Pacific Northwest
NATIONAL LABORATORY

Persistence Control
@PNNL

DOE BSSD Performance Management Metrics Report Q2: 3/31/2023

SFA Laboratory Research Manager: Jon Magnuson[1], Jon.Magnuson@pnnl.gov
SFA Principal Investigator: Robert G. Egbert[1], Robert.Egbert@pnnl.gov
[1]**Pacific Northwest National Laboratory**, Richland, WA 99354

Investigators: Robert G. Egbert[1], Jayde A. Aufrecht[1], Adam M. Deutschbauer[2], Joshua R. Elmore[1], Adam M. Guss[3], Pubudu P. Handakumbura[1], Caroline S. Harwood[6], Vivian S. Lin[1], Ryan S. McClure[1], Jason E. McDermott[1], Ernesto S. Nakayasu[1], William C. Nelson[1], Devin Coleman-Derr[4], Enoch Yeung[5]

Participating Institutions: [1]**Pacific Northwest National Laboratory**, Richland, WA 99354; [2]Lawrence Berkeley National Laboratory, 94720; [3]Oak Ridge National Laboratory, Oak Ridge, TN 37831; [4]University of California, Berkeley, CA 94704; [5]University of California, Santa Barbara, CA 93106; [6]University of Washington, Seattle, WA 98195.

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

Q2 Target: Report on progress towards designing new secure functions into specific microbial isolates in the laboratory.

Executive Summary

As PNNL's contribution to the Secure Biosystems Design program, the Persistence Control (PerCon) Science Focus Area (SFA) aims to use high-throughput genetic and bioinformatic tools to understand, predict, and control plant-microbe interactions in the rhizosphere to benefit the bioenergy crop sorghum. Specifically, we investigate how microbes survive in the rhizosphere in order to engineer metabolic dependence on root exudate compounds for bacterial isolates native to the sorghum rhizosphere. These metabolic addiction phenotypes will promote growth and activity in the plant rhizosphere and prevent proliferation away from the plant. Pursuing the secure biocontainment aims of our research will position the US to responsibly harness the full potential of rhizosphere microbiomes, to support ecosystems that fulfill national goals for a resilient bioeconomy through sustainable bioenergy cropping systems with potential to durably deposit atmospheric carbon in soils.

In the inaugural funding cycle of the SFA, we have advanced understanding of genome engineering and gene expression in taxonomically diverse bacteria[1]. These developments, summarized as follows, will advance the design of secure biosystems:

1. Engineering rhizosphere isolates compatible with high-throughput genome integration.
2. Evaluation of recombinase-assisted genome integration efficiency and accuracy.
3. Antibiotic marker recycling for responsible and iterative genome engineering.
4. High-throughput regulatory element screening for reliable gene expression.

Background

The potential to employ emerging genetic and computational tools to design and securely deploy synthetic rhizosphere microbiomes, defined communities of microbes that colonize plant roots and benefit plant growth, offers tremendous opportunity to realize highly productive and stress-tolerant biomass cropping systems. A critical obstacle to realizing this vision is understanding fundamental principles of microbiome persistence in complex environments, notably amid the dynamics of root exudation and microbial colonization. Further, we need predictive tools to assess the risks associated with the deployment or unintended release of engineered microbes in plant and soil ecosystems and to mitigate those risks through effective strategies to control persistence. To meet these challenges, high-throughput genetic manipulation and bioinformatic tools provide a platform to elucidate the genetic elements underpinning rhizosphere community function, to create beneficial communities, and to control their environmental persistence.

The Persistence Control SFA investigates approaches to reshape the environmental niche of native sorghum rhizosphere bacterial isolates by creating an engineered metabolic addiction to root exudate compounds. We aim to understand the genes and networks that underpin environmental persistence of microbiomes to create dependencies of engineered microbes on plants such that the microbes can no longer persist in the environment in the absence of the target plant (Figure 1). Our strategy to control the environmental persistence of engineered microbes is to establish metabolic addiction to plant root exudate compounds by installing compound-specific catabolic pathways while removing pathways used for scavenging nutrients in soil microbiome environments. This strategy requires exquisite control of genome content to, for instance, eliminate metabolic niches that allow survival cross-feeding of nutrients by other microbiome members as well as to introduce complete pathways to thrive in the plant rhizosphere, but not in the rhizosphere of other plants. It also demands a deep understanding of plant or microbe interactions that could overcome control through complementation.

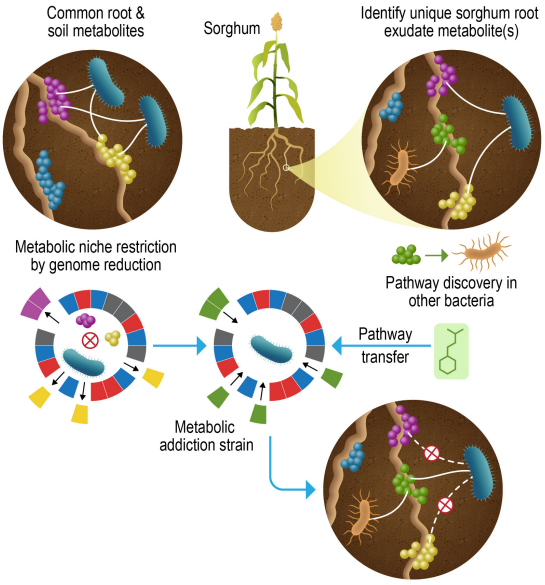

Figure 1. The Persistence Control SFA aims to control the proliferation of engineered microbes by establishing metabolic addiction of microbe to a bioenergy crop supported by the engineered microbiome.

To help accomplish the goals of the Secure Biosystems Design program, the PerCon SFA team draws expertise from across three DOE Office of Science National Laboratories – Pacific Northwest National Laboratory, Lawrence Berkeley National Laboratory, and Oak Ridge National Laboratory – and three Universities: the University of California Berkely, the University of California Santa Barbara, and the University of Washington. To understand the genes and networks that control environmental fitness we have developed and integrated research strategies that draw from our team's expertise in functional genomics, synthetic biology, microbial ecology, chemical biology, bioinformatics, machine learning, and plant-microbe interactions. In this report, we share a selection of our approaches for persistence control to securely design new functions into organisms for bioenergy and bioeconomy applications.

Engineering rhizosphere isolates compatible with high-throughput genome integration

Efficient genome engineering is critical to understand and utilize microbial functions. In the Persistence Control SFA we seek to add non-native functions, such as the ability to consume sorghum-specific root exudate compounds for metabolic addiction, into non-model bacteria that have been isolated from the sorghum rhizosphere. For this, we developed SAGE (serine recombinase-assisted genome engineering). SAGE uses a collection of enzymes from bacteriophages to efficiently insert many different recombinant DNA molecules, each encoding genes encoding non-native functions, into an engineered region of the host chromosome (**Fig. 2**).

We selected five bacteria that represent a range of taxonomic diversity and application spaces (e.g., plant growth promotion, bioremediation, and industrial biotechnology) to demonstrate SAGE. These include three plant growth promoting bacteria, the broadly used model soil microbe *Pseudomonas fluorescens* SBW25 and two novel sorghum rhizosphere isolates: *Pseudomonas frederiksbergensis* TBS10 and *Pseudomonas facilor* TBS28. Additionally, we tested a model purple photosynthetic bacteria used to study nitrogen fixation, hydrogen production, and other phenomena (*Rhodopseudomonas palustris* CGA009), a model actinomycete that is both studied for bioremediation and is related to *Streptomyces* that provide drought resistance to plants and produce antibiotics (*Rhodococcus jostii*), and *Psuedomonas putida*, perhaps the most used bioproduction chassis for active BER biodesign research.

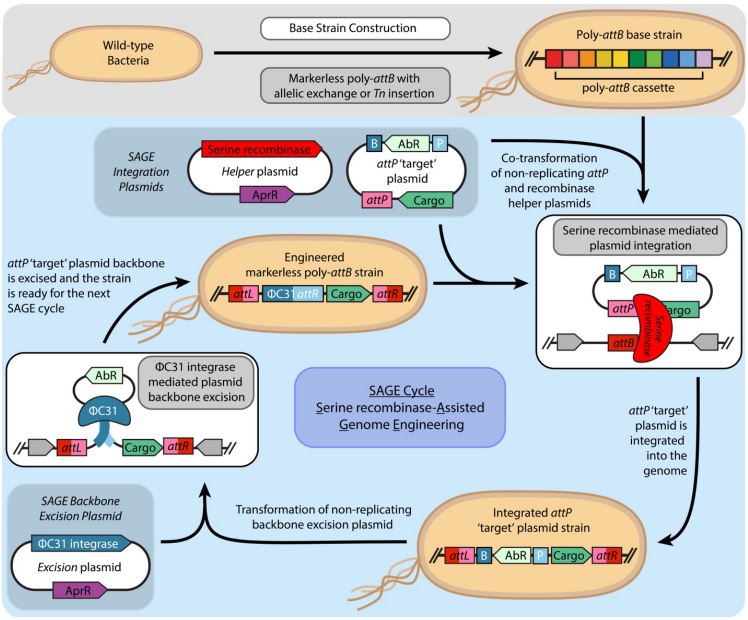

Figure 2. Iterative genome integration of genetic constructs using Serine recombinase-Assisted Genome Engineering (SAGE). The SAGE cycle enables rapid, efficient, and site-specific integration of multiple DNA fragments into bacterial chromosomes. A base strain with *attB* sequences is created using standard methods. The *attB* sequences are landing pads for the integration of non-replicating plasmids through transformation and expression of a matching recombinase. Unnecessary components of the integrated plasmid are excised with a specific recombinase, enabling multiple cycles of SAGE integration.

We prepared each bacteria for SAGE by integrating a 'landing pad', also known as a poly-*attB* attachment site cassette into the strain's chromosome (**Fig. 2** – top right). The landing pad contains ten unique attachment sequences, called *attB* sites, that serve as sites for insertion of DNA molecules by one of ten phage-derived recombinase enzymes. In addition to the *attB* sites, the landing pad includes sequences to screen for insertion of DNA and sequences that block the integrated DNA or neighboring native DNA elements from interference in gene expression. We used one of two different methods to integrate the cassette into each organism. The first,

homologous recombination, uses native host DNA repair machinery to integrate the landing pad into a specific region of the host chromosome. This method requires a known genome sequence – which is not always available – and is not effective in all organisms. So, we used another method, that uses a phage transposase rather than host machinery to integrate the DNA into the chromosome. This method has the advantage of working in organisms lacking a genome sequence. However, it has the disadvantage of integrating the landing pad into a random location in the chromosome, and insertion at some locations can disrupt important host functions. We used the transposase method to insert landing pads into *P. putida* and *P. frederiksbergensis* and homologous recombination in the other organisms.

Another challenge when using transposases for landing pad insertion is that an antibiotic selection marker cassette is co-integrated and must be removed before the antibiotic can be used for SAGE. We addressed this by flanking the antibiotic selection marker with the attachment site sequences recognized by the recombinase enzyme from ϕC31 phage. When ϕC31 recombinase is expressed in these strains, it excises the selection marker from the chromosome, leaving behind a small 'scar' sequence that is no longer recognized by the enzyme. Using a method called electroporation, which uses current to transiently open pores in the walls of bacterial cells, we introduced a DNA plasmid containing a gene for expression of the ϕC31 recombinase into each strain. This plasmid DNA does not replicate in the cell, producing transient recombinase expression. Any cells that take up the plasmid delete the DNA between the ϕC31attachment sites and become sensitive to the antibiotic. Following electroporation of this plasmid, ~1% and ~10% of *P. putida* and *P. frederiksbergensis* cells became kanamycin sensitive, ready for SAGE integration.

Evaluation of recombinase-assisted genome integration efficiency and accuracy

To assess the universality of SAGE, we transformed SAGE plasmids encoding each recombinase in all five bacterial hosts. We used electroporation to simultaneously introduce two plasmids into cells – one non-replicating *helper* plasmid that expresses a SAGE recombinase and the non-replicating *cargo* plasmid pGW60 (**Fig. 3C**), which contains a gene encoding a fluorescent protein, a kanamycin resistance gene, and an array of *attP* sequences, each matched to one of the ten recombinases. Each recombinase recombines DNA between a unique pair of *attB* and *attP* sequences. Transformants, or cells modified to be resistant to the antibiotic by integration of recombinant DNA, are expected to result from recombinase-catalyzed recombination between the *attB* site in the chromosome and the matching *attP* site in the pGW60 plasmid (**Fig. 3**). This process integrates the entire non-replicating plasmid into the chromosome while simultaneously converting the *attB* and *attP* sites into two new sites called *attL* and *attR*. The recombinase cannot perform recombination between the newly generated *attL* and *attR* sequences, and thus DNA integration via SAGE recombinases is irreversible. Notably, unlike many of the most used genome engineering systems in bacteria, the genome-edited strains resulting from a SAGE transformation are free of replicating plasmids. This means that no additional labor is required to remove other recombinant DNA before the strain is ready for use.

Genome integration efficiency was measured by counting colonies resistant to kanamycin after transformation (**Fig. 3D**). While recombinase efficiency varied between organisms, several recombinases efficiently incorporated the pGW60 plasmid into the chromosomes of all tested organisms. In general, the Bxb1, TG1, R4, or BL3 recombinases were the most efficient for

chromosomal integration of pGW60. Notably, we observed no kanamycin resistant colony formation in the absence of recombinase expression.

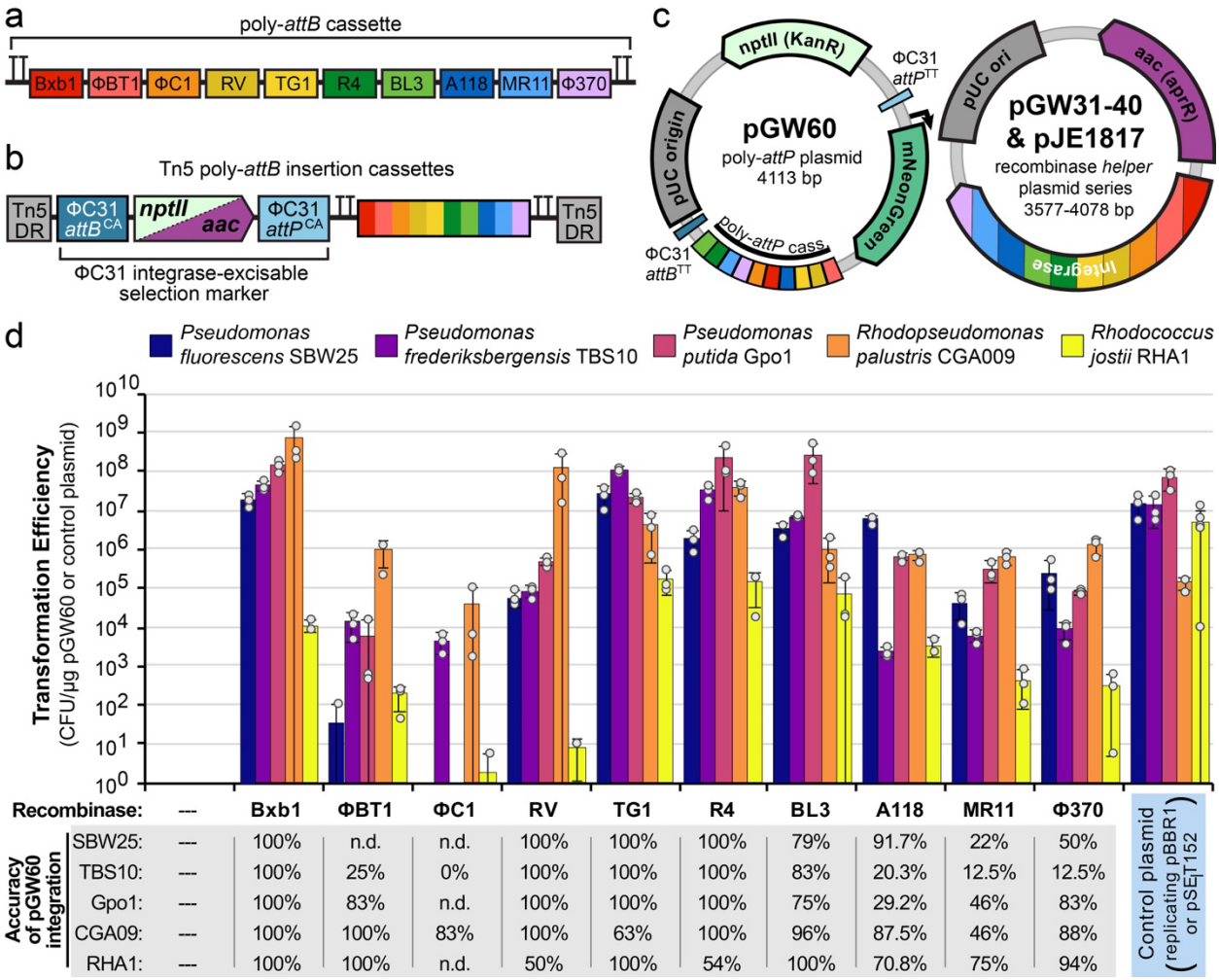

Fig. 3. SAGE enables stable, highly efficient integration of plasmid DNA into the genomes of engineered bacteria. (a) Diagram of genome-integrated 10x poly-attB cassette, including terminators for transcriptional insulation. Each attB sequence is indicated by a color-coded box and is flanked by a random 20 nt DNA spacer sequence. (b) Diagram of Tn5 poly-attB insertion cassettes. Cassettes are flanked by Tn5 direct repeat (DR) sequences and contain an antibiotic resistance cassette (either nptII or aac) upstream of the poly-attB cassette from (a). A cognate pair of ΦC31 att sites flank the resistance cassette, allowing its unidirectional excision by electroporation of the ΦC31 integrase helper plasmid pJE1817. (c) Plasmid maps of SAGE plasmids used for efficiency experiments in panel d. (d) Transformation efficiency when the poly-attP target plasmid pGW60 is transformed with or without an integrase-expressing helper plasmid, or when the positive control plasmid is transformed. Control plasmids are: pJE354 (SBW25, TBS10, Gpo1), pEYF2K (CGA009), or pSET152 (RHA1). Error bars indicate the two-sided standard deviation in three or more biological replicates. Dots indicate individual samples. The accuracy of integration represents the fraction of colonies in which pGW60 recombined into the poly-attB cassette rather than a pseudo-att site, as determined by PCR. Except for BT1 in Gpo1, and RV / Φ370 in RHA1 (which used 12, 20, and 18 samples for screening, respectively) integration accuracy represents the fraction of 24 samples with colony PCR screening results indicating insertion at the intended attB site. n.d. indicates samples not assayed by PCR.

Serine recombinases typically have a strong preference for their native *att* sites, but they are known to target other sites in a host chromosome called pseudo-*att* sites. Integration accuracy is therefore an important metric when assessing the utility of each recombinase as a genetic tool. We screened for integration of pGW60 at the targeted *att* site when using each of the recombinases. Generally, the recombinases that were most efficient at chromosomal integration in each organism also the most accurate. For many recombinases, such as Bxb1, the accuracy is at or near 100%, meaning that screening for accuracy is often not required. Ultimately, even low accuracy (20-30%) does not preclude the use of a given recombinase for many applications. It simply means that screening must be performed to ensure a given colony has accurate integration – a step that is typically required for almost all other competing technologies.

Antibiotic marker recycling for responsible and iterative genome engineering

The removal of antibiotic resistance markers is critical to enabling multiple cycles of SAGE integration (**Fig. 2**) and for strain deployment in operating environments where the spread of antibiotic resistance between organisms is relatively common (e.g. soil, human gut). Accordingly, all SAGE plasmids that integrate recombinant cargo into the chromosome (known as *attP* target plasmids) are designed to allow simple excision of the resistance gene and other extraneous 'non-cargo' DNA from the chromosome (**Fig. 4**). For excision, each plasmid contains a cognate pair of ϕC31 *attP* and *attB* sites that flank the non-cargo region. Transient expression of the ϕC31

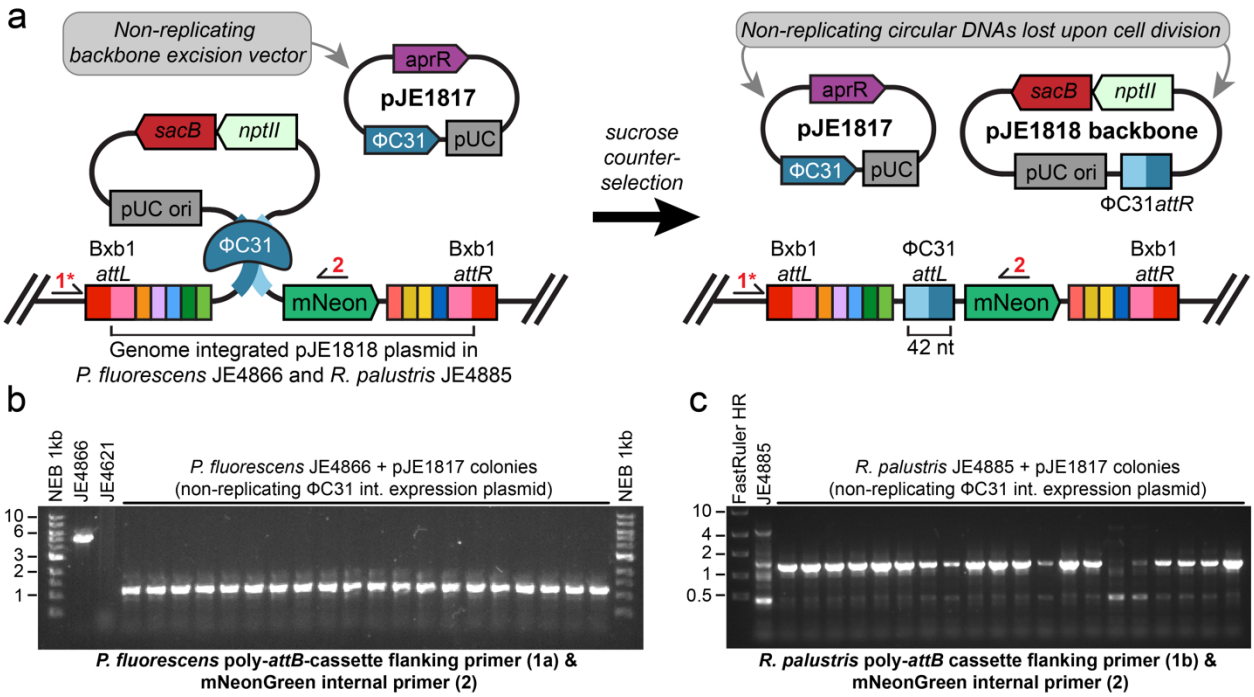

Fig 4. Transient expression of ϕC31 integrase excises the attP plasmid backbone and enables selection marker recycling. (a) Diagram of ΦC31 integrase-mediated excision of an attP target plasmid backbone. Specifically, pJE1818 was inserted at the Bxb1 attB of a chromosomally integrated poly-attB cassette. ΦC31 integrase was expressed from non-replicating helper plasmid pJE1817. Colony PCR primers are indicated by arrows and red numbers. (b-c) Colony PCR validation of plasmid backbone excision in sucrose-resistant (b) P. fluorescens JE4621 and (c) R. palustris JE4632-based strains following SAGE integration and incubation on sucrose-containing medium. Expected band sizes are as follows: JE4866 (4741 bp), JE4866 with backbone excision (1152 bp), JE4621 (no band), JE4885 (4964 bp), JE4855 with backbone excision (1375 bp).

recombinase in these strains (as described above) leads to excision of the DNA that encodes the antibiotic resistance gene through recombination between the *att* sites. To simplify screening for strains with successful excision, we include a counterselection cassette on SAGE plasmids that contains a gene for the levansucrase *sacB*, which confers a sensitivity to sucrose. By plating excision transformants on agar growth media containing sucrose, the resulting sucrose-resistant colonies are almost exclusively successful excision strains. This counterselection process is broadly applicable to different strains, though some bacteria when expressing the *sacB* gene are known to insensitive to sucrose. SAGE is compatible with many other counterselection methods.

High-throughput regulatory element screening for reliable gene expression

Development of genetic elements (e.g., promoters, ribosomal binding sites, terminators) that control expression of synthetic DNA programs genes is a critical aspect of engineering functions in bacteria. To minimize the burden of synthetic gene expression on host health and maximize stability of engineered functions, gene expression from engineered functions must be carefully balanced with core cellular functions. Additionally, bacteria are typically deployed in dynamic environments – and thus it is critical to identify transcriptional promoters that provide consistent gene expression across many conditions. We therefore developed and validated a methodology to assess large libraries of genome-integrated promoters across environmental conditions in SAGE-capable hosts.

Promoter libraries have been developed to control transcriptional activity using reporter protein expression in several organisms. However, most promoter libraries have been developed and characterized using multicopy reporter plasmids, which are likely not representative of the transcription rate from the chromosome. Heterogeneity in copy number among colonies and within subpopulations, plasmid loss within subpopulations, as well as overall higher DNA dosage from plasmids can influence promoter performance. Thus, measurement of gene expression from genome-integrated promoters is likely to be more predictive of activity in deployed environments than expression measurements from replicating plasmids.

To characterize a library of chromosomally integrated promoters, we used a methodology that combines the efficiency of SAGE integration with a modified implementation of an existing high-throughput sequencing transcriptional activity method. We demonstrated this approach to characterize a collection of 287 synthetic and natural promoters in the organisms we used to demonstrate SAGE. In this method, nucleic acid barcode sequencing from both RNA and DNA was used to normalize and quantify both transcript and promoter abundance. In brief, hundreds of promoter elements, each with a unique barcode are introduced in the chromosomes of the host with SAGE. Pools of strains, where each cell contains its own distinct barcoded promoter, are grown together and evaluated at once in a single experiment. After growth in conditions of interest, both the DNA and RNA from each cell is isolated. The barcodes in the RNA and DNA fractions are sequenced, and the ratio of RNA barcodes to DNA barcodes is indicative of the amount of transcription promoted by its corresponding genetic element.

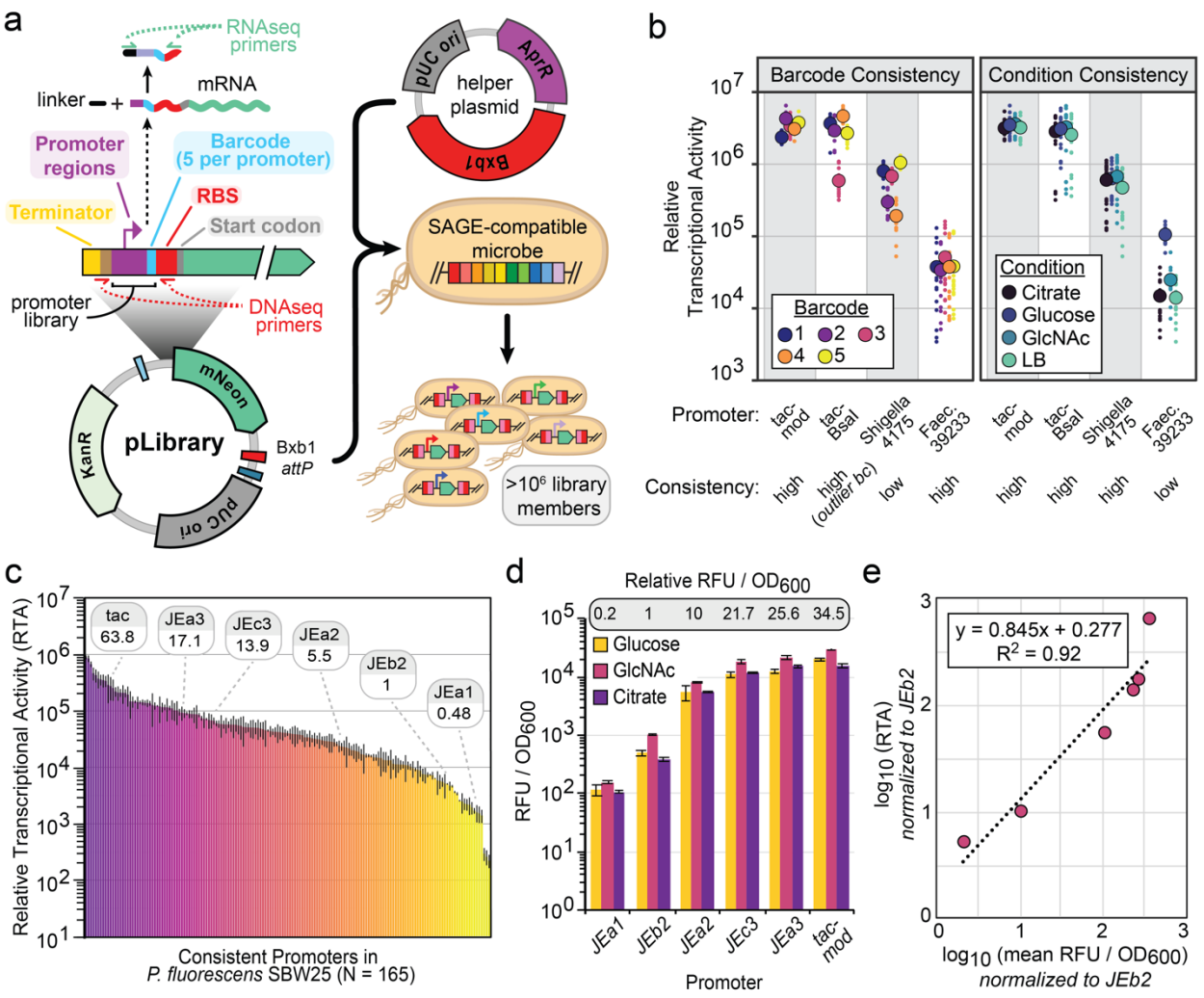

Fig 5. Development and high-throughput analysis of genome-integrated promoter libraries. (a) Overview of promoter library construction and DNA/RNAseq barcode sequencing fragments. (b) Example promoters with different classes of 5' UTR (level of barcode-associated 'noise') and condition sensitivity. Large, outlined circles represent mean values, and small dots represent individual samples. (c) Chart displaying mean relative transcriptional activity for consistent (5'UTR and condition insensitive) promoters in Pseudomonas fluorescens SBW25. Relative strength of the promoters used in panel d is indicated. Error bars represent two-sided standard deviation of between 36-80 samples (see source data file and File D1 for exact numbers). (d) Promoter performance for a small subset of pLibrary promoters in microtiter plate growth assays. Error bars represent two-sided standard deviations in 3 replicates. Relative promoter activity is calculated by comparing mean RFU/OD600 values across all carbon sources. (e) Correlation between relative expression levels determined by RTP and fluorescent protein reporter assay for the set of promoters used in panel d. Linear equation and coefficient of determination between the same promoters using RTA data from (c) and fluorescent plate reader data from (d), as determined by Pearson correlation.

We performed experiments with this method under multiple conditions that are relevant for each of the organisms. For example, we cultivated *P. fluorescens* in media supplemented with one of

three soil-relevant carbon sources: glucose (breakdown product of cellulose), citric acid (compound found in plant root exudate), or N-acetylglucosamine (breakdown product of chitin). The promoters we evaluated provided gene expression ranging four orders of magnitude in each organism. Furthermore, we found that most promoters were insensitive to environmental condition or genetic context.

We validated results from the barcode sequencing method with conventional methods where transcriptional activity is determined by measuring the production of a fluorescent reporter protein. Expression levels measured by each method strongly correlated, suggesting that the high-throughput barcode sequencing provides results that are comparable with fluorescent protein-based methods. While similar trends were observed between methods, barcode sequencing offers several advantages over fluorescent and colorimetric reporter assays: (1) sensitivity is only limited by sequencing depth rather than by physical limitations (e.g., inherent cell fluorescence), (2) thousands of promoters can be tested in a single experiment, (3) measurements are not limited by cell density, and (4) the assay works under anaerobic conditions – thus enabling promoter library development for biotechnologically- or health-relevant organisms (e.g., Clostridia) where reporter systems have been limiting.

References

1. Elmore, J.R., et al., *High-throughput genetic engineering of nonmodel and undomesticated bacteria via iterative site-specific genome integration.* Sci Adv, 2023. **9**(10): p. eade1285.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov