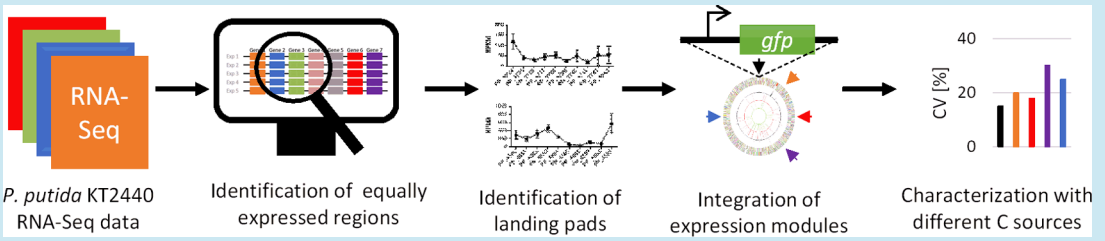
ACS SyntheticBiology

pubs.acs.org/synthbio
Research Article

Reliable Genomic Integration Sites in *Pseudomonas putida* Identified by Two-Dimensional Transcriptome Analysis

Sebastian Köbbing,* Thorsten Lechtenberg, Benedikt Wynands, Lars M. Blank, and Nick Wierckx

Cite This: *ACS Synth. Biol.* 2024, 13, 2060−2072
Read Online



ABSTRACT: Genomic integration is commonly used to engineer stable production hosts. However, so far, for many microbial workhorses, only a few integration sites have been characterized, thereby restraining advanced strain engineering that requires multiple insertions. Here, we report on the identification of novel genomic integration sites, so-called landing pads, for *Pseudomonas putida* KT2440. We identified genomic regions with constant expression patterns under diverse experimental conditions by using RNA-Seq data. Homologous recombination constructs were designed to insert heterologous genes into intergenic sites in these regions, allowing condition-independent gene expression. Ten potential landing pads were characterized using four different *msf GFP* expression cassettes. An insulated probe sensor was used to study locus-dependent effects on recombinant gene expression, excluding genomic read-through of flanking promoters under changing cultivation conditions. While the reproducibility of expression in the landing pads was very high, the msfGFP signals varied strongly between the different landing pads, confirming a strong influence of the genomic context. To showcase that the identified landing pads are also suitable candidates for heterologous gene expression in other Pseudomonads, four equivalent landing pads were identified and characterized in *Pseudomonas taiwanensis* VLB120. This study shows that genomic "hot" and "cold" spots exist, causing strong promoter-independent variations in gene expression. This highlights that the genomic context is an additional parameter to consider when designing integrable genomic cassettes for tailored heterologous expression. The set of characterized genomic landing pads presented here further increases the genetic toolbox for deep metabolic engineering in Pseudomonads.

KEYWORDS: *Pseudomonas putida, genomic integration, heterologous gene expression, synthetic promoters, synthetic biology, metabolic engineering*

INTRODUCTION

Representatives of the genus *Pseudomonas* became favored workhorses in biotechnology in recent years,[1,2] due to their versatile metabolism and high tolerance toward organic compounds.[3−5] To further increase the usability of *Pseudomonas*, the tool development is ever ongoing,[6−8] enabling rapid and efficient metabolic engineering and genome editing.[7]

For the construction of more complex genetically modified microorganisms, different genes and pathways must be combined in a single cell for whole-cell catalysis.[9,10] Here, plasmids reach their limit of usability. Besides varying plasmid copy numbers, resulting in unreliable gene expression, the use of different plasmids in a single cell is difficult since subpopulations can occur, and the fitness of the cells can be negatively affected.[11−13] Genomic integration can circumvent these issues and is commonly used due to the higher genetic stability, lower transcriptional variability, and the independence of selective pressures for maintenance.[14]

Therefore, genomic integration is the method of choice for strain engineering. Several applications for random or targeted integration are available for *Pseudomonas*. The most prominent random integration tool is the transposon Tn5, one of the first identified transposons.[15] Due to its random genomic integration, this transposon is frequently used to construct insertion libraries of microorganisms.[16,17] It allows identifying essential genes or screening for heterologous integrated constructs driven by native promoters.[18,19] Domröse et al. randomly integrated a prodigiosin biosynthesis gene cluster

Received: December 13, 2023
Revised: May 13, 2024
Accepted: May 14, 2024
Published: July 5, 2024

into the genome of *Pseudomonas putida* KT2440 and observed that integration of the constructs in 16S rRNA genes enabled the highest prodigiosin production, likely due to their high transcription and half-life of the mRNA.[19,20] However, such approaches rely on disruptive insertion, making them less suitable for the targeted, knowledge-based generation of whole-cell catalysts. In addition, the best candidates can only be identified through phenotypic screening, limiting the applicability to readily detectible products. Therefore, targeted integration at a single site is a more prominent approach to strain engineering. It enables the avoidance of disruptive insertion of heterologous constructs and the reduction of the variable influence of the unknown genomic context on gene expression. In contrast to the transposon Tn5, the native transposon Tn7 integrates into the Tn7 attachment site (*att*Tn7) of several prokaryotic and eukaryotic species.[21] Since the integration occurs site-specifically, a comparative analysis is feasible and allows, *e.g.*, the calibration and characterization of synthetic promoters.[22,23] Therefore, the *att*Tn7 site is used in many studies and is well described.[22,23] The disadvantage of using this tool is that it can only be used once, as the *att*Tn7 site is present only once in the genome.

Individual homologous recombination-based or recombinase systems are more suitable to facilitate the integration of multiple genes. Elmore et al. developed serine recombinase-assisted genome engineering (SAGE), an integration system using a site-specific recombinase for targeted and markerless genomic integration in *P. putida* KT2440.[24,25] The advantage of this system is a poly-*attB* cassette which includes 10 different sequences, recognized by recombinases for targeted integration. The tool enables the subsequent integration of up to 10 heterologous constructs. However, integration of the poly-*attB* cassette into the genome is necessary.

Martínez-García et al. investigated a homologous recombination-based modification tool for *Pseudomonas*, which enables integration, deletion, and single nucleotide exchanges. The so-called pEMG vector is equipped with two 500−1000 bp target sequences homologous to the up-and downstream regions of the integration site, flanked by I-SceI sites. The plasmid contains an *oriR6K* origin of replication, which *Pseudomonas* does not recognize. To survive antibiotic selection, the plasmid must be incorporated into the genome, guided by the target sequences. To release the formed co-integrate from the genome, the I-SceI sites are used and recognized by the endonuclease I-SceI delivered from a second plasmid. The introduced double-strand breaks inside the I-SceI sites are repaired by homologous recombination by the cell. Successful co-integrate resolution is tested by the loss of co-integrated-mediated antibiotic resistance. Subsequent PCR diagnostics of antibiotic-sensitive clones are used to screen for the desired modification. This method allows the implementation of multiple consecutive scarless and antibiotic marker-free genomic modifications into a single strain and a wide range of vectors containing different antibiotic markers are available.[8]

In addition to suitable molecular tools enabling markerless integration in the genome, suitable integration sites in the genome are also required. However, few publications have reported on integration sites in microorganisms, including *Pseudomonas*. Besides the *att*Tn7 site, Kuhlman and Cox characterized four integration sites in *Escherichia coli*, which were chosen according to their distribution along the genome in the distance to the origin of replication.[26] Thereupon, Englaender et al. compared these four sites by producing different compounds and showed that the chosen integration sites affect product formation.[27] Chaves et al. explicitly looked for integration sites in the genome of *P. putida* KT2440, which were categorized based on the intergenic sequence length.[28] Depending on the integration site, the resulting fluorescence signal used for the characterization varied by up to 27-fold. All authors concluded that the genomic context influences the expression of the integrated gene.[19,26,28] This is due to "hot" and "cold" spots on the genome as not all areas are equally active under distinct conditions.

Although several studies have identified potential genomic integration sites,[20,27,28] relatively little attention has been given to the reliability of such sites across different experimental conditions. We address this using publicly available RNA-Seq data to identify regions across the genome of *P. putida* KT2440 where gene expression has a low coefficient of variation under changing growth conditions. The genomic regions were characterized with four different *msfGFP* expression cassettes under the control of synthetic promoters to evaluate the influence of the landing pads on the expression of the integrated gene. The 10 landing pads across the genome resulted in a wide range of msfGFP expression, depending on the genomic context. Notably, gene expression was relatively insensitive to experimental conditions such as different carbon sources. Based on the findings for *P. putida* KT2440, the equivalent genes and genomic regions were also identified in *Pseudomonas taiwanensis* VLB120 and characterized in its streamlined chassis derivative strain *P. taiwanensis* GRC1.[29] Here, the influence of the integration site on the expression strength could be observed as well. The presented set of characterized genomic landing pads increases the arsenal of genome editing tools, thus enabling deep metabolic engineering of *P. putida* KT2440 and many other Pseudomonades.

RESULTS AND DISCUSSION

Mining *P. putida* KT2440 RNA-Seq Data For Equally Expressed Genes across Multiple Cultivation Conditions. RNA-Seq data representing the entire transcriptional space is rich in information on gene regulation under defined cultivation conditions.[30−33] Depending on the cultivation conditions, gene expression varies, reflecting differences in gene regulation and native promoter activities. We hypothesized that genome regions that showed low gene expression variability over various cultivation conditions would be suitable candidates for integration sites across the genome of *P. putida* KT2440.

For this purpose, RNA-Seq data sets from Frank et al.,[31] Rosa et al.,[32] and Kim et al.[33] (Supporting Table 1) were merged in a single Microsoft Excel file, and missing information, including gene names and orientation on the chromosome, were added for all genes.[34] Values for missing genes in the data sets were excluded from the analysis to avoid negative effects on calculated means and standard deviations. Also, 5S, 23S, and 16S rRNA genes were deleted from the data set since they are highly conserved, and a clear assignment of the determined transcript to one of the seven 16S rRNA operons is difficult. Furthermore, the suitability of the 16S rRNA genes as integration sites for heterologous constructs was previously described by Otto et al. and is, therefore, not intended in this study.[35] At least 5420 open reading frames were analyzed for equal expression levels under different cultivation conditions.

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

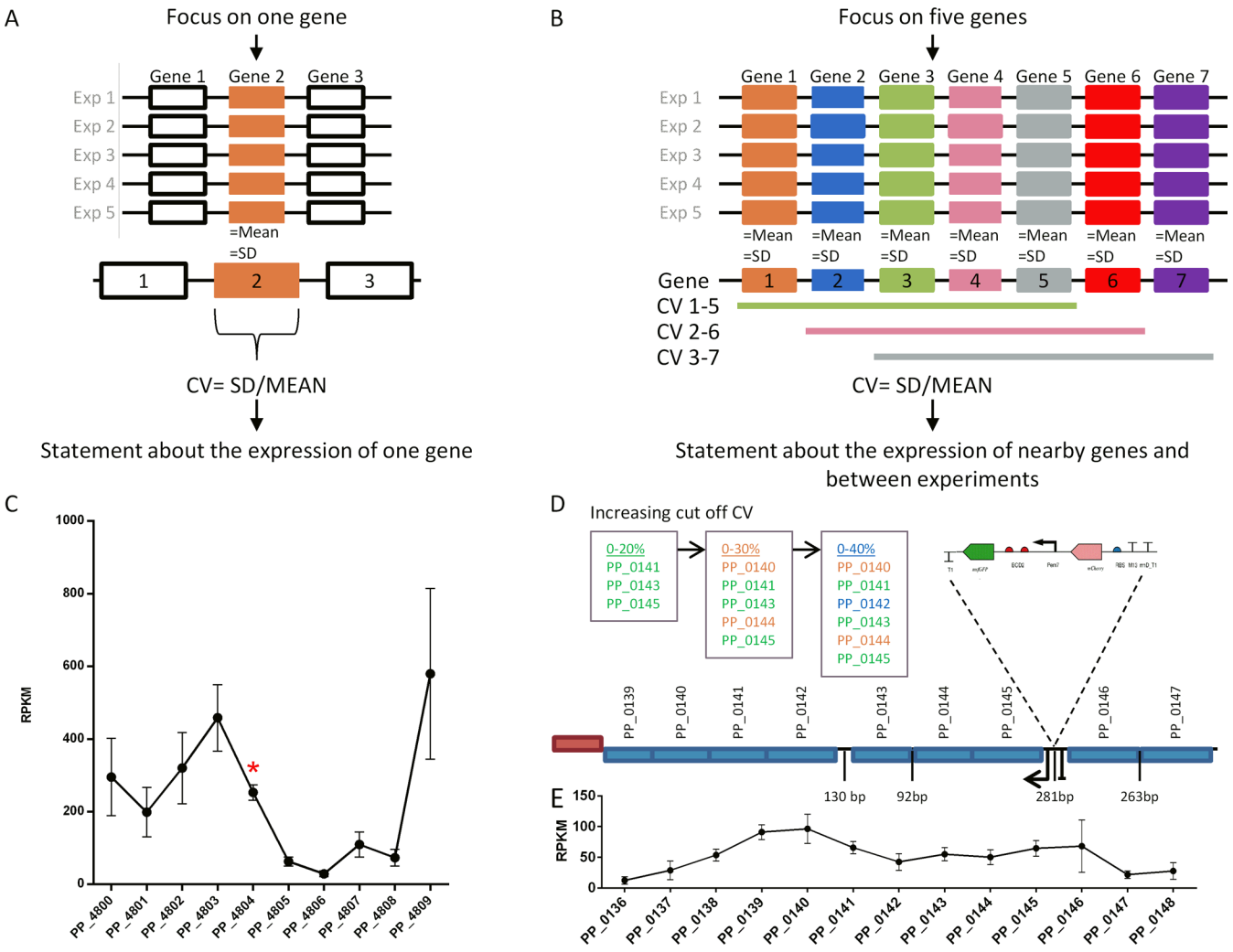


Figure 1. Workflow for identifying genomic regions on the genome of *P. putida* KT2440 with low transcriptional variance. (A) Focus on a single gene for equal expression across different experiments. CV is calculated for each gene across all data sets. (B) Identification of genomic regions to identify equally expressed neighboring genes. For each gene, the mean RPKM from RNA-Seq data sets (Exp1−Exp5) was calculated and used to determine the CV of five consecutive genes (Gene 1−Gene 7). Different RNA-Seq data sets are pictured as gene strings; each gene is represented in the same-colored box for all experiments. (C) Exemplary resulting clusters by using CV for one single gene. Here, PP_4804, indicated with an asterisk, complies with the specifications of a low CV, but nearby genes are differentially expressed. (D) Exemplary identification of integration sites using the example for PP_0145. Low CV values reflect minor differences in expression (green genes). Increasing the cutoff CV stepwise (% variation of CV) increased the number of genes fulfilling the criterion since a higher variation in expression is accepted (orange and blue gene). Nearby genes cluster into larger strings. Genome analysis reveals possible native genomic elements and the orientation of genes, which are used as a guide for the manual selection of integration sites for heterologous constructs. (E) Expression as mean RPKM and standard deviation of genes around PP_0145. Data sets were obtained from the gene expression omnibus (Supporting Table 1, GEO).[36]

In the first step, we identified regions in the genome that exhibit similar expression profiles under different conditions. We purposefully did not focus on single genes but on gene expression profiles of five adjacent genes on the chromosome, covering about 5000 bp (Figure 1). With this focus on regions, we avoid disruptive integration into single genes, which may be used for integrations. Furthermore, genes are often organized in operons, and a targeted integration next to the candidate gene could have undesired polar effects on the native gene expression profiles (Figure 1A). Another problem with selecting a single gene as an integration site is that the expression profiles of the surrounding genes likely vary (Figure 1C). Focusing on the gene's transcription in the middle of the set of five genes and moving the window of five genes one gene further, each gene is analyzed 5 times (Figure 1B). The values of the RPKM (reads per kilobase of transcript per million mapped reads) were used to calculate the means for each gene across the RNA-Seq data sets. To make a statement about the expression profile of the five adjacent genes, especially the central gene, the coefficient of variation (CV) was calculated by dividing the standard deviation by the mean. The CV reflects the variation of the average gene expression of the test set of five genes, and thus, a genomic context with lower and higher variation can be identified. Means and standard deviations were calculated from the five neighboring gene RPKM means (Figure 1B). Subsequently, to identify suitable integration sites from this initial genomic context screen, regions with genes with a low CV across RNA-Seq experiments were selected. The cutoff CV was increased stepwise by 10% to generate a list with neighbored genes. From the resulting list, we have chosen 10 genomic regions for further characterization (Figure 1D,E).

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

Table 1. Transcriptionally Highly Stable Genome Regions in *P. putida* KT2440[a]

region	integration site	gene names	position and orientation on the genome[c]
PP_0136-PP_0148	PP_0145::PP_0146	*yjbB* Na^+/Pi-cotransporter	155,482::155,483
		ygdQ transport protein	PP_0145−; PP_0146+
PP_0340-PP_0347	PP_0340::PP_0341	*gln E* glutamate-ammonia-ligase adenylyltransferase	414,043::414,044
		waaF ADP-heptose:LPS heptosyltransferase II	PP_0340+; PP_0341+
PP_0451-PP_0481	PP_0480::PP_0481	*rplQ* 50S ribosomal protein L17	564,406::564,407
		katA catalase	PP_0480+; PP_0481+
PP_1426-PP_1434	PP_1430::PP_1431	DegP-like serine endoprotease	1,631,110::1,631,111
		lepA elongation factor 4	PP_1430+; PP_1431+
PP_1734-PP_1743	PP_1738::PP_1739	hypothetical protein	1,939,140::1,939,141
		membrane protein	PP_1738−; PP_1739+
PP_2399-PP_3309	PP_3304::PP_3305	Bcr/CflA family multidrug resistance transporter	3,740,869::3,740,870
		TerC family membrane protein	PP_3304+; PP_3305−
PP_3756-PP_3764	PP_3761::PP_3762	two-component system sensor histidine kinase/response regulator	4,292,714::4,292,715
		two-component system response regulator	PP_3761−; PP_3762+
PP_4143-PP_4153	PP_4145::PP_4146	*mltD* membrane-bound lytic murein transglycosylase D	4,683,249::4,683,250
		peptide ABC transporter substrate-binding protein	PP_4145+; PP_4146+
PP_4823-PP_4837	PP_4832::PP_4833	*cobK* precorrin-6x reductase	5,499,953::5,499,954
		SURF1 domain-containing protein	PP_4832+; PP_4833+
PP_4936-PP_4950	PP_4944::PP_4945	carbamoyltransferase	5,627,625::5,627,626
		rlmJ 23S rRNA (adenine(2030)-*N*(6))-methyl-transferase	PP_4944−; PP_4945−
PP_2320-PP_2322[b]	PP_2321::PP_2322	*oprI*, major outer membrane lipoprotein	2,650,189::2,650,190
			PP_2321−; PP_2322−
PP_3805-PP_3808[b]	PP_3806::PP_3807	antibiotic synthesis protein MbtH	4,334,825::4,334,826
			PP_3806−; PP_3807−
PP_4708-PP_4709[b]	PP_4708::PP_4709	*rpsO*, 30S ribosomal protein S15	5,354,344::5,354,345
			PP_4708−; PP_4709−

[a]Regions were identified based on RNA-Seq data and *in silico* regulatory elements analysis. Given are the locus tags and annotation of genes surrounding the integration site, the region in which the gene is located, and the orientation of the genes on the genome. Gene names are taken from *Pseudomonas* Genome DB.[37] [b]Integration site was chosen by high RPKM. [c]Referring to a reference genome (NC_002947.3).

Mapping RPKM values for each gene reveals clusters of neighboring genes with similar expression levels (Supporting Figure 1). Ten genomic regions were identified, showing similar RPKM values across different experiments (Table 1). These regions were further characterized using the genomic sequence and annotation from www.pseudomonas.com to reveal gene names, orientation, and operon structures.[37] Intergenic sequences were analyzed regarding native regulatory elements like promoters and terminators to minimize changes in the native regulation.[38−40] The first identified region includes PP_0136-PP_0148 and contains different transporters and membrane proteins. PP_0340 to PP_0347 includes transferases and kinases primarily for saccharides. A third region is a highly expressed large cluster of genes starting from PP_0451 and ending at PP_0481, containing the genes encoding the 30S and 50S ribosomal proteins. As RNA-Seq data originated from cells at the exponential phase, the protein biosynthesis rate and associated ribosome expression are high.[33,41] The region contained around 17,000 base pairs with high expression levels and low variability between genes, indicating that native promoters driving these genes have similar activity (Supporting Figure 1). All identified genomic regions contain different groups and families of genes and are involved in various metabolic and physiologic processes. The similarity of expression of the clustered genes can often be attributed to operons, but several regions also contain two predicted native promoters depending on the orientation of the surrounding genes. Suitable landing pads were analyzed for the presence of regulatory elements and intergenic sequences long enough to enable genomic integration without affecting the expression of surrounding genes, *e.g.*, by the disruption of native terminator and promoter sequences. Selected intergenic regions used for genome integration featured lengths ranging from 130 to 250 base pairs.

In addition to the rational approach based on low transcriptional variability, we also chose three sites with highly expressed genes, as reflected by high mean RPKM values across cultivation conditions. We have chosen integration sites next to PP_2322, PP_3808, and PP_4709 (Table 1). The highest mean RPKM values were found for 16S rRNA genes, but these sites were omitted for the above-mentioned reasons.

In summary, we mined publicly available RNA-Seq data to identify genomic regions with high expression or low variation in two dimensions: across genomic context and across experimental conditions. Ten of these regions were chosen as landing pads and characterized *in vivo* for heterologous gene expression using genomically modified *P. putida* KT2440 strains.

Characterization of Landing Pads in *P. putida* KT2440. Identified genomic landing pads were characterized *in vivo* using targeted integration in *P. putida* KT2440 applying the pEMG system and glucose as carbon source.[14] We constructed four pEMG integration vectors for each landing pad. The activity of all sites was determined by measuring GFP fluorescence, but the vectors contained different modules to drive GFP expression: a sensor construct, no promoter, promoter P_{em7}, or promoter P_{14g}.[8,22] The sensor construct contains GFP expressed from the constitutive P_{em7} promoter with the bicistronic design element BCD2.[22] Two terminators insulate it to avoid read-through activity from native

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

promoters, and it contains a promoterless *mCherry* with RBS upstream of P_{em7}-GFP as an additional read-through control. A double terminator (M13 central + rrnD T1)[42] upstream of the expression cassette ensures that the measured GFP is only expressed from the P_{em7} promoter and unaffected from the genomic context (Figure 2A).[14,43] The other three constructs

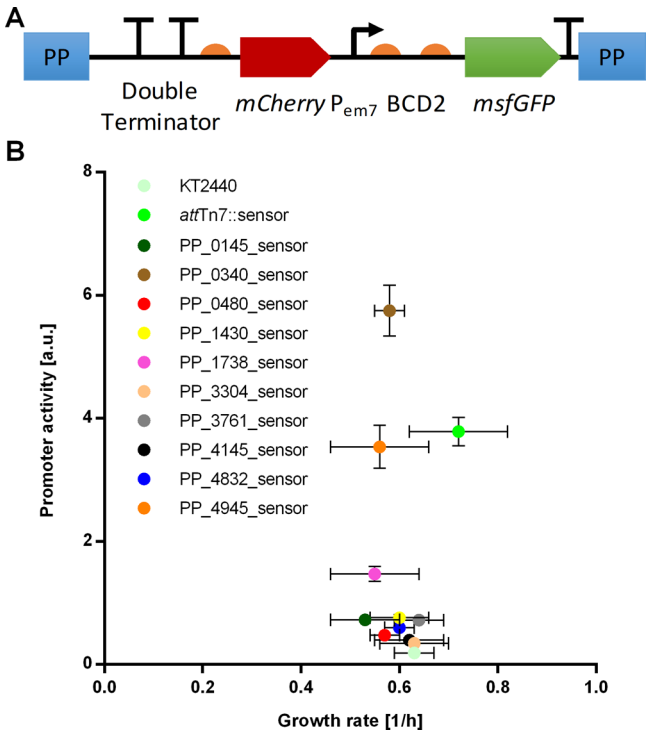


Figure 2. *P. putida* KT2440 carrying the fluorescence sensor in the different landing pads. (A) Graphical presentation of the integrated sensor. A double terminator (consisting of M13 central and rrnD T1) was used for insulation and promoterless *mCherry* as a read-through control. *msfGFP* is driven by the P_{em7}. (B) Determined promoter activities and growth rates for all tested landing pads. The wild type and an integrated sensor in *att*Tn7 were used as controls. Strains were cultured in a BioLector in a mineral salt medium with 20 mM glucose in a 96-well plate and were tested as biological triplicates with three technical replicates each. Error bars indicate the standard error of the mean ($n > 9$).

contained no promoter (BG), P_{em7} (BG13), or the strongest promoter P_{14g} (BG42) from the Zobel et al. collection (Figure 3) and were not protected from genomic read-through by insulation.[22] The P_{em7} and P_{14g} promoters are sigma70-dependent (σ70) and lead to a constitutive expression of the reporter gene during growth. A modular cloning procedure was used to minimize the number of oligonucleotides needed to clone these 13 integration vectors (Supporting Figure 2). We hypothesized that depending on the location of the landing pad in the genome, the genomic context affects the activity. Thus, a range of activities for different sites was expected.

P. putida KT2440 *att*Tn7::sensor contains the sensor sequence integrated into the *att*Tn7 site next to the *glmS* gene (PP_5409) using a mini-Tn7 transposon.[22] This strain is used as a reference for comparing the resulting activities from novel identified landing pads with the commonly used *att*Tn7 site. All strains with the same sensor construct (Figure 2A) integrated into different landing pads across the genome were analyzed for GFP expression, which revealed a strong variation

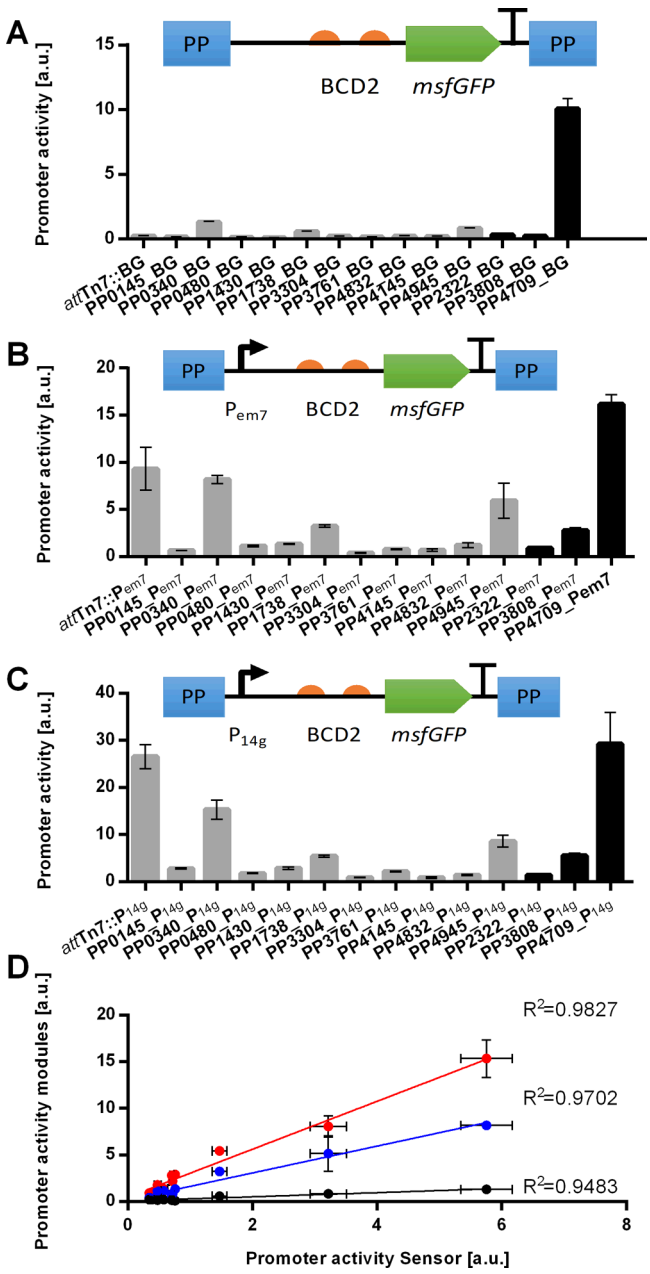


Figure 3. Characterized landing pads after genomic integration of expression cassettes in *P. putida* KT2440. Promoterless control BG (A), synthetic promoters P_{em7} (B), and P_{14g} (C) were genomically integrated into identified landing pads. The corresponding cassette was integrated with Tn7 into the *att*Tn7 site in *P. putida* KT2440 as controls. Black bars indicate landing pads chosen by high RPKM values. (D) Scatter plot of promoter activities. On the *x*-axis, the resulting activities of the sensor for each site are shown, while on the *y*-axis, determined activities for BG (black circles), P_{em7} (blue circles), and P_{14g} (red circles) are plotted. The regression line and corresponding coefficient of determination are shown for each group of promoters. Strains were cultured in a BioLector in mineral salt medium with 20 mM glucose in a 96-well plate and were tested in biological triplicates with three technical replicates each. Error bars indicate the standard error of the mean ($n > 9$).

in activity across the tested landing pads (Figure 2B). Only in landing pad PP_0340 did the sensor show a higher activity than *att*Tn7::sensor, while integration of the sensor construct downstream of PP_4945 showed an activity comparable to the

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

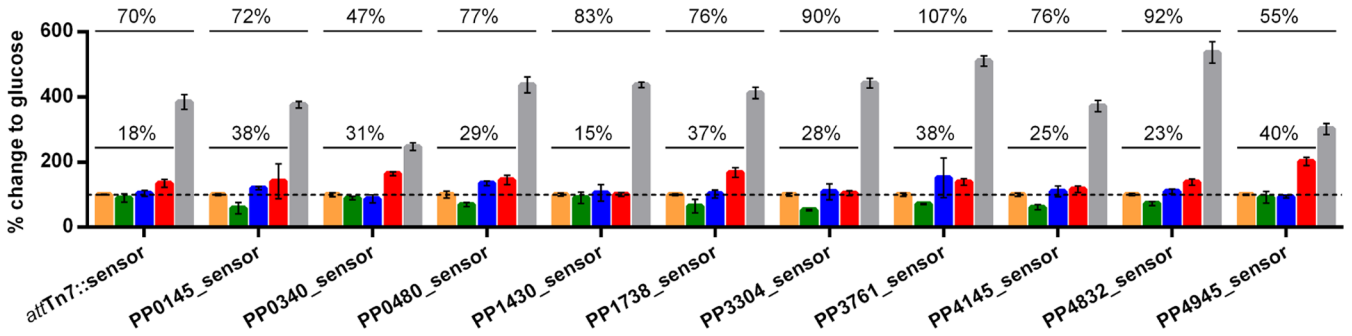


Figure 4. Characterization of all 13 landing pads in *P. putida* KT2440 during growth on alternative carbon sources. Promoter activity values are shown relative to glucose, which was set to 100% (dotted line). CVs were calculated for all carbon sources (upper values) and without glucose with urea (lower values). Strains were cultured in a BioLector using a 96-well plate and mineral salt medium supplemented with 20 mM glucose (orange bars), 20 mM citrate (green bars), 20 mM fructose (blue bars), 30 mM succinate (red bars), or 20 mM glucose with 0.8 M urea (gray bars). Identical strains from at least two different transformations were tested with three biological replicates. Error bars indicate the standard error of the mean ($n > 6$).

benchmark. The remaining landing pads reached less than 50% of the activity of the control *att*Tn7::sensor. The growth rates for each cultivation were determined to investigate the influence of the integrated sensor on the strain's fitness. While the WT reached a growth rate of 0.63 ± 0.04 h^{-1}, the constructed strains reached growth rates between 0.53 ± 0.07 h^{-1} ($p < 0.01$) for PP_0145_sensor and 0.69 ± 0.04 h^{-1} for *att*Tn7::sensor. Notably, the *mCherry* read-through control was negative for all landing pads. This showed that the terminator in front of the sensor is very strong and suitable to prevent genomic read-through into the construct, even from very highly expressed genes (Supporting Figure 3).

Next, all landing pads were also tested with the non-insulated expression cassettes (Figure 3). In these versions of the tested expression cassettes, expression levels were expected to be influenced by read-through activity from the genomic context of the landing pad on the GFP promoters. The strongest promoter sequence used was P_{14g}, which reaches around twice the activity of P_{em7}.[22] As a negative control, a promoterless version (BG) was used. Due to a missing promoter, no expression is expected from this construct. Should this nevertheless occur, this would be an indication of a read-through caused by false terminator determination. Again, the *att*Tn7 site was used as a reference containing the identical expression cassettes. However, the Tn7 constructs are insulated by a terminator located upstream of the expression cassettes and include a gentamicin resistance marker.[22]

The promoterless *msfGFP* expression served as a control for a genomic read-through and yielded relatively low activities in all landing pads and the *att*Tn7 site, indicating that the contribution of read-through is minor for most constructs (Figure 3A). *P. putida* KT2440 PP_0340_BG, PP_1738_BG, and PP_4945_BG reached higher activities compared to *att*Tn7::BG, possibly due to an incorrect terminator analysis of the intergenic sequence used for integration. Integration next to PP_0340 and PP_4945 has the same orientation as the upstream gene on the genome, which could explain the higher read-through activity if transcriptional termination is insufficient. In PP_1738 the upstream gene is orientated in the opposite direction, while the downstream gene is orientated in the same. Maybe the native promoter of PP_1738 is bidirectional and also controls the gene expression of PP_1739.

The expression results for the P_{em7}-containing constructs are in accordance with the results obtained with the promotorless constructs (Figure 3B). Again, PP_0340_P_{em7}, PP_1738_P_{em7}, and PP_4945_P_{em7} reached the highest activities compared to those of the other landing pads. However, *att*Tn7::P_{em7} reached the highest activity in this set, showing that the *att*Tn7 site is very suitable for high gene expression. The synthetic promoter P_{14g} was also used to characterize the landing pads due to its high activity, which is often used for heterologous gene expression.[35,44−46] Interestingly, the expression strength under the control of promoter P_{14g} was much lower in all tested landing pads than in *att*Tn7::P_{14g}. This likely stems from the fact that P_{14g} was obtained by a screening of promoters in the *att*Tn7 site where P_{14g} was selected as strongest promoter,[22] and this observation highlights again that promoter activities are strongly affected by their context.[23]

Overall, landing pads PP_0340, PP_1738, and PP_4945 achieved the highest activities for all of the tested constructs. The strategy presented here enabled us to identify multiple integration sites with a broad range of gene expression, allowing the fine-tuning of such expression while using the same constitutive promoter. Even more importantly, all landing pads showed reproducible activities during different experiments with individual clones derived from separate transformation events, demonstrating that the integration into these sites is very stable, and expression is reliable, one of the essential requirements in synthetic biology.

Integration sites chosen for their high RPKM mean values were tested similarly (Figure 3). While the tested sensor constructs in site PP_2322 and PP_3808 reached similar activities to *P. putida* KT2440 *att*Tn7::sensor, *P. putida* KT2440 PP_4709_sensor achieved a higher activity (Supporting Figure 8) and thus the highest activity found in this study for the sensor construct. Even when no promoter is present, as in strain *P. putida* KT2440 PP_4709_BG, the activity is comparable to that of *P. putida* KT2440 *att*Tn7::P_{em7} (Figure 3). The integration of promoter P_{14g} leads to a slightly higher expression than *P. putida* KT2440 *att*Tn7::P_{14g}. Here, a possible explanation is the incorrect determination of the upstream terminator, as discussed above. The obtained results do not completely match the expectations since all three landing pads were chosen based on their high RPKM, showing no direct correlation between the site expression strength of individual genes and their genomic regions. Indeed, higher

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

activity was expected but only observed using landing pad PP_4709. Including P_{em7} or P_{14g} in the uninsulated constructs leads to higher activities in each site compared to those in the promoterless control. For each construct, the identified landing pads PP_0340, PP_1738, and PP_4945 achieved the highest activities but none of them surpassed the activity of *att*Tn7, again highlighting the suitability of this site for high gene expression. *P. putida* KT2440 *att*Tn7::BG shows relatively low activity, likely due to the larger distance to an upstream promoter and the presence of an upstream terminator in the Tn7 construct.

To demonstrate the correlation between promoter activity and landing pad, the activities for BG, P_{em7}, and P_{14g} were plotted against those of the sensor for each site (Figure 3D). The determination of the coefficient of determination proves our finding with *R* squares between 0.9483 (BG) and 0.9827 (P_{14g}), showing that the genomic context mainly influences the expression strength of the integrated constructs. Promoter P_{em7} achieves around 50% of the activity of P_{14g}, which correlates with the slopes of the regression lines for each set of landing pads of 1.440 and 2.572, respectively.

All 13 landing pads were successfully characterized with different promoters, and a range in activity was observed, while the influence of the genomic context was shown. These landing pads were initially identified to enable a stable expression under varying conditions. Therefore, gene expression under different cultivation conditions was tested in the next experiment.

Characterization of the Influence of Different Carbon Sources on Expression Activity. Low expression variability across different experimental conditions was a fundamental selection criterion for the selection of suitable integration sites. Therefore, we investigated the activity of the selected landing pads, including sites chosen by a high RPKM value, with the integrated sensor construct on various carbon sources, *i.e.*, glucose, citrate, fructose, and succinate, to quantify the variability of the heterologous integrated constructs under changing conditions (Figure 4). Furthermore, urea was supplemented to glucose cultures to test a stress condition since Otto et al. have shown that adding urea greatly impacts growth rate and expression strength.[35] The carbon sources were used in Cmol equivalents to achieve comparable biomass at the end of the experiment. Graphs containing biomass, msfGFP, and mCherry fluorescence can be found in the supplement, as well as all determined growth rates and promoter activities (Supporting Figures 3−7, and Table 2).

Most integration sites show a similar expression strength on the different carbon sources tested. When the promoter activities achieved while cultivated with glucose and urea are not considered, the variability of most integration sites is low (Figure 4). Here again, the reference *att*Tn7 site with integrated mini-Tn7 constructs shows low variability with a CV of 18%, including glucose with urea at 70%. Weaker expression of sites PP_0145, PP_0480, PP_1430, PP_3304, PP_3761, PP_4145, and PP_4832 leads to CV ranging from 15 to 38%. The high expression sites PP_0340 and PP_4945 show low variability while cultivated in different carbon sources, with a CV of 31 and 40%, respectively.

In most of the tested conditions, GFP expression was lower compared to that of the glucose control, although the differences were minor. The *att*Tn7 site, PP_0340, PP_1430, and PP_4945, accomplished the most stable phenotype while cultivated with citrate and fructose, differing only by 13% compared to glucose. Although growth rates obtained on fructose were much lower than those on glucose, activities determined during growth on fructose are comparable to those on glucose. Only PP_0480 and PP_3761 show activity increases of 50 and 34%, respectively. On fructose, a prolonged lag phase of the cultures was observed. Therefore, the fluorescence signal increased much more slowly, influencing the calculation of promoter activities, which is dependent on the fluorescence signal and optical densities during the early exponential phase.

Higher growth rates were achieved during cultivation on succinate, where the fluorescence signal increased earlier than that on glucose. Increased growth rate leads to higher activities of the tested construct due to the use of sigma70-dependent (σ70) promoter. The highest gain of activity was observed for PP_0340 (77%), PP_1738 (68%), and PP_4945 with 109%. Citrate as the carbon source leads to lower promoter activities than glucose. Only *att*Tn7, PP_0340, PP_1430, and PP_4945 reached comparable activities on both substrates.

Growth of the strains with glucose and urea leads to much higher activities of up to 430% compared with glucose alone. Furthermore, the lag phase was much longer than on glucose alone, and the slope was very shallow, especially in the beginning (Suppoting Figure 7). For the cultivation with glucose and urea, the highest optical densities were observed at the end of the exponential phase but were still lower than the final optical density achieved with glucose. Interestingly, the optical density drops to a final value of around 3 while remaining in the glucose cultivations (Supporting Figure 3). Otto et al. postulated that adding urea could influence the morphology of the cells.[35] However, smaller cells would lead to lower measured optical densities, which could be explained by the calibration used to convert scattered light to optical density obtained with glucose precultures. Despite the lower growth rate, the fluorescence signals were increased for all strains, indicating that sigma70 factors were available in the cells, and protein biosynthesis is highly active. Protein stability and activity are affected in the presence of urea, which could be compensated by an overall higher protein expression and probably a higher glucose consumption to facilitate growth.[47] This circumstance is the most likely reason for the strong increased activities, which are, in the end, influenced by the high fluorescence and the low OD. We calculated the CV for each site and carbon source from the resulting promoter activities to characterize and compare the identified landing pads. As discussed above, the presence of urea in the medium highly affected the promoter activity. Therefore, calculations for all tested conditions show a very high CV of 47 to 107% (Figure 4). To obtain a conclusion about the influence of the used carbon sources, the CV was also analyzed without glucose plus urea cultivation (Figure 4). The resulting CVs range from 15% for PP_1430 to 40% for PP_4945, clearly showing that the variation is small. Depending on the CVs, three groups can be described. The first group showed the lowest deviation for PP_1430 (15%), *att*Tn7 (18%), PP_4832 (23%), and PP_4145 (25%). The second group contains the landing pads with a moderate deviation, including PP_3304 (28%), PP_0480 (29%), and PP_0340 (31%). PP_1738 (37%), PP_0145 (38%), PP_3761 (38%), and PP_4945 (40%) form the third group with the highest determined CVs. Sites chosen by high RPKM also achieved stable CVs, especially PP_3808 with 12% and PP_4709 with 24%. PP_2322 resulted in a CV

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

of 45% and did not show a stable expression under changing conditions.

These findings confirmed our initial hypothesis that RNA-Seq data from different experimental conditions could be used to identify low-variability integration sites across the genome of *P. putida* KT2440. However, sites chosen with high RPKM values have shown promising results in our trials and are suitable for integration. The influence of the genomic context is comparable to *att*Tn7 on promoter activities. However, cultivation under urea stress leads to much higher activities at all integration sites, indicating the limit of the approach. In the future, RNA-Seq analysis of different stress conditions might be incorporated into the underlying data set used for the insertion site selection to counteract this effect.

Characterization of Integration Sites in *P. taiwanensis* VLB120. Synthetic biology aims to develop transferable modules and tools. Therefore, selected landing pads identified for *P. putida* KT2440 were also characterized in *P. taiwanensis* VLB120 as a proof of principle to demonstrate their cross-species applicability for heterologous gene expression.

We identified equivalent landing pads for all genomic loci (Supporting Table 3 and Figure 9). Integration plasmids were constructed for *P. taiwanensis* VLB120 for the four landing pads, showing the highest expression in *P. putida* KT2440. The expression constructs remained identical to the uninsulated P_{em7} construct and consisted of synthetic promoter P_{em7}, BCD2, *msf GFP*, and a terminator. The expression constructs were integrated into the genome of *P. taiwanensis* GRC1, which is a genome-reduced chassis of *P. taiwanensis* VLB120.[29] As a control, pBG13 was integrated into *att*Tn7 of *P. taiwanensis* GRC1, resulting in strain *P. taiwanensis* GRC1 *att*Tn7::P_{em7}. The equivalent *P. putida* KT2440 variants were tested as controls. The landing pad PP_0340 from *P. putida* KT2440 corresponds to the genomic locus PVLB_23545 in *P. taiwanensis* VLB120. Site PP_4945 corresponds to PVLB_02480, PP_1738 to PVLB_06360, and PP_1430 to PVLB_05170.

The *att*Tn7::P_{em7} site showed the highest fluorescence signal in both strains, although it was lower in *P. taiwanensis* GRC1 than in *P. putida* KT2440 (Figure 5). PP_0340_P_{em7} and the equivalent PVLB_23545_P_{em7} reached very similar expression levels, like PP_1738_P_{em7} and the PVLB_06360_P_{em7}. PP_4945_P_{em7} is more active than the corresponding site in *P. taiwanensis* GRC1 PVLB_02480_P_{em7}. Inverted observations can be made for PP_1430_P_{em7} and PVLB_05170_P_{em7}. Here, the expression level is higher in *P. taiwanensis*.

The expression levels achieved are not the same in terms of deviation compared with the Tn7 integrated constructs in *P. taiwanensis* VLB120 and *P. putida* KT2440, but the landing pads are suitable in both organisms, even though the ranking cannot necessarily be directly transferred from one species to another. In conclusion, the findings show that the identified landing pads can be transferred to other species of *Pseudomonas* to enable the expression of heterologous genes integrated into the genome.

Multiple Integrations Do Not Affect Each Other's Gene Expression. Often, multiple heterologous genes have to be expressed simultaneously or multiple copies of the same gene are needed to achieve very high expression levels. Here, the use of plasmids can be advantageous due to higher gene dosages, but this also entails the drawbacks discussed above.[11,13] Having a set of genome integration vectors in hand, we integrated the cassette for *msf GFP* at two different genomic loci in one strain to investigate possible interactive effects on promoter activity or gene expression. The measured GFP expression in these combined constructs was compared to the sum of the corresponding single integrations (Figure 6). This yielded only minor differences under 6%, showing that the expression from the two sites is directly cumulative. Likely, the relatively large genomic distance between constructs prevents interference. This distance between two expression cassettes reduces the competition for translational and transcriptional resources, as demonstrated by Qian et al.[48] Therefore, multiple integrations in *P. putida* KT2440 are feasible using the here presented landing pads, with neither detectable effects on heterologous gene expression nor the growth rate of engineered strains.

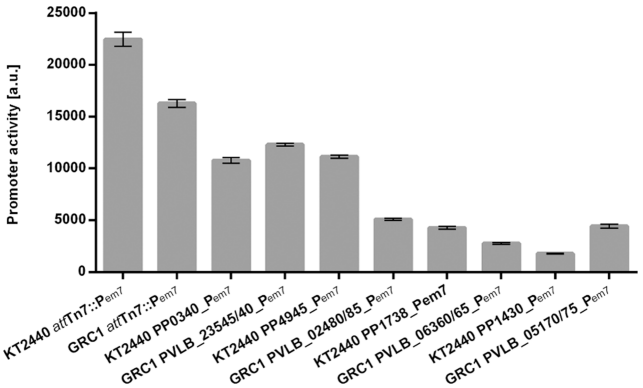


Figure 5. Characterization of homologous landing pads in *P. taiwanensis* GRC1 and equivalent sites in *P. putida* KT2440. The promoter activity, calculated from GFP fluorescence over an optical density, is shown. The measurements used here differ from other determined promoter activities in this study since they were measured on a different device. Strains were cultured in 24-well System Duetz plates in a mineral salt medium with 20 mM glucose. Strains from one transformation were tested as biological triplicates. Error bars indicate the standard error of the mean (n = 3).

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Cultivation Conditions. Strains and plasmids are given in Supporting Table 4. Chemically competent *E. coli* PIR2 cells (Life Technologies, Carlsbad) were used for cloning procedures.[49] *E. coli* strains were grown in lysogeny broth (LB) containing 10 g L^{-1} peptone, 5 g L^{-1} yeast extract, and 5 g L^{-1} sodium chloride.[50] For solid media, 15 g of L^{-1} agar was added to the medium before autoclaving. To maintain mini-Tn7[22] vectors or pEMG[14] plasmid stability in *E. coli* PIR2 cells, 50 mg L^{-1} kanamycin was added to solid or liquid medium. For pRK600[51]-bearing strains, 10 mg L^{-1} chloramphenicol, and for pTNS1,[21] 100 mg L^{-1} ampicillin was added to the medium. pSW-2[14] maintenance was achieved by the addition of 30 mg L^{-1} gentamicin. *E. coli* strains were cultivated at 37 °C. *P. putida* KT2440 and *P. taiwanensis* VLB120 were cultivated at 30 °C. Cultivations of *P. putida* KT2440 and *P. taiwanensis* VLB120 derivatives were performed with mineral salt medium containing 3.88 g L^{-1} K_2HPO_4, 1.63 g L^{-1} NaH_2PO_4, 2 g L^{-1} $(NH_4)_2SO_4$, 0.1 g L^{-1} $MgCl_2{\cdot}6H_2O$, 0.01 g L^{-1} EDTA, 0.002 g L^{-1} $ZnSO_4{\cdot}7H_2O$, 0.001 g L^{-1} $CaCl_2{\cdot}2H_2O$, 0.005 g L^{-1} $FeSO_4{\cdot}7H_2O$, 0.0002 g L^{-1} $Na_2MoO_4{\cdot}2H_2O$, 0.0002 g L^{-1} $CuSO_4{\cdot}5H_2O$, 0.0004 g L^{-1} $CoCl_2{\cdot}6H_2O$, 0.001 g L^{-1} $MnCl_2{\cdot}2H_2O$, and a carbon source.[52] As carbon source 20 mM

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

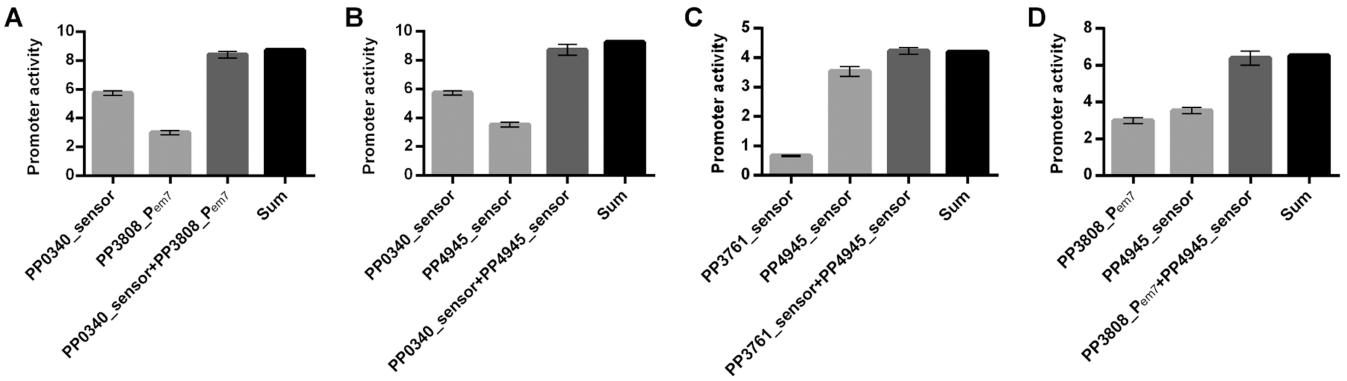


Figure 6. Double genome integration in *P. putida* KT2440. pEMG-based vectors containing the sensor construct or msfGFP driven by P_{em7} were tested. Shown are the activities for corresponding single integrations (light gray bars), double integrations (dark gray bars), and the sum of both single activities (black bars). Results for integration of (A) PP_0340_sensor and PP_3808_P_{em7}, (B) PP_0340_sensor and PP_4945_sensor, (C) PP_3761_sensor and PP_4945_sensor, (D) PP_3808_BG13 and PP_4945_sensor are shown. Strains were cultured in a BioLector in mineral salt medium with 20 mM glucose in a 96-well plate. Identical strains from at least two different transformations were tested with three biological replicates. Error bars indicate the standard error of the mean ($n > 6$).

glucose, 20 mM fructose, 20 mM citrate, 30 mM succinate, or 20 mM glucose with 0.8 M urea were used.[35]

DNA Techniques. All oligonucleotides used in this study are listed in Supporting Table 5 and were synthesized by Eurofins Genomics (Ebersberg, Germany). PCRs for the construction of plasmids were performed with Q5 polymerase (New England Biolabs, Ipswich, Massachusetts), and colony PCRs were performed with lysed cells (60% PEG200, pH 13)[22] and *Taq* 2X Master Mix (New England Biolabs). The isolation of PCR products from reactions or agarose gels and plasmids was performed with respective kits from New England Biolabs. Correctness of plasmid sequences was performed by Sanger sequencing (Eurofins Genomics).

The sensor construct was constructed in several steps using restriction and ligation (Supporting Figure 10). The sensor was first built in pBELK,[18] and the kanamycin resistant marker was inverted (pBELK_invKan). The resistance gene was amplified from pBELK as a template with oligonucleotides SK13 and SK14. The resulting fragments were equipped with *Hind*III and *Aat*II restriction sites and ligated with digested pBELK. Orientation of the resistance gene was confirmed by colony PCR using oligonucleotides BW62 and SK24. The promoter, BCD2 element, and *msf GFP* fluorescence gene were amplified with oligonucleotides SK15 and SK17 from plasmid pBG13. The resulting fragments were equipped with 5′-end *Pac*I and *Nde*I restriction sites and on the 3′-end a *Hind*III. The fragment and pBELK_invKan were digested with *Pac*I/*Hind*III and ligated. The resulting plasmid was called pBELK_-P_{em7}_BCD2_msfGFP_invKan. An additional replication of origin ColE1 for *E. coli* was introduced *via* the *Hind*III restriction site; for origin amplification, pSEVA247Y and oligonucleotides SK20 and SK21 were used. Integration was checked by colony PCR with oligonucleotides SK23 and SK24. The resulting plasmid was called pBELK_-P_{em7}_BCD2_msfGFP_ColE1_invKan. The genomic read-through was prohibited by introducing two strong double terminators from phage M13 and *E. coli* K12 *rrnD1*.[42] The sequences were fused in front of an RBS site and *mCherry* gene and synthesized by Integrated DNA Technologies (Coralville, Iowa). The fragment was equipped with *Pac*I and *Nde*I restriction sites used for the ligation with digested pBELK_P_{em7}_BCD2_msfGFP_ColE1_invKan. The modified Tn5 bearing pBELK was named pTn5_sensor.

The construction of integrable pEMG vectors for the characterization of identified genomic regions in *P. putida* KT2440 was done by Gibson assembly (Supporting Figure 2). Target sequences (TS) for homologous recombination were amplified from isolated genomic DNA of *P. putida* KT2440, and used oligonucleotide combinations are given in Supporting Table 6. The above-described pTn5_sensor was used as a template for amplifying the sensor sequence with oligonucleotides SK190 and SK191. An additional terminator at the 3′-end was amplified from pBG13 with oligonucleotides SK192 and SK193. I-*Sce*I recognition sites-containing plasmid pEMG[14] was digested with *Eco*RI and *Sal*I. NEBuilder Hifi DNA Assembly Master Mix (New England Biolabs) was used for the Gibson assembly reaction. Following the manual, appropriate amounts of digested plasmid, TS elements, the sensor construct, and additional terminator were mixed and assembled. The whole assembly mix was transformed into chemically competent[49] *E. coli* PIR2 (Life Technologies, Carlsbad) and spread on LB agar plates containing 50 mg L^{-1} kanamycin. Arising clones were checked by colony PCR using *Taq* 2X Master Mix (New England BioLabs) and oligonucleotides BW13 and BW14. Correct clones were used for genomic integration by triparental mating, as described above. Multiple integrations were done successively.

Exchange of the sensor against other promoter constructs in pEMG_PPXX_sensor plasmids (XX stands for different genomic regions used for integration) was done by restriction and ligation. Promoterless controls BG_BCD2_*msf GFP*, P_{em7}_BCD2_*msf GFP*, and P_{14g}_BCD2_*msf GFP* were amplified from the corresponding mini-Tn7 plasmids pBG, pBG13, and pBG14g using oligonucleotides SK43 and BW62. pEMG_PPXX_sensor vectors and PCR fragments were digested with *Pac*I and *Spe*I and ligated afterward using T4 ligase (New England Biolabs). Growing clones were tested by colony PCR using oligonucleotides BW13 and BW14. Correct plasmids were sequenced and genomically integrated into *P. putida* KT2440 by mating as described by Wynands et al.[5]

The construction of suitable integration vectors for *P. taiwanensis* VLB120 was performed by Gibson assembly. Identified integration sites for *P. putida* KT2440 were compared with the genome of *P. taiwanensis* VLB120, and appropriate sequences were used. The pEMG backbone was amplified with oligonucleotides TL075 and TL076 from

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

pEMG_PP0340_P_{em7}. Using different oligonucleotide combinations, P_{em7}_BCD2_*msfGFP* was amplified from pEMG_PP0340_P_{em7} (Supporting Table 4). The corresponding TS elements were amplified from isolated genomic DNA of *P. taiwanensis* VLB120 using suitable oligonucleotide combinations (Supporting Table 6). All PCRs were performed with Q5 polymerase (New England Biolabs, Ipswich, Massachusetts) and assembled using the NEBuilder HiFi DNA Assembly 2X Master Mix (NEB).

Genomic Integration Procedures. Integration of mini-Tn7 or pEMG vectors was performed by patch mating on LB agar plates and cultivation overnight at 30 °C. pEMG integration was performed with a pEMG bearing donor strain, acceptor strain *P. putida* KT2440 or *P. taiwanensis* VLB120, and helper strain *E. coli* HB101 pRK600. For the integration of mini-Tn7 plasmid, an additional helper strain, *E. coli* DH5$\alpha\lambda$pir pTNS1,[21] which provides the transposase, was required. For selection and counter-selection of successfully integrated mini-Tn7 bearing *P. putida* KT2440 strains, cetrimide agar plates containing 30 mg L^{-1} gentamicin and 1% glycerol were used. For the selection of *P. taiwanensis* VLB120 LB agar plates containing irgasan (25 mg L^{-1}) and gentamicin (30 mg L^{-1}) were used. The pEMG bearing *P. putida* KT2440 was selected with cetrimide agar plates containing 50 mg L^{-1} kanamycin and 1% glycerol. *P. taiwanensis* VLB120 was selected with LB agar plates containing irgasan (25 mg L^{-1}) and kanamycin (50 mg L^{-1}). Co-integrate release of the pEMG backbone was performed by an additional patch mating consisting of the corresponding *P. putida* KT2440 strain, *E. coli* HB101 pRK600, and *E. coli* DH5 pSW-2 on LB agar plates. Matings were cultivated overnight at 30 °C. Selection for pSW-2 bearing *P. putida* KT2440 strains was done with cetrimide agar plates containing 30 mg L^{-1} gentamicin and 1% glycerol. The proper co-integrate release was verified by transferring single colonies on LB agar plates and cetrimide agar plates containing 50 mg L^{-1} kanamycin and 1% glycerol. Clones that were not able to grow on kanamycin-containing agar plates were subsequently checked by colony PCR to examine the genomic composition.

Measuring Fluorescence and Determination of Promoter Activity. A BioLector (M2P Laboratories, Baesweiler, Germany) was used to measure growth, mCherry, and msfGFP fluorescence of integrated constructs in *P. putida* KT2440 derivatives. Precultures were cultivated overnight at 30 °C and 300 rpm in 24-well System Duetz plates (Enzyscreen, Heemstede, The Netherlands). 1.5 mL of mineral salt medium with 20 mM glucose was used, and optical densities at 600 nm were determined for each strain. 200 μL mineral salt medium with 20 mM glucose was inoculated with an initial optical density at 600 nm of 0.1 and filled in clear bottom 96-well plates (Greiner Bio-One, Kremsmünster, Austria). Cultivation conditions of BioLector experiments were set to 30 °C, 900 rpm, and humidity control of 85%. Three internal filter modules were used for online measurement. msfGFP fluorescence was measured at an excitation wavelength of 488 nm and an emission wavelength of 520 nm. mCherry was measured with an excitation wavelength of 560 nm and an emission wavelength of 610 nm. Biomass was measured at 620 nm with gain 40 as scattered light. A dilution series of stationary phase cultures was used to correlate scattered light to optical density at 620 nm. Promoter activities derived from different lading pads and genetic constructs were determined using Microsoft Excel by calculating the slope of msfGFP fluorescence over optical density during the exponential phase.[22]

The characterization of landing pads in *P. taiwanensis* GRC1 was done in 24-well System Duetz plates and end point measurements. Biomass was measured with an Ultrospec 10 photometer (Biochrom), and fluorescence measurements were performed with an Infinite M1000 PRO (Tecan Austria GmbH) in black bottom 96-well plates (Greiner bio-one). msfGFP fluorescence was measured at an excitation wavelength of 488 nm and an emission wavelength of 510 nm. Precultures were grown in 25 mL of mineral salt medium with 20 mM glucose in 250 mL shake flasks at 30 °C and 200 rpm overnight. The same medium was used for the main cultures. 1.5 mL of medium was used per well in 24-well System Duetz plates. Cultivation was performed overnight at 30 °C and 300 rpm.

Identification of Suitable Integration Sites Using RNA-Seq Data. The identification of suitable integration sites in the *P. putida* KT2440[53] genome was done by analyzing three available sets of RNA-Seq data (Supporting Table 1). All data sets were merged into one data set, missing genes were identified, and values were deleted, to avoid bias of the mean or standard deviation. Furthermore, the start and end of the gene and the genome orientation were added as additional information.[34] Annotation of the merged data was carried out using the genome from *P. putida* KT2440 by Nelson et al.[34] All seven 16sRNA genes were removed from the data set; due to their highly conserved sequences, a false mapping of determined transcripts is highly likely.[41,54]

For each gene, the mean transcription level under the conditions used and the corresponding standard deviation were calculated. Dividing the standard deviation by the mean leads to the coefficient of variation (CV), which reflects the differential expression of one gene for all data sets, *i.e.*, for all different cultivation conditions considered. High CV values represent a high transcriptional change; low values indicate genes expressed comparably under all conditions. The CV of five adjoining genes was determined and compared with the next CV of five genes (CV1 gene 1 to 5, CV2 gene 2−6, and so on, Figure 1C). Starting with a degree of freedom of zero wt %, we found no neighboring genes with the same expression pattern. Increasing the degree of freedom to 10% resulted in genes with the same expression pattern as the neighboring gene. An increasing degree of freedom of up to 40% leads to more candidate genes that are similarly or cotranscribed located next to each other on the genome. We have chosen 10 clustered gene regions from this set (Table 1). To minimize interference with natural regulation, intergenic regions inside the chosen genome regions were used to integrate with a distance of 130 to 250 bp. The intergenic regions were identified using whole-genome information[34] and the *Pseudomonas* Genome DB.[37] To even further minimize disruption of natural regulatory mechanisms, possible integration sites were analyzed for promoter sequences (BPROM,[38] Neural Network Promoter Prediction[39]) and terminators (ARNold[40]).

Statistics. Each landing pad was characterized by 2−3 independent transformations performed on different days. Three clones from each transformant were tested in a Biolector to determine promoter activities, yielding a total of a minimum of nine biological replicates. The mean and standard error for each construct were calculated from these combined biological replicates. The significance of the difference in the activity of

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072

constructs and growth rates was analyzed by one-way ANOVA with Turkey's post hoc comparison.

CONCLUSIONS

RNA-Seq data of *P. putida* KT2440 grown under different conditions was used to identify promising genomic regions that could be exploited for reliable integration sites. A two-dimensional assessment of genomic context under different growth conditions led to the selection of 13 landing pads that were characterized in detail and can now be used to express multiple heterologous constructs. The expression of heterologous gene constructs in these sites did not rely on upstream promoters and is reliable across different cultivation conditions. Integration of the same insulated construct with its constitutive promoter at different genomic sites caused drastic differences in reporter gene expression, confirming the presence of genomic "hot" and "cold" spots that affect expression based on the genomic locus rather than promoter activity. Using uninsulated constructs enabled a wide range of activity, thereby allowing for the tuning of gene expression by a combination of promoters and integration sites. Notably, very high gene expression from selected sites is also possible. Characterization of double integration strains revealed no mutual influences. Four selected landing pads were successfully transferred to *P. taiwanensis*, demonstrating the tool's usefulness for the engineering of Pseudomonads. Besides this, the two-dimensional RNA-Seq data analysis strategy is generically applicable to all microorganisms. In summary, we present an additional tool to the ever-growing toolbox for the deep engineering of a multitude of traits, hereby enabling multiple genomic integrations in a reliable and tunable fashion.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.3c00747.

Information on the used RNA-Seq data sets; graphical presentation of RPKM values for integration sites and neighbored genes; construction of the sensor construct and integrable vectors; determined growth rates and corresponding graphs including biomass, GFP, and mCherry fluorescence using different carbon sources; list of identified landing pads in *P. taiwanensis* VLB120; a complete list of generated vectors and strains; and full list of oligonucleotides and combinations used for amplification (PDF)

AUTHOR INFORMATION

Corresponding Author

Sebastian Köbbing − *Aachen Biology and Biotechnology-ABBt, Institute of Applied Microbiology-iAMB, RWTH Aachen University, 52074 Aachen, Germany;* orcid.org/0000-0002-4606-4961; Email: sebastian.koebbing@rwth-aachen.de

Authors

Thorsten Lechtenberg − *Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich, 52428 Jülich, Germany;* orcid.org/0000-0001-5207-2143

Benedikt Wynands − *Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich, 52428 Jülich, Germany;* orcid.org/0000-0001-8599-3205

Lars M. Blank − *Aachen Biology and Biotechnology-ABBt, Institute of Applied Microbiology-iAMB, RWTH Aachen University, 52074 Aachen, Germany;* orcid.org/0000-0003-0961-4976

Nick Wierckx − *Aachen Biology and Biotechnology-ABBt, Institute of Applied Microbiology-iAMB, RWTH Aachen University, 52074 Aachen, Germany; Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich, 52428 Jülich, Germany;* orcid.org/0000-0002-1590-1210

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.3c00747

Author Contributions

N.W. and S.K. designed the experiments. S.K. identified the landing pads, performed all of the molecular engineering and experiments with *P. putida* KT2440, prepared all figures, and wrote the manuscript. T.L. and B.W. identified the equivalent genomic sites in *P. taiwanensis* VLB120 and conducted the molecular engineering and characterization with *P. taiwanensis* VLB120. N.W. supervised the study, edited the manuscript, and helped acquire funding. L.M.B. discussed data, edited the manuscript, and acquired funding.

Funding

The authors gratefully acknowledge the funding of SK from the European Union Horizon 2020 research and innovation program under grant agreement no. 633962 for the project P4SB. T.L. was supported by the German Federal Ministry of Education and Research *via* the project NO-STRESS [grant number 031B0852A]. N.W. was supported by the German Research Foundation (DFG) through the Emmy Noether program (WI 4255/1−1). The laboratory of L.M.B. is partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the Cluster of Excellence FSC 2186 "The Fuel Science Center" (ID: 390919832).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Anka Sieberichs for helping with characterizing the landing pads in *P. taiwanensis* VLB120, the members of the Institute of Applied Microbiology and the P4SB consortium for fruitful discussions and critical feedback, and Melanie Filbig for her incessant motivating conversations and critical manuscript review.

REFERENCES

(1) Ankenbauer, A.; Schäfer, R. A.; Viegas, S. C.; Pobre, V.; Voß, B.; Arraiano, C. M.; Takors, R. *Pseudomonas putida* KT2440 is naturally endowed to withstand industrial-scale stress conditions. *Microb. Biotechnol.* **2020**, *13* (4), 1145−1161.

(2) Nikel, P. I.; de Lorenzo, V. *Pseudomonas putida* as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. *Metab. Eng.* **2018**, *50*, 142−155.

(3) Jiménez, J. I.; Miñambres, B.; García, J. L.; Díaz, E. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol.* **2002**, *4*, 824−841, DOI: 10.1046/j.1462-2920.2002.00370.x.

(4) Meijnen, J. P.; de Winde, J. H.; Ruijssenaars, H. J. Sustainable production of fine chemicals by the solvent-tolerant *Pseudomonas putida* S12 using lignocellulosic feedstock. *Int. Sugar J.* **2011**, *113*, 24−30.

(5) Wynands, B.; Lenzen, C.; Otto, M.; Koch, F.; Blank, L. M.; Wierckx, N. Metabolic engineering of *Pseudomonas taiwanensis* VLB120 with minimal genomic modifications for high-yield phenol production. *Metab. Eng.* **2018**, *47*, 121−133.
(6) Calero, P.; Nikel, P. I. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* **2019**, *12* (1), 98−124.
(7) Martin-Pascual, M.; Batianis, C.; Bruinsma, L.; Asin-Garcia, E.; Garcia-Morales, L.; Weusthuis, R. A.; van Kranenburg, R.; dos Santos, V. A. P. M. A navigation guide of synthetic biology tools for *Pseudomonas putida*. *Biotechnol. Adv.* **2021**, *49*, No. 107732, DOI: 10.1016/j.biotechadv.2021.107732.
(8) Martínez-García, E.; Aparicio, T.; Goñi-Moreno, A.; Fraile, S.; de Lorenzo, V. SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* **2015**, *43*, D1183−D1189, DOI: 10.1093/nar/gku1114.
(9) Martínez-García, E.; de Lorenzo, V. Molecular tools and emerging strategies for deep genetic/genomic refactoring of *Pseudomonas*. *Curr. Opin. Biotechnol.* **2017**, *47*, 120−132.
(10) Panke, S.; de Lorenzo, V.; Kaiser, A.; Witholt, B.; Wubbolts, M. G. Engineering of a stable whole-cell biocatalyst capable of (S)-styrene oxide formation for continuous two-liquid-phase applications. *Appl. Environ. Microb.* **1999**, *65* (12), 5619−5623.
(11) Lindmeyer, M.; Jahn, M.; Vorpahl, C.; Müller, S.; Schmid, A.; Bühler, B. Variability in subpopulation formation propagates into biocatalytic variability of engineered *Pseudomonas putida* strains. *Front. Microbiol.* **2015**, *6*, No. 1042, DOI: 10.3389/fmicb.2015.01042.
(12) Friehs, K. Plasmid Copy Number and Plasmid Stability. In *New Trends and Developments in Biochemical Engineering*, Advances in Biochemical Engineering; Springer, 2004; Vol. *86*, pp 47−82.
(13) Millan, A. S.; MacLean, R. C.Fitness costs of plasmids: a limit to plasmid transmission *Microbiol. Spectr.* 2017; Vol. *5* 5 DOI: 10.1128/microbiolspec.MTBP-0016-2017.
(14) Martínez-García, E.; de Lorenzo, V. Engineering multiple genomic deletions in gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. *Environ. Microbiol.* **2011**, *13* (10), 2702−2716.
(15) Reznikoff, W. S. Transposon Tn 5. *Annual. Rev. Genet.* **2008**, *42*, 269−286.
(16) Wetmore, K. M.; Price, M. N.; Waters, R. J.; Lamson, J. S.; He, J.; Hoover, C. A.; Blow, M. J.; Bristow, J.; Butland, G.; Arkin, A. P.; Deutschbauer, A. Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *mBio* **2015**, *6* (3), No. e0030615, DOI: 10.1128/mBio.00306-15.
(17) van Opijnen, T.; Lazinski, D. W.; Camilli, A. Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. *Curr. Protoc. Mol. Biol.* **2014**, *106*, 7−16, DOI: 10.1002/0471142727.mb0716s106.
(18) Nikel, P. I.; de Lorenzo, V. Implantation of unmarked regulatory and metabolic modules in gram-negative bacteria with specialised mini-transposon delivery vectors. *J. Biotechnol.* **2013**, *163* (2), 143−154.
(19) Domröse, A.; Klein, A.; Hage-Hülsmann, J.; Thies, S.; Svensson, V.; Classen, T.; Pietruszka, J.; Jaeger, K.-E.; Drepper, T.; Loeschcke, A. Efficient recombinant production of prodigiosin in *Pseudomonas putida*. *Front. Microbiol.* **2015**, *6*, No. 972, DOI: 10.3389/fmicb.2015.00972.
(20) Domröse, A.; Hage-Hülsmann, J.; Thies, S.; Weihmann, R.; Kruse, L.; Otto, M.; Wierckx, N.; Jaeger, K.-E.; Drepper, T.; Loeschcke, A. *Pseudomonas putida* rDNA is a favored site for the expression of biosynthetic genes. *Sci. Rep.* **2019**, *9* (1), No. 7028.
(21) Choi, K. H.; Gaynor, J. B.; White, K. G.; Lopez, C.; Bosio, C. M.; Karkhoff-Schweizer, R. R.; Schweizer, H. P. A Tn7-based broad-range bacterial cloning and expression system. *Nat. Methods* **2005**, *2*, 443−448.
(22) Zobel, S.; Benedetti, I.; Eisenbach, L.; de Lorenzo, V.; Wierckx, N.; Blank, L. M. Tn7-Based device for calibrated heterologous gene expression in *Pseudomonas putida*. *ACS Synth. Biol.* **2015**, *4* (12), 1341−1351.
(23) Köbbing, S.; Blank, L. M.; Wierckx, N. Characterization of Context-Dependent Effects on Synthetic Promoters. *Front. Bioeng. Biotechnol.* **2020**, *8*, No. 551, DOI: 10.3389/fbioe.2020.00551.
(24) Elmore, J. R.; Furches, A.; Wolff, G. N.; Gorday, K.; Guss, A. M. Development of a high efficiency integration system and promoter library for rapid modification of *Pseudomonas putida* KT2440. *Metab. Eng. Commun.* **2017**, *5*, 1−8, DOI: 10.1016/j.meteno.2017.04.001.
(25) Elmore, J. R.; Dexter, G. N.; Baldino, H.; Huenemann, J. D.; Francis, R.; Peabody, G. L.; Martinez-Baird, J.; Riley, L. A.; Simmons, T.; Coleman-Derr, D.; Guss, A. M.; Egbert, R. G. High-throughput genetic engineering of nonmodel and undomesticated bacteria via iterative site-specific genome integration. *Sci. Adv.* **2023**, *9* (10), No. eade1285.
(26) Kuhlman, T. E.; Cox, E. C., Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* **2010**, *38* DOI: 10.1093/nar/gkp1193.
(27) Englaender, J. A.; Jones, J. A.; Cress, B. F.; Kuhlman, T. E.; Linhardt, R. J.; Koffas, M. A. G. Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synth. Biol.* **2017**, *6* (4), 710−720.
(28) Chaves, J. E.; Wilton, R.; Gao, Y.; Munoz, N. M.; Burnet, M. C.; Schmitz, Z.; Rowan, J.; Burdick, L. H.; Elmore, J.; Guss, A.; Close, D.; Magnuson, J. K.; Burnum-Johnson, K. E.; Michener, J. K. Evaluation of chromosomal insertion loci in the *Pseudomonas putida* KT2440 genome for predictable biosystems design. *Metab. Eng. Commun.* **2020**, *11*, No. e00139.
(29) Wynands, B.; Otto, M.; Runge, N.; Preckel, S.; Polen, T.; Blank, L. M.; Wierckx, Nc. Streamlined *Pseudomonas taiwanensis* VLB120 chassis strains with improved bioprocess features. *ACS Synth Biol.* **2019**, *8* (9), 2036−2050.
(30) D'Arrigo, I.; Bojanovic, K.; Yang, X.; Rau, M. H.; Long, K. S. Genome-wide mapping of transcription start sites yields novel insights into the primary transcriptome of *Pseudomonas putida*. *Environ. Microbiol.* **2016**, *18* (10), 3466−3481, DOI: 10.1111/1462-2920.13326.
(31) Frank, S.; Klockgether, J.; Hagendorf, P.; Geffers, R.; Schöck, U.; Pohl, T.; Davenport, C. F.; Tümmler, B. *Pseudomonas putida* KT2440 genome update by cDNA sequencing and microarray transcriptomics. *Environ. Microbiol.* **2011**, *13* (5), 1309−1326.
(32) La Rosa, R.; Nogales, J.; Rojo, F. The Crc/CrcZ-CrcY global regulatory system helps the integration of gluconeogenic and glycolytic metabolism in *Pseudomonas putida*. *Environ. Microbiol.* **2015**, *17* (9), 3362−3378.
(33) Kim, J.; Oliveros, J. C.; Nikel, P. I.; Lorenzo, V.; Silva-Rocha, R. Transcriptomic fingerprinting of *Pseudomonas putida* under alternative physiological regimes. *Environ. Microbiol. Rep.* **2013**, *5* (6), 883−891.
(34) Nelson, K. E.; Weinel, C.; Paulsen, I. T.; Dodson, R. J.; Hilbert, H.; dos Santos, V. A. M.; Fouts, D. E.; Gill, S. R.; Pop, M.; Holmes, M.; Brinkac, L.; Beanan, M.; DeBoy, R. T.; Daugherty, S.; Kolonay, J.; Madupu, R.; Nelson, W.; White, O.; Peterson, J.; Khouri, H.; Hance, I.; Chris Lee, P.; Holtzapple, E.; Scanlan, D.; Tran, K.; Moazzez, A.; Utterback, T.; Rizzo, M.; Lee, K.; Kosack, D.; Moestl, D.; Wedler, H.; Lauber, J.; Stjepandic, D.; Hoheisel, J.; Straetz, M.; Heim, S.; Kiewitz, C.; Eisen, J. A.; Timmis, K. N.; Dusterhoft, A.; Tummler, B.; Fraser, C. M. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **2002**, *4* (12), 799−808, DOI: 10.1046/j.1462-2920.2002.00366.x.
(35) Otto, M.; Wynands, B.; Drepper, T.; Jaeger, K.-E.; Thies, S.; Loeschcke, A.; Blank, L. M.; Wierckx, N. Targeting 16S rDNA for stable recombinant gene expression in *Pseudomonas*. *ACS Synth. Biol.* **2019**, *8* (8), 1901−1912.
(36) Edgar, R.; Domrachev, M.; Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30* (1), 207−210.
(37) Winsor, G. L.; Griffiths, E. J.; Lo, R.; Dhillon, B. K.; Shay, J. A.; Brinkman, F. S. Enhanced annotations and features for comparing

thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res.* **2016**, *44* (D1), D646−53.
(38) Solovyev, V.; Salamov, A. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*; Nova Science Publishers, 2011; pp 61−78.
(39) Reese, M. G. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.* **2001**, *26* (1), 51−56.
(40) Gautheret, D.; Lambert, A. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.* **2001**, *313* (5), 1003−1011.
(41) Kukurba, K. R.; Montgomery, S. B. RNA sequencing and analysis. *Cold Spring Harbor Protoc.* **2015**, *2015* (11), 951−969.
(42) Cambray, G.; Guimaraes, J. C.; Mutalik, V. K.; Lam, C.; Mai, Q.-A.; Thimmaiah, T.; Carothers, J. M.; Arkin, A. P.; Endy, D. Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* **2013**, *41* (9), 5139−5148.
(43) Mutalik, V. K.; Guimaraes, J. C.; Cambray, G.; Lam, C.; Christoffersen, M. J.; Mai, Q. A.; Tran, A. B.; Paull, M.; Keasling, J. D.; Arkin, A. P.; Endy, D. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **2013**, *10* (4), 354−360.
(44) Lenzen, C.; Wynands, B.; Otto, M.; Bolzenius, J.; Mennicken, P.; Blank, L. M.; Wierckx, N. High-yield production of 4-hydroxybenzoate from glucose or glycerol by an engineered *Pseudomonas taiwanensis* VLB120. *Front. Bioeng. Biotechnol.* **2019**, *7*, No. 130, DOI: 10.3389/fbioe.2019.00130.
(45) Li, W.-J.; Jayakody, L. N.; Franden, M. A.; Wehrmann, M.; Daun, T.; Hauer, B.; Blank, L. M.; Beckham, G. T.; Klebensberger, J.; Wierckx, N. Laboratory evolution reveals the metabolic and regulatory basis of ethylene glycol metabolism by *Pseudomonas putida* KT2440. *Environ. Microbiol.* **2019**, *21* (10), 3669−3682.
(46) Ackermann, Y. S.; Li, W. J.; de Hipt, L. O.; Niehoff, P. J.; Casey, W.; Polen, T.; Köbbing, S.; Ballerstedt, H.; Wynands, B.; O'Connor, K. E.; Blank, L. M.; Wierckx, N. Engineering adipic acid metabolism in *Pseudomonas putida*. *Metab. Eng.* **2021**, *67*, 29−40.
(47) Auton, M.; Holthauzen, L. M.; Bolen, D. W. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (39), 15317−15322.
(48) Qian, Y.; Huang, H.-H.; Jiménez, J. I.; Del Vecchio, D. Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* **2017**, *6* (7), 1263−1272.
(49) Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **1983**, *166* (4), 557−580.
(50) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989; p 1546.
(51) Keen, N. T.; Tamaki, S.; Kobayashi, D.; Trollinger, D. Improved broad-host-range plasmids for DNA cloning in gram-negative bacteria. *Gene* **1988**, *70*, 191−197, DOI: 10.1016/0378-1119(88)90117-5.
(52) Hartmans, S.; Smits, J. P.; van der Werf, M. J.; Volkering, F.; de Bont, J. A. M. Metabolism of styrene oxide and 2-phenylethanol in the styrene-degrading *Xanthobacter* strain 124X. *Appl. Environ. Microb.* **1989**, *55*, 2850−2855, DOI: 10.1128/aem.55.11.2850-2855.1989.
(53) Bagdasarian, M.; Lurz, R.; Ruckert, B.; Franklin, F.; Bagdasarian, M.; Frey, J.; Timmis, K. Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene* **1981**, *16*, 237−247.
(54) Chu, Y.; Corey, D. R. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther.* **2012**, *22* (4), 271−274.

https://doi.org/10.1021/acssynbio.3c00747
ACS Synth. Biol. 2024, 13, 2060−2072