

UC Berkeley

UC Berkeley Previously Published Works

Title

Origin of replication discovery for environmentally isolated Pantoea strain enables expression of heterologous proteins, pathways and products

Permalink

<https://escholarship.org/uc/item/7f8558ms>

Journal

iScience, 29(3)

ISSN

2589-0042

Authors

Codik, Alex

Kothari, Ankita

Liu, Hualan

et al.

Publication Date

2026-03-01

DOI

10.1016/j.isci.2026.115031

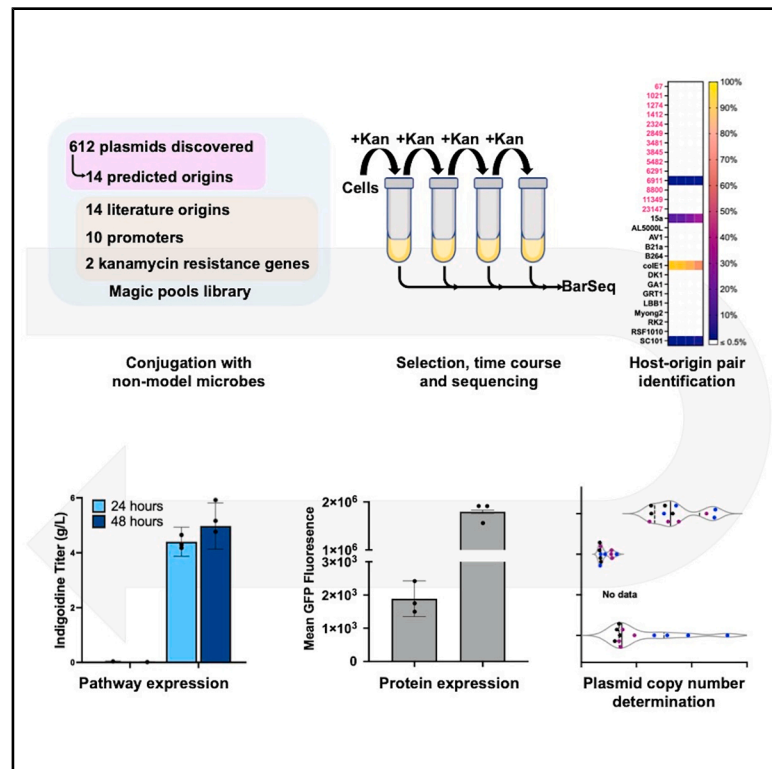
Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Origin of replication discovery for environmentally isolated *Pantoea* strain enables expression of heterologous proteins, pathways and products

Graphical abstract



Authors

Alex Codik, Ankita Kothari, Hualan Liu, ..., Adam P. Arkin, Adam M. Deutschbauer, Aindrila Mukhopadhyay

Correspondence

amukhopadhyay@lbl.gov

In brief

Molecular biology; Molecular genetics; Molecular mechanism of gene regulation; Molecular microbiology

Highlights

- Plasmid library constructed with known and predicted bacterial origins of replication
- Discovery of 3 previously unknown origins that replicate in gram-negative bacteria
- Origin validation in *Pantoea* sp. MT58 through copy number determination
- Bioproduction of GFP, indigoidine, and isoprenol in *Pantoea* sp. MT58



Article

Origin of replication discovery for environmentally isolated *Pantoea* strain enables expression of heterologous proteins, pathways and products

Alex Codik,^{1,2} Ankita Kothari,¹ Hualan Liu,³ Benjamin L. Weinberg,⁴ Trenton K. Owens,⁴ Aparajitha Srinivasan,¹ Alex Rivier,¹ Thomas Eng,¹ Adam P. Arkin,^{2,4,6} Adam M. Deutschbauer,^{4,5} and Aindrila Mukhopadhyay^{1,2,3,7,*}

¹Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

²Comparative Biochemistry Graduate Group, University of California, Berkeley, Berkeley, CA 94720, USA

³US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

⁴Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

⁵Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA

⁶Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA

⁷Lead contact

*Correspondence: amukhopadhyay@lbl.gov

<https://doi.org/10.1016/j.isci.2026.115031>

SUMMARY

Leveraging predicted origin sequences from a previously characterized groundwater plasmidome, we constructed a barcoded plasmid library to screen for previously unknown origins. Testing this library against a panel of representative bacterial strains led to the identification of 3 previously unknown origins that replicate in gram-negative bacteria not previously associated with these origin sequences. Experimental validation confirmed that a plasmid bearing origin 6911 as the sole origin could replicate with a copy number of 9 (± 2) in *Pantoea* sp. MT58, a fast growing and metal tolerant, environmentally important bacterium. Plasmids based on this new origin were used to express the reporter protein GFP, and non-native metabolite pathways for the natural product indigoidine and the terpenoid compound isoprenol. Functional previously unknown origins of replication in such non-model organisms can expand the toolkit for genetic manipulations of both model and less-studied bacteria.

INTRODUCTION

Genetic transformation is a critical step to transition from a genomics-based examination of a bacterial strain to genetics. Transformation of a plasmid capable of replicating in a bacterial host enables an array of genetic approaches that are indispensable in characterizing or engineering microbial strains. In spite of this, it can be a challenge to find and/or construct a plasmid that is able to self-replicate in a microbe of interest.^{1–5} Specifically, the plasmid's origin of replication is a key requirement but predicting origins that will function in a given host is challenging.⁶ However, discovery of new origins that function in diverse, non-model microbes has not been extensively explored.

Recent advances have allowed for the identification of many features of plasmids that are valuable in understanding the functional role of a plasmid as a mobile genetic element in the environment,^{7–12} and to reveal biological parts that can be used in molecular biology approaches.^{11–15} However, identification of the sequence that encodes the origin of replication is challenging and design of origins *de novo* remains elusive. One path forward is the use of existing plasmid sequences that can be analyzed to predict putative origin sequences and can then be experimentally verified. A plasmidome is the total collection of bulk

plasmids present in a given environment that can be measured by whole genome sequencing (WGS).¹⁶ Targeted plasmidome studies, as well as metagenomics studies, reveal a vast range of native plasmids that provide a deep resource to address this challenge.^{17–20}

In this report, we leveraged a previously characterized plasmidome from groundwater samples from the Oak Ridge field research center (ORFRC) containing over 600 unique plasmid sequences.¹⁷ This prior plasmidome dataset revealed key aspects of both the nature and the function of these mobile genetic elements, which included both plasmid¹⁷ and viral sequences.²¹ Earlier studies focused on key features of the native plasmid scaffolds such as the range of mobilization mechanisms and incompatibility groups that are spanned by these plasmids, and the functions annotated on these sequences. One study fully synthesized a native plasmid and tested functions and host ranges experimentally.²² One aspect that remained unexplored was use of these putative plasmid scaffolds for parts discovery. Specifically, we aimed to obtain the origin of replication sequences that could enable us to build genetic tools for new and existing microbial systems. Thus in this study we selected a subset of the groundwater plasmidome sequences from the published plasmidome data and analyzed them for putative predicted origins of replication.



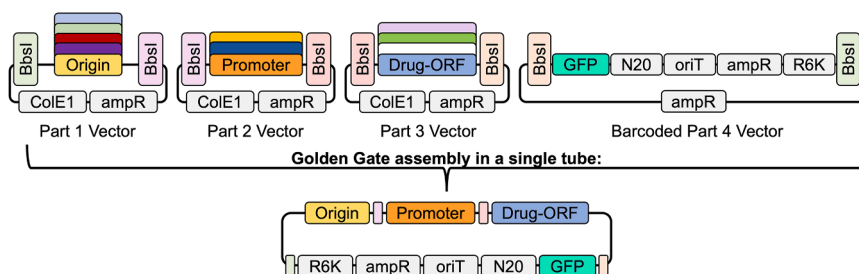


Figure 1. Magic pool construction method
Each colored BstI box represents a unique sticky end cut site that will complement only the BstI boxes with the same color to link plasmid parts together sequentially. N20 is representative of the 20 base pair barcodes.

A primary goal of our study was to establish a workflow that would experimentally validate predicted origin of replication sequences, enable identification of new host-plasmid pairings and advance the usable genetic tools in new environmental and biotechnologically important bacterial strains. To do this, we adopted a magic pool approach to build a DNA barcoded plasmid library containing the predicted origin sequences.²³ A magic pool is a DNA-barcoded plasmid library with different genetic parts, such as promoters and drug resistance genes, which can be used to quickly find a suitable combination of parts for a target bacterium.²³ We experimentally screened the plasmid pool containing a combination of putative and known origins against a well-studied model bacteria (e.g., *E. coli* BW25113) and less-studied environmental strains (e.g., *Pantoea* sp. MT58). The *Pantoea* genus is especially important for further studies; it is a widely distributed genus found both in plant and soil associated communities spanning pathogenic^{24–27} and commensal^{25,28} species. *Pantoea* sp. are also seen as a valuable new platform for biotechnology applications.^{14,25,29–31} In both of these contexts deeper genetic tools and an available model-system, such as the Gammaproteobacteria *Pantoea* sp. MT58 could be valuable. In the ORFRC, *Pantoea* sp. MT58 is a member of the sediment microbial community that is gaining interest due to its metal tolerance capabilities,³² and a recent pangenome study for this bacterium highlighted its facultative anaerobic capabilities being involved in survival in the contaminated groundwater at the ORFRC.³³ In our study we sought to discover additional genetic parts for use in this emerging microbe of interest and demonstrate the use of new plasmids for expression of proteins and pathways.

RESULTS

Environmental plasmids used to predict origins of replication

While there are robust resources for plasmid discovery and analysis, prediction of plasmid origins of replication sequences has fewer tools.^{7,34,35} DoriC, originally developed as a database for *oriC* prediction of bacterial and archaeal chromosomes, was updated with plasmid sequences in DoriC 10.0.³⁶ Thirty two sub-selected circular plasmids from our earlier plasmidome study¹⁷ were assessed using this tool and led to 18 predicted plasmid-based origins of replication that could then be experimentally tested. These 18 origin of replication sequences were synthesized as double strand DNA fragments (see STAR Methods) on a backbone compatible with the magic pool library construction

method (Figure 1). We also examined the 32 native plasmid scaffolds in a recently published plasmid *oriV* predictor, *OriV-finder*.³⁷ The majority of the origins of replication in the initially predicted set were also predicted via this tool, along with additional *oriV* sequences (Data S1). Notably, while naturally occurring plasmids may have multiple origins of replication or replication machinery, these workflows predicted at most one origin for each plasmid.

Origins magic pool library constructed to test origin-host pairings

A magic pool library was constructed using 4 different parts, as outlined in Figure 1. The magic pool was constructed using 38 variants of origins of replication (part 1), 10 variants of kanamycin drug resistance gene promoters (including ribosomal binding site [RBS]) (part 2), and 2 variants of kanamycin drug resistance genes (part 3), per the original design of the magic pools plasmid scaffold.²³ Variants of parts 2 and 3 were chosen to ensure effective kanamycin selection in the target microbe. The construction of this library was performed using golden gate assembly and was sequenced using long-read sequencing. In the mapping of the barcode-part association of the magic pool it was found that the *colE1*-associated plasmids were disproportionately abundant, likely due the high copy number of this origin in *E. coli*.

In the library mentioned above, part 1 is the variable part; the origin sequence and the key component being assessed, represented by one of 18 putative origins or one of 20 origins mined from literature (Data S1). Parts 2 and 3 came from a previous study.²³ Part 4 was the barcoded backbone for all plasmids constructed in the magic pools and contains a conditional R6K origin of replication for plasmid maintenance in *E. coli*, an origin of transfer, and an antibiotic selection marker (carbenicillin). The DNA barcodes are random 20 nucleotide sequences flanked by common PCR priming sites. The barcodes are introduced into the part 4 plasmid backbone via golden gate assembly from a PCR product containing the random 20 mers, as previously described in Liu et al.²³ The sequences of parts 1, 2, and 3 can be found in Data S1.

The magic pool backbone (part 4) has a barcode which can establish a plasmid-barcode association. These associations consist of a single barcode and a combination of parts (a single origin, a single promoter, and a single kanamycin drug resistance gene). Barcodes that linked to multiple plasmids or combinations of parts were eliminated from consideration. This resulted in 8,802 total barcoded plasmids that mapped a unique barcode to a part 1 origin (some of these barcodes could not be

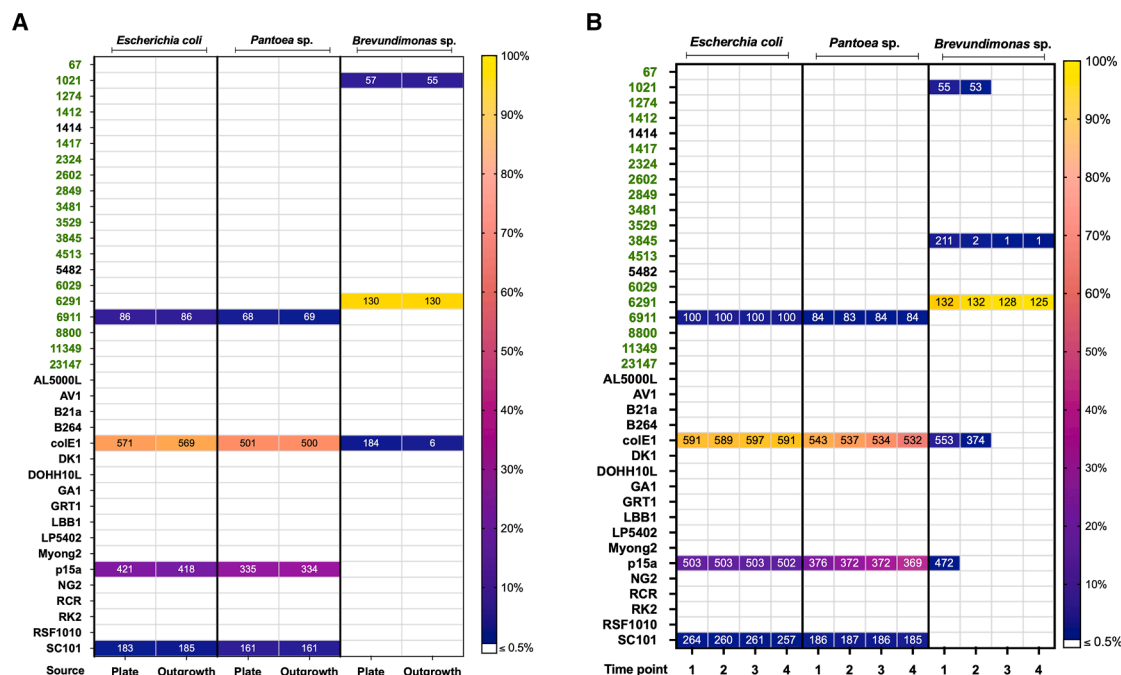


Figure 2. Magic pool BarSeq results for all origin sequences tested

The y axis denotes the names of the origins of replication, with predicted origins in green, and origins found in literature in black. Colors in each time point, or source, square represent the fraction of barcode counts associated with that origin versus the total fraction of barcodes in that sample (see legend). The number present in each box represents the total number of unique barcodes detected for that particular origin of replication at that time point, or for that source.

(A) Plate-based sources for the magic pool conjugations with the 3 hosts. “Plate” is indicative of cells harvested directly from an agar plate, while “Outgrowth” is indicative of an overnight of growth in liquid media after scraping colonies off of an agar plate.

(B) Liquid-based time points for magic pool conjugations with the 3 hosts. Time points 1, 2, 3 and 4 represent 24, 48, 72, and 96 h respectively. Note that the recipient strain *E. coli* BW25113 is part of the recipient set along with *Pantoea* MT58 and *Brevundimonas* sp., and is different from the conjugation strain *E. coli* WM3064, which was used as the donor. Note also that the 3845 hit for *Brevundimonas* sp. was analyzed due to a single bar code enrichment and represents an artifact rather than a real candidate.

confidently assigned a part 2 or part 3). The number of barcodes associated with each origin in the magic pool can be seen in Figure S1.

Barcode sequencing reveals host-origin pairings

To evaluate the compatibility of the various plasmid origins across different bacterial species, the magic pool library was introduced into three distinct hosts: *Escherichia coli* BW25113,³⁸ *Pantoea* sp. MT58,³² and *Brevundimonas* sp. GW460-12-10-14-LB2²³ (Data S2: strain and plasmid table). We then used barcode sequencing of the resulting bacterial populations to quantify the relative enrichment of each plasmid variant, which allowed us to infer the origin compatibility in each host. To determine whether different selection strategies influenced the plasmid populations recovered, we applied two complementary approaches: plate-based selection and serial liquid outgrowth. In both workflows, comparison of barcode abundances against the map of the pooled library enabled us to determine which plasmid variants were retained under each selection condition (Figure 2). Similarly, we analyzed the data for high-confidence part 2 and part 3 sequences that worked effectively for driving kanamycin resistance in each strain (see Figures S2–S4 for part 2 and 3 variants breakdown in all strains).

Across both selection methods *E. coli* BW25113 and *Pantoea* sp. MT58 showed enrichment of four origins (Figure 2A). Three of these were established origins namely p15a, colE1, and SC101 all of which were previously shown to enable plasmid replication in *E. coli*.^{39–41} The previously uncharacterized 6911 origin showed enrichment in both *E. coli* BW25113 and *Pantoea* sp. MT58.

In *Brevundimonas* sp. GW460-12-10-14-LB2 two previously unknown origins—1021 and 6291 were enriched in plate-based selection. We detected barcode counts from colE1-containing plasmids in this strain, but we do not consider this a functioning origin by our analysis, because only a fraction of the colE1-associated barcodes was detected.

On the other hand, during serial liquid passaging, plasmids carrying the 6291 origin outcompeted all others and dominated the *Brevundimonas* population by the final time point. This outcome is consistent with selective dynamics in liquid culture, where origins with higher copy numbers can lead to stronger kanamycin resistance and thus confer a fitness advantage. The disproportionate enrichment of 6291 likely masked the 1021 origin signal, even though 1021 was functional in the plate-based method. Barcode counts attributed to origins 3845, colE1, and p15a were low and declined over time, consistent with carryover

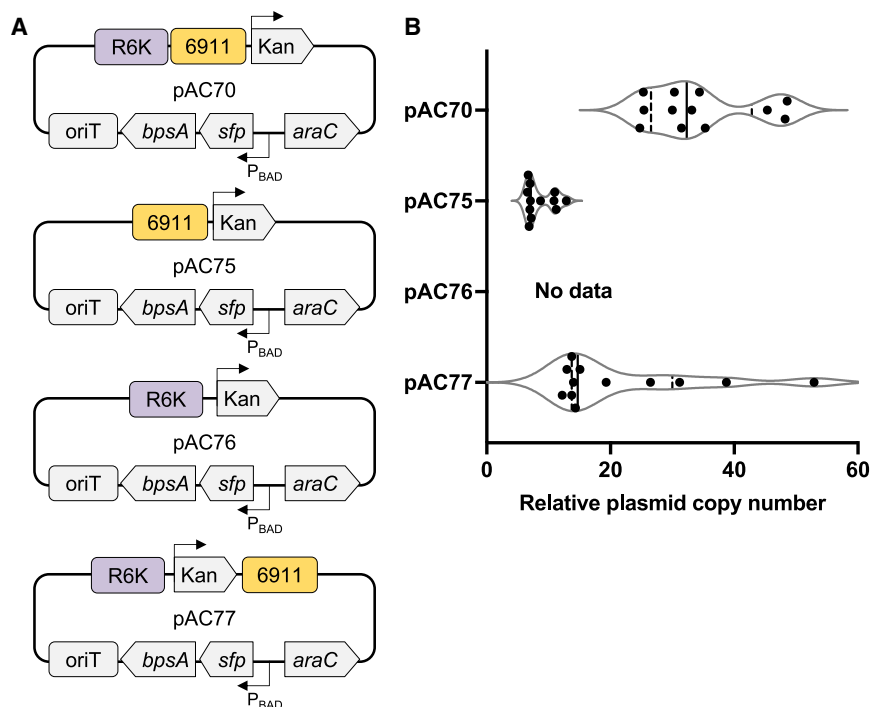


Figure 3. Validation of previously unknown origin 6911 in *Pantoea* sp. MT58

(A) Plasmid maps that correspond to the plasmids being tested in (B).

(B) Violin plot of plasmid copy number. Data for each plasmid is from 3 biological replicates at 4 dilutions. A solid vertical line represents the data median. Vertical dashed lines represent quartiles.

6911 (Figure 3A, plasmid pAC70). This plasmid contains the R6K origin proximal to the origin 6911. The second plasmid constructed contains only the origin 6911 as a replication mechanism (Figure 3A, plasmid pAC75). The third plasmid contains only the R6K origin as a replication mechanism (Figure 3A, plasmid pAC76). The final plasmid was constructed with both origins, R6K and 6911, but has placed these origins distal from each other, to see if R6K affects origin 6911 even when not proximal (Figure 3A, plasmid pAC77).

All four plasmids were individually conjugated into *Pantoea* sp. MT58 and successful conjugations were verified by

from the *E. coli* donor strain. In support of this, these are all high-abundance parts in the magic pool (Figures S1 and S4).

Together, these results demonstrate that the pooled library screening reliably identifies plasmid origins capable of replication in diverse bacteria and highlight host-specific differences in origin compatibility and competitive dynamics. Of note, this approach uncovered previously uncharacterized origin 6911 as broadly functional in both *E. coli* BW25115 and *Pantoea* sp. MT58.

Presence of two origins of replication changes the plasmid copy number in *Pantoea* sp. MT58

Many magic pool generated plasmids containing the previously unknown origin 6911 showed evidence of driving plasmid replication in *Pantoea* sp. MT58 and *E. coli* BW25113, based on the BarSeq results (Figure 2). We further investigated this origin of replication and its copy number in the environmentally isolated strain *Pantoea* sp. MT58.

The backbone of the magic pool contains an R6K origin for cloning and for maintaining all of the plasmids in *E. coli* pir⁺ strains. The existence of this origin of replication in the backbone of a magic pool used to test origins of replication requires the decoupling of the R6K from the origin being tested in the host-origin pair. To decouple the R6K present in the magic pool backbone and the origin 6911, and to test the use of this origin, four plasmids were constructed (Figure 3A). These four plasmids were constructed to contain the parts from the magic pool that were necessary (6911 origin and/or conditional R6K origin, kanamycin promoter and resistance gene, and oriT) but also contained a small two-gene pathway to more accurately represent a real world use and application of this plasmid.

The first plasmid was identical to the plasmid found in the library in regards to the conditional R6K origin and the origin

the appearance of kanamycin resistant colonies whereas the control sample did not exhibit spontaneous kanamycin resistance. All plasmids, except for pAC76, were successfully conjugated into *Pantoea* sp. MT58. This indicates that R6K alone was not sufficient for plasmid replication but origin 6911 alone was sufficient for plasmid replication in *Pantoea* sp. MT58.

Transformants were then examined using colony-based quantitative PCR (qPCR) to obtain plasmid copy number estimates. qPCR was used to compare the genomic DNA amplicon relative to the plasmid amplicon to provide the number of plasmid copies per genome copy, a relative plasmid copy number per cell, for each plasmid. Rounded to the nearest whole number, pAC70 had a relative copy number of 34 (± 9), pAC75 had a relative copy number of 9 (± 2) and pAC77 had a relative copy number of 22 (± 13) (Figure 3B and Data S3). The method was also tested for a plasmid containing colE1, reported to have 25–30 copies per cell.⁴¹ Colony-qPCR determined that, in *E. coli* 10- β , the plasmid had a copy number of 38 (± 11) (Figure S5). While this experiment queried only origin 6911, the data suggests that having an origin of replication on a plasmid that also contains an R6K origin of replication may affect the plasmid copy number and the plasmid's replicative capacity in a host.

Expressing a GFP reporter using the previously unknown origin 6911 in *Pantoea* sp. MT58

The backbone of each plasmid in the magic pool encodes a green fluorescent protein (GFP), a commonly used reporter protein in synthetic biology workflows in multiple microbial systems.^{42,43} In our study, GFP can be used to test functional origins of replication as well as expression of a heterologous protein in a new host. The original plasmid for *Pantoea* sp. MT58 did not show any GFP expression. However, introduction of an inducible

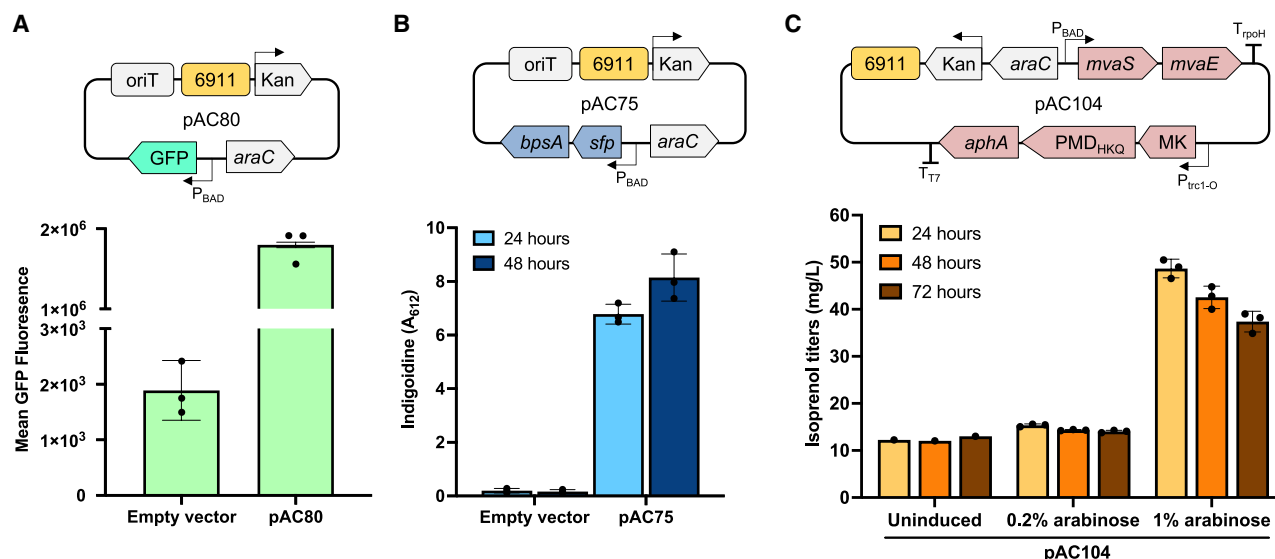


Figure 4. Applications of origin 6911 in *Pantoea* sp. MT58

(A) *Pantoea* sp. MT58 GFP expression (bottom panel) using plasmid containing only origin 6911 (top panel).

(B) Indigoidine absorbance measurements from expression in *Pantoea* sp. MT58 (bottom panel) using a plasmid containing only origin 6911 (top panel).

(C) Isoprenol production levels at different inducer levels (bottom panel) using a 6911-containing plasmid (top panel). Error bars in all three panels represent standard deviation from three biological replicates.

promoter, P_{BAD}, (Figure 4A), led to expression of GFP (Figure 4A). Representative flow cytometry samples can be seen in Figure S5, and data can be seen in Data S4.

Demonstrating the utility of the origin 6911 for plasmid-based heterologous pathway expression in *Pantoea* sp. MT58

The successful expression of a reporter gene, GFP, in *Pantoea* sp. MT58 using the origin 6911 containing plasmid pAC80 motivated us to further investigate the potential to express pathways for non-native and secondary metabolites. To this end, we transformed *Pantoea* sp. MT58 with arabinose inducible (P_{BAD}) multi-gene pathway plasmids, pAC75 and pAC104 in parallel experiments and produced a non-native non-ribosomal peptide product indigoidine, and the hemiterpene compound isoprenol, respectively (Figure S7).

In the engineered *Pantoea* strain heterologously expressing *sfp*_{B. subtilis}::*bpsA*_{S. lavendulae} on the plasmid pAC75, upon induction by 0.3% (w/v) arabinose and extraction of harvested cells with DMSO, we observed an average OD₆₁₂ of 4.4 at 24 h and 5.0 at 48 h (Figure 4B and Data S5). The OD₆₁₂ values are indirect measures of indigoidine titers and confirmed expression of the two-step indigoidine pathway containing the phosphopantetheinyl transferase (*sfp*) and the non-ribosomal peptide synthetase (NRPS, *bpsA*). Successful production of indigoidine represents an ability to express a non-native natural product in *Pantoea* sp. MT58 leveraging the plasmid-based expression system enabled by the origin 6911. Indigoidine is also a valuable bio-based synthetic indigo alternative that has been previously reported in other industrially relevant hosts.^{44–48}

We also validated the use of this previously unknown origin 6911 based plasmid to express products from more complex

multi-gene pathways. For this we used the heterologous isoprenoid pathway for the hemiterpene short chain alcohol isoprenol. A previously reported plasmid pLY670⁴⁹ was converted into pAC104 by swapping the origin RK2 (optimized for expression in *P. putida*) with the origin 6911. Terpenes span a wide array of compounds from those useful in altering the metabolic profile of a microbe to those that serve as bioproduction targets. In this case, isoprenol is also a useful commodity chemical and a well-established biofuel precursor.^{49,50} Short chain alcohols also serve important roles in fitness and interactions in microbial communities^{51–54} and an ability to express them provides additional tools to alter and examine microbial systems in context of their community. In our system, we observed a maximum isoprenol production of 48.67 mg/L in 24 h (Figure 4C and Data S6) in glucose minimal medium. A drop in isoprenol titers (to 37.38 mg/L) is observed by 72 h and consistent with published reports.⁴⁹ *Pantoea* sp. MT58 cannot utilize isoprenol as the sole carbon source (Figure S8A), but this strain encodes many alcohol dehydrogenases⁵⁵ that can degrade isoprenol. Hence, the decrease in titers at 72 h in the engineered *Pantoea* sp. MT58 could be attributed to either the inherent volatility of isoprenol and/or non-catabolic degradation.

Between the two non-native systems, inducer requirement was markedly higher to achieve metabolite production in the isoprenoid system (1% w/v arabinose) vs. that from the NRPS system for indigoidine (0.3% w/v arabinose) (Figures 4B and 4C). While indigoidine is produced from the amino acid precursor-glutamate, isoprenol is derived from acetyl-CoA which is a key central metabolite at the intersection of carbohydrate, lipid, and protein metabolism with higher competing pathway demand. Additionally, the heterologous isoprenol pathway is also cofactor-dependent⁴⁹ compared to the indigoidine pathway

which could impose a metabolic burden. The accumulation of pathway intermediates may impose additional toxicity even though *Pantoea* sp. MT58 is tolerant to up to 7 g/L of the final product-isoprenol (Figure S8B).^{56–58} Furthermore, *Pantoea* sp. MT58 consumes arabinose as a sole carbon source (Figure S9). These results establish this previously unknown origin 6911 as a promising genetic tool that improved the genetic tractability of *Pantoea* sp. MT58, and are an important first step in strain domestication to investigate the ecological role of this bacterium and to develop industrially relevant hosts. However additional optimizations are needed to use these systems to either modulate the levels of these metabolites to examine their role in community interactions, or to facilitate the development of chassis strains.

DISCUSSION

Discovering replicable plasmids is essential to moving from genomics-to genetics-based approaches. We leveraged a large plasmidome dataset from prior reports to specifically focus on plasmid origin sequences as these could be used to develop engineered replicating plasmids. It is interesting to consider that the native plasmid scaffolds used in the present study each have their own host range and plasmid replicating features. Our prior studies have explored the host range of these extra-chromosomal genetic materials (including plasmids and viral genomes), both via direct assay of microbial hosts with synthesized plasmids²² and predicting hosts for viral genomes.²¹ However a non-native replicating plasmid containing only the sequence for such novel origins of replication may have a different host range. While newer prediction tools such as the OriV-finder³⁷ corroborated the sequences selected for building the magic-pools library, the functional validation of both the hosts in which the resulting plasmid can operate, and its characteristics, need to be conducted experimentally. The OriV-finder tool predicted additional sequences relative to our original set, and new AI/ML tools in this domain⁵⁹ may soon provide a large number of putative origin sequences. This further necessitates an experimental workflow as we have developed to empirically validate efficacy in a range of microbial hosts.

Our workflow identified a previously unknown origin for the non-model *Pantoea* sp. MT58, which was confirmed through qPCR experiments by decoupling of the conditional R6K origin and the origin 6911. The variability in copy number seen in the two plasmids that contain both the R6K and the 6911 origin may be due to the competing/commensal origins, a change in plasmid partitioning or replication mechanisms. This is clearly different from the 6911-only plasmid's tight variability. Occasionally multi-origin plasmids are found in nature,⁶⁰ and studying these interactions may help elucidate the specific interaction between the R6K and 6911. The low copy number origin 6911 was then used to express three molecules, each with increasing complexity, in *Pantoea* sp. MT58. While the magic pools library was used to screen a small subset of hosts, both model and non-model, it can be expanded to include additional bacterial hosts for further discovery. The products described in this study are highly relevant to key metabolic pathways, and the production of chemicals and fuels. Now that we have shown basic ca-

pabilities to express genes and pathways to produce heterologous proteins and metabolites, additional optimization of promoter strength, metabolic burden, or host physiology can further enhance production and lead to the use of this microbe as a bioproduction chassis. Overall the development of genetic tools for this microbial host has value in its use as a biotechnology platform as well as in future studies for understanding its role in the environment.

Discovery of novel origins of replication add to the existing genetic toolkit for bacteria, like *Pantoea* sp. MT58, but for others it could provide an entirely new opening for genetics, like in *Brevundimonas* sp. GW460-12-10-14-LB2. In this study we focused on validating previously unknown origin based plasmids in *Pantoea* sp. MT58 due to its importance as a denitrifying microbe in ground water and sediments systems^{32,33,61} as well as a representative microbe from the biologically important *Pantoea* genus.^{24–28} In our study we could also characterize important features of the new plasmid such as copy number. However, other aspects, such as the compatibility group, of these new plasmids remain uncharacterized. It is important to note that many aspects of plasmid function are dependent on the host, presence of other plasmids or replicating units and the functional genes encoded on them. As more plasmids become available for these relatively new genetically tractable bacterial systems, characterization of such features will also be possible. Exploiting these new tools in environmental bacteria requires additional investigation and experimentation, but promises to provide a range of new tools and approaches in many new microbial systems.

Limitations of the study

While this study used DoriC to predict plasmid-based origins of replication, the tool itself is designed to work with genomes. However, recent development of plasmid databases and prediction tools has the potential to alleviate this issue. In our study we examined our native scaffold set using one of these recent tools, OriV-finder, and found useful additional origins of replication sequences that provide targets for future experimental validation.

The conditions under which a transformation occurs can greatly influence the success or efficiency, particularly antibiotic concentrations and media formulations. When transforming a pooled library of plasmids into bacterial strains, these conditions can cause false negative results. Without a significant effort to disprove these false negatives it becomes very difficult to distinguish a true negative and a false negative. As such, the positive results provide useful genetic tools, but the sequences with no transformation under the tested conditions could be tested in additional microbial hosts under alternate selection, recovery, and outgrowth conditions. However, if the goal is to discover plasmids or origins for a specific high-interest microbial target, many antibiotic conditions, and media formulations should be tested.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Aindrita Mukhopadhyay (amukhopadhyay@ibl.gov).

Materials availability

Materials will be made available upon request.

Data and code availability

Data: All sequencing data has been deposited in NCBI and is publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#).

Code: This paper does not report original code. All software and algorithms used for analysis are listed in the [key resources table](#).

Other: This paper does not report any additional resources.

ACKNOWLEDGMENTS

The authors thank the Mukhopadhyay group for their constructive feedback regarding the manuscript. Dr. Shweta Priya (LBNL) provided valuable guidance in conducting the qPCR experiments. We want to thank Dr. Guilherme de Siqueira (LBNL) for help in uploading sequencing data to NCBI. We thank the authors of DoriC and OriV-finder for help with assessing putative origins of replication. We also acknowledge the JBEI (jbei.org) strain collection for one of the strains used in this manuscript. This material is based upon work by the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) (<http://enigma.lbl.gov>) project, a Science Focus Area Program at Lawrence Berkeley National Laboratory (LBNL) supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research under Contract Number DE-AC02-05CH11231.

AUTHOR CONTRIBUTIONS

A.C., A.K., and A.M. developed the study. A.K., H.L., and A.M.D. designed the origins magic pool library. A.C. and to build the library. A.C. and B.L.W. tested the library. A.C. performed qPCR, GFP and indigoidine experiments. A.R. developed electroporation methods for *Pantoea* sp. MT58. A.C. and A.S. performed isoprenol experiments, with A.S. performing GC-FID data analysis. A.C. performed all other data analysis. T.E., A.M.D., and A.M. provided supervisory roles and feedback. A.P.A., A.M.D., and A.M. acquired the funds for the project. A.C. drafted the initial manuscript. All authors have read, provided feedback, and approved the manuscript for publication.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Bacterial strains and growth conditions
- [METHOD DETAILS](#)
 - Origin of replication prediction
 - Construction of the kanamycin magic pool origin libraries
 - Magic pool conjugation
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
 - Analyzing the plasmid fitness within each bacterial strain
 - Colony quantitative PCR
 - GFP quantification in *Pantoea* sp. MT58
 - Indigoidine production
 - Isoprenol production

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2026.115031>.

Received: June 17, 2025

Revised: October 28, 2025

Accepted: February 11, 2026

Published: February 17, 2026

REFERENCES

1. Gilbert, C., Brumwell, S.L., Crits-Christoph, A., Kang, S.C., Martín-Moldes, Z., Alsharif, W., Esmurria, A., Nguyen, M.-A., Lee, H.H., and Ostrov, N. (2023). A scalable framework for high-throughput identification of functional origins of replication in non-model bacteria. Preprint at bioRxiv. <https://doi.org/10.1101/2023.05.19.541510>.
2. Faktorová, D., Nisbet, R.E.R., Fernández Robledo, J.A., Casacuberta, E., Sudek, L., Allen, A.E., Ares, M., Aresté, C., Balestreri, C., Barbrook, A.C., et al. (2020). Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nat. Methods* 17, 481–494. <https://doi.org/10.1038/s41592-020-0796-x>.
3. Rouches, M.V., Xu, Y., Cortes, L.B.G., and Lambert, G. (2022). A plasmid system with tunable copy number. *Nat. Commun.* 13, 3908. <https://doi.org/10.1038/s41467-022-31422-0>.
4. Elston, K.M., Phillips, L.E., Leonard, S.P., Young, E., Holley, J.-A.C., Ahsanullah, T., McReynolds, B., Moran, N.A., and Barrick, J.E. (2023). The Pathfinder plasmid toolkit for genetically engineering newly isolated bacteria enables the study of *Drosophila*-colonizing Orbaceae. *ISME Commun.* 3, 49. <https://doi.org/10.1038/s43705-023-00255-3>.
5. Gilbert, C., Crits-Christoph, A., Ledieu-Dherbécourt, E., Kang, S.C., Brumwell, S.L., Lee, H.H., and Ostrov, N. (2024). Design and Construction towards a Pan-Microbial Toolkit. Preprint at BioRxiv. <https://doi.org/10.1101/2024.02.23.581749>.
6. Bouet, J.-Y., and Funnell, B.E. (2019). Plasmid localization and partition in enterobacteriaceae. *EcoSal Plus* 8. <https://doi.org/10.1128/ecosalplus.ESP-0003-2019>.
7. Camargo, A.P., Call, L., Roux, S., Nayfach, S., Huntemann, M., Palaniappan, K., Ratner, A., Chu, K., Mukherjee, S., Reddy, T.B.K., et al. (2024). IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. *Nucleic Acids Res.* 52, D164–D173. <https://doi.org/10.1093/nar/gkac964>.
8. Dong, M.-J., Luo, H., and Gao, F. (2023). DoriC 12.0: an updated database of replication origins in both complete and draft prokaryotic genomes. *Nucleic Acids Res.* 51, D117–D120. <https://doi.org/10.1093/nar/gkac964>.
9. Lipworth, S., Matlock, W., Shaw, L., Vihta, K.-D., Rodger, G., Chau, K., Barker, L., George, S., Kavanagh, J., Davies, T., et al. (2024). The plasmidome associated with Gram-negative bloodstream infections: A large-scale observational study using complete plasmid assemblies. *Nat. Commun.* 15, 1612. <https://doi.org/10.1038/s41467-024-45761-7>.
10. Androsiuk, L., Shay, T., and Tal, S. (2023). Characterization of the environmental plasmidome of the red sea. *Microbiol. Spectr.* 11, e0040023. <https://doi.org/10.1128/spectrum.00400-23>.
11. Shin, G.-Y., Asselin, J.A., Smith, A., Aegerter, B., Coutinho, T., Zhao, M., Dutta, B., Mazzone, J., Neupane, R., Gugino, B., et al. (2025). Plasmids encode and can mobilize onion pathogenicity in *Pantoea* agglomerans. *ISME J.* 19, wraf019. <https://doi.org/10.1093/ismejo/wraf019>.
12. Shetty, S., Kamble, A., and Singh, H. (2024). Insights into the Potential Role of Plasmids in the Versatility of the Genus *Pantoea*. *Mol. Biotechnol.* 66, 3398–3414. <https://doi.org/10.1007/s12033-023-00960-3>.
13. Lai, L., Xin, R., and Cui, T. (2024). Gene editing technology combined with response surface optimization to improve the synthesis ability of lycopene in *Pantoea dispersa* MSC14. *J. Appl. Microbiol.* 135, lxae272. <https://doi.org/10.1093/jambio/lxae272>.
14. Nitta, N., Tajima, Y., Yamamoto, Y., Moriya, M., Matsudaira, A., Hoshino, Y., Nishio, Y., and Usuda, Y. (2021). Fermentative production of enantiopure (S)-linalool using a metabolically engineered *Pantoea ananatis*. *Microb. Cell Fact.* 20, 54. <https://doi.org/10.1186/s12934-021-01543-0>.

15. Tao, L., Jackson, R.E., and Cheng, Q. (2005). Directed evolution of copy number of a broad host range plasmid for metabolic engineering. *Metab. Eng.* 7, 10–17. <https://doi.org/10.1016/j.ymben.2004.05.006>.
16. Stalder, T., Press, M.O., Sullivan, S., Liachko, I., and Top, E.M. (2019). Linking the resistome and plasmidome to the microbiome. *ISME J.* 13, 2437–2446. <https://doi.org/10.1038/s41396-019-0446-4>.
17. Kothari, A., Wu, Y.-W., Chandonia, J.-M., Charrier, M., Rajeev, L., Rocha, A.M., Joyner, D.C., Hazen, T.C., Singer, S.W., and Mukhopadhyay, A. (2019). Large Circular Plasmids from Groundwater Plasmidomes Span Multiple Incompatibility Groups and Are Enriched in Multimetal Resistance Genes. *mBio* 10, e02899-18. <https://doi.org/10.1128/mBio.02899-18>.
18. Acman, M., van Dorp, L., Santini, J.M., and Balloux, F. (2020). Large-scale network analysis captures biological features of bacterial plasmids. *Nat. Commun.* 11, 2452. <https://doi.org/10.1038/s41467-020-16282-w>.
19. Ji, Y., Shang, J., Tang, X., and Sun, Y. (2023). HOTSPOT: hierarchical host prediction for assembled plasmid contigs with transformer. *Bioinformatics* 39, btad283. <https://doi.org/10.1093/bioinformatics/btad283>.
20. Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E.P.C., Garcillán-Barcia, M.P., and de la Cruz, F. (2020). Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat. Commun.* 11, 3602. <https://doi.org/10.1038/s41467-020-17278-2>.
21. Kothari, A., Roux, S., Zhang, H., Prieto, A., Soneja, D., Chandonia, J.-M., Spencer, S., Wu, X., Altenburg, S., Fields, M.W., et al. (2021). Ecogenomics of groundwater phages suggests niche differentiation linked to specific environmental tolerance. *mSystems* 6, 101128. <https://doi.org/10.1128/mSystems.00537-21>.
22. Kothari, A., Soneja, D., Tang, A., Carlson, H.K., Deutschbauer, A.M., and Mukhopadhyay, A. (2019). Native Plasmid-Encoded Mercury Resistance Genes Are Functional and Demonstrate Natural Transformation in Environmental Bacterial Isolates. *mSystems* 4, e00588-19. <https://doi.org/10.1128/mSystems.00588-19>.
23. Liu, H., Price, M.N., Waters, R.J., Ray, J., Carlson, H.K., Lamson, J.S., Chakraborty, R., Arkin, A.P., and Deutschbauer, A.M. (2018). Magic pools: parallel assessment of transposon delivery vectors in bacteria. *mSystems* 3, e00143-17. <https://doi.org/10.1128/mSystems.00143-17>.
24. Coutinho, T.A., and Venter, S.N. (2009). *Pantoea ananatis*: an unconventional plant pathogen. *Mol. Plant Pathol.* 10, 325–335. <https://doi.org/10.1111/j.1364-3703.2009.00542.x>.
25. Walterson, A.M., and Stavrinides, J. (2015). *Pantoea*: insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiol. Rev.* 39, 968–984. <https://doi.org/10.1093/femsre/fuv027>.
26. Lv, L., Luo, J., Ahmed, T., Zaki, H.E.M., Tian, Y., Shahid, M.S., Chen, J., and Li, B. (2022). Beneficial effect and potential risk of *pantoea* on rice production. *Plants* 11, 2608. <https://doi.org/10.3390/plants11192608>.
27. Mani, S., and Nair, J. (2021). *Pantoea* infections in the neonatal intensive care unit. *Cureus* 13, e13103. <https://doi.org/10.7759/cureus.13103>.
28. Pérez-Carrascal, O.M., Choi, R., Massot, M., Pees, B., Narayan, V., and Shapira, M. (2022). Host Preference of Beneficial Commensals in a Microbially-Diverse Environment. *Front. Cell. Infect. Microbiol.* 12, 795343. <https://doi.org/10.3389/fcimb.2022.795343>.
29. Liu, N., and Cui, T. (2024). Using Omics Techniques to Analyze the Effects of Gene Mutations and Culture Conditions on the Synthesis of β -Carotene in *Pantoea dispersa*. *Fermentation* 10, 83. <https://doi.org/10.3390/fermentation10020083>.
30. Lai, L., Li, S., Zhang, S., Liu, M., Xia, L., Ren, Y., and Cui, T. (2024). Enhancing Benzo[a]pyrene Degradation by *Pantoea dispersa* MSC14 through Biostimulation with Sodium Gluconate: Insights into Mechanisms and Molecular Regulation. *Microorganisms* 12, 592. <https://doi.org/10.3390/microorganisms12030592>.
31. Hoshino, Y., Moriya, M., Matsudaira, A., Katashkina, J.I., Nitta, N., Nishio, Y., and Usuda, Y. (2020). Stereospecific linalool production utilizing two-phase cultivation system in *Pantoea ananatis*. *J. Biotechnol.* 324, 21–27. <https://doi.org/10.1016/j.jbiotec.2020.09.021>.
32. Thorgersen, M.P., Ge, X., Poole, F.L., Price, M.N., Arkin, A.P., and Adams, M.W.W. (2019). Nitrate-Utilizing Microorganisms Resistant to Multiple Metals from the Heavily Contaminated Oak Ridge Reservation. *Appl. Environ. Microbiol.* 85, e00896-19. <https://doi.org/10.1128/AEM.00896-19>.
33. Thorgersen, M.P., Goff, J.L., Trotter, V.V., Poole, F.L., Arkin, A.P., Deutschbauer, A.M., and Adams, M.W.W. (2024). Fitness factors impacting survival of a subsurface bacterium in contaminated groundwater. *ISME J.* 18, wrae176. <https://doi.org/10.1093/ismejo/wrae176>.
34. Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Møller Aarestrup, F., and Hasman, H. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. <https://doi.org/10.1128/AAC.02412-14>.
35. Tang, X., Shang, J., Ji, Y., and Sun, Y. (2023). PLASMe: a tool to identify PLASMid contigs from short-read assemblies using transformer. *Nucleic Acids Res.* 51, e83. <https://doi.org/10.1093/nar/gkad578>.
36. Luo, H., and Gao, F. (2019). DoriC 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res.* 47, D74–D77. <https://doi.org/10.1093/nar/gky1014>.
37. Li, Y., and Gao, F. (2025). OriV-Finder: a comprehensive web server for bacterial plasmid replication origin analysis. *Nucleic Acids Res.* 53, W451–W456. <https://doi.org/10.1093/nar/gkaf341>.
38. Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* 97, 6640–6645. <https://doi.org/10.1073/pnas.120163297>.
39. Thompson, M.G., Sedaghatian, N., Barajas, J.F., Wehrs, M., Bailey, C.B., Kaplan, N., Hillson, N.J., Mukhopadhyay, A., and Keasling, J.D. (2018). Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Sci. Rep.* 8, 1590. <https://doi.org/10.1038/s41598-018-20016-w>.
40. Selzer, G., Som, T., Itoh, T., and Tomizawa, J. (1983). The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. *Cell* 32, 119–129. [https://doi.org/10.1016/0092-8674\(83\)90502-0](https://doi.org/10.1016/0092-8674(83)90502-0).
41. Hersfield, V., Boyer, H.W., Yanofsky, C., Lovett, M.A., and Helinski, D.R. (1974). Plasmid ColEI as a molecular vehicle for cloning and amplification of DNA. *Proc. Natl. Acad. Sci. USA* 71, 3455–3459. <https://doi.org/10.1073/pnas.71.9.3455>.
42. Guizliou, S., Sauveplane, V., Chang, H.-J., Clerté, C., Declerck, N., Jules, M., and Bonnet, J. (2016). A part toolbox to tune genetic expression in *Bacillus subtilis*. *Nucleic Acids Res.* 44, 7495–7508. <https://doi.org/10.1093/nar/gkw624>.
43. Kushwaha, M., and Salis, H.M. (2015). A portable expression resource for engineering cross-species genetic circuits and pathways. *Nat. Commun.* 6, 7832. <https://doi.org/10.1038/ncomms8832>.
44. Wehrs, M., Pahl, J.-P., Moon, J., Li, Y., Tanjore, D., Keasling, J.D., Pray, T., and Mukhopadhyay, A. (2018). Production efficiency of the bacterial non-ribosomal peptide indigoidine relies on the respiratory metabolic state in *S. cerevisiae*. *Microb. Cell Fact.* 17, 193. <https://doi.org/10.1186/s12934-018-1045-1>.
45. Wehrs, M., Gladden, J.M., Liu, Y., Platz, L., Pahl, J.-P., Moon, J., Papa, G., Sundstrom, E., Geiselman, G.M., Tanjore, D., et al. (2019). Sustainable bioproduction of the blue pigment indigoidine: Expanding the range of heterologous products in *R. toruloides* to include non-ribosomal peptides. *Green Chem.* 21, 3394–3406. <https://doi.org/10.1039/C9GC00920E>.
46. Xu, F., Gage, D., and Zhan, J. (2015). Efficient production of indigoidine in *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* 42, 1149–1155. <https://doi.org/10.1007/s10295-015-1642-5>.
47. Ghiffary, M.R., Prabowo, C.P.S., Sharma, K., Yan, Y., Lee, S.Y., and Kim, H.U. (2021). High-Level Production of the Natural Blue Pigment Indigoidine from Metabolically Engineered *Corynebacterium glutamicum* for

- Sustainable Fabric Dyes. *ACS Sustain. Chem. Eng.* 9, 6613–6622. <https://doi.org/10.1021/acssuschemeng.0c09341>.
48. Eng, T., Banerjee, D., Menasalvas, J., Chen, Y., Gin, J., Choudhary, H., Baidoo, E., Chen, J.H., Ekman, A., Kakumanu, R., et al. (2023). Maximizing microbial bioproduction from sustainable carbon sources using iterative systems engineering. *Cell Rep.* 42, 113087. <https://doi.org/10.1016/j.celrep.2023.113087>.
49. Banerjee, D., Yunus, I.S., Wang, X., Kim, J., Srinivasan, A., Menchavez, R., Chen, Y., Gin, J.W., Petzold, C.J., Martin, H.G., et al. (2024). Genome-scale and pathway engineering for the sustainable aviation fuel precursor isoprenol production in *Pseudomonas putida*. *Metab. Eng.* 82, 157–170. <https://doi.org/10.1016/j.ymben.2024.02.004>.
50. Lim, H., Srinivasan, A., Menchavez, R., Yunus, I.S., Noh, M.H., White, M., Chen, Y., Gin, J.W., Palsson, B.O., Lee, T.S., et al. (2026). Evolution guided tolerance engineering of *Pseudomonas putida* KT2440 for production of the sustainable aviation fuel precursor isoprenol. *Metab. Eng.* 97, 322–335. <https://doi.org/10.1016/j.ymben.2025.05.007>.
51. Taylor, M., Tuffin, M., Burton, S., Eley, K., and Cowan, D. (2008). Microbial responses to solvent and alcohol stress. *Biotechnol. J.* 3, 1388–1397. <https://doi.org/10.1002/biot.200800158>.
52. Junker, R.R., and Tholl, D. (2013). Volatile organic compound mediated interactions at the plant-microbe interface. *J. Chem. Ecol.* 39, 810–825. <https://doi.org/10.1007/s10886-013-0325-9>.
53. Hammerbacher, A., Coutinho, T.A., and Gershenzon, J. (2019). Roles of plant volatiles in defence against microbial pathogens and microbial exploitation of volatiles. *Plant Cell Environ.* 42, 2827–2843. <https://doi.org/10.1111/pce.13602>.
54. Schalck, T., Bergh, B.V.D., and Michiels, J. (2021). Increasing solvent tolerance to improve microbial production of alcohols, terpenoids and aromatics. *Microorganisms* 9, 249. <https://doi.org/10.3390/microorganisms9020249>.
55. Thompson, M.G., Incha, M.R., Pearson, A.N., Schmidt, M., Sharpless, W.A., Eiben, C.B., Cruz-Morales, P., Blake-Hedges, J.M., Liu, Y., Adams, C.A., et al. (2020). Fatty acid and alcohol metabolism in *Pseudomonas putida*: functional analysis using random barcode transposon sequencing. *Appl. Environ. Microbiol.* 86, e01665-20. <https://doi.org/10.1128/AEM.01665-20>.
56. Jones, J.A., Toparlak, Ö.D., and Koffas, M.A.G. (2015). Metabolic pathway balancing and its role in the production of biofuels and chemicals. *Curr. Opin. Biotechnol.* 33, 52–59. <https://doi.org/10.1016/j.copbio.2014.11.013>.
57. George, K.W., Thompson, M.G., Kim, J., Baidoo, E.E.K., Wang, G., Benites, V.T., Petzold, C.J., Chan, L.J.G., Yilmaz, S., Turhanen, P., et al. (2018). Integrated analysis of isopentenyl pyrophosphate (IPP) toxicity in isoprenoid-producing *Escherichia coli*. *Metab. Eng.* 47, 60–72. <https://doi.org/10.1016/j.ymben.2018.03.004>.
58. de Siqueira, G.M.V., Srinivasan, A., Chen, Y., Gin, J.W., Petzold, C.J., Lee, T.S., Guazzaroni, M.-E., Eng, T., and Mukhopadhyay, A. (2025). Alternate routes to acetate tolerance lead to varied isoprenol production from mixed carbon sources in *Pseudomonas putida*. *Appl. Environ. Microbiol.* 97, e0212324. <https://doi.org/10.1128/aem.02123-24>.
59. Irvine, J., Arora, J., Martinson, J.N.V., Patel, J.R., Hasham, S.I., Cress, B.F., and Rubin, B.E. (2025). Generating functional plasmid origins with OriGen. *Nucleic. Acids Res.* 53, gkaf1198. <https://doi.org/10.1093/nar/gkaf1198>.
60. Pagotto, F., and Dillon, J.A. (2001). Multiple origins and replication proteins influence biological properties of beta-lactamase-producing plasmids from *Neisseria gonorrhoeae*. *J. Bacteriol.* 183, 5472–5481. <https://doi.org/10.1128/JB.183.19.5472-5481.2001>.
61. Thorgersen, M.P., Lancaster, W.A., Vaccaro, B.J., Poole, F.L., Rocha, A.M., Mehlhorn, T., Pettenato, A., Ray, J., Waters, R.J., Melnyk, R.A., et al. (2015). Molybdenum availability is key to nitrate removal in contaminated groundwater environments. *Appl. Environ. Microbiol.* 81, 4976–4983. <https://doi.org/10.1128/AEM.00917-15>.
62. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115. <https://doi.org/10.1093/nar/gks596>.
63. Zhao, S., and Fernald, R.D. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J. Comput. Biol.* 12, 1047–1064. <https://doi.org/10.1089/cmb.2005.12.1047>.
64. Ham, T.S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N.J., and Keasling, J.D. (2012). Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res.* 40, e141. <https://doi.org/10.1093/nar/gks531>.
65. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. <https://doi.org/10.1038/nmeth.1318>.
66. Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3, e3647. <https://doi.org/10.1371/journal.pone.0003647>.
67. Huang, Y.Y., Price, M.N., Hung, A., Gal-Oz, O., Ho, D., Carion, H., Deutschbauer, A.M., and Arkin, A.P. (2022). Functional screens of bar-coded expression libraries uncover new gene functions in carbon utilization among gut Bacteroidales. Preprint at bioRxiv. <https://doi.org/10.1101/2022.10.10.511384>.
68. David, B., Li, J., Masood, F., Blassick, C., Jensen, P., and Jensen, K. (2021). Laboratory Exercise to Measure Plasmid Copy Number by qPCR. *J. Microbiol. Biol. Educ.* 22, e00125-21. <https://doi.org/10.1128/jmbe.00125-21>.
69. Ruijter, J.M., Barnewall, R.J., Marsh, I.B., Szentirmay, A.N., Quinn, J.C., van Houdt, R., Gunst, Q.D., and van den Hoff, M.J.B. (2021). Efficiency correction is required for accurate quantitative PCR analysis and reporting. *Clin. Chem.* 67, 829–842. <https://doi.org/10.1093/clinchem/hvab052>.
70. Linger, J.G., Vardon, D.R., Guarnieri, M.T., Karp, E.M., Hunsinger, G.B., Franden, M.A., Johnson, C.W., Chupka, G., Strathmann, T.J., Pienkos, P.T., and Beckham, G.T. (2014). Lignin valorization through integrated biological funneling and chemical catalysis. *Proc. Natl. Acad. Sci. USA* 111, 12013–12018. <https://doi.org/10.1073/pnas.1410657111>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial strains		
<i>Escherichia coli</i> BW25113	JBEI isolate collection	N/A
<i>Escherichia coli</i> TransforMax EC100D <i>pir</i> -116	Biosearch Technologies	Cat#EC100D
<i>Escherichia coli</i> WM3064	Biosearch Technologies	Custom order
10-beta Competent <i>Escherichia coli</i>	New England Biolabs	Cat#C3019H
<i>Pantoea</i> sp. MT58	ENIGMA isolate collection	N/A
<i>Brevundimonas</i> sp. GW460-12-10-14-LB2	ENIGMA isolate collection	N/A
Chemicals, peptides and recombinant proteins		
LB Broth, Miller	BD Difco	Cat#244620
Carbenicillin (100 mg/mL)	Teknova	Cat#C2135
Kanamycin (50 mg/mL)	Teknova	Cat#K2127
Diaminopimelic acid (DAP)	Sigma-Aldrich	Cat#D1377
NEBuilder HiFi DNA Assembly Master Mix	New England Biolabs	Cat#M5520AA
Bpil (BbsI) (10 U/μL)	Thermo Fisher Scientific	Cat#ER1011
Buffer G (10X)	Thermo Fisher Scientific	Cat#BG5
T4 DNA Ligase (2,000,000 U/mL)	New England Biolabs	Cat#M0202M
Sodium phosphate dibasic	Sigma-Aldrich	Cat#S9763
Potassium phosphate monobasic	Sigma-Aldrich	Cat#P0662
Sodium chloride	Sigma-Aldrich	Cat#S9888
Glucose solution	Teknova	Cat#G2020
Ammonium sulfate	Sigma-Aldrich	Cat#A4915
Magnesium sulfate	Sigma-Aldrich	Cat#208094
Calcium chloride	Sigma-Aldrich	Cat#C4901
Trace metal solution	Teknova	Cat#T1001
Arabinose	Sigma-Aldrich	Cat#A3256
Critical commercial assays		
QIAprep Spin Miniprep Kit	Qiagen	Cat#27106
Q5 High-Fidelity DNA Polymerase	New England Biolabs	Cat#M0491S
Zymo Clean & Concentrator	Zymo Research	Cat#D4004
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad	Cat#1725270
Oligonucleotides		
AC259	5' gggagcttatcgatgcgtca 3'	
AC260	5' agtccttcgccagcgtaa 3'	
AC265	5' aggcttcaatgtcgccaaga 3'	
AC266	5' ctctactggcagggcatagc 3'	
AC271	5' gcggtttcgggtcgataagc 3'	
AC272	5' tactggcacaccttctgctg 3'	
AC273	5' tgtttgccggatcaagagct 3'	
AC274	5' ccctgacgagcatcacaaa 3'	
Recombinant DNA		
pAC70	This paper	This paper
pAC85	This paper	This paper

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pAC76	This paper	This paper
pAC77	This paper	This paper

Software and algorithms

SoftMax Pro 7.1.2	Molecular Devices	https://shop.moleculardevices.com/collections/software-and-instruments/products/softmax-pro-standard
BD Accuri™ C6 Software 1.0.264.21	BD Biosciences	https://www.bdbiosciences.com/en-us/products/software/flowjo-software?tab=flowJo-v11-software
OpenLab software v3.6	Agilent	https://www.agilent.com/en/product/software-informatics/analytical-software-suite/chromatography-data-systems/openlab-cds
CFX Manager 3.1	Bio-Rad	https://www.bio-rad.com/en-us/sku/1845000-cfx-manager-software?ID=1845000
Primer3Plus version: 3.3.0	Untergasser et al. 2012 ⁶²	https://www.primer3plus.com/index.html
PCR miner	Zhao and Fernald, 2005 ⁶³	http://miner.ewindup.cn/miner/
DoriC 10.0	Luo and Gao, 2019 ³⁶	https://tubic.org/doric10/public/index.php
OriV-Finder	Li and Gao, 2025 ³⁷	https://tubic.org/OriV-Finder/OriV-Finder/

Other

Veriti 96-well Thermal Cycler	Thermo Fisher Scientific	Cat#4375305
CFX96 Touch Real-Time PCR Detection System	Bio Rad	Cat#184-5096
8890 GC System	Agilent	Cat#G7110A
Flame Ionization Detector	Agilent	Cat#G3540A
Auto sampler	Agilent	Cat#G4513A
BD C6 Accuri™ flow cytometer	BD Biosciences	Cat#7693A
SpectraMax M2	Molecular Devices	Cat#0112-0102

Deposited Data

PacBio sequencing data	NCBI BioProject: PRJNA1400610	SAMN54496139
Nanopore sequencing data	NCBI BioProject: PRJNA1400610	SAMN54496138
<i>Escherichia coli</i> plate	NCBI BioProject: PRJNA1400610	SAMN54496120
<i>Escherichia coli</i> outgrowth	NCBI BioProject: PRJNA1400610	SAMN54496123
<i>Escherichia coli</i> liquid time point 1	NCBI BioProject: PRJNA1400610	SAMN54496126
<i>Escherichia coli</i> liquid time point 2	NCBI BioProject: PRJNA1400610	SAMN54496129
<i>Escherichia coli</i> liquid time point 3	NCBI BioProject: PRJNA1400610	SAMN54496132
<i>Escherichia coli</i> liquid time point 4	NCBI BioProject: PRJNA1400610	SAMN54496135
<i>Pantoea</i> sp. plate	NCBI BioProject: PRJNA1400610	SAMN54496121
<i>Pantoea</i> sp. outgrowth	NCBI BioProject: PRJNA1400610	SAMN54496124
<i>Pantoea</i> sp. liquid time point 1	NCBI BioProject: PRJNA1400610	SAMN54496127
<i>Pantoea</i> sp. liquid time point 2	NCBI BioProject: PRJNA1400610	SAMN54496130
<i>Pantoea</i> sp. liquid time point 3	NCBI BioProject: PRJNA1400610	SAMN54496133
<i>Pantoea</i> sp. liquid time point 4	NCBI BioProject: PRJNA1400610	SAMN54496136
<i>Brevundimonas</i> sp. plate	NCBI BioProject: PRJNA1400610	SAMN54496122
<i>Brevundimonas</i> sp. outgrowth	NCBI BioProject: PRJNA1400610	SAMN54496125
<i>Brevundimonas</i> sp. liquid time point 1	NCBI BioProject: PRJNA1400610	SAMN54496128
<i>Brevundimonas</i> sp. liquid time point 2	NCBI BioProject: PRJNA1400610	SAMN54496131
<i>Brevundimonas</i> sp. liquid time point 3	NCBI BioProject: PRJNA1400610	SAMN54496134
<i>Brevundimonas</i> sp. liquid time point 4	NCBI BioProject: PRJNA1400610	SAMN54496137

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacterial strains and growth conditions

The bacterial strains and plasmids used in this study are listed in [Data S2](#). *E. coli* BW25113 was acquired from an LBNL public registry (<https://public-registry.jbei.org/>).⁶⁴ All other strains listed in [Data S2](#) were isolated from the Oak Ridge Field Research Center (ORFRC); (https://public.ornl.gov/orfc/orfc3_site.cfm).

Competent cells of *E. coli* TransforMax EC100D *pir*-116 cloning strain and the *E. coli* WM3064 conjugation donor strain were purchased from Lucigen (Biosearch Technologies, Alexandria, MN). *E. coli* 10- β competent cells were purchased from New England BioLabs (NEB, Ipswich, MA, USA). All bacterial strains were grown in liquid and solid medium using a Luria-Bertani (LB) base (Becton Dickinson, Milpitas, CA). Antibiotic concentrations for selection were carbenicillin (carb, 50 μ g/mL) and kanamycin (kan, 50 μ g/mL). When the *E. coli* WM3064 strain was used, media was supplemented with diaminopimelic acid (DAP) at a concentration of 300 μ M.

All non-magic pool plasmids were constructed using standard Gibson cloning methods.⁶⁵ Plasmid DNA isolation was performed using a miniprep plasmid isolation kit according to the manufacturer's instructions (Qiagen, Redwood City, CA). General microbiological manipulations, PCR amplification, golden gate assembly, cloning, and transformation procedures were described previously.²³ Enzymes were purchased from New England Biolabs (NEB, Ipswich, MA, USA) and Thermo Fisher Scientific (Waltham, MA, USA). Oligonucleotide primers were obtained from Integrated DNA Technologies (IDT, Coralville, IA, USA). Double strand DNA gene fragments were ordered for synthesis from Genewiz Inc (Azenta Life Sciences, South Plainfield, NJ). All plasmids constructed, not including magic pool library construction, were verified by whole plasmid sequencing by Plasmidsaurus Inc (South San Francisco, CA). Unless noted, all reagents and media components were purchased from Sigma-Aldrich (St. Louis, MO, USA).

METHOD DETAILS

Origin of replication prediction

From our earlier plasmidome study, 32 circular plasmid sequences were analyzed using DoriC for prediction of the origin of replication⁸ and the resulting 18 predicted origins of replication were used in the subsequent experimental validation workflow. For this study, DoriC 10.0 was used (details for DoriC 10 are provided at <https://tubic.org/doric10/public/index.php>).³⁶ DoriC has since been updated to Dori 12.0⁸ and a new plasmid focused OriV-finder³⁷ tool was also released recently. All 32 scaffolds examined in this study were also analyzed using the OriV-finder tool. Details of OriV-finder are available at <https://tubic.org/OriV-Finder/OriV-Finder/>. Examined scaffold sequences, original DoriC predictions and the most recent OriV-finder predictions are provided in supplementary data ([Data S1](#)).

Construction of the kanamycin magic pool origin libraries

Part 1 origin plasmids were synthesized into a pJW52 backbone⁶⁵ by Genscript USA Inc (Piscataway, NJ). Parts 2 (promoter for drug resistance marker) and 3 (drug resistance coding sequence (CDS)) (also in pJW52 backbones) plasmids were constructed previously.²³ All plasmids containing parts 1, 2, and 3 can be found in [Data S1](#). For all part 1 origins, 100 ng of each plasmid were added to a part 1 mix. This was also done for parts 2 and 3, separately. 100 ng of each mix was then added into the golden gate assembly.⁶⁶ The origin magic pool was constructed by Golden Gate assembly using methods described previously.²³

The magic pool was built in two dilutions, "1x" and "2x" based on the number of *E. coli* *pir*⁺ colonies added to 100 mL of LB medium supplemented with DAP and carbenicillin. The magic pool was allowed to grow overnight, followed by a glycerol stock and four plasmid DNA isolations per dilution. An amplification of the barcoded region of each plasmid assembly was performed to check the number of barcodes per assembly. For the magic pool the "1x" dilution was selected due to its barcode coverage relative to the total combination of possible vectors (~10x). Utilizing the DNA isolated previously, the magic pool was transformed into the *E. coli* WM3064 conjugation donor strain. The entirety of the transformation was then grown in 100 mL of media supplemented with DAP and stored via glycerol stock after an overnight incubation.

The plasmid DNA that was isolated from the *E. coli* *pir*⁺ cells, and subsequently used to transform into the *E. coli* WM3064 conjugation donor, was sequenced by Plasmidsaurus using long-read sequencing (once with PacBio sequencing technology and once with in-house using Nanopore sequencing) with custom analysis and annotation. Sequencing data was analyzed by searching through each long read and comparing the sequence to a list of parts that were used in the construction. The output of this search was a list of barcodes and the parts (1,2 and 3) associated with the barcode. For us to confidently assign a unique barcode to an origin part, we had to observe at least two unique long-reads that supported this observation. In some cases a single barcode was associated with multiple parts from each grouping of parts (i.e., multiple part 1s can be associated with a single barcode) (see [Figure S1](#) and [Data S7](#) for part 1 barcode-origin associations). These non-unique barcodes were not used for any further analysis.

Part 2.4 showed a 160 bp sequence similarity to the part 4 backbone (promoter driving *bla* gene). This caused the software that analyzed the long read sequencing data to not be able to differentiate between part 2.4 and the backbone. In [Figures S2–S4](#) "unmapped part 2" represents either unmapped part 2s or part 2.4.

Magic pool conjugation

The following protocol was adapted from Liu et al.²³ Recipient strains were streaked out on LB agar for isolated colonies. A single colony of the recipient was used to inoculate a 10 mL volume of LB and grown overnight at 30 °C and shaking at 200 rpm. A 2 mL glycerol stock of the *E. coli* WM3064 donor strain, containing the magic pool library, was inoculated into 50 mL of LB supplemented with carbenicillin and DAP and allowed to grow at 37 °C, while shaking at 200 rpm, for 4 h. One OD of donor cells was pelleted in a microcentrifuge tube and washed twice with LB. One OD of recipient cells was pelleted and washed once with LB. Each conjugation contained a 1:1 OD ratio of donor:recipient cells. A donor pellet and a recipient pellet were resuspended in the same 200 µL of LB and plated onto an LB plate supplemented with DAP. Conjugations were incubated at 30 °C overnight. Cells were harvested from the conjugation plates and resuspended in 4 mL of LB with 15% glycerol. These conjugation stocks were frozen at –80°C.

To determine the number of colonies per mL, a tube of conjugation stock was thawed and serially diluted down to 1×10^{-6} . Dilutions 1×10^0 to 1×10^{-4} were plated on kanamycin selective media while dilutions 1×10^{-5} and 1×10^{-6} were plated on non-selective media to check for transformation efficiency and for survival of the recipient. Colony counts from these plates were used to estimate the volume of conjugation mix would be needed to get ~5,000 colonies. This volume was then either inoculated directly into liquid selection, and grown for the time course experiments, or plated onto LB supplemented with kanamycin and incubated overnight at 30 °C to form small colonies.

For the liquid selection time course experiment, ~5,000 CFUs were inoculated into 10 mL of LB supplemented with kanamycin and allowed to grow overnight at 30 °C and shaking at 200 rpm. 1 mL of the culture was pelleted and plasmid DNA was extracted for sequencing. 500 µL of the grown culture was then used to back dilute into 10 mL of fresh LB supplemented with kanamycin. This back dilution happened 3 times resulting in 4 total time points.

For plate-based selection, plates were incubated overnight at 30 °C. Colonies were then harvested and resuspended in LB. The resuspension was diluted to an OD of 3 and pelleted. The plasmid DNA from this pellet was extracted and used for sequencing. With the same resuspension, 50 mL cultures were started at an OD of 0.5. These cultures were incubated overnight at 30 °C and shaken at 200 rpm and pellets were collected the next day for DNA extraction and sequencing.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analyzing the plasmid fitness within each bacterial strain

For each sample PCR amplification of the DNA barcodes, and BarSeq, was performed as described previously.^{23,67} On average, 4 million reads were generated per sample. For the plate- and liquid-based experiments, we typically collected ~5000 colonies per experiment resulting in a read coverage per colony of 800x.

A list of the count of barcodes per sample was generated. Each file, containing barcode sequencing data for one strain's time point, was filtered to remove all barcode counts that were below 6. After filtering, an average of 98.64% ($\pm 0.36\%$) of the barcode counts remained. The filtered barcodes were then compared to the barcode-part association generated and the resulting data was used to generate all origin-related data, as well as the analysis of each recipient bacterium's preferred part 2s and part 3s. The barcodes for each time point were also investigated for barcode diversity. This was done by determining the number of barcodes representing each origin present at each time point.

To classify an origin as functional we considered both the plate-based and liquid outgrowth assays. In both instances, we required multiple barcodes associated with a single origin to show high barcode counts. For the purely liquid transfer experiments, we also expected that long-term selection in the presence of kanamycin would lead to instances of some origins "out-performing" others, due potentially to copy number variation in the kanamycin resistance gene. Relatedly, we found that the part 2 promoter sequence had a large impact on driving kanamycin resistance, which could also lead to barcodes associated with functional origins going up or down over multiple transfers. So, for the liquid experiments with multiple transfers we also looked at the behavior and barcode counts of the different origins and their kanamycin resistance cassettes holistically (Figures S2–S4).

Colony quantitative PCR

SsoAdvanced Universal SYBR Green Supermix was ordered from BioRad (BioRad Inc, Hercules, CA). All qPCR primers were selected by using Primer3Plus (<https://www.primer3plus.com/index.html>) using the qPCR settings. Primers for genomic DNA were ordered for gene IAI47_12350 (a glycosyl transferase/lysophospholipid acyltransferase), a gene found only in the *Pantoea* sp. MT58 genome and not on the introduced plasmid. This primer pair amplifies a region of 528 bp. Primers for plasmid amplification were ordered to amplify a region within the origin 6911. This primer pair amplifies a region of 598 bp. Melt curves utilizing all primer pairs show only one amplicon. Plates were run on a CFX96 Real-Time System using Bio-Rad CFX Manager 3.1 software. A fresh single colony of *Pantoea* sp. MT58 with pAC70, *Pantoea* sp. MT58 with pAC75, and *Pantoea* sp. MT58 with pAC77 was picked into 40 µL of Ambion nuclease-free water, separately, and resuspended by pipetting up and down. 20 µL of the resuspended colony was then diluted into 20 µL of water. This was repeated for a total of 3 dilutions (undiluted, 1:2, 1:4, 1:8). 2 µL of each dilution was used for each qPCR reaction, which was performed in triplicate. All colony qPCR data and primers used for all colony qPCR experiments can be found in Data S3. Plasmids will be made available upon request.

The thermocycler protocol and qPCR mastermix were executed as described in David et al.⁶⁸

Copy numbers (CN) were calculated using the following formula⁶⁸ (Equation 1):

$$CN = 2^{-(Ct_{plasmid} - Ct_{gDNA})} \quad (\text{Equation 1})$$

Copy number was then corrected by primer efficiency (E) by the following formula (Equation 2):

$$CN_{adjusted} = CN \times (E_{gDNA \text{ primers}} / E_{plasmid \text{ primers}}) \quad (\text{Equation 2})$$

Primer efficiency was calculated using PCR Miner (<http://miner.ewindup.cn/miner/>), which calculates the primer efficiency of each well by performing a linear regression of the fluorescence values during the exponential phase of amplification.

The final equation (Equation 3) used to determine the plasmid copy numbers was the following⁶⁹:

$$CN_{adjusted} = (2^{-(Ct_{plasmid} - Ct_{gDNA})}) \times (E_{gDNA \text{ primers}} / E_{plasmid \text{ primers}}) \quad (\text{Equation 3})$$

GFP quantification in *Pantoea* sp. MT58

The original magic pool plasmid design utilized a promoter that was non-functional in driving expression of GFP in *Pantoea* sp. MT58. Plasmid pAC80 was constructed combining origin 6911, part 2.4 and part 3.8, as well as an inducible promoter, P_{BAD}, to drive expression of the GFP. Plasmid pAC80 for GFP expression was conjugated into *Pantoea* sp. MT58 using standard protocols as described above. The empty vector, pAC87, was constructed by removing the GFP from pAC80.

Both the GFP-containing plasmid and the empty vector were conjugated in *Pantoea* sp. MT58 and successful conjugations were observed through colonies resistant to kanamycin. Three colonies of *Pantoea* sp. MT58 containing pAC80 and three colonies of *Pantoea* sp. MT58 containing the empty vector were each inoculated into a 24 deep well plate containing 3 mL of LB with kanamycin and 0.3% (w/v) arabinose and grown overnight at 30 °C while shaking at 200 rpm. After growing overnight, each culture was split into three technical replicates and back diluted (150 µL into 3 mL) in 3 mL of LB with kanamycin and 0.3% (w/v) arabinose for another overnight growth.

1 µL was sampled from the dense overnight culture and added to 99 µL of phosphate buffered saline (PBS) in a 96 well flat bottom plate. GFP was quantified by using a BD C6 Accuri flow cytometer by reading 30,000 events using a medium flow rate (35 µL/min, core size = 16 µm). One wash cycle was performed in between each well. Raw data can be found in Data S4. Representative flow cytometry data can be seen in Figure S6.

Indigoidine production

Indigoidine production plasmid pAC75 and empty vector pAC87 were conjugated into *Pantoea* sp. MT58 using standard protocols as described above. A single colony of each strain was inoculated into 5 mL of LB with kanamycin and grown overnight at 30 °C while shaking at 200 rpm. The following day, 300 µL of the seed culture was back diluted into a 24 deep well plate, in triplicate, containing 3 mL of LB with kanamycin and 0.3% (w/v) arabinose. This plate was incubated in a shaker at 30 °C and shaking at 900 rpm. After 24 h, 200 µL of each sample was taken from the plate and spun down at 14,000 rpm for 5 min. The 24 DWP was then put back into the shaker for another 24 h in which the final time point would be taken. Extraction occurred by removing the supernatant of spun down cultures, resuspending the pellet in 500 µL of DMSO and shaking at 3000 rpm for 20 min. Due to the high absorbance of the sample, 20 µL of the extracted material was diluted into 80 µL DMSO and then put into a clear bottom 96 well plate and absorbance was measured at 612 nm (Data S5).

Isoprenol production

Experiments to determine isoprenol toxicity, catabolism and production were performed using minimal salt (M9) medium composed of 1x M9 salts (6.8 g/L Na₂HPO₄, 3 g/L KH₂PO₄, 0.5 g/L NaCl), 1% (w/v) glucose, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 mM CaCl₂ and trace metal solution (500 µL per 1 L medium, Product No. 1001, Teknova Inc, Hollister, CA).^{49,70} All inoculations were grown at 30 °C and shaken at 200 rpm. Isoprenol production plasmid pAC104 was electroporated into *Pantoea* sp. MT58. Electrocompetent cells were prepared by spinning down 2.25 mL of an overnight culture at stationary phase at 8000 rpm for 3 min. Supernatant was removed and the pellet was washed with 750 µL of 10% glycerol three times, centrifuging at 8000 rpm for 3 min each wash. Washed pellet was resuspended in 250 µL of 10% glycerol. 50 µL of cells was mixed with 50 ng of DNA and electroporated at 2.5 kV/cm in a 1 cm cuvette. Cells were recovered in SOC and shaking at 900 rpm for 1 h. 200 µL of cells were plated on LB supplemented with kanamycin and grown overnight.

To determine isoprenol toxicity and catabolism, three single colonies of *Pantoea* sp. MT58 were inoculated in 5 mL LB in glass culture tubes and allowed to grow overnight at 30 °C and shaking at 200 rpm. Cultures were then back diluted in 5 mL M9 with 1% glucose in glass culture tubes and allowed to grow overnight at 30 °C and shaking at 200 rpm. This overnight culture was then used as an inoculum for further experiments as follows.

For the isoprenol consumption assay, cultures were then inoculated again in M9 (either with or without 1% glucose) either with, or without, 1 g/L isoprenol in a 24 deep well plate at an OD of 0.01, which were sealed and incubated at 30 °C with agitation in a microplate reader for 24 h.

For the isoprenol toxicity assay, cultures were then inoculated again in M9 with 1% glucose and varying concentrations of isoprenol in a 24 deep well plate at an OD of 0.01, which were sealed and incubated at 30 °C with agitation in a microplate reader for 24 h.

All plates were sealed with a semi-permeable film (Breathe Easy Film, USA Scientific, Ocala, FL) and incubated in a microplate reader (Molecule Devices M2 Plate Reader, San Jose, CA). Readings were taken every 15 min with continuous agitation at 30 °C in between readings (Figure S8).

For isoprenol production, three single colonies of *Pantoea* sp. MT58 were inoculated in 5 mL of LB supplemented with kanamycin in glass culture tubes and allowed to grow overnight at 30 °C while shaking at 200 rpm. Cultures were then back diluted in 5 mL M9 supplemented with 1% glucose and kanamycin in glass culture tubes and allowed to grow overnight at 30 °C while shaking at 200 rpm. Cultures were then back diluted in 5 mL of M9 supplemented with 1% glucose, kanamycin and either 0.2% (w/v), 1% (w/v) or no arabinose in glass culture tubes and were allowed to grow overnight at 30 °C while shaking at 200 rpm. At 24, 48, and 72 h time points samples were collected to measure cell density (Figure S10) and isoprenol according to Banerjee et al.⁴⁹ Samples were mixed with equal volumes of ethyl acetate and vortexed at max speed for 15 min. Samples were then centrifuged and 80 µL of the upper organic layer was used for analysis using Gas Chromatography- Flame Ionization Detector (GC-FID). 80 µL of the ethyl acetate layer was transferred to a GC vial containing an insert, and 1 µL was analyzed using an Agilent Technologies GC 8890 system (USA) with a FID and a DB-WAX capillary column (15 m × 0.25 mm × 0.25 µm, Agilent Technologies, USA) for quantification. Helium served as the carrier gas at a flow rate of 2.2 mL/min, and the injection volume was 1 µL in splitless mode. The injector and detector temperatures were set to 250 °C and 300 °C, respectively. The GC oven followed a programmed temperature gradient: an initial hold at 40 °C, a first ramping to 100 °C at 15 °C/min and then to 230 °C at 30 °C/min followed by a final hold at 230 °C for 1 min. Data collection and analysis were performed using OpenLab software (Agilent Technologies, USA). Analytical grade standards obtained from Sigma-Aldrich (St. Louis, MO, USA) were used to determine isoprenol concentration.