatctagaa
atatttatctgattaataagatgatcttcttgagatcgtttttggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgcagggc
ggttttttcgaaggttctctgagctaccaactctttgaaccgaggtaactggcttggaggagcgcagtcaccaaaacttgtcctttcagttta
gccttaaccggcgcatgacttcaagactaactcctctaaatcaattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttgg
actcaagacgatagttaccggataaggcgcagcggtcggactgaacggggggttcgtgcatacagtccagcttggagcgaactgcc
tacccggaactgagtgtcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaaggcagga
acaggagagcgcacgagggagccgccaggggggaaacgcctggtatctttatagtcctgtcgggtttcgccaccactgatttgagcgt
cagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttgccgcggccctctcacttccctgttaagtatcttcctggc
atcttccaggaaatctccgccccgttcgtaagccatttccgctcgccgcagtcgaacgaccgagcgtagcgagtcagtgagcgagga
agcggaatatatcctgtatcacatattctgctgacgcaccggtgcagcctttttttctcctgccacatgaagcacttcactgacaccctcatc
agtgccaacatagtaagccagtatacactccgctagcgctgatgtccggcggtgcttttgccgttacgcaccaccccgtcagtagctga
acaggagggacagagttgcgtttctacaaactcttcctgtcgtcatatctacaagccggcgcgccaaattgacaattactcatccggctc
gaataatgtgtggaacttaaacacacacaggaggaaaacatatggaaaaaaaaatcaccggctacaccaccgttgacatctctca
gtggcaccgtaaagaacactttgaagcgttccagtctgtcgcgcagtgtacctacaaccagaccgttcagctagacatcaccgcgttc
ctgaaaaccgttaaaaaaaacaaacacaaattctacccggcgttcattcacatcctggcgcgtctgatgaacgcgcacccggaatttc
gtatggcgatgaaagacggtgaactggttatctgggactctgttcacccgtgctacaccgttttccacgaacagaccgaaaccttctcttc
tctgtggtctgaataccacgacgacttccgtcagttcctgcacatctactctcaggacgttgcgtgctacggtgaaaacctggcgtacttc
ccgaaaggtttcatcgaaaacatgttcttcgtttctgcgaacccgtgggtttctttcacctctttcgacctgaacgtggcgaacatggacaa
cttcttcgcgccggttttcactatgggtaaatactacacccagggtgacaaagttctgatgccgctggcgatccaggttcaccacgcggtt
tgcgacggtttccacgttggtcgtatgctgaacgaactccagcagtattgcgacgaatggcagggtggtgcgtaaactcactcctagcc
cgcctaataagagctcatgaagttcctattccgaagttccgcgaacgcgtaaaggatctaggtgaagatc

pSLG033:

cgcattaaaatctagcgaggatccgagatctctttcctttccatcaaaaaaatattgatgaaatgagctgttgacaattaatcatcggctc
gtataatgtgtggaattgtcacacaggaaacagaattcccggggatctgggggatcatcgatggttgtaaaagaattcggtggagcgg
tagttcagtcggttagaatacctgcctgtcacgcaggggggtcgcgggttcgagtcccgtccgttccgccaccctaattagggggcgtagtt
caattggtagagcaccggtctccaaaaccgggtgttgggagttcgagtctctccgcccctgccagaaatcatccttgtcgatgggagca
gtaaaacctctacaggcttgtagctcaggtggttagagcgcacccctgataagggtgaggtcggtggttcaagtccactcaggcctac
caaatttgcacggcaaatttgaagaggttttaactacatgttatggggctatagctcagctgggagagcgcctgctttgcacgcaggag
gtctgcggttcgatcccgcatagctccaccatctctgtagtgattaagagcgtgataagcaattttcgtgtcccccttcgtctagaggcccag
gacaccgccctttcacggcggtaacaggggttcgaatcccctaggggacgccactctaggaaatccgccataaaacaaaaggctc
agtcggaagactgggcctttttgtttatgtcgacgggcataaataggtttaattttgctacgggggcgttatttaggttttttcttctttcgaaaa
aatctttctttatgaagttaaaagctatgtattcaatagcatattttgaatatggacatagaatagtgcttatcactattgcatatagcatcttat
ctgacacaaggaaataatacccttcgctgtttttttgttataaggtatatatatataagtgtgcagtacaggccaaataaaatattttttatgta
gtatcttaaatcccgcaagaggcccggcagtaccggcataaccaagcctatgcctacagcatccagggtgacggtgccgaggatga
cgatgagcgcattgttagatttcatacacggtgcctgactgcgttagcaatttaactgtgataaactaccgcattaaagcttatcgatgata
agctgtcaaacatgagaattacaacttatatcgtatgggggctgacttcaggtgctacatttgaagagataaaattgcactgaaatctagaa
atatttatctgattaataagatgatcttcttgagatcgtttttggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgcagggc
ggttttttcgaaggttctctgagctaccaactctttgaaccgaggtaactggcttggaggagcgcagtcaccaaaacttgtcctttcagttta
gccttaaccggcgcatgacttcaagactaactcctctaaatcaattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttgg

actcaagacgatagttaccggataaggcgcagcggtcggactgaacgggggggttcgtgcatacagtccagcttggagcgaactgcc
tacccggaactgagtgtcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaaggcagga
acaggagagcgcacgagggagccgccaggggggaaacgcctggtatctttatagtcctgtcgggtttcgccaccactgatttgagcgt
cagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttgccgcggccctctcacttccctgttaagtatcttcctggc
atcttccaggaaatctccgccccgttcgtaagccatttccgctcgccgcagtcgaacgaccgagcgtagcgagtcagtgagcgagga
agcggaatatatcctgtatcacatattctgctgacgcaccggtgcagcctttttttctcctgccacatgaagcacttcactgacaccctcatc
agtgccaacatagtaagccagtatacactccgctagcgctgatgtccggcggtgcttttgccgttacgcaccaccccgtcagtagctga
acaggagggacagagttgcgtttctacaaactcttcctgtcgtcatatctacaagccggcgcgccaaattgacaattactcatccggctc
gaataatgtgtggaacttaaacacacacaggaggaaaacatatggaaaaaaaaatcaccggctacaccaccgttgacatctctca
gtggcaccgtaaagaacactttgaagcgttccagtctgtcgcgcagtgtacctacaaccagaccgttcagctagacatcaccgcgttc
ctgaaaaccgttaaaaaaaacaaacacaaattctacccggcgttcattcacatcctggcgcgtctgatgaacgcgcacccggaatttc
gtatggcgatgaaagacggtgaactggttatctgggactctgttcacccgtgctacaccgttttccacgaacagaccgaaaccttctcttc
tctgtggtctgaataccacgacgacttccgtcagttcctgcacatctactctcaggacgttgcgtgctacggtgaaaacctggcgtacttc
ccgaaaggtttcatcgaaaacatgttcttcgtttctgcgaacccgtgggtttctttcacctctttcgacctgaacgtggcgaacatggacaa
cttcttcgcgccggttttcactatgggtaaatactacacccagggtgacaaagttctgatgccgctggcgatccaggttcaccacgcggtt
tgcgacggtttccacgttggtcgtatgctgaacgaactccagcagtattgcgacgaatggcagggtggtgcgtaaactcactcctagcc
cgcctaataagagctcatgaagttcctattccgaagttccgcgaacgcgtaaaggatctaggtgaagatccacatttccccgaaaagt
gccacctgcatcgatttattatgacaacttgacggctacatcattcacttttttcttcacaaccggcacggaactcgctcgggctggccccg
gtgcattttttaaatacccgcgagaaatagagttgatcgtcaaaaccaacattgcgaccgacggtggcgataggcatccgggtggtgc
tcaaaagcagcttcgcctggctgatacgttggtcctcgcgccagcttaagacgctaatccctaactgctggcggaaaagatgtgacag
acgcgacggcgacaagcaaacatgctgtgcgacgctggcgatatcaaaattgctgtctgccaggtgatcgctgatgtactgacaagc
ctcgcgtacccgattatccatcggtggatggagcgactcgttaatcgcttccatgcgccgcagtaacaattgctcaagcagatttatcgc
cagcagctccgaatagcgcccttcccccttgcccggcgttaatgatttgcccaaacaggtcgctgaaatgcggctggtgcgcttcatccg
ggcgaaagaaccccgtattggcaaatattgacggccagttaagccattcatgccagtaggcgcgcggacgaaagtaaacccactg
gtgataccattcgcgagcctccggatgacgaccgtagtgatgaatctctcctggcgggaacagcaaaatatcacccggtcggcaaac
aaattctcgtccctgatttttcaccaccccctgaccgcgaatggtgagattgagaatataacctttcattcccagcggtcggtcgataaaa
aaatcgagataaccgttggcctcaatcggcgttaaacccgccaccagatgggcattaaacgagtatcccggcagcaggggatcattt
tgcgcttcagccatacttttcatactcccgccattcagagaagaaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttac
tggctcttctcgctaaccaaaccggtaaccccgcttattaaaagcattctgtaacaaagcgggaccaaagccatgacaaaaacgcgt
aacaaaagtgtctataatcacggcagaaaagtccacattgattatttgcacggcgtcacactttgctatgccatagcattttttatccataag
attagcggatcctacctgacgctttttatcgcaactctctactgtttctccatacccgtttttttgggaattcgagctctaaggaggttataaaa
aatgccacaatttgatatattatgtaaaacaccacctaaggtgcttgttcgtcagtttgtggaaaggtttgaaagaccttcaggtgagaaa
atagcattatgtgctgctgaactaacctatttatgttggatgattacacataacggaacagcaatcaagagagccacattcatgagctat
aatactatcataagcaattcgctgagtttcgatattgtcaataaatcactccagtttaaatacaagacgcaaaaagcaacaattctggaa
gcctcattaaagaaattgattcctgcttgggaatttacaattattccttactatggacaaaaacatcaatctgatatcactgatattgtaagta
gtttgcaattacagttcgaatcatcggaagaagcagataagggaaatagccacagtaaaaaaatgcttaaagcacttctaagtgagg
gtgaaagcatctgggagatcactgagaaaatactaaattcgtttgagtatacttcgagatttacaaaaacaaaaactttataccaattcct
cttcctagctactttcatcaattgtggaagattcagcgatattaagaacgttgatccgaaatcatttaaattagtccaaaataagtatctggg
agtaataatccagtgtttagtgacagagacaaagacaagcgttagtaggcacatatacttctttagcgcaaggggtaggatcgatcca
cttgtatatttggatgaatttttgaggaattctgaaccagtcctaaaacgagtaaataggaccggcaattcttcaagcaataaacaggaa
taccaattattaaaagataacttagtcagatcgtacaataaagctttgaagaaaaatgcgccttattcaatctttgctataaaaaatggcc
caaaatctcacattggaagacatttgatgacctcatttctttcaatgaagggcctaacggagttgactaatgttgtgggaaattggagcga
taagcgtgcttctgccgtggccaggacaacgtatactcatcagataacagcaatacctgatcactacttcgcactagtttctcggtactat
gcatatgatccaatatcaaaggaaatgatagcattgaaggatgagactaatccaattgaggagtggcagcatatagaacagctaaa
gggtagtgctgaaggaagcatacgatacccgcatggaatgggataatatcacaggaggtactagactacctttcatcctacataaat

agacgcatataaccatggcatgcatggtatcgagatggcacatagccttgctcaaattggaatcaggtttgtgccaataccagtagaa
acagacgaagaatccatgggtatggacagttttccctttgatatgtaacg

5. References

1.   Quan, S.; Skovgaard, O.; McLaughlin, R. E.; Buurman, E. T.; Squires, C. L., Markerless *Escherichia coli rrn* Deletion Strains for Genetic Determination of Ribosomal Binding Sites. *G3: Genes|Genomes|Genetics* **2015,** *5* (12), 2555-2557.

2.   Carlson, E. D.; d'Aquino, A. E.; Kim, D. S.; Fulk, E. M.; Hoang, K.; Szal, T.; Mankin, A. S.; Jewett, M. C., Engineered ribosomes with tethered subunits for expanding biological function. *Nature Communications* **2019,** *10* (1), 3920.

3.   Orelle, C.; Carlson, E. D.; Szal, T.; Florin, T.; Jewett, M. C.; Mankin, A. S., Protein synthesis by ribosomes with tethered subunits. *Nature* **2015,** *524*, 119.

4.   Fried, S. D.; Schmied, W. H.; Uttamapinant, C.; Chin, J. W., Ribosome Subunit Stapling for Orthogonal Translation in E. coli. *Angewandte Chemie International Edition* **2015,** *54* (43), 12791-12794.

5.   Kim, D. S.; Watkins, A.; Bidstrup, E.; Lee, J.; Topkar, V.; Kofman, C.; Schwarz, K. J.; Liu, Y.; Pintilie, G.; Roney, E.; Das, R.; Jewett, M. C., Three-dimensional structure-guided evolution of a ribosome with tethered subunits. *Nature Chemical Biology* **2022**.

6.   Rackham, O.; Chin, J., A network of orthogonal ribosome‐mRNA pairs. *Nature chemical biology* **2005,** *1*, 159-66.

7.   Wang, K.; Fredens, J.; Brunner, S. F.; Kim, S. H.; Chia, T.; Chin, J. W., Defining synonymous codon compression schemes by genome recoding. *Nature* **2016,** *539* (7627), 59-64.

8.   Aleksashin, N. A.; Leppik, M.; Hockenberry, A. J.; Klepacki, D.; Vázquez-Laslop, N.; Jewett, M. C.; Remme, J.; Mankin, A. S., Assembly and functionality of the ribosome with tethered subunits. *Nature communications* **2019,** *10* (1), 930.

9.   Aleksashin, N. A.; Szal, T.; d'Aquino, A. E.; Jewett, M. C.; Vázquez-Laslop, N.; Mankin, A. S., A fully orthogonal system for protein synthesis in bacterial cells. *Nature Communications* **2020,** *11* (1), 1858.

10.   Jewett, M. C.; Fritz, B. R.; Timmerman, L. E.; Church, G. M., In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol Syst Biol* **2013,** *9*, 678.

11.   Hammerling, M. J.; Fritz, B. R.; Yoesep, D. J.; Kim, D. S.; Carlson, E. D.; Jewett, M. C., In vitro ribosome synthesis and evolution through ribosome display. *Nature Communications* **2020,** *11* (1), 1108.

12.   Kofman, C.; Lee, J.; Jewett, M. C., Engineering molecular translation systems. *Cell Syst* **2021,** *12* (6), 593-607.

13.   Lee, J.; Schwarz, K. J.; Kim, D. S.; Moore, J. S.; Jewett, M. C., Ribosome-mediated polymerization of long chain carbon and cyclic amino acids into peptides in vitro. *Nature Communications* **2020,** *11* (1), 4304.

14.   Kruger, A.; Watkins, A.; Wellington-Oguri, R.; Romano, J.; Kofman, C.; DeFoe, A.; Kim, Y.; Anderson-Lee, J.; Fisker, E.; Townley, J.; d'Aquino, A.; Das, R.; Jewett, M., *Community science designed ribosomes with beneficial phenotypes*. 2021.

15.   Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.-R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J., Genomically Recoded Organisms Expand Biological Functions. **2013,** *342* (6156), 357-360.

16.   Mandell, D. J.; Lajoie, M. J.; Mee, M. T.; Takeuchi, R.; Kuznetsov, G.; Norville, J. E.; Gregg, C. J.; Stoddard, B. L.; Church, G. M., Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **2015,** *518* (7537), 55-60.

17.   Rovner, A. J.; Haimovich, A. D.; Katz, S. R.; Li, Z.; Grome, M. W.; Gassaway, B. M.; Amiram, M.; Patel, J. R.; Gallagher, R. R.; Rinehart, J.; Isaacs, F. J., Recoded organisms engineered to depend on synthetic amino acids. *Nature* **2015,** *518* (7537), 89-93.

18.   Fredens, J.; Wang, K.; de la Torre, D.; Funke, L. F. H.; Robertson, W. E.; Christova, Y.; Chia, T.; Schmied, W. H.; Dunkelmann, D. L.; Beránek, V.; Uttamapinant, C.; Llamazares, A. G.; Elliott, T. S.; Chin, J. W., Total synthesis of Escherichia coli with a recoded genome. *Nature* **2019,** *569* (7757), 514-518.

19.   Robertson Wesley, E.; Funke Louise, F. H.; de la Torre, D.; Fredens, J.; Elliott Thomas, S.; Spinck, M.; Christova, Y.; Cervettini, D.; Böge Franz, L.; Liu Kim, C.; Buse, S.; Maslen, S.; Salmond George, P. C.; Chin Jason, W., Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science (New York, N.Y.)* **2021,** *372* (6546), 1057-1062.

20.   Katoh, T.; Iwane, Y.; Suga, H., tRNA engineering for manipulating genetic code. *RNA Biol* **2018,** *15* (4-5), 453-460.

21.   Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G., Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proceedings of the National Academy of Sciences* **1997,** *94* (19), 10092-10097.

22.   Yesselman, J. D.; Eiler, D.; Carlson, E. D.; Gotrik, M. R.; d'Aquino, A. E.; Ooms, A. N.; Kladwang, W.; Carlson, P. D.; Shi, X.; Costantino, D. A.; Herschlag, D.; Lucks, J. B.; Jewett, M. C.; Kieft, J. S.; Das, R., Computational design of three-dimensional RNA structure and function. *Nature Nanotechnology* **2019,** *14* (9), 866-873.

23.   Kim, D. S., Watkins, A., Bidstrup, E., Lee, J., Topkar, V., Schwarz, K.J., Kofman, C.,Schwarz, K.J., Liu, Y., Pintilie, G., Roney, E., Das, R., and Jewett, M.C. , 3D-structure-guided evolution of a ribosome with tethered subunits. *Nature Chemical Biology* **In press**.

24.   Kofman, C.; Lee, J.; Jewett, M. C., Engineering molecular translation systems. *Cell systems* **2021,** *12* (6), 593-607.

25.   Datsenko, K. A.; Wanner, B. L., One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS* **2000,** *97* (12), 6640-6645.

26.   Bassalo, M. C.; Garst, A. D.; Halweg-Edwards, A. L.; Grau, W. C.; Domaille, D. W.; Mutalik, V. K.; Arkin, A. P.; Gill, R. T., Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synthetic Biology* **2016,** *5* (7), 561-568.

27.   Des Soye, B. J.; Gerbasi, V. R.; Thomas, P. M.; Kelleher, N. L.; Jewett, M. C., A Highly Productive, One-Pot Cell-Free Protein Synthesis Platform Based on Genomically Recoded Escherichia coli. *Cell chemical biology* **2019,** *26* (12), 1743-1754 e9.

28.   Ma, L.; Li, Y.; Chen, X.; Ding, M.; Wu, Y.; Yuan, Y.-J., SCRaMbLE generates evolved yeasts with increased alkali tolerance. *Microbial Cell Factories* **2019,** *18* (1), 52.

29.   Dymond, J.; Boeke, J., The Saccharomyces cerevisiae SCRaMbLE system and genome minimization. *Bioeng Bugs* **2012,** *3* (3), 168-171.

30.   Jia, B.; Wu, Y.; Li, B.-Z.; Mitchell, L. A.; Liu, H.; Pan, S.; Wang, J.; Zhang, H.-R.; Jia, N.; Li, B.; Shen, M.; Xie, Z.-X.; Liu, D.; Cao, Y.-X.; Li, X.; Zhou, X.; Qi, H.; Boeke, J. D.; Yuan, Y.-J., Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nature communications* **2018,** *9* (1), 1933-1933.

31.   Luo, Z.; Wang, L.; Wang, Y.; Zhang, W.; Guo, Y.; Shen, Y.; Jiang, L.; Wu, Q.; Zhang, C.; Cai, Y.; Dai, J., Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES. *Nature communications* **2018,** *9* (1), 1930.

32.   Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L., Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009,** *25* (11), 1422-1423.

33.   Ringrose, L.; Lounnas, V.; Ehrlich, L.; Buchholz, F.; Wade, R.; Stewart, A. F., Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. *Journal of Molecular Biology* **1998,** *284* (2), 363-384.

34.   Deatherage, D. E.; Barrick, J. E., Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* **2014,** *1151*, 165-188.
35.   Wang, H. H.; Isaacs, F. J.; Carr, P. A.; Sun, Z. Z.; Xu, G.; Forest, C. R.; Church, G. M., Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **2009,** *460*, 894.
36.   Rangarajan, A. A.; Yilmaz, C.; Schnetz, K., Deletion of FRT-sites by no-SCAR recombineering in Escherichia coli. *Microbiology* **2022,** *168* (4).
37.   Terasaka, N.; Hayashi, G.; Katoh, T.; Suga, H., An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. *Nat Chem Biol* **2014,** *10* (7), 555-7.
38.   Ostrov, N.; Landon, M.; Guell, M.; Kuznetsov, G.; Teramoto, J.; Cervantes, N.; Zhou, M.; Singh, K.; Napolitano, M. G.; Moosburner, M.; Shrock, E.; Pruitt, B. W.; Conway, N.; Goodman, D. B.; Gardner, C. L.; Tyree, G.; Gonzales, A.; Wanner, B. L.; Norville, J. E.; Lajoie, M. J.; Church, G. M., Design, synthesis, and testing toward a 57-codon genome. *Science* **2016,** *353* (6301), 819-822.
39.   Yassin, A.; Fredrick, K.; Mankin, A. S., Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (46), 16620.