Volume 23

Synthetic biology is the engineering of complex, biologically based systems with unnatural functions. Research is starting at small molecules up to cells, tissues, and organisms by applying a rational and systematic synthesis route. Synthetic biology combines different disciplines to pursuing the vision of building up artificial systems, hierarchically designed and inspired by nature, while using standardized and characterized parts. However, a prediction of the outcome is rarely possible because standardization in different biological disciplines is missing, like promoter activity, media composition, and growth rate calculation.

This thesis aimed to increase the number of rationally designed and well-characterized genetic tools for *Pseudomonas putida* KT2440 to foster synthetic biology. This thesis contributes to the reliability of synthetic biology that often proved as a major issue. The construction of sigma-70 dependent promoters for heterologous expression is discussed in detail, and different approaches are shown. The generation of a stacked promoter library, with the characterization of context-dependent controls, enabled us to predict the resulting promoter activity. Besides synthetic promoters, the frequently used promoters P_{Lac}, P_{Trp}, and their derivatives were characterized and compared. Also, the core promoter sequence of a sigma-38 dependent promoter was characterized using a genome walking method, resulting in different length fragments from the upstream sequence of the candidate gene.

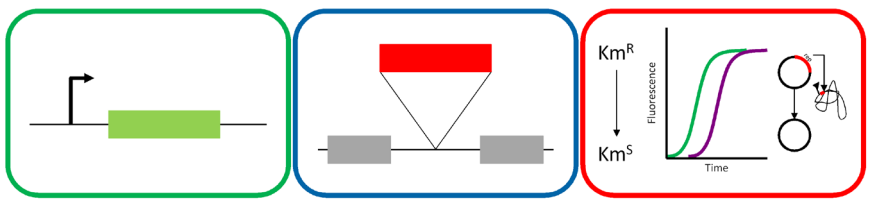
The use of plasmids is disadvantageous. Therefore, suitable genomic integrations sites are needed. Here we applied to approaches to identify them. A Tn5 library of *P. putida* KT2440 was generated and characterized for reporter gene expression. In parallel, we analyzed RNA-Seq data from different cultivation conditions to identify regions along the genome, which are equally expressed. Ten integration sites were characterized in detail using different genetic constructs to determine the influence of the genomic context.

In summary, this thesis contributes to *P. putida* KT2440 synthetic biology. The increased number of well-characterized tools will further support the many efforts to establish this microbe as a workhorse in the bioeconomy and biotechnological applications. Getting the tough challenges, like the use of aromatic compounds from plastic or lignin as substrates, will distinguish *P. putida* KT2440 from the well-established microbes.

ISBN 978-3-86359-899-0

Sebastian Köbbing

Development of synthetic biology tools for *Pseudomonas putida*



Development of synthetic biology tools for *Pseudomonas putida*
Sebastian Köbbing

iAMB Institute of Applied Microbiology | ABBt Aachen Biology and Biotechnology | RWTH AACHEN UNIVERSITY
Edition Wissenschaft Apprimus

Development of synthetic biology tools for *Pseudomonas putida*

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Sebastian Köbbing

Master of Science in der Biologie

aus

Münster, Deutschland

Berichter:

Univ.-Prof. Dr.-Ing. Lars M. Blank

Univ.-Prof. Dr. Nick Wierckx

Tag der mündlichen Prüfung: 25. August 2020

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://portal.dnb.de abrufbar.

Sebastian Köbbing:

Development of synthetic biology tools for *Pseudomonas putida*

1. Auflage, 2020

Gedruckt auf holz- und säurefreiem Papier, 100% chlorfrei gebleicht.

Apprimus Verlag, Aachen, 2020
Wissenschaftsverlag des Instituts für Industriekommunikation und Fachmedien an der RWTH Aachen
Steinbachstr. 25, 52074 Aachen
Internet: www.apprimus-verlag.de, E-Mail: info@apprimus-verlag.de

Printed in Germany

ISBN 978-3-86359-899-0

D 82 (Diss. RWTH Aachen University, 2020)

Eidesstattliche Erklärung

Hiermit erkläre ich, Sebastian Köbbing, dass diese Dissertation und die darin dargelegten Inhalte die eigenen sind und selbstständig, als Ergebnis der eigenen originären Forschung, generiert wurden.

Hiermit erkläre ich an Eides statt

1. Diese Arbeit wurde vollständig oder größtenteils in der Phase als Doktorand dieser Fakultät und Universität angefertigt;

2. Sofern irgendein Bestandteil dieser Dissertation zuvor für einen akademischen Abschluss oder eine andere Qualifikation an dieser oder einer anderen Institution verwendet wurde, wurde dies klar angezeigt;

3. Wenn immer andere eigene- oder Veröffentlichungen Dritter herangezogen wurden, wurden diese klar benannt;

4. Wenn aus anderen eigenen- oder Veröffentlichungen Dritter zitiert wurde, wurde stets die Quelle hierfür angegeben. Diese Dissertation ist vollständig meine eigene Arbeit, mit der Ausnahme solcher Zitate;

5. Alle wesentlichen Quellen von Unterstützung wurden benannt;

6. Wenn immer ein Teil dieser Dissertation auf der Zusammenarbeit mit anderen basiert, wurde von mir klar gekennzeichnet, was von anderen und was von mir selbst erarbeitet wurde;

7. Ein Teil oder Teile dieser Arbeit wurden zuvor veröffentlicht und zwar in:

S. Köbbing, L. M. Blank, N. Wierckx. “Characterization of context-dependent effects on synthetic promoters”, *Front. Bioeng. Biotechnol*. (2020), 8(551), doi: 10.3389/fbioe.2020.00551

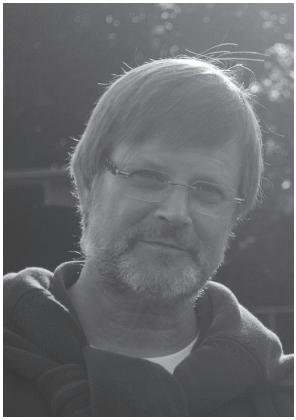
Aachen, den 26.08.2020

Sebastian Köbbing

Funding

This work was performed as a part of the project “P4SB – From Plastic waste to Plastic value using *Pseudomonas putida* Synthetic Biology”, which was funded by the European Union´s Horizon 2020 research and innovation program under grant agreement no. 633962.

Danksagung



Klaus Köbbing

*10.08.1956 †5.08.2019

Zuerst möchte ich mich bei meinem Vater Klaus dafür bedanken, dass er mich immer unterstützt und mir dadurch mein Abitur und das Studium in Münster und Aachen ermöglicht hat. Leider wird Papa diese Dissertation nicht mehr lesen können, obwohl er bis zuletzt gesagt hat, dass er diesen Tag noch erleben möchte. Daher kann ich mich nur für alles, was er mir ermöglicht hat, bedanken. Meiner Mutter Michaela möchte ich dafür danken, dass sie sich die ganze Zeit so gut um Papa gekümmert hat und immer bei ihm war. Entschuldige bitte, dass ich so selten zuhause war und stattdessen im Labor stand, statt euch und dich zu unterstützen. Meinen Schwestern Stefanie und Jana möchte ich an dieser Stelle herzlichst dafür danken, dass ihr zwei Mama und Papa all die Jahre geholfen habt. Danke dafür, dass ihr all die Jahre an mich geglaubt und mich so motiviert habt, auch wenn mich Nachfragen zum Stand meiner Dissertation irgendwann genervt haben.

Ein großer Dank gilt meinem Doktorvater Professor Dr. Lars M. Blank und meinem Betreuer Professor Dr. Nick Wierckx. Danke für die Möglichkeit meine Dissertation am Institut für Angewandte Mikrobiologie anfertigen zu können. Für die großartige Unterstützung und das Vertrauen, das ihr mir in den letzten Jahren entgegengebracht habt.

Nick, ich danke dir für die anregenden wissenschaftlichen Diskussionen und die neuen Ideen, die wir entwickelt und so weitere Forschungsansätze gefunden haben. Die immer wieder erzeugte Motivation während meiner Dissertation, sei es im Labor oder beim Schreiben. Dafür, dass deine Tür am iAMB immer offenstand und du mich beim Schreiben der Dissertation und der Publikation

viel unterstützt hast. Deine Betreuung war super und hat mir dabei geholfen, die Ziele zu erreichen. Die Zusammenarbeit hat mir sehr viel Freude bereitet.

In diesem Zuge danke ich auch der EU, die mir im Rahmen des Projekts P4SB viele Reisen innerhalb Europas ermöglicht hat. Wie sagte meine Schwester Jana: „wenn du schon nicht in den Urlaub fährst, dann schicken die dich halt von der Arbeit überall hin“. Danke an alle beteiligten des Projekts für die schönen Treffen und interessanten Diskussionen. Ein großer Dank gilt meiner Reisegruppe bestehend aus Wing-Jin, Till, Lars und Nick, der meist mittels App nach den besten Restaurants Ausschau gehalten hat.

Für die Unterstützung meiner Arbeit möchte ich meiner Studentin Atscharah Panyot, den Praktikanten Justinas Valiulis und Raika Schäfer danken. Ihr habt mir viel Arbeit abgenommen und so sehr geholfen.

Dem ganzen Team des iAMB möchte ich für die tolle Zeit Danke sagen. Für die Feierabendbiere, Betriebsausflüge, Grillabende, Doktoranden Feiern und FIFA Runden am Institut, die mir immer sehr viel Spaß gemacht haben und mich den Frust aus dem Labor vergessen lassen haben.

Birthe, Hamed und Isabel möchte ich besonders danken. Danke, dass ihr immer für mich da wart. Auch für die vielen Diskussionen rund um Arbeit und die vielen Abende, die wir zusammen verbracht haben, sei es im Köpi, Büro, Stadion oder auf Schalke beim Darts.

All meinen Freunden aus der Heimat danke ich dafür, dass sie mich immer unterstützt und Interesse gezeigt haben, auch wenn ich mal länger nicht Daheim war. Ich bin meinen Kegelbrüdern auch nicht böse, dass die meisten dachten, ich hätte Chemie studiert. Euch danke ich für die regelmäßigen Abende und die Kegeltouren der letzten Jahre.

Ich möchte mich bei allen Menschen, die mich in den letzten Jahren im Rahmen meiner Dissertation und neben dem Universitätsleben unterstützt haben, danken. Ich wünsche allen eine erfolgreiche Zukunft und viel Spaß im neuen Job oder in der noch anstehenden Promotionszeit.

Danke!

Table of contents

Summary

Biotechnological applications become a forward-looking alternative to chemical production of fine and bulk chemicals. The envisaged circular (bio)economy is aiming to replace fossil resources by renewable biomass, CO2, and other carbon-rich streams like plastic. Microbial hosts with an expanded range of substrates are used to produce all required products. Here, synthetic biology is used for the tailoring of hyperproducing strains, starting with the prominent chassis strain *Pseudomonas putida*. *P. putida* KT2440 has a versatile metabolism and high resistance towards oxidative stress.

This thesis aimed to increase the number of rationally designed and well-characterized genetic tools for *P. putida* KT2440 to foster synthetic biology. This thesis contributes to the reliability of synthetic biology that often proved as a major issue.

We constructed and characterized several sigma-70 factor dependent synthetic promoters and combinations for heterologous gene expression. A library was constructed with single nucleotide exchanges, which reveals that the -35 and -10 regions are crucial for efficient promoter activity. Combined promoters, so-called stacked promoters, allowed, after detailed characterization of the genetic context, prediction of the resulting expression strength.

Stable genomic integration is often used for metabolic engineering, but characterized sites, so-called landing pads, across the genome of *P. putida* KT2440 are missing. We analyzed RNA-Seq data towards regions that are equally expressed to identify suitable integration sites across the genome of *P. putida* KT2440. These landing pads enabled high heterologous expression with low variability. With well-characterized promoters and landing pads, fine tuning of gene expression can be conducted on two levels, promoter and integration site. Additionally, the thesis delivers tools for marker recycling and an alternative sigma dependent promoter.

In summary, this thesis contributes to *P. putida* KT2440 synthetic biology. The increased number of well-characterized tools will further support the many efforts to establish this microbe as a workhorse in the bioeconomy and biotechnological applications. Getting the tough challenges, like the use of aromatic compounds from plastic or lignin as substrates, will distinguish *P. putida* KT2440 from the well-established microbes.

Zusammenfassung

Biotechnologische Anwendungen sind eine zukunftsweisende Alternative, um die chemische Produktion von Fein- und Massenchemikalien zu ersetzen. Die zirkulare (Bio)Ökonomie zielt darauf ab, fossile Rohstoffe durch nachwachsende Biomasse, CO2 und Kohlenstoffreiche Abfallprodukte zu ersetzen. Dabei werden Wirtsorganismen mit einem erweiterten Substratspektrum verwendet, um dieselben Produkte in einem biologischen System zu produzieren. Die synthetische Biologie wird genutzt, um maßgeschneiderte Überproduktionsstämme für die Anwendungen zu entwickeln. Ein prominenter Vertreter ist *Pseudomonas putida* KT2440. Ein vielseitiger Stoffwechsel und eine hohe Resistenz gegenüber oxidativem Stress zeichnen diesen Organismus aus.

Das Ziel dieser Arbeit war die Anzahl gut charakterisierter und rational konstruierter genetischer Werkzeuge für *P. putida* KT2440 zu erweitern. Diese Doktorarbeit leistet einen Beitrag dazu, die Zuverlässigkeit in der synthetischen Biologie zu steigern, welche sich häufig als problematisch erweist.

Wir haben mehrere Sigma-70-Faktor-abhängige synthetische Promotoren und Kombinationen für die heterologe Genexpression konstruiert und charakterisiert. Eine Promoter-Bibliothek mit einzelnen Basenaustauschen wurde erzeugt, welche zeigte, dass die -35 und -10 Regionen für eine effiziente Promotoraktivität entscheidend sind. Kombinierte Promotoren erlaubten uns, die resultierende Expressionsstärke vorherzusagen, nachdem der genetische Kontext genau charakterisiert wurde.

Eine stabile genomische Integration wird häufig für das metabolische Engineering benötigt, es fehlen jedoch charakterisierte Stellen im Genom von *P. putida* KT2440. Wir haben RNA-Seq-Daten analysiert, um geeignete Integrationsstellen im Genom von *P. putida* KT2440 zu identifizieren, die gleichmäßig exprimiert werden. Die Charakterisierung der Integrationsstellen zeigte eine hohe Expressionsstärke mit geringer Variabilität, während die Stämme mit verschiedenen Kohlenstoffquellen kultiviert wurden. Die Verwendung von charakterisierten Promoter Sequenzen und Integrationsstellen erlaubt die Genexpression auf zwei Ebenen zu steuern und abzustimmen. Des Weiteren wurden in dieser Doktorarbeit Werkzeuge für das Recycling von Resistenzgenen entwickelt und ein Promoter der von einem alternativen Sigma-Faktor abhängig ist.

Zusammenfassend trägt diese Arbeit zur synthetischen Biologie für *P. putida* KT2440 bei. Die steigende Zahl an verfügbaren und insbesondere charakterisierten Werkzeugen unterstützt die Anstrengungen, *P. putida* KT2440 als Modelorganismus sowohl in der Bioökonomie als auch in der biotechnologischen Anwendung zu etablieren. Die Verwendung von aromatischen Verbindungen aus Kunststoffen oder Lignin als Substrat stellt eine große Herausforderung dar. Hier kann *P. putida* KT2440 Vorteile gegenüber anderen etablierten Mikroorganismen nutzen.

List of abbreviations

%	percentage
°C	degree Celsius
3MB	3-methylbenzoate
A or a	nucleotide adenine
ALE	adoptive laboratory evolution
Ap^R	ampicillin resistance
attTn7	transposon 7 attachment side
a.u.	arbitrary units
bp	base pair
BCD2	bicistronic design 2
BLAST	basic local alignment search tool
C or c	nucleotide cytosine
$CaCl_2$	calcium chloride
$CaCl_2x2H_2O$	calcium chloride dihydrate
cDNA	complementary deoxyribonucleic acid
$CoCl_2x6H_2O$	cobalt(II) chloride hexahydrate
Cm^R	chloramphenicole resistance
Ct	cycle threshold
$CuSO_4x5H_2O$	copper(II) sulfate pentahydrate
CV	coefficient of variation
DNA	deoxyribonucleic acid
DNase	desoxyribonuclease
dNTPs	deoxynucleotide Mix
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
et al.	et alii, and others
$FeSO_4x7H_2O$	iron(II) sulfate heptahydrate
FLP	flippase
for	forward oligonucleotide
FRT	flippase recognition target
G or g	nucleotide guanine
g	gram
gDNA	genomic deoxyribonucleic acid

GMO	genetically modified organisms
Gm^R	gentamycin resistance
H	nucleotide adenine, cytosine or thymine
h	hour
HCl	hydrogen chloride
IF1	initiation factor 1
IPTG	isopropyl β-d-1-thiogalactopyranoside
I-*Sce*I	*Saccharomyces cerevisiae* endonuclease
K_2HPO_4	dipotassium hydrogen phosphate
KOH	potassium hydroxide
Km^R	kanamycin resistance
kV	kilovoltage
L	liter
LB	lysogeny broth
M	nucleotide adenine or cytosine
M	molar
max	maximum
mCherry	member of the of monomeric red fluorescent proteins
MCS	multiple cloning site
mg	milligram
$MgCl_2x6H_2O$	magnesium chloride hexahydrate
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate
µ	growth rate
µL	microliter
µF	microfarad
µm	micrometer
mL	milliliter
Min	minute
mM	millimolar
mm	millimeter
$MnCl_2x2H_2O$	mangan chloride dihydrate
mob	mobilization gene
msfGFP	monomeric super fold green fluorescence protein
N or n	nucleotide

n	number of samples
Na_2CO_3	sodium carbonate
NaH_2PO_4	sodium dihydrogen phosphate
ng	nanogram
NGS	next generation sequencing
$(NH_4)_2SO_4$	ammonium sulfate
nm	nanometer
NaOH	sodium hydroxide
OD_{600}	optical density at 600 nm
ONPG	ortho-nitrophenyl-β-galactoside
*ori*R6K	origin of replication R6K
oriT	origin of transfer
oriV	origin of replication
PCN	plasmid copy number
P. putida	*Pseudomonas putida*
pH	negative logarithm of the concentration of hydronium ions
pmol	picomole
PCR	polymerase chain reaction
qPCR	real time quantitative polymerase chain reaction
RBS	ribosome binding site
rev	reverse oligonucleotide
RNA	ribonucleic acid
RNase	ribonuclease
RNA-Seq	ribonucleic acid sequencing
RPKM	reads per kilobase per million reads
rpm	revolutions per minute
S	nucleotide cytosine or guanine
S. cerevisiae	*Saccharomyces cerevisiae*
SD	Shine Dalgarno sequence
SDS	sodium dodecyl sulfate
SNP	single nucleotide polymorphism
T or t	nucleotide thymine
TAG	Triacyl glyceride
Tc^R	tetracycline resistance
Tn5	transposon 5

Tn7	transposon 7
Tn7L	transposon Tn7 left border
Tn7R	transposon Tn7 right border
tnpA	transposase for transposon Tn5
tnsABC+D	transposase
tra	transfer gene
TS1	target sequence 1
TS2	target sequence 2
TSS	transcriptional start site
USA	United States of America
5´UTR	5´untranslated region
V	voltage
xylS	transcriptional regulator XylS of XylS/Pm regulator/promoter system
$ZnSO_4x7H_2O$	zinc sulfate heptahydrate

List of figures

List of tables

Chapter 1
General introduction

Contributions

This chapter was written by Sebastian Köbbing and reviewed by Nick Wierckx and Lars M. Blank.

1 General introduction

1.1 Pseudomonads

The Pseudomonads are a promising group of bacteria for industrial applications and became common workhorses in biotechnology in recent years.[1-3] A versatile metabolism enables them to grow on several carbon sources like glucose and glycerol, but also on a wide range of aliphatics and aromatics.[4-6] *Pseudomonas putida* KT2440 is a non-pathogenic representative of this versatile group of bacteria. Recently the substrate spectrum of *P. putida* KT2440 was enlarged by applying molecular biology applications and tools. Bator *et al.*[7] integrated rhamnolipid production genes into the genome of *P. putida* KT2440 and achieved growth on xylose while transforming plasmids encoding different metabolic pathways. Further improvement of physiological properties was accomplished by adaptive laboratory evolution (ALE) like a shorter lag phase.[8] ALE was also used by Li *et al.*[9] to obtain strains that can utilize and grow on plastic monomers as sole carbon sources like ethylene glycol and 1,4-butanediol.[10] Both monomers can be derived from depolymerized polyethylene terephthalate (PET) or polyurethane (PU).[9-11] Identified genomic mutations originated from ALE experiments were tested and verified by reverse engineering. Different *Pseudomonas* strains have been engineered for the production of chemical products with industrial importance from different renewable carbon sources, like furandicarboxylic acid, rhamnolipids, and aromatics.[1, 12-18] *Pseudomonas* is highly tolerant to chemical stresses and can survive harmful conditions caused by oxidative stress.[1, 16, 19, 20] Some strains can thrive under a second phase of toxic hydrophobic solvents such as toluene or styrene.[21, 22].

To increase the usability of *P. putida* KT2440, genetic tool development is in full progress[23, 24], enabling rapid and efficient metabolic engineering and genome editing.[3, 25-30] Besides frequently discussed CRISPR/Cas9 system for targeted genome modifications, also MAGE is available.[3, 28, 29] Both methods focus on genome modifications of the host by introducing deletions, insertions, and single nucleotide exchanges at multiple sites in parallel. But at the moment, still, elaborate screening efforts are necessary to identify favored phenotypes. Especially while using MAGE a huge genetic diversity can be generated. State of the art is the use of homologous recombination-based gene integration and/or deletion.[16, 26] Also, this enables rapid and efficient insertion of heterologous constructs virtually anywhere in the genome. The chassis design is a common approach to streamline a cell towards different properties. Genome reduction can be achieved by deletion of non-essential genes, flagella apparatus, and prophages.[16, 31] For this purpose, the pEMG system is ideal due to its homologous based specific integration, which leads to scarless modifications.[26] Resulting strains possess a lower metabolic burden and higher growth rates.[32]

Hence a lower competition for transcriptional and translational resources leads to a more stable expression of heterologous genes.

The importance of *Pseudomonas* in industry and research leads to an adaptation of several molecular methods, which are often initially developed for *Escherichia coli*. The group of Pseudomonads counts numerous different microorganisms. Still, a lot of tools are characterized, especially in *P. putida* KT2440, emphasizing its significance.[3, 27, 29] Introduction of the SEVA collection enlarged the usability of available vectors with different properties and promoted scientific exchange.[33, 34]

1.2 From Plastic waste to Plastic value using *Pseudomonas putida* Synthetic Biology

Sustainability of production of chemical goods and everyday items is now prominent in the political and public discussion. The “green deal” of the EU commission is searching for a practical solution to enormous tasks ahead. The most public discussion emerged around plastic products and their future recycling strategies. Crude oil depending production of plastics increased during the last decades all over the world to over 350 million tonnes per year, with often insufficient waste management, leaving waste plastic in landfills, rivers and oceans.[35-38] Due to low recycling rates for everyday used plastics[37], we have to investigate new ways to replace these articles against more sustainable and environmentally friendly ones.[38-40] Here often, bioplastics are mentioned as a promising alternative, ideally produced from renewable feedstocks and being potentially biodegradable.[41] A bioplastic can originate from renewable resources, but at the same time doesn't have to be biodegradable. Also, biodegradable ones from fossil fuels or non-biodegradables from biobased production are feasible. Different combinations obliterate the idea of a real bioplastic and promote 'greenwashing', due to the lack of uniform designation.[42] Biopolymers like polyhydroxy-alkanoates (PHA), polyhydroxybutyrate (PHB), polylactic acid (PLA), and polyethylene furandicarboxylate (PEF) can be produced from renewable feedstocks like biomass and are potential products to substitute common plastic in the future.[43-47] A variety of bacteria can produce PHA and PHB from different carbon sources, like glucose, glycerol, xylose, and vegetable oils.[48-51] But here the next problem occurs, oil deposits on earth are coming in the next centuries to an end, while oil consumption was still increasing before the Corona crisis, due to higher wealth worldwide.[52] Fuels have to be replaced by sustainable products in the near future. Possible candidates are biodiesel and bioethanol.[53] These fuels are bio-based and can be produced from sugars, vegetable oils, or animal fats. However, these carbon sources directly compete with food supply. The rising world population needs more food and, therefore, larger crop areas for cultivation to fed humans and animals alike.[54] Since these areas are shrinking, increasing

competition for this resource is inescapable and at least not favored. This leads to the need for alternative raw materials for the production of bioplastics.

To reduce or even overcome this dilemma, we have to avoid the necessity of crude oil and biomass use in many areas. One way to circumvent the need is to investigate metabolic pathways for bio-conversion and biotransformation by bacterial hosts in biotechnological applications.[1-3, 39, 43] The most prominent plastic in public is PET, an everyday used plastic. One hundred forty-one million tonnes of plastics are used for packaging worldwide per year.[35] The plastic bottles made from PET are one product we associate the change of technology, as glass bottles are also still in use. Packaging of food or also, in general, is often used once, hence coined single-use plastic. Recycling is often more expensive than new material production.[39] In 2015 the recycling rate of PET was less than 30 %.[55] Increasing plastic production worldwide and growing waste streams are asking for efficient recycling systems and enzymatic solution for depolymerization.[38, 56] European Unions determined rates for 2020 proposed several years ago were to recycle 50 % of household and 70 % of construction waste.[39] At that time, the discussion about plastic waste started and the European Union's Horizon 2020 research and innovation program funded research projects in Europe. Here P4SB originated, a consortium of universities and companies across Europe. P4SB was focussing on the biotransformation of plastic waste into value-added products. The plastic transformation proposed was a classic case of up-cycling, as not the same material, but material of higher value where aimed at. The main topics of P4SB were depolymerization of PET and PU to their monomers and engineering strains, which are capable of using these monomers as sole carbon and energy source (Figure 1). Using molecular approaches like ALE or expression of heterologous genes enabled *P. putida* KT2440 to grow on different monomers derived from PET and PU.[9, 10, 57] Chosen strain *P. putida* KT2440 is naturally capable to produced PHA, a storage polymer as a source of carbon in time of starvation.[44] Since the formation of PHA is not dependent on the carbon source, Li *et al*.[10] was able to show that genetically modified *P. putida* KT2440 can grow on 1,4-butanediol and produce PHA. Here a bioplastic was produced from a former plastic derived monomer. They are demonstrating the power of this research to solve future questions regarding plastic recycling. Next to PHA production, the purification of PHA was addressed. Since PHA is stored inside the bacteria, the whole-cell biocatalyst has to be destroyed to release the polymer. While state-of-the-art, cell lysis could decrease the cost of PHA production. In P4SB self-lytic systems were tested to obtain timed self-destruction of the cell without losing the entire population of cells. P4SB aimed to construct a robust whole-cell catalyst. Synthetic biology plays a key role in this effort, while the tools available have to be extended.[58]

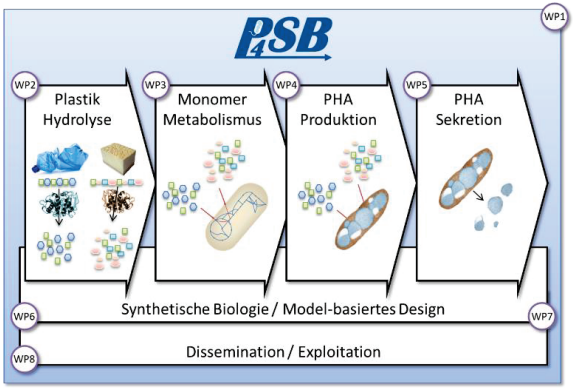


Figure 1: Strategy and organization of P4SB (https://www.p4sb.eu). The aim was to generate a robust whole-cell catalyst, which can depolymerize PET and PU, metabolize resulting monomers, and secrete the bioplastic PHA.

1.3 Synthetic Biology

The term synthetic biology is not unknown to the community. In 1978 Szybalksi and Skalka[59] wrote about the new era of synthetic biology while working on restriction nucleases. They reported being able to construct recombinant DNA and new gene arrangements. Nowadays, these methods are state of the art, if not even outdated.[60] Synthetic biology is the synthesis of complex, biologically based systems with unnatural functions. They are starting from small molecules to cells, tissues, and organisms by applying a rational and systematic synthesis route.[61, 62] Synthetic biology is often related to systems biology, but the difference is that synthetic biology also includes engineering disciplines of synthetic systems. Systems biology is mainly focussing on modeling approaches and tries to understand the fundamentals of existing biological systems. Not only the heterologous expression of a gene of interest but also artificial genomes and genetic networks are generated and used in this field.[63, 64] Engineering starts with an idea of design and further how to build this and test resulting outcomes. Strain modeling is an ongoing research field. However, a prediction of the outcome is rarely possible because standardization in different biological disciplines is missing. Synthetic biology combines different disciplines, like biology, chemistry, engineering, and informatics.[61] The idea is also to speed-up development and increases safety by using standardized parts, sometimes copied from nature.[65] Comparable to standardized screws with a normed thread that fits a standardized nut. Each engineer knows the properties of this screw and knows what he can do with it. Another example is paper sheets with standardized sizes given by the German Institute for Standardization. Accessories in an office are fitted to DinA4 (210 x 297 mm), like folders, hole puncher, envelopes, trays, and printers. Standardization makes life more comfortable because you know what to expect. Alternatively, you need different tools for at least the same use, caused by various properties. Synthetic biology is trying to be as predictable as

a screw.[66] But a screw is not a living organism and standardization is difficult since cellular life is affected by many factors, from biotic and abiotic nature.

Standards only emerged in some fields like DNA sequencing, microarray data[67], enzyme nomenclature, and systems biology models.[68] For most aspects of synthetic biology, even some basic ones, no standard is available, like promoter activity, media composition, and growth rate calculation. SEVA (Standard European Vector Architecture) and 'Registry of Standard Biological Parts' are pushing the efforts to standardize cloning procedures by describing guidelines and sharing genetic resources.[33, 34, 69] These rules are increasing the interchangeability of parts between plasmids and are a small step towards standardization in this field. Since synthetic biology is based on the rational engineering of systems, a lot of targets can be adjusted.[61] Synthetic genomes can be synthesized with fewer genes to avoid side products or unnecessary properties, like flagella and prophages.[16, 32, 58, 70] Resulting strains can have advantageous characteristics like a shorter lag phase or higher growth rates compared to the original strain. Containing fewer genes can have an impact on heterologous gene expression because competition for resources needed for protein biosynthesis is reduced.[71] Next to complete genomes also single proteins can be re-designed by changing the corresponding DNA sequence.[72] The codon usage of species is not constant, resulting primarily from the GC-content of genomes and also from the regulation of different tRNAs that evolved during evolution.[73] The issue of alternative codon usage during gene synthesis can be overcome by the adjustment of the DNA sequence. The procedure is called codon optimization and enables a more stable protein expression by adaptation of included triplets more suiting to tRNAs more prominent in the host cell. Changing the encoding nucleotide sequence of a certain protein can also be used to optimize protein properties.[56] Resulting altered amino acid composition leads to variations of the active site and steric configuration.[72] The protein sequence is directly linked to the steric organization, the latter being difficult to determine from sequence alone. However, this year a huge step towards a priori protein structure predictions was published. Using artificial intelligence allowed the prediction of a protein structure at a quality not seen before, which could complement laborious X-ray crystallography and NMR spectroscopy.[74, 75]

The vision of synthetic biology is to build up artificial systems, hierarchically designed and inspired by nature, while using standardized and characterized parts.[58] Combining applications from molecular biology, biochemistry, platform technologies, and engineering on different levels could make it come true that someday parts can be changed between species with predicted properties. The idea of interchangeable parts from different species with known functions is driving the field of synthetic biology for years. But the function of parts has to be understood first and interactions known before scientists can create synthetic life that supports Darwinian evolution.[58, 76] Since now, we are not able to predict properties from DNA sequence only. It all

depends on the context.[77] Alvarez *et al*.[77] compared context-dependent transcription with politics, everything is dependent on the location. Szybalksi and Skalka[59] used the term synthetic biology in 1978, 40 years have passed since then and we are still not able to predict the phenotype of genetically engineered strains.[78] The long time that has passed reflects maybe the complexity of life.

1.4 History on molecular tools and microbiology

For thousands of years, humankind is using microorganisms for the production of groceries, more or less subconsciously. Already 7,000 years before Christ, people used yeast to produce beverages like beer and wine.[79] Bread production is much older and also used yeast to generate a fluffy consistency. Already 5,000 years ago in Egypt, distinct yeast strains were cultivated for the production.[80] Other popular and early developed foods are cheese and sauerkraut. In the past, used ingredients contained random yeast or bacterial strains on their surface and these caused the favored effect, but sometimes it went wrong. Today we know which organisms are responsible for the production of alcohol or how holes inside a cheese wheel arise. These strains are now pure cultures and production is controlled at many levels to ensure food safety. For this progress, a lot of research was necessary, but not only in food production, also our medical understand increased. By focusing on these fields, often new findings come up, which leads to revolutionary findings. In 1928 Alexander Fleming found penicillin producing fungi that were grown on a petri dish, while the bacterium S*taphylococcus* was not growing next to the fungi.[81] Penicillin revolutionized the treatment of infections and further antibiotics were identified.

At that time also the history of molecular biology began, including disciplines from biochemistry, genetics, microbiology, and virology.[82] Physicists and chemists wanted to uncover the fundamental aspects of life. In 1953 James Watson and Francis Crick, with the help of Rosalind Franklin**,** decoded and reconstructed the structure of the DNA double helix and thus laying the foundation stone for molecular genetics.[83] These developments allowed further milestones, like replication of recombinant DNA and the synthetic production of human insulin in genetically engineered microorganisms.[84, 85] The arguably most powerful tool in DNA research, which is a standard method nowadays, is polymerase chain reaction (PCR) and was developed by Kary Mullis in 1983.[86] Sanger and Coulson had developed a method to determine the DNA sequence, which was called Sanger sequencing.[87] This technique was used for decades as sole technology and is still used while many new DNA sequencing techniques are nowadays available. Next generation sequencing (NGS) is more prominently used today and enables high-throughput sequencing reactions. Here different methods were developed and are still used to determine foreign sequences, like Illumina sequencing, SOLiD sequencing, Ion Torrent, and 454

pyrosequencing.[88-91] Fostered by more powerful and cheaper computational possibilities NGS emerged in this field and lowers the price for a complete genome sequencing reaction following Moore's Law.[92, 93] Next to DNA also RNA was possible to get sequenced and allowed researcher a deeper insight on transcript levels. In combination with a third pedestal method, which is consolidated under -omics, containing upon metabolites, proteins, genome, and transcripts, nearly all information on molecules present in a cell at a certain time point is available. Gathering results from all this method reveal the central dogma of molecular biology, which referred to Crick from 1958 and was re-stated by himself in 1970 (Figure 2).[94, 95]

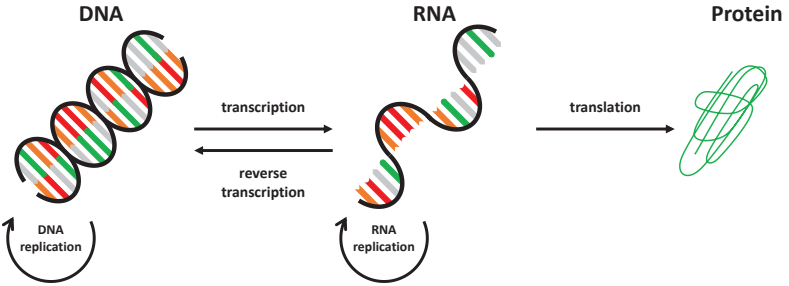


Figure 2: Central dogma of molecular biology.

1.5 Heterologous expression systems

Several microorganisms own non-essential extra-chromosomal DNA, which replicates autonomously.[96] These molecules are called plasmids. Copy number and size can vary from one to hundreds and up to several hundred kilobases, respectively. Smaller modified plasmids are used for decades as a heterologous expression system to characterize gene function, regulate pathways, and to engineer microbes for the production of valuable products. Depending on the origin of replication, the number of plasmids can fluctuate very strongly.[97] This number is called plasmid copy number (PCN) and is controlled by the strength of the origin of replication.[98] But varying copy numbers of the plasmid during the cellular life cycle might lead to unreliable results.[99-101] Furthermore, reduced growth rates and high fitness costs can be a result of plasmids, especially of high copy plasmids, which can reach up to 700 copies per cell.[102, 103] Using high copy plasmids are bearing the risk of promoting genetic instability, caused by unequal distribution of the plasmids during cell division and multimerization.[104] Plasmids with lower copy numbers are often more stable.[105] A minimal plasmid contains an origin of replication, a resistance marker, and a multiple cloning site (MCS; Figure 3). MCS features different recognition sites for endonucleases, which are only one time present in the plasmid and can be used to integrate foreign genes for heterologous expression. Often MCS is embedded into a part of the *lacZ* gene to perform a blue/white screening.[106, 107] Using α-complementation allows transforming modified *E. coli* strains

(*lacZ*ΔM15) with these plasmids and a visual selection for clones bearing a plasmid with insertion by using agar plates with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; Figure 3).[108] Further features can be an origin of transfer (*tra*) for plasmid transfer between cells via conjugation and inducible or constitutive promoters. Proper plasmid propagation during cell division needs the addition of antibiotics or other selective agents.[109] Without selection pressure, cells are no longer constrained to keep and replicate the plasmid. But cells can also be forced to replicate plasmids without antibiotic pressure. Hägg *et al*. constructed an initiation factor 1 (IF1) deficient *E. coli* mutant and replaced this gene on a plasmid. Afterwards only plasmid carrying cells were able to grow and the number of plasmid-free cells didn't accumulate during cultivation. Using an essential gene instead of an antibiotic resistance marker brings the cell into a dependency on plasmid stability. In contrast to commonly used plasmid systems, the need of antibiotics to stabilize plasmid propagation in large scale fermentation processes is unnecessary. For more complex genetically modified microorganism different genes and pathways have to be combined in a cell. The ultimate goal is the generation of a whole-cell catalyst, combining all needed steps in one cell. Here plasmids reach their limit of usability. Propagation of multiple different plasmids in a single cell is hardly accomplishable and often limited to one plasmid. This is caused by competition for the same molecular replication machinery and necessity of different antibiotics.

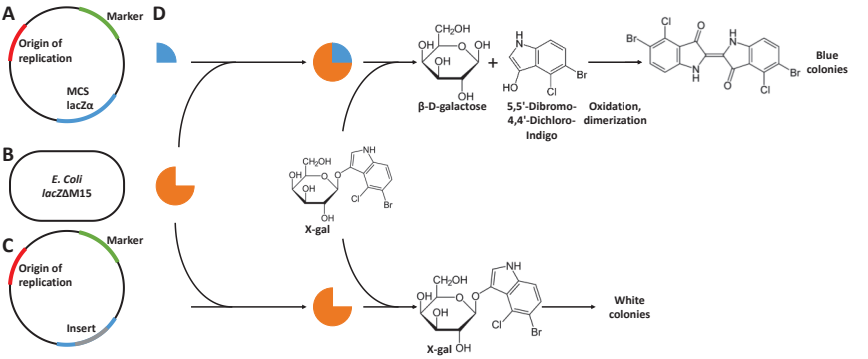


Figure 3: Structural organization of a plasmid, including blue-white screening. (A) A minimal plasmid is consisting of an origin of replication (red), a resistance gene used as a marker (green), and a multiple cloning site (MCS, blue) containing an α subunit of LacZ (β-galactosidase) for cloning purposes. (B) Schematic presentation of an *E. coli* cell with a mutated *lacZ* gene missing the α subunit. (C) Plasmid with an insertion (grey) into the MCS. As a result, the α subunit of LacZ can't be synthesized. (D) Blue-white screening is based on β-galactosidase activity and cleavage of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) into 5,5'-dibromo-4,4'-dichloro-indigo and β-D-galactose. Active α subunit (blue quarter circle) is produced by protein biosynthesis when no insertion takes place. Combined mutated (orange three-quarter circle) LacZ protein and α subunit lead to an active enzyme. This cleaves X-gal into 5,5'-dibromo-4,4'-dichloro-indigo and β-D-galactose, further 5,5'-dibromo-4,4'-dichloro-indigo is oxidized and by dimerization a blue dye occurs, which indicates that no integration happened. Instead, white colonies mean an integration takes place. For selection on plasmid bearing strains with an insertion, agar plates supplemented with corresponding antibiotics and X-gal are used.

Next to plasmids also transposable elements (TE) were found in many living organisms. In 1945 Barbara McClintock discovered the first one in maize. Unfortunately, her work was ignored at the beginning, but 30 years later, she was awarded a Nobel prize in physiology or medicine.[110, 111] In the 1960s, these elements were also found in bacterial cells and from than research on TE was continued. TE are mobile DNA fragments, which are jumping inside the genome by transposition. This is executed by conservative or replicative transposition. The structural organization of transposons is differing, but one common pattern are inverted repeats flanking them. These sequences are recognized by the transposase and enable them to jump. TE owns a transposase gene and can hence perform transposition autonomously. The most prominent transposon is arguably Tn5, which integrates randomly into the genome of different microorganisms (Figure 4).[112, 113] Transposon Tn7, in contrast, integrates targeted into the *attTn7* site of many microorganisms. Due to drawbacks by using plasmids development of transposon based expression system increased.[27, 114, 115] In contrast to natural transposons, no transposase in encoded in the element himself, therefore no subsequent transposition is possible. Thereby stability is raised and reliable experiments are possible.

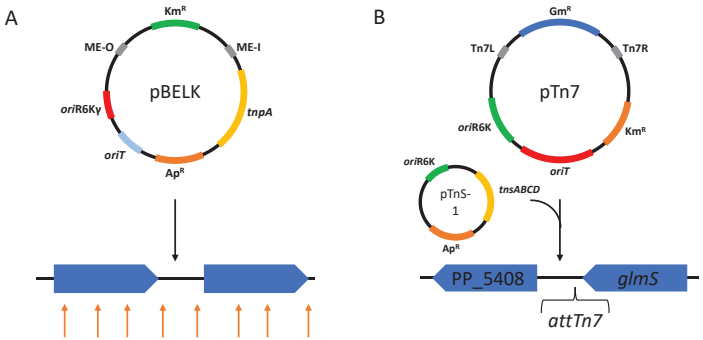


Figure 4: Transposon bearing suicide vectors for targeted and random integration of heterologous constructs. (A) pBELK[114] contains a transposon Tn5, which integrates as a single copy in random positions of the genome, highlighted by arrows (orange). The plasmid contains a *tnsA* gene for transposition, ampicillin marker, the origin of replication *ori*R6Kγ, and origin of transfer *oriT*. Tn5 here only contains a kanamycin resistance gene and is flanked by ME-O and ME-I. (B) pTn7[27] is bearing a transposon Tn7, which integrates as a single copy in the *attTn7* site of several bacteria.[115] In *P. putida* KT2440, surrounding genes are PP_5408 and PP_5409. The plasmid has a kanamycin resistance marker, the origin of replication *ori*R6K, and an origin of transfer o*riT*. Tn7 is flanked by Tn7L and Tn7R. For the selection of proper transposition, a gentamycin resistance marker is included. pTnS-1 provides a transposase since pTn7 himself has no own one.

All the mentioned tools have to be integrated into the host cell by transformation. But some bacteria like *Bacillus subtilis* and *Streptococcus pneumoniae* can take up DNA from their environment. This is called natural competence. These strains own machinery to detect, take up, and integrate foreign DNA.[116, 117] Integration into the genome is done by homologous recombination, while antibiotics are not necessary for maintenance. Martínez-García *et al.*[26] developed a vector that enables targeted genomic integration in *Pseudomonas* by homologous

recombination (Figure 5). Applying this tool allows us to target each site in the genome and can be used for insertions, deletions, and generation of point mutations. Since the integration occurs by homologous recombination, the complete plasmid is inserted. To reach a modification a second step is following to release unwanted parts of the plasmids by expressing I-*Sce*I. This is an exonuclease from *S. cerevisiae* and caused double-strand breaks by cutting I-*Sce*I sites included in the plasmid. The cell afterwards repairs the breaks by homologous recombination and the modification is remaining in the genome.

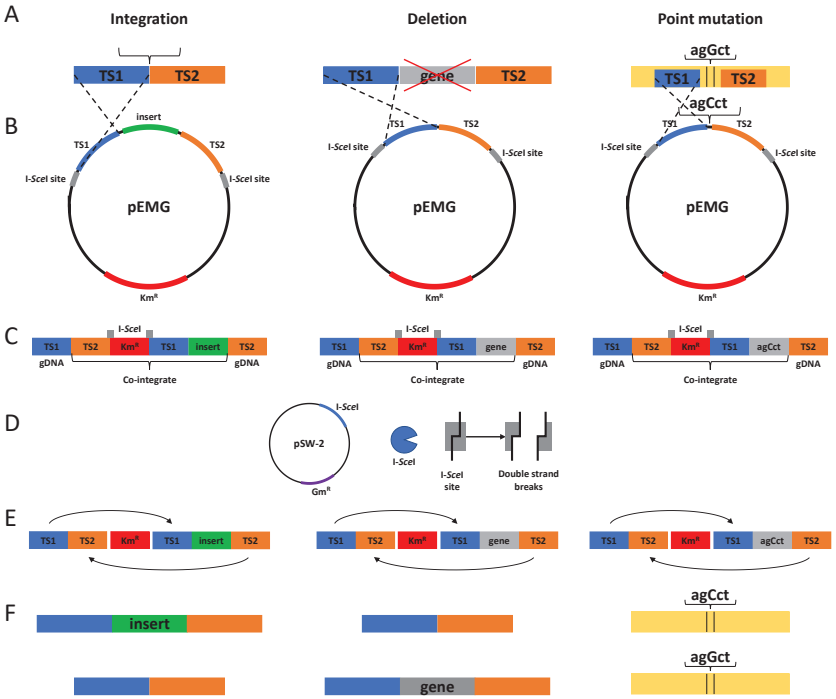


Figure 5: pEMG approach for targeted markerless integration in *P. putida* developed by Martínez-García *et al.*[26] (A) Different modifications are feasible by applying the pEMG system. Depending on the modification, TS elements have to be chosen. (B) pEMG vector contains a kanamycin resistant marker for selection, I-*Sce*I sites flanking the MCS (in empty vectors), *ori*R6K as the origin of replication which can't be used by *P putida* for replication, two TS elements for homologous recombination and in between sequences for modification. (C) pEMG after genomic integration by triparental mating. The whole pEMG vector is integrated over TS1 or TS2, here for simplification integration via TS1 is shown. (D) I-*Sce*I encoding pSW-2 plasmid with gentamycin resistance marker and origin of replication *ori*RK2. I-*Sce*I expression is controlled by 3-methyl benzoate inducible promoter system XylS/P_m. I-*Sce*I recognizes I-*Sce*I sites and is introducing double-strand breaks. (E) Repair of double-strand breaks by homologous recombination. Recombination can be performed via TS1 or TS2. Depending on the TS element used, different cases can occur. (F) Modification can either be generated or wild type situation can be restored. Integration events and results of co-integrate release can be checked by colony PCR.

1.6 Gene expression

Not only the choice of a suitable expression system for heterologous genes is important, also how to control gene expression itself. Here two major screws can be adjusted to tune and adapt to a

certain situation, apart from codon optimization.[118, 119] Promising tools are native or synthetic promoters, which are presenting the initial point of transcription.[120] These short sequences consist in the case of a sigma 70 (σ^{70}) promoter of a highly conserved -35 and -10 regions and an interspacing part (Figure 6).[121]. For *E. coli,* the -35 region consensus sequence is TTGACA and the -10 region TATAAT. Similar sequences can be found in all microorganisms.[122]

A holo-enzyme is consisting of a RNA polymerase with the catalytic activity of RNA polymerization and a sigma (σ) factor. The sigma factor is responsible for the recognition of a promoter sequence.[123] RNA polymerase can be attached to different, so-called sigma factors resulting in sequence-specific promoter recognition.[124] In *E. coli,* seven sigma factors are known, which are controlled by environmental or physiological changes.[125] Same sigma factors are identified in *Pseudomonas* and due to highly conserved sequences, they can be used to generate phylogenetic trees. In research, mainly sigma 70 dependent promoters find use. Because this sigma factor is responsible for the expression of housekeeping genes and transcribes most of the genes, which leads to constitutive expression.[126] The sigma factor can recognize -35 and -10 sequences and positions the holo-enzyme to the DNA. RNA polymerase helicase activity unwinds the double-strand by breaking the hydrogen bonds and uncovers the single-stranded template strand. RNA polymerase and sigma factor separate from each other, while the one is further elongating the generated mRNA molecule and the other detaches from the DNA sequence. Ongoing elongation opens the double-strand for RNA polymerization.

Most synthetic promoter libraries are based on sigma 70 promoters. By exchanging -35 and -10 regions large effects are described. At the same time, the interspaced sequence shows influences by changing the sequence and length.[120, 127] The optimal length of the interspacing sequence depends on the host, but 17 bp is described being favorable .[122, 128] Therefore, consensus sequences are often maintained. In contrast, the rest of the sequence is randomized.[27, 128] Thereby, millions of different promoter sequences are generated, which leads to a wide range in activity, as Solem *et al*.[128] and Zobel *et al*.[27] have shown. Applying high throughput technologies for the characterization of thousands or millions of different plasmids is possible. Manual identification of suitable promoters is laborious and restricted to a small pool of sequences. The creation of low redundancy libraries, in contrast, has a limited outcome and characterization is less elaborate. Figure 6 shows different approaches for the generation of diversity in a promoter library, normally a long oligonucleotide containing the randomized core promoter sequence is used in a PCR reaction followed by cloning procedures. Alternatively, existing promoters can be used in error-prone PCR methods to generate genetic diversity.

Between the promoter and translational start site (TSS), a ribosome binding site (RBS) is located. Otherwise, the ribosome is not able to recognize the start point on the mRNA and no protein is

synthesized (Figure 6). The RBS consensus sequence 5'-AGGAGG-3´ is present in prokaryotes and part of the 5' untranslated region (5'UTR) of mRNA.[129] This sequence is called Shine-Dalgarno (SD) sequence, referred to the discoverers, and is about 9 bp upstream from TSS in *E. coli*.[130] While transcription of the complementary mRNA of the DNA template is formed, this leads to an anti-Shine-Dalgarno sequence 3'-UCCUCC-5´. Sequence UCCU is recognized by 16S rRNA inside the 30S small ribosomal subunit and guides the mRNA to the P site. Amino acids carrying tRNAs are recruited and bind to the corresponding codon on the mRNA inside the ribosome. Chemical bonds between amino acids lead to a primary protein structure. Varying the RBS sequence has an influence on mRNA translation, caused by better or worse binding of 16S rRNA to the anti-Shine-Dalgarno sequence.[131]

Changing promoter and RBS sequences are powerful tools to fine tune gene expression on transcriptional and translational level.[131-134] Avoiding bottlenecks in pathways or bypassing metabolites is feasible and can be controlled. But the use of accurately characterized modules in the desired host is needed, here synthetic biology plays an important role.[58, 135]

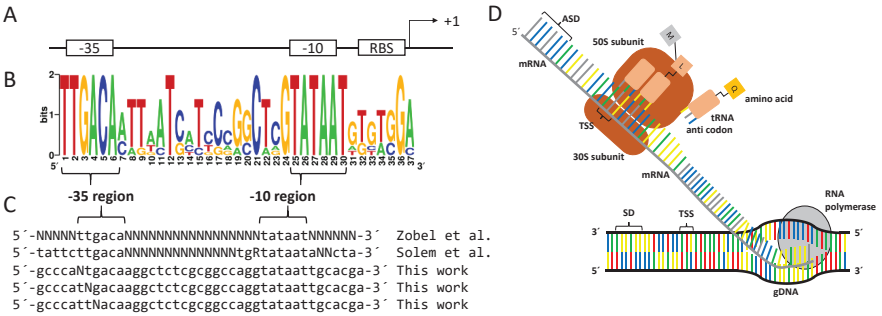


Figure 6: Promoter organization and protein biosynthesis as a major component of cellular life. (A) Structural organization of a core promoter sequence. -35 region is located around 35 bp upstream of the translation start site, -10 region around 10 bp. Next to the promoter, a ribosome binding site (RBS) is present for the initiation of the translation process. The translational start site (TSS) is indicated by an arrow and +1. (B) Sequence logo from P_{trp}, P_{lac}, P_{lacUV5}, P_{tac}, P_{trc}, P_{tic}, and P_{14g} visualized with an online tool provided by the University of California, Berkeley.[136, 137] X-axis is showing different positions along with the promoter sequences. Y-axis shows bits, which are by the number of same nucleotides present at the same position in the promoter sequence. A maximum value of $\log_2(4) = 2$ is possible. (C) Different sequences used for the generation of sequence diversity in promoter libraries. These are parts from longer oligonucleotide and author names are given as a reference. Randomized nucleotides are highlighted with capitalized letter N, standing for thymine, adenine, cytosine, or guanine. For (A) and (B) -35 and -10 regions are indicated with curly brackets. (D) Schematic presentation of protein biosynthesis existing of transcription and translation. RNA polymerase and sigma factor recognize promoter sequence (A), and transcription is initiated, followed by RNA polymerization. Ribosomes are guided to the TSS after the detection of the anti-SD sequence by the 30S subunit. Amino acids carrying tRNA can bind inside the ribosome to the corresponding codon and amino acids are associated with each other via peptide bonds. A red letter is highlighting nucleotide thymine, green adenine, blue cytosine, yellow guanine, and grey uracil, which is only present in mRNA molecules replacing thymine.

1.7 Scope of the thesis

This study aimed to enlarge the existing toolbox for Pseudomonads and especially for *P. putida* KT2440. Here the focus was on the characterization of modified transposons with beneficial properties, characterization of different genomic integrations sites across the genome, and the increasing availability of synthetic promoters with a wide range of activities in small nuances. Well-characterized modules are necessary for reliable research and to pursue the goal of synthetic biology.

Chapter 1 provides a view on the development of molecular biology and microbiology, by given milestone accomplished in the last hundred years. The conception of synthetic biology is described by mentioning its basic idea and ultimate goals. Available heterologous expression systems are explained and the generation of synthetic promoter libraries clarified. The project P4SB is described and the reasons are given why this project initiated many others and such has a huge impact for the next decades. This project pushes ahead synthetic biology in Europe by uniting researchers from different disciplines to tackle the problem of plastic waste together.

Chapter 2 describes the materials and methods used in this study. All approaches are given as well constructed strains, plasmids, and sequences generated.

Chapter 3.1 is introducing stacked promoters and their outcome. Synthetic promoters interspaced with different length inert sequences were combined and activities determined. Using context-dependent controls and a combination of different promoters, a semi-empirical correlation was developed. This enables us to predict the resulting activity of the here used promoters.

Chapter 3.2 presents the identification of suitable integration sites across the genome of *P. putida* KT2440. Genomic read through activity caused by native promoters is often used for transposon Tn5 approaches to utilize native activity of a distinct genomic context. Here we excluded the upstream effects with a terminator and measured resulting fluorescence signals of a Tn5 library consisting of 500,000 mutants by FACS. This reveals “hot” and “cold” spots and further identification of integration sites was made by analyzing RNA-Seq data sets. Equally expressed regions under different conditions were found and sites characterized. They were resulting in low variability while cultivated with different carbon sources. We were confirming our approach that RNA-Seq data can be used to identify suitable integration sites.

Chapter 3.3 presents a collection of various tools that are characterized in this study. Varying plasmid copy numbers often throwback the reliability of such experiments. In this study, the replication gene was replaced into the genome under the control of a synthetic promoter and simultaneously deleted from the plasmid. The new plasmid was not able to replicate in *P. putida* KT2440 wild type but in strains that are expressing the replication gene. Since plasmids have to be stabilized by antibiotics, corresponding markers are rare. Thus, a marker recycling system for

transposon Tn7 plasmids was generated based on FRT-FLP recombination. Markerless strains show low variation compared with strains bearing the original construct or if the resistance gene is still present. Generation of many new Tn7 plasmids with formerly characterized pathways is laborious, here a pEMG based plasmid was constructed for targeted integration into Tn7 in *attTn7* site to disrupt the gentamycin resistance gene. This system can be used for already created strains bearing a transposon Tn7 integration. The characterization of an effect by recycling of a marker was done with constitutive synthetic promoters. Most synthetic promoters are sigma 70 factor dependent. Achieving a delayed gene expression can be ensured by using alternative sigma factors. We analyzed RNA-Seq data from different growth stages to identify genes exclusively expressed in the stationary phase. The upstream sequence from one gene was characterized to localize the core promoter.

The results of this study are discussed in chapter 4 and an attempt is made to illuminate them as a whole. The importance of well-characterized genetic modules is underlined here and ideas on how to improve synthetic biology are mentioned.

Chapter 2
Material and methods

Partially published as:

S. Köbbing, L. M. Blank, N. Wierckx. "Characterization of context-dependent effects on synthetic promoters", *Front. Bioeng. Biotechnol*. (2020), 8(551), doi: 10.3389/fbioe.2020.00551

Contributions

This chapter was written by Sebastian Köbbing and reviewed by Nick Wierckx and Lars M. Blank.

2 Material and Methods

2.1 Cultivation conditions

E. coli strains were grown in lysogeny broth (LB) medium containing10 g L^{-1} peptone, 5 g L^{-1} yeast extract, and 5 g L^{-1} sodium chloride.[108] For solid media 15 g L^{-1} agar was added to the medium before autoclaving. Selection for *Pseudomonas* strains was done by using cetrimide agar plates prepared as described by the manufacturer. As a carbon source, 1 % (w/v) glycerol (99 %) was added to the medium before autoclaving.

Cultivation of *E. coli* was done at 37°C and *P. putida* KT2440 at 30°C. System Duetz® (24 or 96 well plates, Enzyscreen, Netherlands) with a volume of 1.5 mL or 0.5 mL of liquid medium were used (d = 50 mm, n = 300 rpm, T = 30 °C and Φ = 80 %). Cultivation of *E. coli* strains for plasmid isolation was either performed in test tubes with 5 mL LB medium at 37°C or in System Duetz® (24 well plates, Enzyscreen, Netherlands) at 30°C with 2 mL LB medium.

For plasmid stability or selection on strains with an integration, antibiotics were added to liquid and solid media. 50 mg L^{-1} kanamycin, 30 mg L^{-1} gentamycin, 10 mg L^{-1} tetracycline, 10 mg L^{-1} chloramphenicol, 50 mg L^{-1}streptomycin, or 500 mg L^{-1} ampicillin were added to the medium after autoclaving. New plates tested with wild type and resistance strains to check them. Induction of I-*Sce*I expression by P_m Promoter activated via *xylS* was ensured by the addition of 15 mM 3-methyl benzoate (3MB; solved in 70% (w/v) ethanol) to the medium after autoclaving.[26]

Promoter characterization in *P. putida* KT2440 was done with minimal medium described by Hartmans *et al*.[138] The medium contained 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} $MgCl_2x6H_2O$, 0.01 g L^{-1} EDTA, 0.002 g L^{-1} $ZnSO_4x7H_2O$, 0.001 g L^{-1} $CaCl_2x2H_2O$, 0.005 g L^{-1} $FeSO_4x7H_2O$, 0.0002 g L^{-1} $Na2MoO_4x2H_2O$, 0.0002 g L^{-1} $CuSO_4x5H_2O$, 0.0004 g L^{-1} $CoCl2x6H_2O$, 0.001 g L^{-1} $MnCl_2x2H_2O$, and a carbon source. Here 20 mM glucose, 20 mM fructose, 20 mM citric acid, 30 mM succinate, or 20 mM glucose with 0.8 M urea were used. If necessary, antibiotics were added as described above. Measuring endpoint fluorescence was done with either minimal medium or LB medium, as described above. Addition of antibiotics as stated.

2.2 DNA techniques and cloning procedures

The following sections are describing how plasmids and strains were generated and characterization was performed. Oligonucleotides, strains, and plasmids for each section are given in separate tables, depending on each section.

2.2.1 Enzymes and PCR reactions

Enzymes for digestion and PCR reaction components were ordered at New England Biolabs (Ipswich, Massachusetts, USA), as well as corresponding buffer solutions. Standard digestions were done with 1 µL of each enzyme, 2 µL buffer, 100 to 500 ng DNA, and filled up with water to 20 µL. Digestion with three different enzymes required a total volume of 30 µL, to avoid inhibitory effects of glycerol on enzyme activity. Incubation was done at 37°C for one hour. Prevention of self-ligation of the plasmid backbones was obtained by alkaline phosphatase (Fermentas, Waltham, Massachusetts, USA), which was added for an additional hour to the reaction. Purification was done by agarose gel electrophoresis or using PCR & DNA Cleanup Kit (New England Biolabs, Ipswich, Massachusetts, USA). PCR with Q5 High Fidelity polymerase was done in a total volume of 50 µL. 2.5 µL of each oligonucleotide, 10 µL buffer, 5 µL dNTPs, 1 µL template, and 1 µL Q5 polymerase. If needed 10 µL GC enhancer was added. Oligonucleotides used in this study were ordered at Eurofins Genomics (Ebersberg, Germany) and are listed in the following tables. Colony PCRs were performed with *Taq* 2X Master Mix (New England BioLabs, Ipswich, Massachusetts, USA) in a total volume of 12.5 µL. 6.25 µL Master Mix, 0.25 µL of each oligonucleotide, and 1 µL template. Cell lysis of single colonies was done in 30 µL lysis buffer containing 60 % alkaline PEG 200 (pH adjusted to 13-13.5 with 2M KOH) for 15 minutes at room temperature.[16] Colonies were picked with a pipette tip or toothpick and cell material transferred to lysis buffer containing reaction tubes. FleyCycler2 from Analytik Jena (Jena, Germany) was used for PCR. DNA fragments were separated by agarose gel electrophoresis (Voltage 100 V, resistance 500 ohms, 20 minutes) using 1 % agarose gels (Bio-Budget, Krefeld, Germany) with Roti® Safe Stain (50 µL L^{-1}, Carlroth, Karlsruhe, Germany). Fragments were isolated from agarose gels by applying DNA Gel Extraction kit (New England Biolabs, Ipswich, Massachusetts, USA). For direct PCR fragment isolation PCR & DNA Cleanup Kit (New England Biolabs, Ipswich, Massachusetts, USA) was used. Plasmid preparation was done with Plasmid Miniprep Kit (New England Biolabs, Ipswich, Massachusetts, USA). Sequencing reactions were done by Eurofins Genomics (Ebersberg, Germany).

2.2.2 Transformation

Chemical competent *E. coli* cells were prepared in our lab, as described by Hanahan.[139] Pre-cultures with LB medium were cultivated overnight at 37°C. 50 mL main cultures with SOB medium were inoculated with an initial OD_{600} of 0.1 and cultivated at 30°C and 250 rpm. SOB medium contained 5 g L^{-1} yeast extract, 20 g L^{-1} peptone, 0.6 g L^{-1} sodium chloride, and 0.2 g L^{-1} potassium chloride. After autoclaving, 10 mL L^{-1} 2 M Mg^{2+} was added. The solution consists of 1 M $MgSO_4$ and 1 M $MgCl_2$. Cells were grown to an OD_{600} of 0.5, reaching this point, cells were

incubated on ice for 10 minutes. Cells were centrifuged for 15 minutes at 5,000 rpm at 4°C. The pellet was resuspended in 10 mL sterile 0.1°M $CaCl_2$ solution and chilled on ice for 30 minutes. An additional centrifugation step, as described before, followed by resuspending the cells in 3 mL 0.1 $MCaCl_2$ and 1 mL 50% (w/v) glycerol. Aliquots with different volumes were stored at -80°C. For transformation 100 µL cells were thawed on ice. DNA to be transformed either given to the cells or inverse and chilled on ice for 30 minutes. Transformation by heat shock was done for 90 seconds at 42°C. Afterwards directly 300 µL prewarmed LB medium was added to the cells and regeneration was done at 37°C for 90 minutes. Cells spun down and 200 µL of the supernatant was discarded, resting volume was used for resuspension and 100 µL plated on appropriate LB agar plates.

Electroporation was done with quick generated competent *P. putida* KT2440 cells. Strains were stroked on agar plates and growing cells collected in a reaction tube. Resuspended with 10 % (w/v) glycerol and centrifugated for 5 minutes at 16,000 rpm. The resulting pellet was resuspended again with 500 µL of 10% (w/v) glycerol and centrifugated for 5 minutes at 16,000 rpm. This procedure was repeated five times. In the last step, cells were resuspended in 100 µL 10% (w/v) glycerol and used for electroporation. 100 ng of plasmid was added to competent cells and electroporated with a Gene Pulser Xcell Microbial System (Bio-Rad Laboratories, Hercules, USA; capacity of 25 µF, resistance of 200 ohms, and 2.5 kV) using 2 mm gap electroporation cuvettes (Bio-Budget, Krefeld, Germany). After electroporation 500 µL LB medium was added to the cells. After one hour of regeneration at 30°C, cells were pelleted (1 minute at 16,000 rpm) and 200 µL supernatant discarded. The resting volume was used for resuspension and 100 µL plated on appropriate agar plates.

2.2.3 Genomic integration procedures

Integration of the mini Tn7 transposon was performed by patch mating on LB agar plates and cultivation overnight at 30°C.[27] This was done with a mini Tn7 suicide plasmid-bearing donor strain, acceptor strain *P. putida* KT2440, and two helper-strains. *E. coli* HB101[140] is bearing plasmid pRK600[141] with mobilization genes. *E. coli* DH5αλpir[142] is bearing pTnS-1[115] encodes a transposase for transposition of the mini Tn7 transposon. For selection and counter-selection of Tn7-bearing *P. putida* KT2440, cetrimide agar plates containing 30 mg L^{-1} gentamycin and 1 % glycerol were used. The description of the helper strains is given in Table 1.

The integration of pEMG vectors was also done by triparental mating. As donor *E. coli* PIR2 with the appropriate plasmid, helper strain *E. coli* HB101 pRK600, and *P. putida* KT2440 were used. Selection on pEMG integration was done with selective cetrimide agar plates containing 50 mg L^{-1} kanamycin. *P. putida* KT2440 cells with an integrated pEMG vector can grow and can be picked

on additional plates. Release of the co integrate, caused by full genomic integration of the pEMG vector by homologous recombination, was done in an additional mating using plasmid pSW-2 and pRK600 as helper strains. The plasmid encodes an endonuclease from *Saccharomyces cerevisiae*, which recognizes and cuts I-*Sce*I sites inside the integrated construct and causes double-strand breaks, which are repaired by homologous recombination of the target sequences which are complementary to the genome. Matings were stroked on cetrimide agar plates containing 30 mg×L-1 gentamycin and 1 % glycerol as sole carbon source. Single colonies were picked twice on LB and cetrimide agar plates containing 50 mg×L-1 kanamycin and 1 % glycerol as the sole carbon source to identify clones without kanamycin resistance gene. After co integrate release, two genotypes are possible Recombination leads to the wild type situation or a genomic alteration takes place. Correct integration was verified by colony PCR using oligonucleotides flanking the integration site. Named helper strains are given in Table 1.

2.2.4 Stacking promoters

All strains used and generated in this section are listed in Table 1. Oligonucleotides used in this section are given in Table 2. For the generation of different length spacer sequences, PCR was used. Up to 40 bp length, a forward oligonucleotide (SK11, SK34, SK36, or SK38) and reverse oligonucleotide SK2 were used with plasmid pBG14g as a template (Table 3). Q5 polymerase (New England Biolabs, Ipswich, Massachusetts, USA) with proofreading activity was used for amplification and reactions were prepared as described by the manufacturer. Longer spacer sequences from 50 to 100 bp were assembled by PCR with two long oligonucleotides with complementary 3´-ends and an initial annealing step in the PCR (Table 4). Spacer sequences are given in Table 5. Dimer formation of 3´-ends was checked *in silico* (www.thermofisher.com) to ensure that stacked promoter constructs can be formed by annealing of two oligonucleotides. Promoterless plasmid pBG was used as a backbone. Generated spacer fragments were digested by *Pac*I and *Avr*II (New England Biolabs, Ipswich, Massachusetts, USA) at 37°C. Cloning procedures are following the rules of SEVA and are thus compatible with other constructs.[34] Plasmid pBG was additionally treated with alkaline phosphatase (Fermentas, Waltham, Massachusetts, USA) to circumvent self-ligation. Digested fragments were purified using PCR & DNA Cleanup Kit (New England Biolabs, Ipswich, Massachusetts, USA). DNA concentrations of eluted DNA were measured with a NanoDrop One (Thermo Scientific, Waltham, Massachusetts, USA). Fragments and backbone with adjusted concentrations were ligated using T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA) at room temperature for 30 minutes. The transformation of ligated plasmids was performed by heat shock into chemically competent *E. coli* PIR2 cells.[139] Plasmid isolation was done with Plasmid Miniprep Kit (New England Biolabs,

Ipswich, Massachusetts, USA) and sequences were confirmed by Sanger sequencing with SK43 (Eurofins Genomics, Ebersberg, Germany). Oligonucleotide combinations and detailed PCR protocols are described in Table 6.

For the construction of promoter positions controls, previously cloned plasmids containing stacked promoters were used as a template (Table 7). The construction of the promoterless control is shown in Table 8. Construction of the single nucleotide polymorphism (SNP) library was done by PCR with plasmid pBG14g as template and primers SK63 to SK92 containing single degenerate nucleotides and resulting in different promoter sequences (Table 9). We used colony PCR with oligonucleotides SK4 and SK5 to verify the correct and full-length genomic integration of Tn7 at the attTn7 site in *P. putida* KT2440.

Table 1: Strains and plasmids used and generated to characterize stacking promoters.

Strain	Description	Reference	iAMB strain collection
E. coli			
HB101	*F⁻ mcrB mrr hsdS20(rB⁻ mB⁻) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^R) *gln V44λ⁻*	Boyer and Roulland-Dussoix[140]	2073
CC118λpir	Δ(*ara-leu*) *araD* Δ*lacX74 galE galK phoA20 thi-1 rpsE rpoB argE*(*Am*) *recA1,* lysogenized with λpir phage	Herrero *et al.*[143]	2072
PIR2	F- Δ*lac169 rpoS* (Am) *robA1 creC510 hsdR514 endA reacA1 uidA* (Δ*Mlui*)::*pir*	Life Technologies	3222
DH5αλpir	*endA1 hsdR17 glnV44* (= *supE44*) *thi-1 recA1 gyrA96 relA1 φ80dlac*Δ(*lacZ*)*M15* Δ(*lacZYA-argF*)*U169 zdg-232*::Tn10 *uidA*::*pir*+	de Lorenzo lab	2673
P. putida			
KT2440	Wild-type strain derived of *P. putida* mt-2 cured of the pWW0 plasmid	Bagdasarian *et al.*[144]	2058
BG	Gm^R, *P. putida* KT2440 with genomic insertion of pBG	Zobel *et al.*[27]	3234
BG13	Gm^R, *P. putida* KT2440 with genomic insertion of pBG13	Zobel *et al.*[27]	3235
BG14a 28	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a	Zobel *et al.*[27]	3228
BG14b 35	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b	Zobel *et al.*[27]	3230
BG14c 37	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14c	Zobel *et al.*[27]	3231
BG14d 51	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d	Zobel *et al.*[27]	3233
BG14e 17	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e	Zobel *et al.*[27]	3225
BG14f 25	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f	Zobel *et al.*[27]	3227
BG14g 42	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g	Zobel *et al.*[27]	3232
BG14f_##_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f_##_14g, spacer with varying length from ten to 100 bp	This work	6226-6235
BG_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_80i	This work	6236
BG_80new	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_80new	This work	6237
BG14a_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a_80i_14g	This work	6239
BG14b_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b_80i_14g	This work	6240
BG14c_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14c_80i_14g	This work	6241
BG14d_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d_80i_14g	This work	6242
BG14e_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e_80i_14g	This work	6243

Strain	Description	Reference	iAMB strain collection
BG14f_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f_80i_14g	This work	6244
BG14g_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_80i_14g	This work	6245
BG14f_80i_14f_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f_80i_14f_80i_14g	This work	6245
BG14g_80i_14a	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_80i_14a	This work	6247
BG14a_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a_80i	This work	6248
BG14b_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b_80i	This work	6249
BG14c_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14c_80i	This work	6250
BG14d_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d_80i	This work	6251
BG14e_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e_80i	This work	6252
BG14f_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f_80i	This work	6253
BG14g_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_80i	This work	6254
BG_80i_14a	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_80i_14a	This work	6255
BG_80i_14c	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_80i_14c	This work	6256
BG_80i_14d	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_80i_14d	This work	6257
BG_80i_14e	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_80i_14e	This work	6258
BG_80i_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_80i_14g	This work	6259
BG14f_80new	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f_80new	This work	6260
BG_80new_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_80new_14f	This work	6261
BG14f_80new_14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f_80new_14g	This work	6238
BG14g_SPL_PosZZ_N	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_SPL_PosZZ_N	This work	6271-6319
BG14g_1a_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_1a_80i	This work	6262
BG14g_1c_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_1c_80i	This work	6263
BG14g_1g_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_1g_80i	This work	6264
BG14g_2a_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_2a_80i	This work	6265
BG14g_2c_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_2c_80i	This work	6266
BG14g_2g_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_2g_80i	This work	6267
BG14g_26a_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_26a_80i	This work	6268
BG14g_26c_80i	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_26c_80i	This work	6269
BG14g_26g_80	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_26g_80i	This work	6270
Plasmids			
pRK600	Cm^R, oriColE1, *tra* + *mob* + of RK2	Keen *et al.*[141]	2073

Strain	Description	Reference	iAMB strain collection
pTnS-1	Ap^R, oriR6K, *TnSABC+D* operon	Choi *et al.*[115]	3221
pBG	Km^R, Gm^R,oriR6K, Tn7L and Tn7R extremes, BCD2−*msfgfp* fusion	Zobel *et al.*[27]	3210
pBG13	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter P_{em7}, *msfGFP*	Martínez-Garcia *et al.*[33]	3211
pBG14a	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14a, *msfGFP*	Zobel *et al.*[27]	3214
pBG14b	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14b, *msfGFP*	Zobel *et al.*[27]	3217
pBG14c	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14c, *msfGFP*	Zobel *et al.*[27]	3218
pBG14d	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14d, *msfGFP*	Zobel *et al.*[27]	3220
pBG14e	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14e, *msfGFP*	Zobel *et al.*[27]	3212
pBG14f	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14f, *msfGFP*	Zobel *et al.*[27]	3213
pBG14g	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14g, *msfGFP*	Zobel *et al.*[27]	3219
pBG14f_##_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14f/14g, spacer with varying length from ten to 100 bp	This work	6440-6449
pBG_80i	Km^R, Gm^R, oriR6K, pBG-derived, promoterless control, reverse complement spacer sequence with 80 bp length	This work	6450
pBG_80new	Km^R, Gm^R, oriR6K, pBG-derived, promoterless control, *msfGFP*, new spacer sequence with 80 bp length	This work	6451
pBG14f_80new_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14f/14g, *msfGFP*, new spacer sequence with 80 bp length	This work	6452
pBG14a_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14a/14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6453
pBG14b_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14b/14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6454
pBG14c_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14c/14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6455
pBG14d_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14d/14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6456
pBG14e_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14e/14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6457
pBG14f_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14f_80i_14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6458
pBG14g_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14g_80i_14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6459
pBG14f_80i_14f_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14f 80i 14f 80i 14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6460
pBG14g_80i_14a	Km^R, Gm^R, oriR6K, pBG-derived, stacked promoter 14g_80i_14a, *msfGFP*, inverted spacer with a length of 80 bp	This work	6461
pBG14a_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14a, *msfGFP*, inverted spacer with a length of 80 bp	This work	6462
pBG14b_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14b, *msfGFP*, inverted spacer with a length of 80 bp	This work	6463
pBG14c_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14c, *msfGFP*, inverted spacer with a length of 80 bp	This work	6464
pBG14d_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14d, *msfGFP*, inverted spacer with a length of 80 bp	This work	6465
pBG14e_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14c, *msfGFP*, inverted spacer with a length of 80 bp	This work	6466
pBG14f_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14f, *msfGFP*, inverted spacer with a length of 80 bp	This work	6467
pBG14g_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6468
pBG_80i_14a	Km^R, Gm^R, oriR6K, pBG-derived, second position promoter control 14a, *msfGFP*, inverted spacer with a length of 80 bp	This work	6469
pBG_80i_14c	Km^R, Gm^R, oriR6K, pBG-derived, second position promoter control 14c, *msfGFP*, inverted spacer with a length of 80 bp	This work	6470
pBG_80i_14d	Km^R, Gm^R, oriR6K, pBG-derived, second position promoter control 14d, *msfGFP*, inverted spacer with a length of 80 bp	This work	6471

Strain	Description	Reference	iAMB strain collection
pBG_80i_14e	Km^R, Gm^R, oriR6K, pBG-derived, second position promoter control 14e, inverted spacer with a length of 80 bp	This work	6472
pBG_80i_14g	Km^R, Gm^R, oriR6K, pBG-derived, second position promoter control 14g, *msfGFP*, inverted spacer with a length of 80 bp	This work	6473
pBG14f_80new	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14f, *msfGFP*, new spacer sequence with 80 bp length	This work	6474
pBG_80bp_14g_new	Km^R, Gm^R, oriR6K, pBG-derived, second position promoter control 14g, *msfGFP*, new spacer sequence with 80 bp length	This work	6475
pBG14g_SPL_PosZZ_N	Km^R, Gm^R, oriR6K, pBG-derived, single nucleotide promoter library with specific positions changes, library is based on 14g, *msfGFP*	This work	6476-6533
pBG14g_1a_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14g_1a with change at first nucleotide in -35 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6476
pBG14g_1c_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14c_1c with change at first nucleotide in -35 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6477
pBG14g_1g_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14g 1g with change at first nucleotide in -35 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6478
pBG14g_2a_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14g 2a with change at second nucleotide in -35 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6479
pBG14g_2c_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14c 2c with change at second nucleotide in -35 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6480
pBG14g_2g_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14g 2g with change at second nucleotide in -35 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6481
pBG14g_26a_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14c 26a with change at third nucleotide in -10 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6482
pBG14g_26c_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14c_26a with change at third nucleotide in -10 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6483
pBG14g_26g_80i	Km^R, Gm^R, oriR6K, pBG-derived, first position promoter control 14c 26a with change at third nucleotide in -10 element, *msfGFP*, inverted spacer with a length of 80 bp	This work	6484

distance between two promoters in base pairs
ZZ stands for position 1 to 30 of BG14g based SPL
N stands for nucleotides a, c, g or t

Table 2: Oligonucleotides used in for the generation of stacked promoter and their controls. Promoter sequences and restriction sites are indicated by different colors or are underlined.

Name	Sequence 5´-3´[a]	Reference
SK2	ACACCATAGGTCAGGGTAGTC	This work
SK4	AGTCAGAGTTACGGAATTGTAGG	Zobel *et al.*[27]
SK5	GTCGAGAAAATTGCCGAGCT	Zobel *et al.*[27]
SK11	CGCTTAATTAAGCCCGTTGACATGACATGGTTTTGAGGGTATAATGTGGCGAC CTTAATTAAGCCCATTGACAAG	This work
SK34	CGCTTAATTAAGCCCGTTGACATGACATGGTTTTGAGGGTATAATGTGGCGAG GACGAGTCACCATGTGCCAGCCCATTGACAAGGCTCTCG	This work
SK36	CGCTTAATTAAGCCCGTTGACATGACATGGTTTTGAGGGTATAATGTGGCGAG GACGAGTCACCATGTGCCAGGGGCGATAAGCCCATTGACAAGGCTCTCG	This work
SK38	CGCTTAATTAAGCCCGTTGACATGACATGGTTTTGAGGGTATAATGTGGCGAG GACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGCCCATTGACAAG GCTCTCG	This work
SK43	ATCAAACATCGACCCACGGCGTAAC	This work

Name	Sequence 5´-3´[a]	Reference
SK50	GCGTTAATTAAGCCCGTTGACATGACATGGTTTTGAGGGTAT5AATGTGGCGA GGACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCAT	This work
SK51	GCGCCTAGGTCGTGCAATTATACCTGGCCGCGAGAGCCTTGTCAATGGGCATG AATACTCCCACCGATCG	This work
SK52	GCGCCTAGGTCGTGCAATTATACCTGGCCGCGAGAGCCTTGTCAATGGGCCTT CACCGCGATGAATACTCCCACCGATCG	This work
SK53	GCGCCTAGGTCGTGCAATTATACCTGGCCGCGAGAGCCTTGTCAATGGGCAAC CCAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work
SK54	GCGCCTAGGTCGTGCAATTATACCTGGCCGCGAGAGCCTTGTCAATGGGCTTC CCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work
SK55	GCGCCTAGGTCGTGCAATTATACCTGGCCGCGAGAGCCTTGTCAATGGGCACA AGCACCTTTCCCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCG ATCG	This work
SK56	GCGCCTAGGTCGTGCAATTATACCTGGCCGCGAGAGCCTTGTCAATGGGCCTG CTGGGACACAAGCACCTTTCCCACGCGAACCCAGCGCCTTCACCGCGATGAAT ACTCCCACCGATCG	This work
SK57	GCGTTAATTAAGCCCGTTG	This work
SK58	GCGCCTAGGTCGTGCAATTATAC	This work
SK63	CGCTTAATTAAGCCCA**N**TGACAAGGC	This work
SK64	CGCTTAATTAAGCCCAT**N**GACAAGGCT	This work
SK65	CGCTTAATTAAGCCCATT**N**ACAAGGCTC	This work
SK66	CGCTTAATTAAGCCCATTG**N**CAAGGCTCT	This work
SK67	CGCTTAATTAAGCCCATTGA**N**AAGGCTCTC	This work
SK68	CGCTTAATTAAGCCCATTGAC**N**AGGCTCTCG	This work
SK69	CGCTTAATTAAGCCCATTGACA**N**GGCTCTCGC	This work
SK70	CGCTTAATTAAGCCCATTGACAA**N**GCTCTCGCG	This work
SK71	CGCTTAATTAAGCCCATTGACAAG**N**CTCTCGCGG	This work
SK72	CGCTTAATTAAGCCCATTGACAAGG**N**TCTCGCGGC	This work
SK73	CGCTTAATTAAGCCCATTGACAAGGC**N**CTCGCGGCC	This work
SK74	CGCTTAATTAAGCCCATTGACAAGGCT**N**TCGCGGCCA	This work
SK75	CGCTTAATTAAGCCCATTGACAAGGCTC**N**CGCGGCCAG	This work
SK76	CGCTTAATTAAGCCCATTGACAAGGCTCT**N**GCGGCCAGG	This work
SK77	CGCTTAATTAAGCCCATTGACAAGGCTCTC**N**CGGCCAGGT	This work
SK78	CGCTTAATTAAGCCCATTGACAAGGCTCTCG**N**GGCCAGGTA	This work
SK79	CGCTTAATTAAGCCCATTGACAAGGCTCTCGC**N**GCCAGGTAT	This work
SK80	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCG**N**CCAGGTATA	This work
SK81	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGG**N**CAGGTATAA	This work
SK82	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGC**N**AGGTATAAT	This work
SK83	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCC**N**GGTATAATT	This work
SK84	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCA**N**GTATAATTG	This work
SK85	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAG**N**TATAATTGC	This work
SK86	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAGG**N**ATAATTGCA	This work
SK87	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAGGT**N**TAATTGCAC	This work
SK88	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAGGTA**N**AATTGCACG	This work
SK89	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAGGTAT**N**ATTGCACGA	This work
SK90	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAGGTATA**N**TTGCACGAC	This work
SK91	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAGGTATAA**N**TGCACGAC C	This work
SK92	CGCTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAGGTATAAT**N**GCACGAC CT	This work
SK93	GCGTTAATTAAGCCCGTTGACATGACATGGTTTTGAGGGTATAATGTGGCGAT TCCCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work
SK94	GCGCCTAGGTCGTGCAATTATACCTGGCCGCGAGAGCCTTGTCAATGGGCGGA CGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCAT	This work
SK97	GCGTTAATTAACTAGGTTGACATGGATATAATGTATGTATTCCCACGCGAACC CAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work
SK98	GCGTTAATTAATTTATTTGACATGCGTGATGTTTAGAATTATAATTTGGGGATT CCCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work
SK99	GCGTTAATTAAGTGAATTGACATGTCAATTTTTATGTTGTATAATATAACTATT CCCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work
SK100	GCGTTAATTAATCTACTTGACATCCGACATTCGCGACTGTATAATAAGTTGATT CCCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work

Name	Sequence 5´-3´[a]	Reference
SK101	GCGTTAATTAACGGAGTTGACAACACTCGAAAAGCCGAGTATAATCAGATGATT CCCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work
SK102	GCGTTAATTAAGCCCATTGACAAGGCTCTCGCGGCCAGGTATAATTGCACGATT CCCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCGATCG	This work
SK103	GCGTTAATTAACTAGGTTGAC	This work
SK104	GCGTTAATTAATTTATTTGACATG	This work
SK105	GCGTTAATTAAGTGAATTGACAT	This work
SK106	GCGTTAATTAATCTACTTGACAT	This work
SK107	GCGTTAATTAACGGAGTTGAC	This work
SK108	GCGTTAATTAAGCCCATTGACA	This work
SK122	GCGCCTAGGTAGTTATATTATACAACATAAAAATTGACATGTCAATTCACGGA CGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCAT	This work
SK123	GCGCCTAGGTACATACATTATATCCATGTCAACCTAGGGACGAGTCACCATGT GCCAGGGGCGATAACGATCGGTGGGAGTATTCAT	This work
SK124	GCGCCTAGGTAGTTATATTA	This work
SK125	GCGCCTAGGTACATACATTATATC	This work
SK135	GCGCCTAGGTCATCTGATTATACTCGGCTTTTCGAGTGTTGTCAACTCCCGGGAC GAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCAT	This work
SK136_	GCGCCTAGGTCAACTTATTATACAGTCGCGAATGTCGGATGTCAAGTAGAGGGA CGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCAT	This work
SK137	GCGCCTAGGTCATCTG	This work
SK138	GCGCCTAGGTCAACTTA	This work
SK145	CGCTTAATTAATTCCCACGCGAACCC	This work
SK146	CGCCCTAGGGGACGAGTCACCATGTG	This work
SK272	GCGTTAATTAAGCCCGTTGACATGACATGGTTTTGAGGGTATAATGTGGCGAA ACTGCCTTGCTTCTCGGTGTGATACCCCCGTCAGCCGCCCGCGGTGGTG	This work
SK273	GCGCCTAGGTCGTGCAATTATACCTGGCCGCGAGAGCCTTGTCAATGGGCGTG CCCAATACTCCAATCGGCTTTCACGTGCACCACCGCGGGCGGCTGAC	This work
SK274	CGCTTAATTAAAACTGCCTTGCTTCTCGGTG	This work
SK275	CGCCCTAGGGTGCCCAATACTCCAATCGG	This work
SK334	AGGCATTCGTGAAGTCATGG	This work
SK335	ATGTAACCGCTGAGAACGTC	This work
SK336	ACTGGGTGGACAGGTAGTGG	This work
SK337	CACCGCAGACAAACAGAAGA	This work

a) Restriction sites *Pac*I and *Avr*II are underlined. Depending on the included promoter, oligonucleotides are colored differently: light blue letters presenting promoter 14a, orange 14b, purple 14c, green 14d, grey 14e, red 14f, and blue 14g. Furthermore, highlighted bolt typed letters are indicating random nucleotides. N stands for nucleotides a, c, g, or t.

Table 3: Oligonucleotide combinations used for the generation of spacer sequences with a length of up to 40 bp.

Distance [bp]	Promoter	Forward oligonucleotide	Reverse oligonucleotide	Template
10	14f_10_14g	SK11	SK2	pBG14g
20	14f_20_14g	SK34	SK2	pBG14g
30	14f_30_14g	SK36	SK2	pBG14g
40	14f_40_14g	SK38	SK2	pBG14g

Table 4: Oligonucleotide combinations used for the generation of spacer sequences with a spacer length from 50 to100 bp. PCR was used to generate dsDNA fragment from two long oligonucleotides containing restrictions sites and promoter sequences. Block forming is achieved by complementary 3´-ends.

Distance [bp]	Promoter	1st PCR Forward oligonucleotide 1st promoter	Reverse oligonucleotide 2nd promoter	2nd PCR Forward oligonucleotide 1st promoter	Reverse oligonucleotide 2nd promoter
50	14f_50_14g	SK50	SK51	SK57	SK58
60	14f_60_14g	SK50	SK52	SK57	SK58
70	14f_70_14g	SK50	SK53	SK57	SK58
80	14f_80_14g	SK50	SK54	SK57	SK58
90	14f_90_14g	SK50	SK55	SK57	SK58
100	14f_100_14g	SK50	SK56	SK57	SK58

Distance [bp]	Promoter	1st PCR Forward oligonucleotide 1st promoter	Reverse oligonucleotide 2nd promoter	2nd PCR Forward oligonucleotide 1st promoter	Reverse oligonucleotide 2nd promoter
80i	14f_80i_14g	SK93	SK94	SK57	SK58
80new	14f_80new_14g	SK272	SK273	SK57	SK58

Table 5: Spacer sequences with different lengths characterized in this study.

Distance [bp]	Sequence
10	CCATAATTAA
20	GGACGAGTCACCATGTGCCA
30	GGACGAGTCACCATGTGCCAGGGGCGATAA
40	GGACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGG
50	GGACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCAT
60	GGACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCATCGCGGTGAAG
70	GGACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCATCGCGGTGAAGGC GCTGGGTT
80	GGACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCATCGCGGTGAAGGC GCTGGGTTCGCGTGGGAA
80i	TTCCCACGCGAACCCAGCGCCTTCACCGCGATGAATACTCCCACCGATCGTTATCGCCCCTGG CACATGGTGACTCGTCC
80new	AACTGCCTTGCTTCTCGGTGTGATACCCCCGTCAGCCGCCCGCGGTGGTGCACGTGAAAGCCG ATTGGAGTATTGGGCAC
90	GGACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCATCGCGGTGAAGGC GCTGGGTTCGCGTGGGAAAGGTGCTTGT
100	GGACGAGTCACCATGTGCCAGGGGCGATAACGATCGGTGGGAGTATTCATCGCGGTGAAGGC GCTGGGTTCGCGTGGGAAAGGTGCTTGTGTCCCAGCAG

Table 6: Oligonucleotide combinations used for the generation of stacked promoters with a spacer distance of 80 bp. PCR was used to generate dsDNA fragments from two long oligonucleotides containing restrictions sites and promoter sequences. Block forming is achieved by complementary 3´-ends.

Promoter combination	1st PCR Forward oligonucleotide 1st promoter	Reverse oligonucleotide 2nd promoter	2nd PCR Forward oligonucleotide 1st promoter	Reverse oligonucleotide 2nd promoter
14a_80i_14g	SK97	SK94	SK103	SK58
14b_80i_14g	SK98	SK94	SK104	SK58
14c_80i_14g	SK99	SK94	SK105	SK58
14d_80i_14g	SK100	SK94	SK106	SK58
14e_80i_14g	SK101	SK94	SK107	SK58
14f_80i_14g	SK50	SK94	SK57	SK58
14g_80i_14g	SK102	SK94	SK108	SK58
14g_80i_14a	SK102	SK123	SK108	SK125
14b_80i_14c	SK98	SK122	SK104	SK124
14e_80i_14d	SK101	SK107	SK136	SK138
14d_80i_14e	SK100	SK106	SK135	SK137

Table 7: Oligonucleotides used for the generation of promoter position controls for two positions and different promoters. Stacking promoter containing vectors were used as a template.

1st and 2nd position promoter control	Forward oligonucleotide	Reverse oligonucleotide	Template
14a_80i	SK43	SK146	pBG14a/14g
14b_80i	SK43	SK146	pBG14b/14c
14c_80i	SK43	SK146	pBG14c/14e
14d_80i	SK43	SK146	pBG14e/14d
14e_80i	SK43	SK146	pBG14d/14e
14f_80i	SK43	SK146	pBG14f/14g

1st and 2nd position promoter control	Forward oligonucleotide	Reverse oligonucleotide	Template
14g_80i	SK43	SK146	pBG14g/14g
14g_1x[a]_80i	SK63	SK146	pBG14g_80
14g_2x[a]_80i	SK64	SK146	pBG14g_80
14g_26x[a]_80i	SK88	SK146	pBG14g_80
14f_80new	SK43	SK275	pBG14f/80new/14g
80_14a	SK145	SK2	pBG14g/14a
80i_14c	SK145	SK2	pBG14b/14c
80i_14d	SK145	SK2	pBG14e/14d
80i_14e	SK145	SK2	pBG14d/14e
80i_14g	SK145	SK2	pBG14g/14g
80new_14g	SK274	SK2	pBG14f/80new/14g

Table 8: Oligonucleotide combinations used for the generation of spacer control without promoters. As a template previously, generated vectors with distinct spacer sequences were used.

Distance [bp] control	Forward oligonucleotide	Reverse oligonucleotide	Template
pBG_80i	SK145	SK146	pBG14f_80i_14g
pBG_80new	SK274	SK275	pBG14f_80new_14g

Table 9: Complete list of tested single nucleotide polymorphism and Zobel *et al.*[27] promoter sequences. 14g represents the original sequence, whereas 14G promoters include changes indicated as lowercase letters. Relative activity to 14G is given in percentage.

Promoter name	Core sequence 5´-3´	Activity relative to 14g [%]	iAMB strain collection
14g	TTGACAAGGCTCTCGCGGCCAGGTATAATT	100±2,7	3232
14a	TTGACATGGATATAATGTATGTA	1±0,1	3228
14G_3_a	TTaACAAGGCTCTCGCGGCCAGGTATAATT	3±0,4	6271
14G_1_c	cTGACAAGGCTCTCGCGGCCAGGTATAATT	4±0,1	6272
14G_1_a	aTGACAAGGCTCTCGCGGCCAGGTATAATT	7±0,4	6273
14G_23_t	TTGACAAGGCTCTCGCGGCCAGtTATAATT	8±0,5	6274
14G_5_g	TTGAgAAGGCTCTCGCGGCCAGGTATAATT	9±0,3	6275
14G_1_g	gTGACAAGGCTCTCGCGGCCAGGTATAATT	14±0,6	6276
14G_4_g	TTGgCAAGGCTCTCGCGGCCAGGTATAATT	17±1,5	6277
14G_2_c	TcGACAAGGCTCTCGCGGCCAGGTATAATT	21±0,4	6278
14b	TTGACATGCGTGATGTTTAGAATTATAATT	23±0,4	3230
14G_25_g	TTGACAAGGCTCTCGCGGCCAGGTgTAATT	30±2,2	6279
14G_2_g	TgGACAAGGCTCTCGCGGCCAGGTATAATT	32±0,8	6280
14c	TTGACATGTCAATTTTTATGTTGTATAATA	35±0,5	3231
14G_2_a	TaGACAAGGCTCTCGCGGCCAGGTATAATT	39±0,5	6281
14G_30_g	TTGACAAGGCTCTCGCGGCCAGGTATAATg	42±1,1	6282
14G_6_c	TTGACcAGGCTCTCGCGGCCAGGTATAATT	45±1,9	6283
14G_24_a	TTGACAAGGCTCTCGCGGCCAGGaATAATT	45±1,9	6284
14G_6_t	TTGACtAGGCTCTCGCGGCCAGGTATAATT	51±2,4	6285
14d	TTGACATCCGACATTCGCGACTGTATAATA	53±0,8	3233
14G_21_t	TTGACAAGGCTCTCGCGGCCtGGTATAATT	54±1,2	6286
14G_9_t	TTGACAAGtCTCTCGCGGCCAGGTATAATT	54±0,5	6287
14G_8_a	TTGACAAaGCTCTCGCGGCCAGGTATAATT	55±0,5	6288
14G_29_g	TTGACAAGGCTCTCGCGGCCAGGTATAAgT	60±0,9	6289
14G_8_t	TTGACAAtGCTCTCGCGGCCAGGTATAATT	62±1,2	6290
14G_20_a	TTGACAAGGCTCTCGCGGCaAGGTATAATT	64±0,9	6291
14G_30_a	TTGACAAGGCTCTCGCGGCCAGGTATAATa	64±2,8	6292
14G_19_a	TTGACAAGGCTCTCGCGGaCAGGTATAATT	65±0,1	6293
14G_10_a	TTGACAAGGaTCTCGCGGCCAGGTATAATT	66±1,2	6294
14G_16_t	TTGACAAGGCTCTCGtGGCCAGGTATAATT	66±1,6	6295
14G_12_a	TTGACAAGGCTaTCGCGGCCAGGTATAATT	67±0,4	6296
14G_12_g	TTGACAAGGCTgTCGCGGCCAGGTATAATT	67±1,1	6297
14G_20_g	TTGACAAGGCTCTCGCGGCgAGGTATAATT	67±1,1	6298

Promoter name	Core sequence 5´-3´	Activity relative to 14g [%]	iAMB strain collection
14G_14_t	TTGACAAGGCTCTtGCGGCCAGGTATAATT	68±0,6	6299
14G_11_g	TTGACAAGGCgCTCGCGGCCAGGTATAATT	68±2,2	6300
14G_12_t	TTGACAAGGCTtTCGCGGCCAGGTATAATT	68±2,5	6301
14G_22_t	TTGACAAGGCTCTCGCGGCCAtGTATAATT	68±3,8	6302
14G_13_g	TTGACAAGGCTCgCGCGGCCAGGTATAATT	68±1,5	6303
14G_21_c	TTGACAAGGCTCTCGCGGCCcGGTATAATT	69±2,6	6304
14G_22_a	TTGACAAGGCTCTCGCGGCCAaGTATAATT	69±2,4	6305
14G_17_c	TTGACAAGGCTCTCGCcGCCAGGTATAATT	70±1,8	6306
14G_10_g	TTGACAAGGgTCTCGCGGCCAGGTATAATT	70±1	6307
14G_21_g	TTGACAAGGCTCTCGCGGCCgGGTATAATT	70±2	6308
14G_17_a	TTGACAAGGCTCTCGCaGCCAGGTATAATT	71±1,5	6309
14G_15_t	TTGACAAGGCTCTCtCGGCCAGGTATAATT	71±0,7	6310
14G_16_a	TTGACAAGGCTCTCGaGGCCAGGTATAATT	72±2,2	6311
14G_7_t	TTGACAtGGCTCTCGCGGCCAGGTATAATT	74±2,8	6312
14G_14_g	TTGACAAGGCTCTgGCGGCCAGGTATAATT	75±3	6313
14G_28_t	TTGACAAGGCTCTCGCGGCCAGGTATAtTT	81±0,5	6314
14G_28_g	TTGACAAGGCTCTCGCGGCCAGGTATAgTT	83±0,9	6315
14G_27_g	TTGACAAGGCTCTCGCGGCCAGGTATgATT	84±0,9	6316
14G_26_a	TTGACAAGGCTCTCGCGGCCAGGTAaAATT	84±1,4	6317
14G_28_c	TTGACAAGGCTCTCGCGGCCAGGTATAcTT	84±4,3	6318
14G_26_c	TTGACAAGGCTCTCGCGGCCAGGTAcAATT	89±1,3	6319
14e	TTGACAACACTCGAAAAGCCGAGTATAATC	91±1,4	3225
14f	TTGACATGACATGGTTTTGAGGGTATAATG	93±2,6	3227

2.2.5 Tn5 library

Sensor construction was done in multiple single steps and was formerly based on plasmid pBELK (Figure 7).[114] Strains used and generated are given in Table 10. Table 11 is listing all used oligonucleotides for this section. In the first step, the kanamycin resistance gene was inverted to avoid a negative effect by the transcription of the gene. Therefore, the kanamycin gene was amplified with pBELK as a template with oligonucleotides SK13 and SK14 using Q5 polymerase (New England Biolabs, Ipswich, Massachusetts, USA). The resulting fragment was equipped with restriction sites for *Hind*III and *Aat*II and contained two FRT sites flanking the kanamycin gene. pBELK and modified resistance gene were digested with *Hind*III and *Aat*II (New England Biolabs, Ipswich, Massachusetts, USA) and ligated using T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA). The orientation of the kanamycin gene in resulting plasmids was checked by colony PCR Taq 2X Master Mix (New England BioLabs, Ipswich, Massachusetts, USA) and oligonucleotides BW62 and SK24. The new plasmid is named pBELK_invKan and was used for further cloning procedures. pBG13 was used as a template for the amplification of P_{em7}_BCD2_*msfGFP* using Q5 polymerase and oligonucleotides SK15 and SK17. New restriction sites on both ends were introduced. At 5´-end *Pac*I and *Nde*I were added, at 3´-end a *Hind*III site. pBELK and the new fragment containing a promoter and reporter gene were digested with *Pac*I and *Hind*III (New England Biolabs, Ipswich, Massachusetts, USA) and ligated with T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA). Resulting colonies were visually characterized by green fluorescence using a blue light illuminator. Glowing clones were sequenced

using oligonucleotide SK24 and the new plasmid is named pBELK_P_{em7}_BCD2_msfGFP_invKan. For a possible plasmid rescue, a replication origin ColE1 for in *E. coli* was introduced by restriction and ligation. pSEVA247Y was used as a template for PCR with oligonucleotides SK20 and SK21. The resulting fragment is modified by two *Hind*III restriction sites at both ends. pBELK_P_{em7}_BCD2_msfGFP_invKan and the PCR product were digested by *Hind*III and ligated with T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA). Arising clones were checked by colony PCR for the presence of ColE1 using Taq 2X Master Mix (New England BioLabs, Ipswich, Massachusetts, USA) and oligonucleotides SK23 and SK24. Positive clones were sequenced and the new plasmid is called pBELK_P_{em7}_BCD2_msfGFP_ColE1_invKan. A fragment containing a double terminator M13_*rrnD1*[78] fused with MCS_RBS_mCherry was synthesized by IDT (Integrated DNA Technologies) and blunt ligated with pJet (Thermo Fisher Scientific), resulting in pJet::M13rrnD_MCS_mCherry (Thermo Fisher Scientific). Isolated plasmids were digested with *Pac*I and *Nde*I (New England Biolabs, Ipswich, Massachusetts, USA), alike transposon containing target plasmid pBELK_P_{em7}_BCD2_msfGFP_ColE1_invKan. After ligation with T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA) growing clones were checked by sequencing and the new transposon plasmid is called pTn5_sensor. All PCR fragments used were cut out from agarose gels and purified using DNA Gel Extraction kit (New England Biolabs, Ipswich, Massachusetts, USA). Concentrations were measured with a NanoDrop One (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

pTn5_sensor vector was integrated by two types of triparental mating using *P. putida* KT2440 as recipient strain and *E. coli* HB101 pRK600 as helper strain. To test whether the integration of our pTn5_sensor construct into the genome is capable, we firstly performed a patch mating. Here *P. putida* KT2440, *E. coli* HB101 pRK600, and *E. coli* PIR2 pTn5_sensor were patched on a LB agar plate and cultivated overnight at 30°C. Growing patch matings were stroked on cetrimide with 1 % glycerol as carbon source and 50 mg×L^{-1} kanamycin. Arising colonies were picked on cetrimide with 1 % glycerol as carbon source and 50 mg×L^{-1} kanamycin and further analyzed. The generation of a Tn5 library was done by triparental mating using liquid cultures of all three strains. LB liquid precultures of each strain were cultivated overnight at the appropriate temperature. Main cultures of 50 mL LB liquid medium, plasmid bearing strains with additional antibiotics, were inoculated with an initial optical density of 0.1 measured at 600 nm. The optical density was measured at regular intervals until all strains reach an OD_{600} of 3. Strains were adjusted regarding the OD_{600} value to gain the same cell amount of each partner for the matings and were merged in 50 mL tubes. Pelleting of the cells was done by centrifugation at 5,000 rpm for 5 minutes. The supernatant was discarded and the resulting pellet resuspended in 2.5 mL LB liquid medium.

500 µL of the cell mixture was dropped on dried LB agar plates and incubated overnight at 30°C. In total 20 matings were done following the procedure described above. Mating spots were scratched from the plate with a sterile inoculation loop and five spots sampled in a single 50 mL tube. Cells were resolved in 5 mL 1xKPi puffer (3.88g K_2HPO_4, NaH_2PO_4, pH 7).[138] Cells were earned by centrifugation at 5,000 rpm for 5 minutes, the supernatant was discarded and the pellet with 2 mL 1xKPi puffer[138] resuspended. 400 µL cells were spread out on 145 mm Petri dishes containing cetrimide with 1 % glycerol as carbon source and 50 mg×L^{-1} kanamycin for selection on Tn5 carrying *P. putida* KT2440 strains. After overnight cultivation at 30°C, the cells were washed from the plates with 4 mL 1xKPi puffer[138] and all liquids collected in a 50 mL tube. Before washing up, colonies on several Petri dishes in an area of 6.25 cm^2 were counted to determine the number of Tn5 carrying cells. Cells were pelleted by centrifugation at 5,000 rpm for 10 minutes. The supernatant was discarded and the pellet resuspended with 2.5 mL 1x KPi puffer. 1.4 mL of the cell mixture was mixed with 600 µL 50 % glycerol and stored at -80°C. This procedure was done with at least 43 matings in parallel and was done twice to reach enough mutants. Mutant number and background were determined by dilution series spread on LB agar plates and cetrimide agar plate containing 1 % glycerol as carbon source and 50 mg×L^{-1} kanamycin.

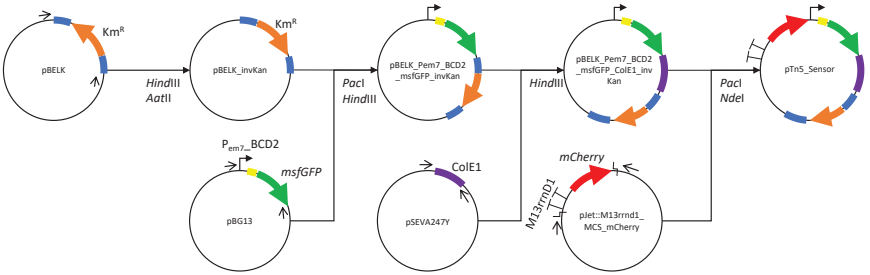


Figure 7: Schematic presentation of cloning procedures to generate a sensor construct in transposon Tn5 bearing pBELK. Parts that were used for cloning are shown in different colors. Blue boxes highlighting FRT sites for FRT/FLP recombination and orange indicates a kanamycin resistance gene from pBELK. From pBG13, we amplified P_{em7} promoter (standing arrow) followed by BCD2 shown as yellow box and reporter gene *msfGFP* in green. Terminators are indicated as "T" and are directly fused with MCS, RBS, and *mCherry* shown in red, the fragment was synthesized by IDT and subcloned into pJET. From pSEVA247Y, we amplified ColE1 (purple) for a possible plasmid rescue. Small arrows on the left and right site of highlighted genes showing oligonucleotides for amplification of desired sequences. Serrated strokes showing restriction sites used to gain a fragment directly from a plasmid. The resulting plasmid is named pBELK_M13rrnD1_MCS_RBS_mCherry_P_{em7}_BCD2_msfGFP_ColE1_invKan. In this study, the shorter name pTn5_sensor was used.

Chosen Tn5_sensor carrying strains were tested for the site of integration by arbitrary-primed PCR as describe by Nikel *et al.*[114]. For the first PCR reaction, a randomized oligonucleotide arb6 and a normal oligonucleotide SK59 were used. The second step contained a nested PCR reaction and was performed with arb2 and SK60 (Table 12). PCR reactions were done with Q5 polymerase (New England Biolabs, Ipswich, Massachusetts, USA) and resulting fragment cut out from agarose

gels. Purification was done with DNA Gel Extraction kit (New England Biolabs, Ipswich, Massachusetts, USA) and sequenced with oligonucleotide arb2.

Table 10: Strains and plasmids used and generated in this section for the construction of a Tn5 based sensor and resulting in Tn5 bearing strains.

Strain	Description	Reference	iAMB strain collection
E. coli			
HB101	*F⁻ mcrB mrr hsdS20*(*rB⁻ mB⁻*) *recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^R) *gln V44λ⁻*	Boyer and Roulland-Dussoix[140]	2073
CC118λpir	Δ(*ara-leu*) *araD ΔlacX74 galE galK phoA20 thi-1 rpsE rpoB argE*(*Am*) *recA1,* lysogenized with λpir phage	Herrero *et al.*[143]	2072
PIR2	F- Δ*lac169 rpoS* (Am) *robA1 creC510 hsdR514 endA reacA1 uidA* (Δ*Mlui*)::*pir*	Life Technologies	3222
P. putida			
KT2440	Wild-type strain derived of *P. putida* mt-2 cured of the pWW0 plasmid	Bagdasarian *et al.*[144]	2058
BG_sensor	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_sensor at *attTn7* site	This work	6320
Tn5_1	Tn5 integration strain in 5736423-5736644, inside *hutU*, Km^R	This work	TBA
Tn5_2	Tn5 integration strain in 6113184-6113031, inside *clsA*, Km^R	This work	TBA
Tn5_3	Tn5 integration strain in one of the rRNA 23S ribosomal RNA genes, Km^R	This work	TBA
Tn5_10	Tn5 integration strain in one of the rRNA 16S ribosomal RNA genes, Km^R	This work	TBA
Tn5_12	Tn5 integration strain in one of the rRNA 23S ribosomal RNA genes, Km^R	This work	TBA
Plasmids		This work	
pRK600	Cm^R, oriColE1, *tra* + *mob* + of RK2	Keen *et al.*[141]	2073
pBELK	Mini-Tn5 delivery vector; *tnpA oriV*(R6Kγ) *oriT*(RK2), *lacIQ,* P_{trc}, *bla,* FRT-*aphA*-FRT, Ap^R, Km^R	Nikel *et al.* [114]	2255
pBELK_invKan	Mini-Tn5 delivery vector; *tnpA oriV*(R6Kγ) *oriT*(RK2), *lacIQ,* P_{trc}, *bla,* FRT-inverted_*aphA*-FRT, Ap^R, Km^R	This work	6534
pBELK_P_{em7}_BCD2_ msfGFP_invKan	Mini-Tn5 delivery vector; *tnpA oriV*(R6Kγ) *oriT*(RK2), *lacIQ,* promoter P_{em7} with BCD2 and *msfGFP*, *FRT-inverted_aphA-FRT*, Ap^R, Km^R	This work	6535
pSEVA247Y	pRO1600/ColE1, oriT, *YFP*, Km^R,	Silva-Rocha *et al.*[34]	2393
pBELK_P_{em7}_BCD2_ msfGFP_ColE1_invKan	Mini-Tn5 delivery vector; *tnpA oriV*(R6Kγ) *oriT*(RK2), *lacIQ,* promoter P_{em7} with BCD2 and *msfGFP*, ColE1, FRT-inverted_*aphA*-FRT, Ap^R, Km^R	This work	6536
pJet::M13rrnD_MCS_ mCherry	pMB1, Ap^R, M13rrnD1 double terminator, MCS, RBS mCherry	This work	6537
pTn5_sensor	Mini-Tn5 delivery vector; *tnpA oriV*(R6Kγ) *oriT*(RK2), *lacIQ,* M13rrnD1 double terminator, MCS, RBS mCherry, promoter P_{em7} with BCD2 and *msfGFP*, ColE1, *FRT-inverted_aphA-FRT*, Ap^R, Km^R	This work	6538
pBG_sensor	Km^R, Gm^R, *ori*R6K, pBG-derived, M13rrnD1 double terminator, MCS, RBS mCherry, P_{em7} BCD2 *msfGFP*,	This work	6539

Table 11: Oligonucleotides used for the construction of Tn5_sensor in this section. Restriction sites are underlined.

Name	Sequence 5´-3´[a]	Reference
SK13	CGCGACGTCCTTGGACTCCTGTTGATAGAT	This work
SK14	CGCAAGCTTGAAGTTCCTATACTTTCTAGAGAATA	This work
SK15	CGCAAGCTTGCAATTCTTATTTGTAGAGTTCATCCATG	This work

Name	Sequence 5´-3´[a]	Reference
SK17	GCGTTAATTAAAACATATGTTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCCCTAGGCCCAAGTTCACTTAAAAAG	This work
SK19	AAGTAGTCGGGGATGTCGG	This work
SK20	CGCCGCAAGCTTCGATAATCTCATGACCAAAA	This work
SK21	CGCGCGAAGCTTCCGCGTTGCTGGCGTTTTTC	This work
SK23	TCCTCTAGAGTCGACCTGCA	This work
SK24	TGCAGTTTGTTCATATCCGGAT	This work
SK60	GCCGGAACTGTTTCTGAAACAT	This work
arb6	GGCACGCGTCGACTAGTACNNNNNNNNNNACGCC	Nikel *et al.*[114]
arb2	GGCACGCGTCGACTAGTAC	Nikel *et al.*[114]
BW62	CATCCGGCTCGTATAATG	Provided by B. Wynands

Table 12: PCR program used for the amplification of flanking sequences to identify transposon Tn5 integration sites in *P. putida* KT2440

PCR step	Program	Oligonucleotide
1st	94°C 5:00 Min 6x 94°C 0:20 Min; 30°C 0:30 Min; 72°C 1:00 Min 30x 94°C 0:30 Min; 30°C 0:30 Min; 72°C 2:00 Min 72°C 5:00 Min	arb6 + 59
2nd	94°C 1:00 Min 30x 94°C 0:30 Min; 50°C 0:30 Min; 72°C 2:00 Min 72°C 5:00 Min	arb2 + 60

2.2.6 Genomic integration sites

All strains used and generated are listed in Table 13. Oligonucleotides are given in Table 14. The construction of integrable pEMG vectors was done with a minimal set of oligonucleotides by using the same overhangs (Table 15, Figure 8). All forward and reverse oligonucleotides of corresponding target sequence elements, the forward oligonucleotide for the sensor and the reverse one for the terminator are sharing overhang sequences for Gibson assembly. Target sequences (TS elements) for the construction of pEMG integration vectors were amplified from *P. putida* KT2440 genomic DNA (oligonucleotide combinations are listed in the supplement). Amplified TS elements were purified using PCR & DNA Cleanup Kit (New England Biolabs, Ipswich, Massachusetts, USA). The sensor was amplified from pTn5_sensor using Q5 polymerase (New England Biolabs, Ipswich, Massachusetts, USA) and oligonucleotides SK190 and SK191. An additional terminator was also amplified from pTn5_sensor as a template using Q5 polymerase (New England Biolabs, Ipswich, Massachusetts, USA) and oligonucleotides SK192 and SK193. Both fragments were cut out of agarose gels and purified with DNA Gel Extraction kit (New England Biolabs, Ipswich, Massachusetts, USA). pEMG backbone was digested with *Eco*RI and *Sal*I (New England Biolabs, Ipswich, Massachusetts, USA). Concentrations of TS elements, digested pEMG backbone, sensor, and terminator fragment were measured with a NanoDrop One (Thermo Scientific, Waltham, Massachusetts, USA). Assembly of the different pEMG plasmids was done by Gibson Assembly. Following the manual of the NEBuilder Hifi DNA Assembly Master Mix (New England Biolabs, Ipswich, Massachusetts, USA). The total volume of the

reaction was reduced to 10 µL. The whole assembly mix was transformed to chemically competent[139] *E. coli* PIR2 (Life Technologies, Carlsbad, USA) cells and spread on LB agar plates containing 50 mg×L^{-1} kanamycin. Arising clones were picked and checked by colony PCR using *Taq* 2X Master Mix (New England BioLabs, Ipswich, Massachusetts, USA) 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 vectors (XX stand for different genomic regions used for integration) was done by restriction and ligation. Promoterless control BG_BCD2_*msfGFP*, P_{em7}_BCD2_*msfGFP* and P_{14g}_BCD2_*msfGFP* were amplified from corresponding mini Tn7 plasmids pBG. pBG13 and pBG14g using Q5 polymerase (New England Biolabs, Ipswich, Massachusetts, USA) and oligonucleotides 43 and BW62. The resulting fragment was cut out of an agarose gel and purified using DNA Gel Extraction kit (New England Biolabs, Ipswich, Massachusetts, USA). pEMG vectors and PCR fragments were digested with *Pac*I and *Spe*I and ligated afterwards using T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA). Growing clones were tested by colony PCR using oligonucleotides BW13 and BW14. Correct clones were sequenced and genomically integrated in *P. putida* KT2440 by triparental mating.

Table 13: Strains and plasmids used and generated for the characterization of genomic integration sites in this section.

Strain	Description	Reference	iAMB strain collection
E. coli			
HB101	*F⁻ mcrB mrr hsdS20(rB⁻ mB⁻) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^R) *gln V44λ⁻*	Boyer and Roulland-Dussoix[140]	2073
CC118λpir	Δ(*ara-leu*) *araD* Δ*lacX74 galE galK phoA20 thi-1 rpsE rpoB argE*(*Am*) *recA1,* lysogenized with λpir phage	Herrero *et al.*[143]	2072
PIR2	F- Δ*lac169 rpoS* (Am) *robA1 creC510 hsdR514 endA reacA1 uidA* (Δ*Mlui*)::*pir*	Life Technologies	3222
DH5αλpir	*endA1 hsdR17 glnV44* (= *supE44*) *thi-1 recA1 gyrA96 relA1* φ*80dlac*Δ(*lacZ*)*M15* Δ(*lacZYA-argF*)*U169 zdg-232*::Tn10 *uidA*::*pir+*	de Lorenzo lab	2673
P. putida			
KT2440	Wild-type strain derived of *P. putida* mt-2 cured of the pWW0 plasmid	Bagdasarian *et al.*[144]	2058
BG	Gm^R, *P. putida* KT2440 with genomic insertion of pBG at *attTn7* site	Zobel *et al.*[27]	3234
BG13	Gm^R, *P. putida* KT2440 with genomic insertion of pBG13 at *attTn7* site and *msfGFP*	Zobel *et al.*[27]	3235
BG13_mCherry	Gm^R, *P. putida* KT2440 with genomic insertion of pBG13 at *attTn7* site and *mCherry*	This study	6321
BG14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g at *attTn7* site	Zobel *et al.*[27]	3232
BG_sensor	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_sensor at *attTn7* site	This work	6320
PP0145_sensor	Marker-less genomic integration of pEMG_PP0145_sensor in *P. putida* KT2440, PP0146::sensor::PP0145	This work	6322

Strain	Description	Reference	iAMB strain collection
PP0145_BG	Marker-less genomic integration of pEMG_PP0145_BG *P. putida* KT2440, PP0146::BG::PP0145	This work	6323
PP0145_P$_{em7}$	Marker-less genomic integration of pEMG_PP0145_P$_{em7}$ in *P. putida* KT2440, PP0146::P$_{em7}$::PP0145	This work	6324
PP0145_P$_{14g}$	Marker-less genomic integration of pEMG_PP0145_P$_{14g}$ in *P. putida* KT2440, PP0146::P$_{14g}$::PP0145	This work	6325
PP0340_sensor	Marker-less genomic integration of pEMG_PP0340_sensor in *P. putida* KT2440, PP0340::sensor::PP0341	This work	6326
PP0340_BG	Marker-less genomic integration of pEMG_PP0340_BG between in *P. putida* KT2440, PP0340::BG::PP0341	This work	6327
PP0340_P$_{em7}$	Marker-less genomic integration of pEMG_PP0340_P$_{em7}$ in *P. putida* KT2440, PP0340::P$_{em7}$::PP0341	This work	6328
PP0340_P$_{14g}$	Marker-less genomic integration of pEMG_PP0340_P$_{14g}$ in *P. putida* KT2440, PP0340::P$_{14g}$::PP0341	This work	6329
PP0480_sensor	Marker-less genomic integration of pEMG_PP0480_sensor in *P. putida* KT2440, PP0480::sensor::P0481	This work	6330
PP0480_BG	Marker-less genomic integration of pEMG_PP0480_BG in *P. putida* KT2440, PP0480::BG::P0481	This work	6331
PP0480_P$_{em7}$	Marker-less genomic integration of pEMG_PP0480_P$_{em7}$ in *P. putida* KT2440, PP0480::P$_{em7}$::P0481	This work	6332
PP0480_P$_{14g}$	Marker-less genomic integration of pEMG_PP0480_P$_{14g}$ in *P. putida* KT2440, PP0480::P$_{14g}$::P0481	This work	6333
PP1430_sensor	Marker-less genomic integration of pEMG_PP1430_sensor in *P. putida* KT2440, PP1430::sensor::PP1431	This work	6334
PP1430_BG	Marker-less genomic integration of pEMG_PP1430_BG in *P. putida* KT2440, PP1430::BG::PP1431	This work	6335
PP1430_P$_{em7}$	Marker-less genomic integration of pEMG_PP1430_P$_{em7}$ in *P. putida* KT2440, PP1430::P$_{em7}$::PP1431	This work	6336
PP1430_P$_{14g}$	Marker-less genomic integration of pEMG_PP1430_P$_{14g}$ in *P. putida* KT2440, PP1430::P$_{14g}$::PP1431	This work	6337
PP1738_sensor	Marker-less genomic integration of pEMG_PP1738_sensor in *P. putida* KT2440, PP1738::sensor::PP1739	This work	6338
PP1738_BG	Marker-less genomic integration of pEMG_PP1738_BG in *P. putida* KT2440, PP1738::BG::PP1739	This work	6339
PP1738_P$_{em7}$	Marker-less genomic integration of pEMG_PP1738_P$_{em7}$ in *P. putida* KT2440, PP1738::P$_{em7}$::PP1739	This work	6340
PP1738_P$_{14g}$	Marker-less genomic integration of pEMG_PP1738_P$_{14g}$ in *P. putida* KT2440, PP1738::P$_{14g}$::PP1739	This work	6341
PP3304_sensor	Marker-less genomic integration of pEMG_PP3304_sensor in *P. putida* KT2440, PP3304::sensor::PP3305	This work	6342
PP3304_BG	Marker-less genomic integration of pEMG_PP3304_BG in *P. putida* KT2440, PP3304::BG::PP3305	This work	6343
PP3304_P$_{em7}$	Marker-less genomic integration of pEMG_PP3304_P$_{em7}$ in *P. putida* KT2440, PP3304::P$_{em7}$::PP3305	This work	6344
PP3304_P$_{14g}$	Marker-less genomic integration of pEMG_PP3304_P$_{14g}$ in *P. putida* KT2440, PP3304::P$_{14g}$::PP3305	This work	6345
PP3761_sensor	Marker-less genomic integration of pEMG_PP3761_sensor in *P. putida* KT2440, PP3761::sensor::PP3762	This work	6346
PP3761_BG	Marker-less genomic integration of pEMG_PP3761_BG between in *P. putida* KT2440, PP3761::BG::PP3762	This work	6347
PP3761_P$_{em7}$	Marker-less genomic integration of pEMG_PP3761_P$_{em7}$ in *P. putida* KT2440, PP3761::P$_{em7}$::PP3762	This work	6348
PP3761_P$_{14g}$	Marker-less genomic integration of pEMG_PP3761_P$_{14g}$ in *P. putida* KT2440, PP3761::P$_{14g}$::PP3762	This work	6349
PP4145_sensor	Marker-less genomic integration of pEMG_PP4145_sensor in *P. putida* KT2440, PP4145::sensor::PP4146	This work	6350
PP4145_BG	Marker-less genomic integration of pEMG_PP4145_BG in *P. putida* KT2440, PP4145::BG::PP4146	This work	6351
PP4145_P$_{em7}$	Marker-less genomic integration of pEMG_PP4145_P$_{em7}$ in *P. putida* KT2440, PP4145::P$_{em7}$::PP4146	This work	6352

Strain	Description	Reference	iAMB strain collection
PP4145_P_{14g}	Marker-less genomic integration of pEMG_PP4145_P_{14g} in *P. putida* KT2440, PP4145::P_{14g}::PP4146	This work	6353
PP4832_sensor	Marker-less genomic integration of pEMG_PP4832_sensor in *P. putida* KT2440, PP4832::sensor::PP4833	This work	6354
PP4832_BG	Marker-less genomic integration of pEMG_PP4832_BG in *P. putida* KT2440, PP4832::BG::PP4833	This work	6355
PP4832_P_{em7}	Marker-less genomic integration of pEMG_PP4832_P_{em7} in *P. putida* KT2440, PP4832::P_{em7}::PP4833	This work	6356
PP4832_P_{14g}	Marker-less genomic integration of pEMG_PP4832_P_{14g} in *P. putida* KT2440, PP4832::P_{14g}::PP4833	This work	6357
PP4945_sensor	Marker-less genomic integration of pEMG_PP4945_sensor in *P. putida* KT2440, PP4945::sensor::PP4944	This work	6358
PP4945_BG	Marker-less genomic integration of pEMG_PP4945_BG in *P. putida* KT2440, PP4945::BG::PP4944	This work	6359
PP4945_P_{em7}	Marker-less genomic integration of pEMG_PP4945_P_{em7} in *P. putida* KT2440, PP4945::P_{em7}::PP4944	This work	6360
PP4945_P_{14g}	Marker-less genomic integration of pEMG_PP4945_P_{14g} in *P. putida* KT2440, PP4945::P_{14g}::PP4944	This work	6361
PP2322_sensor	Marker-less genomic integration of pEMG_PP2322_sensor in *P. putida* KT2440, PP2322::sensor::PP2321	This work	6366
PP2322_BG	Marker-less genomic integration of pEMG_PP2322_BG in *P. putida* KT2440, PP2322::BG::PP2321	This work	6367
PP2322_P_{em7}	Marker-less genomic integration of pEMG_PP2322_P_{em7} in *P. putida* KT2440, PP2322::P_{em7}::PP2321	This work	6368
PP2322_P_{14g}	Marker-less genomic integration of pEMG_PP2322_P_{14g} in *P. putida* KT2440, PP2322::P_{14g}::PP2321	This work	6369
PP3808_sensor	Marker-less genomic integration of pEMG_PP3808_sensor in *P. putida* KT2440, PP3808::sensor::PP3807	This work	6370
PP3808_BG	Marker-less genomic integration of pEMG_PP3808_BG in *P. putida* KT2440, PP3808::BG::PP3807	This work	6371
PP3808_P_{em7}	Marker-less genomic integration of pEMG_PP3808_P_{em7} in *P. putida* KT2440, PP3808::P_{em7}::PP3807	This work	6372
PP3808_P_{14g}	Marker-less genomic integration of pEMG_PP3808_P_{14g} in *P. putida* KT2440, PP3808::P_{14g}::PP3807	This work	6373
PP4709_sensor	Marker-less genomic integration of pEMG_PP4709_sensor in *P. putida* KT2440, PP4709::sensor::PP4708	This work	6374
PP4709_BG	Marker-less genomic integration of pEMG_PP4709_BG in *P. putida* KT2440, PP4709::BG::PP4708	This work	6375
PP4709_P_{em7}	Marker-less genomic integration of pEMG_PP4709_P_{em7} in *P. putida* KT2440, PP4709::P_{em7}::PP4708	This work	6376
PP0340_sensor+ PP3808_P_{em7}	Marker-less genomic integration of pEMG_PP0340_sensor, PP0340::sensor::PP0341, and marker-less genomic integration of pEMG_PP3808_P_{em7}, PP3808::P_{em7}::PP3807, in *P. putida* KT2440	This work	6413
PP0340_sensor+ PP4945_sensor	Marker-less genomic integration of pEMG_PP0340_sensor, PP0340::sensor::PP0341, and marker-less genomic integration of pEMG_PP4945_sensor, PP4945::sensor::PP4944, in *P. putida* KT2440	This work	6414
PP3761_sensor+ PP4945_sensor	Marker-less genomic integration of pEMG_PP3761_sensor, PP3761::sensor::PP3762, and marker-less genomic integration of pEMG_PP4945_sensor, PP4945::sensor::PP4944, in *P. putida* KT2440	This work	6415
PP3808_P_{em7}+ PP4945_sensor	Marker-less genomic integration of pEMG_PP3808_P_{em7}, PP3808::P_{em7}::PP3807, and marker-less genomic integration of pEMG_PP4945_sensor, PP4945::sensor::PP4944, in *P. putida* KT2440	This work	6416
PP0340_sensor+ BG13_mCherry	Marker-less genomic integration of pEMG_PP0340_sensor, PP0340::sensor::PP0341 and genomic insertion of pBG13_mCherry at *attTn7* site, Gm^R, in *P. putida* KT2440	This work	6417
Plasmids			

Strain	Description	Reference	iAMB strain collection
pRK600	Cm^R, *ori*ColE1, *tra* + *mob* + of RK2	Keen *et al.*[141]	2073
pTnS-1	Ap^R, *ori*R6K, *TnSABC+D* operon	Choi *et al.*[115]	3221
pSW-2	Gm^R, *ori*RK2, *xylS*, $P_m \rightarrow$ I-*Sce*I (transcriptional fusion of I-*Sce*I to Pm)	Martínez-Garcia *et al.*[26]	2404
pBBFLP	Helper plasmid used for antibiotic markers excision in *P. putida* strains; *oriV*(pBBR1) *oriT*(RK2) RK2 mob^+ λP_R::FLP λ(cI857) *sacB tet*, Tc^R	De las Heras *et al.*[145]	3607
pEMG	Km^R, *ori*R6K, *lacZα* with two flanking I-*Sce*I sites	Martínez-Garcia *et al.*[26]	2075
pBG	Km^R, Gm^R, *ori*R6K, Tn7L and Tn7R extremes, BCD2−*msfgfp* fusion	Zobel *et al.*[27]	3210
pBG13	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter P_{em7}, *msfGFP*	Martínez-Garcia *et al.*[26]	3211
pBG13_mCherry	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter P_{em7}, *mCherry*	This study	6540
pBG14g	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter P_{14g}, *msfGFP*	Zobel *et al.*[27]	3219
pEMG_PP0145_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP0145 and PP0146, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in – strand orientation,	This work	6541
pEMG_PP0145_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP0145 and PP0146, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in – strand orientation,	This work	6542
pEMG_PP0145_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP0145 and PP0146, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in – strand orientation,	This work	6543
pEMG_PP0145_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP0145 and PP0146, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in – strand orientation,	This work	6544
pEMG_PP0340_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP0340 and PP0341, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter /P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6545
pEMG_PP0340_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP0340 and PP0341, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6546
pEMG_PP0340_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP0340 and PP0341, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6547
pEMG_PP0340_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP0340 and PP0341, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6548
pEMG_PP480_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP480 and PP481, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation	This work	6549

Strain	Description	Reference	iAMB strain collection
pEMG_PP480_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP480 and PP481, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6550
pEMG_PP480_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP480 and PP481, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6551
pEMG_PP480_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP480 and PP481, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6552
pEMG_PP1430_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP1430 and PP1431, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6553
pEMG_PP1430_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP1430 and PP1431, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6554
pEMG_PP1430_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP1430 and PP1431, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6555
pEMG_PP1430_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP1430 and PP1431, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6556
pEMG_PP1738_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP1738 and PP1739, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6557
pEMG_PP1738_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP1738 and PP1739, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6558
pEMG_PP1738_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP1738 and PP1739, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6559
pEMG_PP1738_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP1738 and PP1739, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6560
pEMG_PP3304_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3304 and PP3305, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoterP_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6561
pEMG_PP3304_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3304 and PP3305, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6562

Strain	Description	Reference	iAMB strain collection
pEMG_PP3304_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3304 and PP3305, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6563
pEMG_PP3304_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3304 and PP3305, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6564
pEMG_PP3761_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3761 and PP3762, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6565
pEMG_PP3761_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3761 and PP3762, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6566
pEMG_PP3761_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3761 and PP3762, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6567
pEMG_PP3761_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3761 and PP3762, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6568
pEMG_PP4145_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4145 and PP4146, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6569
pEMG_PP4145_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4145 and PP4146, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6570
pEMG_PP4145_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4145 and PP4146, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6571
pEMG_PP4145_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4145 and PP4146, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6572
pEMG_PP4832_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4832 and PP4833, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6573
pEMG_PP4832_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4832 and PP4833, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in + strand orientation	This work	6574
pEMG_PP4832_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4832 and PP4833, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6575

Strain	Description	Reference	iAMB strain collection
pEMG_PP4832_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4832 and PP4833, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in + strand orientation,	This work	6576
pEMG_PP4945_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4945 and PP4944, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6577
pEMG_PP4945_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4945 and PP4944, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6578
pEMG_PP4945_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4945 and PP4944, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6579
pEMG_PP4945_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4945 and PP4944, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6580
pEMG_PP2322_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP2322 and PP2321, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6585
pEMG_PP2322_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP2322 and PP2321, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6586
pEMG_PP2322_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP2322 and PP2321, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6587
pEMG_PP2322_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP2322 and PP2321, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6588
pEMG_PP3808_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3808 and PP3807, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6589
pEMG_PP3808_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3808 and PP3807, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6590
pEMG_PP3808_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3808 and PP3807, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6591
pEMG_PP3808_P_{14g}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP3808 and PP3807, TS elements flanked by I-*Sce*I sites, promoter P_{14g} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6592

Strain	Description	Reference	iAMB strain collection
pEMG_PP4709_sensor	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4709 and PP4708, TS elements flanked by I-*Sce*I sites, containing a sensor construct with a combined terminator from M13 and *E. coli* S12 *rrnD1*, promoterless *mCherry*, promoter P_{em7} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6593
pEMG_PP4709_BG	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4709 and PP4708, TS elements flanked by I-*Sce*I sites, promoterless with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6594
pEMG_PP4709_P_{em7}	Km^R, *ori*R6K, carrying TS1 and TS2 for interspacing sequence of PP4709 and PP4708, TS elements flanked by I-*Sce*I sites, promoter P_{em7} with BCD2 element and *msfGFP*, integration in - strand orientation,	This work	6595

Table 14: Oligonucleotides used for the construction of pEMG vectors for targeted integration.

Name	Sequence 5´-3´[a]	Reference
SK2	ACACCATAGGTCAGGGTAGTC	This work
SK4	AGTCAGAGTTACGGAATTGTAGG	Zobel *et al.*[27]
SK5	GTCGAGAAAATTGCCGAGCT	Zobel *et al.* [27]
SK13	CGCGACGTCCTTGGACTCCTGTTGATAGAT	This work
SK25	TCTCTGATGTTACATTGCACAAG	This work
SK59	CCGCGTCTGAACAGCAAC	This work
SK60	GCCGGAACTGTTTCTGAAACAT	This work
BW13	TTTGCACTGCCGGTAGAAC	Provided by B. Wynands
BW14	AATACGCAAACCGCCTCTC	Provided by B. Wynands
BW62	CATCCGGCTCGTATAATG	Provided by B. Wynands
BW63	CCGAGCGTTCTGAACAAATC	Provided by B. Wynands
SK190	CTTTAATTAAAAAGCAAGCTGATAAACCGATACAATT	This work
SK191	TGAAGTTCCTCTTGCAATTCTTATTTGTAGAGTTCATC	This work
SK192	GAATTGCAAGAGGAACTTCACTAGTCTGGATTCTCAC	This work
SK193	ATCTGACGTCCTTGGACTCCTGTTGATAG	This work
SK198	GGAGTCCAAGGACGTCAGATGCTTGCCCGCGAAGCAGC	This work
SK199	GAAGCTTGCATGCCTGCAGGCAGGTCGAAGGTCAGTACCCG	This work
SK200	AGGGATAACAGGGTAATCTGAGGTGGCAAGTTCTTCTACAC	This work
SK201	AGCTTGCTTTTTAATTAAAGCACCACAGTGACAACGCG	This work
SK240	CCAGGCTGGAATAGGAAATC	This work
SK241	TCCTCGGTATCGACAACATC	This work
SK202	AGGGATAACAGGGTAATCTGGGCCCGCGACCTGGAAAA	This work
SK203	AGCTTGCTTTTTAATTAAAGCACTGCTCAGGGCCTGTG	This work
SK204	GGAGTCCAAGGACGTCAGATTGCAATCCCTGTGGGAGC	This work
SK205	GAAGCTTGCATGCCTGCAGGTTGCGCACATCGTTCAGC	This work
SK242	TCGGTTACGGCAAGATGGG	This work
SK243	ACCGGGTGATCGTTCTTCG	This work
SK206	AGGGATAACAGGGTAATCTGCGACCGCCTGATCGTCGT	This work
SK207	AGCTTGCTTTTTAATTAAAGATAGATGTATAAACAAAAAACCGGGCACTAG	This work
SK208	GGAGTCCAAGGACGTCAGATTGTTTATTTCTATTGATGTAAGTCATG	This work
SK209	GAAGCTTGCATGCCTGCAGGCTGGAGTGTTGTTACCGAC	This work
SK244	TCGACCTCAAAGGTGACAG	This work
SK245	GTCTTGAAGTGCCACTTGAC	This work
SK210	AGGGATAACAGGGTAATCTGTCGTCATGTCGGCCGATC	This work
SK211	AGCTTGCTTTTTAATTAAAGTTCATGAAAAAGGGCAGCC	This work
SK212	GGAGTCCAAGGACGTCAGATTCTTCACAATTGCCGGAATG	This work
SK213	GAAGCTTGCATGCCTGCAGGATGGCGGTGTAGCAGTTG	This work

Name	Sequence 5´-3´[a]	Reference
SK246	CCCGCTGTTCAACATGAAAG	This work
SK247	ACCACACCCAGGTAATTGTC	This work
SK214	AGGGATAACAGGGTAATCTGAGCTGGTGGGCGATGACAG	This work
SK215	AGCTTGCTTTTTAATTAAAGCAAATGCCACCGCAGCGG	This work
SK216	GGAGTCCAAGGACGTCAGATGCGTACTTGCTATCTGCAAC	This work
SK217	GAAGCTTGCATGCCTGCAGGATTTCTGAAAGCTGGCGC	This work
SK248	TTCCAGCATCAGCTCGTAG	This work
SK249	CGCTGGTTTATGTGCTGTG	This work
SK218	AGGGATAACAGGGTAATCTGAACTGTCTTCGTGGTCTTTGCCGCTGGC	This work
SK219	AGCTTGCTTTTTAATTAAAGCCGCTGCTGCCCGGGCCA	This work
SK220	GGAGTCCAAGGACGTCAGATAGGCGACCGCTTTCACGC	This work
SK221	GAAGCTTGCATGCCTGCAGGAGAATTGCACGCTCAACGCTTTG	This work
SK250	TCCTTATCGCGCTGTTCTAC	This work
SK251	ATGCCAGCATGGACTATGTG	This work
SK222	AGGGATAACAGGGTAATCTGCCTGCTGCAAGGCCATCAG	This work
SK223	AGCTTGCTTTTTAATTAAAGACTGGCTCAATGCCCTGC	This work
SK224	GGAGTCCAAGGACGTCAGATATTTGAAGGTAGCGTTCGGC	This work
SK225	GAAGCTTGCATGCCTGCAGGCGCGCCTACGAGCGAAAC	This work
SK252	GCGTAGTTGGCCGAAAGTG	This work
SK253	ATCCCTGGCATCAGCAGTG	This work
SK226	AGGGATAACAGGGTAATCTGAAAGCAGCGTTTGCCCGGGCC	This work
SK227	AGCTTGCTTTTTAATTAAAGTTCGCGGGCACGCCCGCT	This work
SK228	GGAGTCCAAGGACGTCAGATTGGGCCTGAAAACCCAGC	This work
SK229	GAAGCTTGCATGCCTGCAGGGAACACCACCCAGCTGCG	This work
SK254	TCCACGCCAGAAGCCTATG	This work
SK255	AGGTGGTGGCCTTGAAGTC	This work
SK230	AGGGATAACAGGGTAATCTGCGCAATGCCGCCAGCGCC	This work
SK231	AGCTTGCTTTTTAATTAAAGTGAAAGGGCAGCGCCTAAAGGG	This work
SK232	GGAGTCCAAGGACGTCAGATCAACACAAAGGCACTCCTACAAGGATTTCAGA CTTTTCTTCC	This work
SK233	GAAGCTTGCATGCCTGCAGGCGCTCGCCGGCGGCGATC	This work
SK256	GTGCACGTGGAAGTGTTCG	This work
SK257	CGCGTTGTCCTTGGTGATG	This work
SK234	AGGGATAACAGGGTAATCTGCGCGCGTGACCGTGCACC	This work
SK235	AGCTTGCTTTTTAATTAAAGCAGGCAACGCGGTGGATGG	This work
SK236	GGAGTCCAAGGACGTCAGATCTTCGCGGGTGAACCCGC	This work
SK237	GAAGCTTGCATGCCTGCAGGGGGTCGAAGCCCAGTTGC	This work
SK258	CGGTGTTGATGACCAACTC	This work
SK259	AGCCGTTCGCCTATATCG	This work
SK286	AGGGATAACAGGGTAATCTGCGTGTGCGACAACGTCGAC	This work
SK287	AGCTTGCTTTTTAATTAAAGGGCCGGTTAAAACCCAGG	This work
SK288	GGAGTCCAAGGACGTCAGATCTGAAGGCCGCTCCTGCG	This work
SK289	GAAGCTTGCATGCCTGCAGGCGGCGACCCTTGCATTTTC	This work
SK302	CAACCCGTACGACGCTATC	This work
SK303	TGCTGTTCACCCTGCTCAAG	This work
SK290	AGGGATAACAGGGTAATCTGGATCGACAGTTAAAGTCACCC	This work
SK291	AGCTTGCTTTTTAATTAAAGTATTACTTGCGGCTGGCTTTG	This work
SK292	GGAGTCCAAGGACGTCAGATATCCTCACGGATTGTTAAGAAGCC	This work
SK293	GAAGCTTGCATGCCTGCAGGCGCACCAGTCACGGTTGC	This work
SK304	ACCGCATGTACTACTTCCC	This work
SK305	CCACGTTGGTCTCGAACTG	This work
SK294	AGGGATAACAGGGTAATCTGGGTATCGGTGAAGTGAGCGAAG	This work
SK295	AGCTTGCTTTTTAATTAAAGCACCACTGACCAGGCCGC	This work
SK296	GGAGTCCAAGGACGTCAGATTGCGTGGGGTGTCGCAGG	This work
SK297	GAAGCTTGCATGCCTGCAGGCAGCGTCTCTTTCTCGGAAGGAC	This work
SK306	GATATCCTTGGCCCAACCC	This work
SK307	AAGCGGCCTGAAGAAGGAC	This work
SK298	AGGGATAACAGGGTAATCTGGCCACAGCCTGGGTGCAC	This work
SK299	AGCTTGCTTTTTAATTAAAGCAATCAGGCGGACAGCCG	This work
SK300	GGAGTCCAAGGACGTCAGATACGTCACCGCTTGCCTGGGGG	This work
SK301	GAAGCTTGCATGCCTGCAGGCGCGCCGGCGGCACGCAT	This work
SK308	GGCAGCCAGTTGCTCATAG	This work

Name	Sequence 5´-3´[a]	Reference
SK309	ATATCGGTGAAGCCCTTGG	This work

a) Restriction sites *Pac*I, *Hind*III, *Spe*I, *Nde*I, *Aat*II, *Eco*RI, *Sal*I, and *Avr*II are underlined. Coloured nucleotides indicate homologous sequences used for the Gibson Assembly reaction.

Table 15: Oligonucleotide combinations used for the construction of pEMG_PPXX_sensor vectors. Given are combinations for TS elements, sensor, terminator, and mapping of integration events.

Vector	TS1		TS2		Reference		Terminator		Mapping	
	for	rev	for	rev	for	rev	for	rev	for	rev
pEMG_PP0042	194	195	196	197	190	191	192	193	238	239
pEMG_PP0145	198	199	200	201	190	191	192	193	240	241
pEMG_PP0340	202	203	204	205	190	191	192	193	242	243
pEMG_PP0480	206	207	208	209	190	191	192	193	244	245
pEMG_PP1430	210	211	212	213	190	191	192	193	246	247
pEMG_PP1738	214	215	216	217	190	191	192	193	248	249
pEMG_PP3304	218	219	220	221	190	191	192	193	250	251
pEMG_PP3761	222	223	224	225	190	191	192	193	252	253
pEMG_PP4145	226	227	228	229	190	191	192	193	254	255
pEMG_PP4833	230	231	232	233	190	191	192	193	256	257
pEMG_PP4945	234	235	236	237	190	191	192	193	258	259

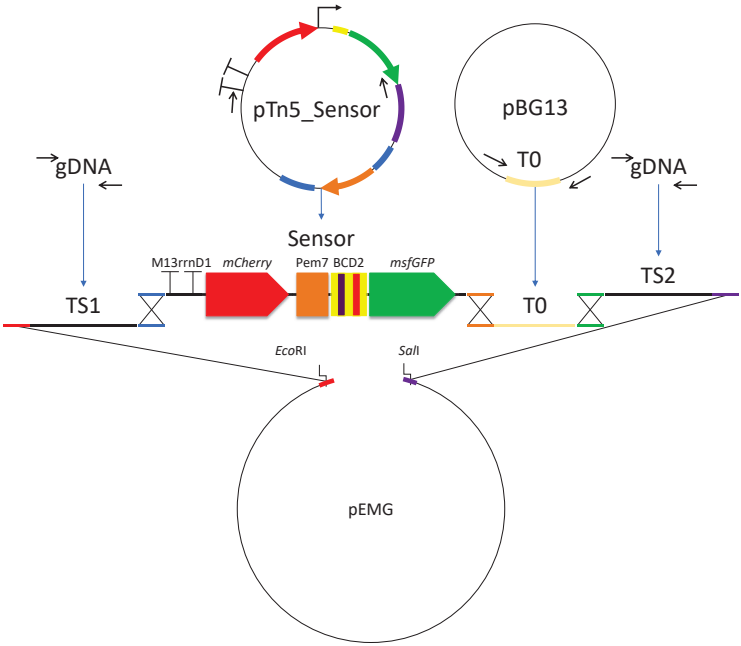


Figure 8: Construction of integrable pEMG vectors based on RNA-Seq data analysis containing the sensor from pTn5_sensor. Target elements for homologous recombination were amplified from *P. putida* KT2440 genomic DNA. The sensor was amplified from pTn5_sensor. An additional downstream terminator T1 was used from pBG13. Overhangs used for Gibson assembly are highlighted in different colors. Oligonucleotides for target elements are joining the same overhangs for easier and cheaper assembly procedures. The same overhang sequences are indicated in red, blue, orange, green, and purple. pEMG backbone was digested with *Eco*RI and *Sal*I.

2.2.7 Late stage promoter

RNA-Seq data sets were merged in a Microsoft Excel file, separated by carbon source. Data sets were sorted for low RPKM values (0, 0.5, and 1) during exponential phase and genes with higher RPKM values stationary phase filtered. Genes with the highest RPKM for stationary phase were collected and overlapping genes identified. Calculation of a fold change or ratio was not possible because filtered genes can show a RPKM value of zero. It was caused by the specification during data processing. Therefore, we subtracted the stationary phase value minus the exponential phase. Here gene PP1640 showed the highest difference and was further characterized.

Strains used and generated here are listed in Table 16. Oligonucleotides used in this section are given in Table 17. Using promoter prediction tools, three possible promoter sequences were found. Further, the full intergenomic region upstream of PP_1640 was tested. Genomic DNA of *P. putida* KT2440 was isolated from overnight cultures grown in lysogeny broth medium as described by the manufacturer (High Pure PCR Template Preparation Kit, Roche, Mannheim, Germany). gDNA was used as a template to amplify certain upstream regions of PP1640 using oligonucleotide SK156 and SK157. The resulting fragment was digested with *Pac*I and *Avr*II and following subcloning in pBG vector. The plasmid was called pBG_1640. Oligonucleotide SK158 was used to generate P2, P4 with SK159, P6 with SK160, and as revers oligonucleotide SK2 was used. Forward oligonucleotide contains predicted promoter sequences. Resulting fragments were purified and digested with *Pac*I and *Nco*I. Further ligated with digested pBG backbone and transferred into *E. coli* PIR2. Correctness was verified by Sanger sequencing (Eurofins Genomics, Ebersberg, Germany) with oligonucleotide SK43.

The generation of an upstream sequence-based library was performed in three different approaches. gDNA was used as a template and a complete upstream sequence was amplified using oligonucleotide SK178 and SK178. The purified fragment was afterwards used in PCRs to generate the library. Different length with the constant back was done with oligonucleotides SK166 and SK178, and the reverse length was performed with SK168 and SK179. Mixed length was generated by using oligonucleotides SK166 and SK168. PCRs consists of two steps, similar to the stacking promoter approach. The first round includes previously mentioned oligonucleotide combinations. Here the randomized oligonucleotide was able to bind at several sites in parallel. The second round was used for rapid amplification of former build fragments. Approach for forward analysis used SK157 and SK167, reverse SK156 and SK169, and mixed SK167 and SK169 (Figure 30, Table 18). Fragments were digested with *Pac*I and *Avr*II and ligated with the pBG backbone. Determination of resulting promoter activities performed as described following.

Table 16: Strains and plasmids used are generated test native promoter activity from PP_1640 upstream sequence.

Strain	Description	Reference	iAMB strain collection
E. coli			
HB101	*F⁻ mcrB mrr hsdS20(rB⁻ mB⁻) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^R) *gln V44λ⁻*	Boyer and Roulland-Dussoix[140]	2073
CC118λpir	Δ(*ara-leu*) *araD* Δ*lacX74 galE galK phoA20 thi-1 rpsE rpoB argE*(*Am*) *recA1,* lysogenized with λpir phage	Herrero *et al.*[143]	2072
PIR2	F- Δ*lac169 rpoS* (Am) *robA1 creC510 hsdR514 endA reacA1 uidA* (Δ*Mlui*)::*pir*	Life Technologies	3222
DH5αλpir	*endA1 hsdR17 glnV44* (= *supE44*) *thi-1 recA1 gyrA96 relA1 φ80dlac*Δ(*lacZ*)*M15* Δ(*lacZYA-argF*)*U169 zdg-232*::Tn10 *uidA*::*pir+*	de Lorenzo lab	2673
P. putida			
KT2440	Wild-type strain derived of *P. putida* mt-2 cured of the pWW0 plasmid	Bagdasarian *et al.*[144]	2058
BG14a	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a	Zobel *et al.*[27]	3228
PP_1640_P2	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_PP_1640_P2	This work	6377
PP_1640_P4	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_PP_1640_P4	This work	6378
PP_1640_P6	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_PP_1640_P6	This work	6379
PP_1640_Full	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_PP_1640_Full	This work	6380
PP_1640_for	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_PP_1640_for	This work	6381-6383
PP_1640_rev	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_PP_1640_rev	This work	6384-6385
PP_1640_mix	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_PP_1640_mix	This work	6387-6389
Plasmids			
pRK600	Cm^R, oriColE1, *tra* + *mob* + of RK2	Keen *et al.*[141]	2073
pTnS-1	Ap^R, oriR6K, *TnSABC+D* operon	Choi *et al.*[115]	3321
pBG	Km^R, Gm^R, oriR6K, Tn7L and Tn7R extremes, BCD2−*msfgfp* fusion	Zobel *et al.*[27]	3210
pBG14a	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14a, *msfGFP*	Zobel *et al.*[27]	3217
pBG_PP_1640_P2	Km^R, Gm^R, oriR6K, pBG-derived, predicted promoter of PP_1640, *msfGFP*	This work	6596
pBG_PP_1640_P4	Km^R, Gm^R, oriR6K, pBG-derived, predicted promoter of PP_1640, *msfGFP*	This work	6597
pBG_PP_1640_P6	Km^R, Gm^R, oriR6K, pBG-derived, predicted promoter of PP_1640, *msfGFP*	This work	6598
pBG_PP_1640_Full	Km^R, Gm^R, oriR6K, pBG-derived, upstream sequence of PP_1640, *msfGFP*	This work	9599
pBG_PP_1640_for	Km^R, Gm^R, oriR6K, pBG-derived, upstream sequence of PP_1640 used as template for randomized PCRs approach, constant reverse oligonucleotide, *msfGFP*	This work	6600-6602
pBG_PP_1640_rev	Km^R, Gm^R, oriR6K, pBG-derived, upstream sequence of PP_1640 used as template for randomized PCRs approach, constant forward oligonucleotide, *msfGFP*	This work	6603-6605
pBG_PP_1640_mix	Km^R, Gm^R, oriR6K, pBG-derived, upstream sequence of PP_1640 used as template for randomized PCRs approach, both oligonucleotides contained randomized nucleotides, *msfGFP*	This work	6606-6608

Table 17: Oligonucleotides used in this section to identify the core promoter of PP1640.

Name	Sequence 5´-3´[a]	Reference
SK4	AGTCAGAGTTACGGAATTGTAGG	Zobel *et al.*[27]

Name	Sequence 5´-3´[a]	Reference
SK5	GTCGAGAAAATTGCCGAGCT	Zobel *et al.* [27]
SK43	ATCAAACATCGACCCACGGCGTAAC	This work
SK156	CGCTTAATTAACATGACCTTGTCGCCATCC	This work
SK157	CGCCCTAGGTGCCAAACGCCCCCCAATATG	This work
SK158	GACTTAATTAATGGATCCGCGAGGTGTGACAAAAGCCGGCTATTTTACGCACA AACGCGACCCTAGGGCCCAAGTTCACTTA	This work
SK159	GACTTAATTAAAAGCGCGGTTTATGTAGGTAACGGCTTCCTATACTGCGACAT AAGGCCACCCTAGGGCCCAAGTTCACTTA	This work
SK160	GACTTAATTAAGGTGTTAGGGCCTCGGGCCGCTTACGCTCCATATTGGGGGGC GTTTGGCACCTAGGGCCCAAGTTCACTTA	This work
SK166	CGCATTACACGGCTTAATTAANNNNN	This work
SK167	CGCATTACACGGCTTAATTAA	This work
SK168	CGCATTACACGGCCCTAGGNNNNN	This work
SK169	CGCATTACACGGCCCTAGG	This work
SK178	TGCCAAACGCCCCCCAATATG	This work
SK179	CATGACCTTGTCGCCATCC	This work

a) Restriction sites *Pac*I and AvrII are underlined. N stands for nucleotides a, c, g, or t.

Table 18: Oligonucleotide combinations for the generation of an upstream sequence-based library of PP_1640.

Approach	1st PCR 0.075µL (10pmol) oligo	2nd PCR 2.5µL (10pmol) oligo
Forward Analyse	SK178+SK166	SK157+SK167
Revers Analyse	SK179+SK168	SK156+SK169
Mix Analyse	SK166+SK168	SK167+SK169

2.2.8 Tic Tac Lac

Characterization of standard promoter sequences was performed after the genomic integration of transposon Tn7. Strains and plasmids used and generated are listed in Table 19. Oligonucleotides are given in Table 20.

pSEVA234_pLac_O was generated via restriction and ligation. Promoter P_{Lac} is integrated in the forward oligonucleotide SK144 and complementary to the start sequence of *msfGFP*. As template pBG13 was used with reverse oligonucleotide SK62. PCR fragment and pSEVA234, containing the LacIq gene and corresponding promoter, were digested with *Pac*I and *Kpn*I and ligated using T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA). Resulting plasmid is called pSEVA234_pLac_O_*msfGFP*. Cultivation was done with minimal medium described above with 20 mM glucose as carbon source and 50 mg L^{-1} kanamycin. Induction of LacIq was performed by the addition of 1 mM IPTG and resulting msfGFP fluorescence measured with a synergyMX plate reader (Biotek, Bad Friedrichshall, Germany).

Core promoter sequences of P_{Lac}, P_{Trp}, P_{LacUV5}, P_{Tac}, P_{Trc}, and P_{Tic} were directly included in the corresponding oligonucleotides SK310, SK311, SK312, SK313, SK314, and SK315. The back was complementary to the BCD2 element in pBG13 and used as a template. Reverse oligonucleotide SK2 was used. Resulting fragment and pBG were digested with *Pac*I and *Nco*I and ligated using T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA). Constructed plasmids were called pBG_P_{Lac}, pBG_P_{Trp}, pBG_P_{LacUV5}, pBG_P_{Tac}, pBG_P_{Trc}, and pBG_P_{Tic}.

Sanger sequencing reaction verified the correctness of cloned sequences (Eurofins Genomics, Ebersberg, Germany) with oligonucleotide SK43.

Table 19: Strains and plasmids used and generated in this section for the characterization of standard promoter sequences.

Strain	Description	Reference	iAMB strain collection
E. coli			
HB101	*F⁻ mcrB mrr hsdS20(rB⁻ mB⁻) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^R) *gln V44λ⁻*	Boyer and Roulland-Dussoix[140]	2073
CC118λpir	Δ(*ara-leu*) *araD* Δ*lacX74 galE galK phoA20 thi-1 rpsE rpoB argE*(*Am*) *recA1,* lysogenized with λpir phage	Herrero *et al.*[143]	2072
PIR2	F- Δ*lac169 rpoS* (Am) *robA1 creC510 hsdR514 endA reacA1 uidA* (Δ*Mlui*)*::pir*	Life Technologies	3222
DH5αλpir	*endA1 hsdR17 glnV44* (= *supE44*) *thi-1 recA1 gyrA96 relA1 φ80dlac*Δ(*lacZ)M15* Δ(*lacZYA-argF)U169 zdg-232*::Tn10 *uidA::pir+*	de Lorenzo lab	2673
P. putida			
KT2440	Wild-type strain derived of *P. putida* mt-2 cured of the pWW0 plasmid	Bagdasarian *et al.*[144]	2058
BG13	Gm^R, *P. putida* KT2440 with genomic insertion of pBG13	Zobel *et al.*[27]	3235
BG14a	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a	Zobel *et al.*[27]	3228
BG14b	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b	Zobel *et al.*[27]	3230
BG14c	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14c	Zobel *et al.*[27]	3231
BG14d	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d	Zobel *et al.*[27]	3233
BG14e	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e	Zobel *et al.*[27]	3225
BG14f	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f	Zobel *et al.*[27]	3227
BG14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g	Zobel *et al.*[27]	3232
BG_pLac	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_pLac	This work	6390
BG_pTac	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_pTac	This work	6391
BG_pTic	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_pTic	This work	6392
BG_pLacUV5	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_pLacUV5	This work	6393
BG_pTrp	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_pTrp	This work	6394
BG_Trc	Gm^R, *P. putida* KT2440 with genomic insertion of pBG_pTrc	This work	6395
Plasmids			
pRK600	Cm^R, oriColE1, *tra* + *mob* + of RK2	Keen *et al.*[141]	2073
pTnS-1	Ap^R, oriR6K, *TnSABC+D* operon	Choi *et al.*[115]	3221
pBG13	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter P_{em7}, *msfGFP*	Martínez-Garcia *et al.*[33]	3211
pBG14a	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14a, *msfGFP*	Zobel *et al.*[27]	3214
pBG14b	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14b, *msfGFP*	Zobel *et al.*[27]	3217
pBG14c	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14c, *msfGFP*	Zobel *et al.*[27]	3218
pBG14d	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14d, *msfGFP*	Zobel *et al.*[27]	3220
pBG14e	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14e, *msfGFP*	Zobel *et al.*[27]	3212
pBG14f	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14f, *msfGFP*	Zobel *et al.*[27]	3213
pBG14g	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14g, *msfGFP*	Zobel *et al.*[27]	3219
pBG_pLac	Km^R, Gm^R, oriR6K, pBG-derived, promoter P_{Lac}, *msfGFP*	This work	6609
pBG_pTac	Km^R, Gm^R, oriR6K, pBG-derived, promoter P_{Tac}, *msfGFP*	This work	6610
pBG_pTic	Km^R, Gm^R, oriR6K, pBG-derived, promoter P_{Tic}, *msfGFP*	This work	6611
pBG_pLacUV5	Km^R, Gm^R, oriR6K, pBG-derived, promoter P_{LacUV5}, *msfGFP*	This work	6612
pBG_pTrp	Km^R, Gm^R, oriR6K, pBG-derived, promoter P_{Trp}, *msfGFP*	This work	6613
pBG_pTrc	Km^R, Gm^R, oriR6K, pBG-derived, promoter P_{Trc}, *msfGFP*	This work	6614
pSEVA234	Km^R, pBBR1, LacIq-Ptrc	Silva-Rocha *et al.*[34]	2566
pSEVA234_pLac_*msfGFP*	Km^R, pBBR1, promoter P_{Lac}, *msfGFP*	This work	6615
pSEVA234_pLac_O_*msfGFP*	Km^R, pBBR1promoter P_{Lac}, operator sequence, LacIq, *msfGFP*	This work	6616

Table 20: Oligonucleotides used in this section to characterize standard promoters.

Name	Sequence 5´-3´[a]	Reference
SK2	ACACCATAGGTCAGGGTAGTC	This work
SK4	AGTCAGAGTTACGGAATTGTAGG	Zobel *et al.*[27]
SK5	GTCGAGAAAATTGCCGAGCT	Zobel *et al.* [27]
SK43	ATCAAACATCGACCCACGGCGTAAC	This work
SK62	TAGTCGCCAGGGTTTTCC	This work
SK144	TTAATTAATTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGACCTAGGAAT TGTGAGCGGATAACAATTAGGAGGTTTTCTAATGATCATGGGAATTCATAAAG	This work
SK310	GACTTAATTAATTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGACCTAGG CCCAAGTTCACTTA	This work
SK311	GACTTAATTAATTGACAATTAATCATCGAACTAGTTAACTAGTACGCCCTAGG CCCAAGTTCACTTA	This work
SK312	GACTTAATTAATTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGACCTAGG CCCAAGTTCACTTA	This work
SK313	GACTTAATTAATTGACAATTAATCATCGGCTCGTATAATGTGTGGACCTAGGCC CAAGTTCACTTA	This work
SK314	GACTTAATTAATTGACAATTAATCATCCGGCTCGTATAATGTGTGGACCTAGGC CCAAGTTCACTTA	This work
SK315	GACTTAATTAATTGACAATTAATCATCCCGGCTCGTATAATGTGTGGACCTAGG CCCAAGTTCACTTA	This work

a) Restriction site *Pac*I is underlined.

2.2.9 FRT

All strains used and generated in this section are listed in Table 21. Oligonucleotides used in this section are given in Table 22. Plasmids containing FRT-FLP marker recycling were generated in the same manner. Plasmid pBG13 was used as a template for the origin of transfer *oriT* and origin of replication *oriR6K* containing fragment (oligonucleotides SK264 and SK265). FRT flanked kanamycin gene, including promoter region, were amplified with oligonucleotides SK266 and SK267 while pBELK was used as a template. Promoter 14a to 14g, BCD2, *msfGFP,* and terminator T0 fragment was amplified from appropriate plasmids pBG14a to pBG14g using oligonucleotides SK268 and SK269. P_{em7} containing fragment was amplified from pBG13. Fragments were assembled via Gibson assembly following the manual of the NEBuilder Hifi DNA Assembly Master Mix (New England Biolabs, Ipswich, Massachusetts, USA). All fragments were cut out of agarose gels and purified with DNA Gel Extraction kit (New England Biolabs, Ipswich, Massachusetts, USA). Concentrations of purified fragments were measured with a NanoDrop One (Thermo Scientific, Waltham, Massachusetts, USA).

The integration of the novel mini-Tn7 vector was done by mating, as described above. The release of the kanamycin resistance cassette was done by flippase activity. pBBFLP was isolated from *E. coli* strains and transformed via electroporation into BGX_Kan_FRT bearing *P. putida* KT2440 strains (X stand for different promoters). A quick protocol for the generation of electrocompetent cells was used. Afterwards, cells were plated on LB agar plates containing 30 mg L^{-1} tetracycline. The growth of clones needed up to two days. Colonies were then picked on LB agar plates and LB agar plates with 50 mg L^{-1} kanamycin to identify clones that were no longer resistant against

kanamycin. Verification was done by colony PCR using *Taq* 2X Master Mix (New England BioLabs, Ipswich, Massachusetts, USA) and oligonucleotides SK4 and SK5.

Table 21: Strains and plasmids used and generated in this section for marker recycling using the FRT-FLP system.

Strain	Description	Reference	iAMB strain collection
E. coli			
HB101	*F⁻ mcrB mrr hsdS20(rB⁻ mB⁻) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^R) *gln V44λ⁻*	Boyer and Roulland-Dussoix[140]	2073
CC118λpir	Δ(*ara-leu*) *araD* Δ*lacX74 galE galK phoA20 thi-1 rpsE rpoB argE*(*Am*) *recA1,* lysogenized with λpir phage	Herrero *et al.*[143]	2072
PIR2	F⁻ Δ*lac169 rpoS* (Am) *robA1 creC510 hsdR514 endA reacA1 uidA* (Δ*Mlui*)::*pir*	Life Technologies	3222
DH5αλpir	*endA1 hsdR17 glnV44* (= *supE44*) *thi-1 recA1 gyrA96 relA1* φ*80dlac*Δ(*lacZ*)*M15* Δ(*lacZYA-argF*)*U169 zdg-232*::Tn10 *uidA*::*pir+*	de Lorenzo lab	2673
P. putida			
KT2440	Wild-type strain derived of *P. putida* mt-2 cured of the pWW0 plasmid	Bagdasarian *et al.*[144]	2058
BG13	Gm^R, *P. putida* KT2440 with genomic insertion of pBG13	Zobel *et al.*[27]	3235
BG14b	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b	Zobel *et al.*[27]	3230
BG14c	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14c	Zobel *et al.*[27]	3231
BG14d	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d	Zobel *et al.*[27]	3233
BG14e	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e	Zobel *et al.*[27]	3225
BG14f	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14f	Zobel *et al.*[27]	3227
BG14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g	Zobel *et al.*[27]	3232
BG13_KanRFRT	Km^R with FRT flanking sequences, *P. putida* KT2440 with genomic insertion of pBG13_FRT_Kan	This work	6404
BG14b_FRT_Kan	Km^R with FRT flanking sequences, *P. putida* KT2440 with genomic insertion of pBG14b_FRT_Kan	This work	6398
BG14c_FRT_Kan	Km^R with FRT flanking sequences, *P. putida* KT2440 with genomic insertion of pBG14c_FRT_Kan	This work	6399
BG14d_FRT_Kan	Km^R with FRT flanking sequences, *P. putida* KT2440 with genomic insertion of pBG14d_FRT_Kan	This work	6400
BG14e_FRT_Kan	Km^R with FRT flanking sequences, *P. putida* KT2440 with genomic insertion of pBG14e_FRT_Kan	This work	6401
BG14f_FRT_Kan	Km^R with FRT flanking sequences, *P. putida* KT2440 with genomic insertion of pBG14f_FRT_Kan	This work	6402
BG14g_FRT_Kan	Km^R with FRT flanking sequences, *P. putida* KT2440 with genomic insertion of pBG14g_FRT_Kan	This work	6403
BG13_ΔKan	*P. putida* KT2440 with genomic insertion of pBG13_FRT_Kan with kanamycin gene deletion caused by flippase activity	This work	6405
BG14b_ΔKan	*P. putida* KT2440 with genomic insertion of pBG14b_FRT_Kan with kanamycin gene deletion caused by flippase activity	This work	6406
BG14c_ΔKan	*P. putida* KT2440 with genomic insertion of pBG14c_FRT_Kan with kanamycin gene deletion caused by flippase activity	This work	6407
BG14d_ΔKan	*P. putida* KT2440 with genomic insertion of pBG14d_FRT_Kan with kanamycin gene deletion caused by flippase activity	This work	6408
BG14e_ΔKan	*P. putida* KT2440 with genomic insertion of pBG14e_FRT_Kan with kanamycin gene deletion caused by flippase activity	This work	6409
BG14f_ΔKan	*P. putida* KT2440 with genomic insertion of pBG14f_FRT_Kan with kanamycin gene deletion caused by flippase activity	This work	6410
BG14g_ΔKan	*P. putida* KT2440 with genomic insertion of pBG14g_FRT_Kan with kanamycin gene deletion caused by flippase activity	This work	6411

Strain	Description	Reference	iAMB strain collection
Plasmids			
pRK600	Cm^R, oriColE1, *tra* + *mob* + of RK2	Keen *et al.*[141]	2073
pTnS-1	Ap^R, oriR6K, *TnSABC+D* operon	Choi *et al.*[115]	3221
pBBFLP	Helper plasmid used for antibiotic markers excision in *P. putida* strains; *oriV*(pBBR1) *oriT*(RK2) RK2 *mob*$^+$ λP_R::FLP λ(cI857) *sacB tet*, Tc^R	De las Heras et al.[145]	3607
pBG13	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter P_{em7}, *msfGFP*	Martínez-Garcia *et al.*[33]	3211
pBG13_KanRFRT	Km^R flanked with FRT sites, oriR6K, pBG-derived, promoter P_{em7}, *msfGFP*	This work	6624
pBG14a	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14a, *msfGFP*	Zobel *et al.*[27]	3214
pBG14b	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14b, *msfGFP*	Zobel *et al.*[27]	3217
pBG14c	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14c, *msfGFP*	Zobel *et al.*[27]	3218
pBG14d	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14d, *msfGFP*	Zobel *et al.*[27]	3220
pBG14e	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14e, *msfGFP*	Zobel *et al.*[27]	3212
pBG14f	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14f, *msfGFP*	Zobel *et al.*[27]	3213
pBG14g	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14g, *msfGFP*	Zobel *et al.*[27]	3219
pBG14a_FRT_Kan	Km^R flanked with FRT sites, oriR6K, pBG-derived, promoter 14a, *msfGFP*	This work	6617
pBG14b_FRT_Kan	Km^R flanked with FRT sites, oriR6K, pBG-derived, promoter 14b, *msfGFP*	This work	6618
pBG14c_FRT_Kan	Km^R flanked with FRT sites, oriR6K, pBG-derived, promoter 14c, *msfGFP*	This work	6619
pBG14d_FRT_Kan	Km^R flanked with FRT sites, oriR6K, pBG-derived, promoter 14d, *msfGFP*	This work	6620
pBG14e_FRT_Kan	Km^R flanked with FRT sites, oriR6K, pBG-derived, promoter 14e, *msfGFP*	This work	6621
pBG14f_FRT_Kan	Km^R flanked with FRT sites, oriR6K, pBG-derived, promoter 14f, *msfGFP*	This work	6622
pBG14g_FRT_Kan	Km^R flanked with FRT sites, oriR6K, pBG-derived, promoter 14g, *msfGFP*	This work	6623

Table 22: Oligonucleotides used for reengineering mini Tn7 vector in this section.

Name	Sequence 5´-3´	Reference
SK2	ACACCATAGGTCAGGGTAGTC	This work
SK4	AGTCAGAGTTACGGAATTGTAGG	Zobel *et al.*[27]
SK5	GTCGAGAAAATTGCCGAGCT	Zobel *et al.* [27]
SK43	ATCAAACATCGACCCACGGCGTAAC	This work
SK264	AATCTCTGATAATTGGACAAGGGTCCTTTTC	This work
SK265	TAAAAAACGCAATTGGACGTCGGCATCAAATAAAAC	This work
SK266	ACGTCCAATTGCGTTTTTTATTGGTGAG	This work
SK267	GTTATGGAGCATTTTGGTCATGAGATTATCG	This work
SK268	TGACCAAAATGCTCCATAACATCAAACATC	This work
SK269	TTGTCCAATTATCAGAGATTTTGAGACAC	This work

2.2.10 pEMG Gent KO

Different spacer sequences were analyzed to create a second *attTn7* site. Strains and plasmids are given inTable 23. Table 24 listed oligonucleotides used in this section.

pEMG backbone was digested with *Eco*RI and *Sal*I and purified. Three different gentamycin knock out vectors were constructed with different length sequences reflecting the *attTn7* site. TS2 element was amplified with SK319 and SK320 from *P. putida* KT2440 gDNA and used for all vectors. TS1 was generated for each individual. TS 2 element containing 60 bp was amplified with

oligonucleotides SK317 and SK322. Resulting vector is called pEMG_GentKO_60. 80 bp fragment was amplified with oligonucleotide SK317 and SK325, and the generated vector is called pEMG_GentKO_80. pEMG_GentKO_100 needed to fragments amplified from gDNA. TS1 element with SK319 and SK320, while 120 bp interspacing sequence was amplified with SK326 and SK327. All fragments were purified and assemble by Gibson assembly, according to manufacturers guidelines (New England Biolabs, Ipswich, Massachusetts, USA). Correctness was verified by Sanger sequencing with oligonucleotide BW13 and BW14. By triparental mating, the gentamycin knockout vectors were integrated in *P. putida* KT2440 genome. The co-integrate release was done by a second mating with pSW-1 plasmid, which was induced by 3MB. Afterwards, growth on kanamycin- and gentamycin-containing agar plates was tested to observe the event. No growth on both media highlights successfully generated deletion mutants.

TnsD recognition sequence was introduced in pEMG_GentKO_TnsD. TS elements were amplified with oligonucleotide combinations SK317 and SK340 for TS1 and SK319 and SK320 for TS2. Resulting fragments were assembled with digested pEMG backbone via Gibson assembly (New England Biolabs, Ipswich, Massachusetts, USA).

Table 23: Strains and plasmids used and generated to enable gentamycin knockouts in Tn7 bearing strains.

Strain	Description	Reference	iAMB strain collection
E. coli			
HB101	*F⁻ mcrB mrr hsdS20(rB⁻ mB⁻) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^R) *gln V44λ⁻*	Boyer and Roulland-Dussoix[140]	2073
CC118λpir	Δ(*ara-leu*) *araD* Δ*lacX74 galE galK phoA20 thi-1 rpsE rpoB argE*(*Am*) *recA1,* lysogenized with λpir phage	Herrero *et al.*[143]	2072
PIR2	F- Δ*lac169 rpoS* (Am) *robA1 creC510 hsdR514 endA reacA1 uidA* (Δ*Mlui*)::*pir*	Life Technologies	3222
DH5αλpir	*endA1 hsdR17 glnV44* (= *supE44*) *thi-1 recA1 gyrA96 relA1* φ*80dlac*Δ(*lacZ*)*M15* Δ(*lacZYA-argF*)*U169 zdg-232*::Tn10 *uidA*::*pir+*	de Lorenzo lab	2673
P. putida			
KT2440	Wild-type strain derived of *P. putida* mt-2 cured of the pWW0 plasmid	Bagdasarian *et al.*[144]	2058
BG14g	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g	Zobel *et al.*[27]	3232
BG14_ΔGent_80	Gm^R, *P. putida* KT2440 with genomic insertion of pEMG_GentKO_80, co-integrate released by pSW-1	This work	6412
Plasmids			
pRK600	Cm^R, oriColE1, *tra* + *mob* + of RK2	Keen *et al.*[141]	2073
pSW-1	Ap^R, *ori*RK2, *xylS*, transcriptional fusion of I-*Sce*I to Pm	Martínez-Garcia *et al.*[26]	2076
pEMG	Km^R, *ori*R6K, *lacZα* with two flanking I-*Sce*I sites	Martínez-Garcia *et al.*[26]	2075
pBG14g	Km^R, Gm^R, oriR6K, pBG-derived, promoter 14g, *msfGFP*	Zobel *et al.*[27]	3219
pEMG_GentKO_60	Km^R, *ori*R6K, TS elements targeting integrated Tn7 with two flanking I-*Sce*I sites, TS element complementary to genomic region and start of gentamycin, interspaced by 60 bp including stop codon formation after integration	This work	6625

Strain	Description	Reference	iAMB strain collection
pEMG_GentKO_80	Km^R, *ori*R6K, TS elements targeting integrated Tn7 with two flanking I-*Sce*I sites, TS element complementary to genomic region and start of gentamycin, interspaced by 80 bp including stop codon formation after integration	This work	6626
pEMG_GentKO_100	Km^R, *ori*R6K, TS elements targeting integrated Tn7 with two flanking I-*Sce*I sites, TS element complementary to genomic region and start of gentamycin, interspaced by 100 bp including stop codon formation after integration	This work	6627
pEMG_GentKO_TnsD	Km^R, *ori*R6K, TS elements targeting integrated Tn7 with two flanking I-*Sce*I sites, TS element complementary to genomic region and start of gentamycin, interspaced by TnsD recognition sequence and stop codon formation after integration	This work	6628

Table 24: Oligonucleotides used for the generation of gentamycin knock out vectors.

Name	Sequence 5´-3´	Reference
SK2	ACACCATAGGTCAGGGTAGTC	This work
SK4	AGTCAGAGTTACGGAATTGTAGG	Zobel *et al.*[27]
SK5	GTCGAGAAAATTGCCGAGCT	Zobel *et al.* [27]
SK43	ATCAAACATCGACCCACGGCGTAAC	This work
SK317	AGGGATAACAGGGTAATCTGGTGAAGACGTCGCCAATAAG	This work
SK318	AGAATTAAAGCTGTCGCAGGAG	This work
SK319	TTAACGCGGCTGCTCAAACTTGG	This work
SK320	GAAGCTTGCATGCCTGCAGGAAACTGTAATGCAAGTAGCGTATGCG	This work
SK322	AGTTTGAGCAGCCGCGTTAAGTCGAGTAACGGGTAGTC	This work
SK325	AGTTTGAGCAGCCGCGTTAAGTCGGTGACTGTCGAGTAAC	This work
SK326	CTGCGACAGCTTTAATTCTCTAG	This work
SK327	AGTTTGAGCAGCCGCGTTAAATCTGGCCAAGTCGGTGAC	This work
SK340	AGTTTGAGCAGCCGCGTTAACGTTGACCAGCCTCGTAATC	This work

2.2.11 Genomic control of plasmid copy number

All strains used and generated in this section are listed in Table 25. Oligonucleotides used here are listed in Table 26.

The replication gene of pRO1600 was amplified from pSEVA241 with oligonucleotides TT15 and TT16. Ligation with pBG vectors containing different synthetic promoters and BCD2 elements was performed after restriction with *Avr*II and *Kpn*I of both fragments. Resulting plasmid were called pBG14a_rep241, pBG14b_rep241, pBG14d_rep241, pBG14e_rep241, and pBG14g_rep241. Instead of pRO1600, we named resulting plasmids only with rep241, to indicate the origin of the replication gene. Constructed plasmids were verified by PCR using oligonucleotides SK43 and SK62 and sequencing reactions with SK43. Correct plasmids were genomically integrated in *P. putida* KT2440 in the *attTn7* site, as described above. To obtain proper propagation of origin of replication missing plasmids in *E. coli*, we used *E. coli* TOP10 for transformation procedures. We integrated even mentioned pBG-derived plasmids in *E. coli* TOP10 and clones with desired Tn7 integration were selected on LB-agar plates containing gentamycin

and streptomycin. Successful integration was verified by colony PCR. For *P. putida* KT2440 SK4 and SK5 and for *E. coli* TOP10 oligonucleotides SN15 and SN16 were used.

The deletion of the replication gene of ori pRO1600 from pSEVA241 was directly done from our test construct containing vectors. Therefore, we amplified P_{em7}_BCD2_msfGFP from pBG13 and BCD2_msfGFP from pBG using oligonucleotides SK43 and BW63. After the restriction of pSEVA241 and PCR fragments containing P_{em7}_BCD2_msfGFP or BCD2_msfGFP with *Pac*I and *Spe*I ligation reaction was performed and resulting plasmid transferred into *E. coli* PIR2. Resulting plasmids were named pSEVA241_P_{em7}_BCD2_msfGFP and pSEVA241_BCD2_msfGFP. The deletion was performed by PCR using previously generated plasmids as a template. Oligonucleotides SK45 and SK281 were used to amplify the whole vector excluded the origin of replication. PCR & DNA Cleanup Kit (New England Biolabs, Ipswich, Massachusetts, USA) was used to get rid of the PCR reaction compound and to concentrate PCR fragments. The isolate was then treated with *Dpn*I (New England Biolabs, Ipswich, Massachusetts, USA) to digest the remaining template plasmid and *Aat*II to create overhangs for further ligation. After digestion, the PCR fragment was purified again and then used for ligation. Resulting plasmids were transformed into *E. coli* TOP10 BG14d_rep241 and named pSEVA201_BCD2_msfGFP and pSEVA201_P_{em7}_BCD2_msfGFP. Successful deletion of the origin of replication was tested by colony PCR using oligonucleotides SK47 and SK126 and Sanger sequencing reactions with SK47.

Plasmids missing an own origin of replication were transferred by mating into *P. putida* KT2440 BG14x_rep241 strain, x stand here for promoter 14a, 14b, 14d, 14e, or 14g. As control pSEVA201_BCD2_msfGFP and pSEVA201_Pem7_BCD2_msfGFP we also tried to transfer them in *P. putida* KT2440 wild type.

Fluorescence measurements and determination of promoter activities were performed as described later.

Table 25: Strains and plasmids used and generated to enable genomic controlled plasmid replication.

Strain	Description	Reference	iAMB strain collection
E. coli			
HB101	*F⁻ mcrB mrr hsdS20(rB⁻ mB⁻) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20*(Sm^R) *gln V44λ⁻*	Boyer and Roulland-Dussoix[140]	2073
CC118λpir	Δ(*ara-leu*) *araD* Δ*lacX74 galE galK phoA20 thi-1 rpsE rpoB argE*(*Am*) *recA1,* lysogenized with λpir phage	Herrero *et al.*[143]	2072
PIR2	F- Δ*lac169 rpoS* (Am) *robA1 creC510 hsdR514 endA reacA1 uidA* (Δ*MluI*)::*pir*	Life Technologies	3222
DH5αλpir	*endA1 hsdR17 glnV44* (= *supE44*) *thi-1 recA1 gyrA96 relA1* φ*80dlac*Δ(*lacZ*)*M15* Δ(*lacZYA-argF*)*U169 zdg-232*::Tn10 *uidA*::*pir+*	de Lorenzo lab	2673

Strain	Description	Reference	iAMB strain collection
TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ^-	Invitrogen	3089
P. putida			
KT2440	Wild-type strain derived of *P. putida* mt-2 cured of the pWW0 plasmid	Bagdasarian *et al.*[144]	2058
BG14a_rep241	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a_rep241 in *attTn7*	This work	6418
BG14a_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_h7	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a_rep241 in *attTn7*	This work	6423
BG14a_rep241pSEVA201_P_{em7}_msfGFP_Δrep_i6	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a_rep241 in *attTn7*	This work	6428
BG14a_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_g3	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14a_rep241 in *attTn7*	This work	6433
BG14b_rep241	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b_rep241 in *attTn7*	This work	6419
BG14b_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_h7	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b_rep241 in *attTn7*	This work	6424
BG14b_rep241pSEVA201_P_{em7}_msfGFP_Δrep_i6	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b_rep241 in *attTn7*	This work	6429
BG14b_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_g3	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14b_rep241 in *attTn7*	This work	6434
BG14d_rep241	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d_rep241 in *attTn7*	This work	6420
BG14d_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_h7	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d_rep241 in *attTn7*	This work	6422
BG14d_rep241pSEVA201_P_{em7}_msfGFP_Δrep_i6	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d_rep241 in *attTn7*	This work	6430
BG14d_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_g3	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14d_rep241 in *attTn7*	This work	6435
BG14e_rep241	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e_rep241 in *attTn7*	This work	6421
BG14e_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_h7	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e_rep241 in *attTn7*	This work	6426
BG14e_rep241pSEVA201_P_{em7}_msfGFP_Δrep_i6	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e_rep241 in *attTn7*	This work	6431
BG14e_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_g3	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14e_rep241 in *attTn7*	This work	6436
BG14g_rep241	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_rep241 in *attTn7*	This work	6422
BG14g_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_h7	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_rep241 in *attTn7*	This work	6427
BG14g_rep241pSEVA201_P_{em7}_msfGFP_Δrep_i6	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_rep241 in *attTn7*	This work	6432
BG14g_rep241 pSEVA201_P_{em7}_msfGFP_Δrep_g3	Gm^R, *P. putida* KT2440 with genomic insertion of pBG14g_rep241 in *attTn7*	This work	6437
pRK600	Cm^R, oriColE1, *tra* + *mob* + of RK2	Keen *et al.*[141]	2073
pTnS-1	Ap^R, oriR6K, *TnSABC+D* operon	Choi *et al.*[115]	3221

Strain	Description	Reference	iAMB strain collection
pSEVA241	Km^R, pRO1600/ColE1, mcs	Silva-Rocha *et al.*[34]	2567
pBG	Km^R, Gm^R,oriR6K, Tn7L and Tn7R extremes, BCD2−*msfgfp* fusion	Zobel *et al.*[27]	3210
pBG13	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter P_{em7}, *msfGFP*	Martínez-Garcia *et al.*[33]	3211
pSEVA241_BCD2_msfGFP	Km^R, pRO1600/ColE1, promoterless	This work	6629
pSEVA241_P_{em7}_BCD2_msfGFP	Km^R, pRO1600/ColE1, promoter P_{em7}, BCD2, *msfGFP*	This work	6630
pSEVA201_P_{em7}_BCD2_msfGFP_i6	Km^R, promoter P_{em7}, BCD2, *msfGFP*, deletion of the replication gene of ori pRO1600, isolated from clone i6 during screening	This work	6639
pSEVA201_P_{em7}_BCD2_msfGFP_h7	Km^R, promoter P_{em7}, BCD2, *msfGFP*, deletion of the replication gene of ori pRO1600, isolated from clone h7 during screening	This work	6640
pSEVA201_P_{em7}_BCD2_msfGFP_g3	Km^R, promoter P_{em7}, BCD2, *msfGFP*, deletion of the replication gene of ori pRO1600, isolated from clone g3 during screening	This work	6641
pBG14a_rep241	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter 14a, BCD2, replication gene of pRO1600	This work	6631
pBG14b_rep241	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter 14b, BCD2, replication gene of pRO1600	This work	6632
pBG14d_rep241	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter 14d, BCD2, replication gene of pRO1600	This work	6633
pBG14e_rep241	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter 14e, BCD2, replication gene of pRO1600	This work	6634
pBG14g_rep241	Km^R, Gm^R, *ori*R6K, pBG-derived, promoter 14g, BCD2, replication gene of pRO1600	This work	6635

Table 26: Oligonucleotides used in this section for the genomic controlled plasmid replication.

Name	Sequence 5´-3´	Reference
SK2	ACACCATAGGTCAGGGTAGTC	This work
SK4	AGTCAGAGTTACGGAATTGTAGG	Zobel *et al.*[27]
SK5	GTCGAGAAAATTGCCGAGCT	Zobel *et al.* [27]
SK43	ATCAAACATCGACCCACGGCGTAAC	This work
TT15	GACCTAGGGCCCAAGTTCACTTAAAAAGGAGATCAACAATGAAAGCAATTTTC GTACTGAAACATCTTAATCATGCTAAGGAGGTCAAGGAATGTCGAACGTGGC	This work
TT16	GCGGTACCGGCCTCTAGGCCAGATCCAG	This work
SK45	CGCGACGTCTACGTGAGACGTGTCCAGCGGTTGCCCAAGTAAAAGGTTTAAAA GGGAAAAGGA	This work
SK281	CGCGACGTCGCTCGACTAACCCAGATGCC	This work
SN15	GCTGTCTCGCCTGAAAGGTC	Provided by S. Nies
SN16	GTGGCGGTCAGTTGTATG	Provided by S. Nies
SK62	TAGTCGCCAGGGTTTTCC	This work
BW63	CCGAGCGTTCTGAACAAATC	Provided by B. Wynands
SK47	TAACCCTGCTTCGGGGTCATTAT	This work
SK126	GCGATAGGGGACTGGTAAA	This work

2.3 Measuring fluorescence and determination of promoter activity

msfGFP fluorescence was measured with an excitation wavelength of 488 nm and an emission wavelength of 520 nm using a synergyMX plate reader (Biotek, Bad Friedrichshall, Germany) or a Biolector (M2P Labs, Baesweiler, Germany). Tested strains of *E. coli* PIR2 are bearing plasmids,

whereas *P. putida* KT2440 bearing genomically integrated expression cassettes or plasmids. Cultivation was done as described previously.

For fluorescence endpoint measurements, a synergyMX plate reader (Biotek, Bad Friedrichshall, Germany) was used. Overnight cultures were adjusted with water to an OD_{600} of 1.0. OD_{600} dilution series were used to calibrated absorbance measured at 600 nm to optical density. 200 µL of adjusted cells were measured in black bottom 96 well plates for fluorescence (Greiner Bio-One, Kremsmünster, Austria). Absorption was measured at 600 nm in clear bottom 96 well plates (Greiner Bio-One, Kremsmünster, Austria).

Growth and fluorescence measurements of integrated promoter constructs in *P. putida* KT2440 were performed with a Biolector (M2P Labs, Baesweiler, Germany) in clear bottom 96 well plates (Greiner Bio-One) with a filling volume of 200 µL. Cultures were inoculated to an optical density at 600 nm of 0.1 for each strain from precultures cultivated at 30 °C at 300 rpm in 24 well System Duetz plates (Enzyscreen, Heemstede, The Netherlands) containing 1.5 mL of previously described minimal medium. The Biolector was set to 30°C, 900 rpm, and humidity control of 85 %. Two internal filter modules of the device were used for online measurement. Fluorescence of msfGFP was measured at an excitation wavelength at 488 nm and an emission wavelength of 520 nm with gain 50 or 60. Biomass was determined at 620 nm with gain 40 as scattered light. Scattered light was correlated to OD600 with a dilution series of a stationary phase culture. The determination of promoter activity was done with Microsoft Excel by calculating the slope of GFP fluorescence to optical density during the exponential phase.

2.4 Determination of transcript levels by quantitative real-time PCR

Transcription levels of *msfGFP* were determined by quantitative real-time PCR. RNA was isolated from chosen strains grown on minimal medium containing 20 mM glucose as a sole carbon source in 24 well System Duetz plates at 30°C and 300 rpm.[138] Biological duplicates of each strain were cultivated until an optical density of 1.0 was reached. 1 mL of cell cultures were harvested, the supernatant discarded and the resulting pellet resuspended with 1 mL RNA*later*™ Stabilization Solution (ThermoFisher Scientific). Afterwards the cells were resuspended in 700 µL lyse solution (Monarch™ Total RNA Miniprep Kit, New England Biolabs, Ipswich, Massachusetts, USA) and transferred to bead beating tubes containing glass bead with a size of 0.5 mm (Zymo Research, Irvine, USA). Tubes were beaten for 1 minute to destroy the cells (Mini-Beadbeater-16, Biospec Products, Bartlesville, USA). Cell debris was removed by centrifugation at 13,000 rpm for 2 minutes. The supernatant was transferred to an RNase-free tube and used for further works. RNA isolation from lysed samples followed the manual from the kit Monarch Total RNA Miniprep Kit (New England Biolabs, Ipswich, Massachusetts, USA). The elution of RNA was done with 50 µL

RNase free water. RNA concentration was measured with a NanoDrop One (Thermo Scientific, Waltham, Massachusetts, USA) at 260 nm. Samples were adjusted to a final RNA concentration of 280 ng in a total volume of 40 µL, dilution was done with RNase-free water. An additional DNase treatment was done by adding 5 µL DNaseI and 5 µL DNase I reaction buffer (New England Biolabs, Ipswich, Massachusetts, USA) to the RNA isolates. Digestion was done at 37°C for 10 minutes and DNase inactivation at 75°C for 10 minutes. For cDNA synthesis, LunaScript RT SuperMix Kit (New England Biolabs, Ipswich, Massachusetts, USA) was performed as described in the manual.

The determination of primer efficiencies was done with diluted cDNA from BG14f_80. cDNA was diluted 1:10, 1:20, 1:40, 1:80, 1:160, 1:320 and 1:640. 1.25 µL of each used for the qRT-PCR reaction. A total volume of 10 µL containing 5 µL Universal qPCR Master Mix (New England Biolabs, Ipswich, Massachusetts, USA), 0.25 µL of each oligonucleotide, 1.25 µL sample and 3.25 µL RNase-free water were used. We tested oligonucleotide combinations for the target gene *msfGFP* and housekeeping gene *rpoD* in a CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, USA) using a protocol described in the manual of Universal qPCR Master Mix (New England Biolabs, Ipswich, Massachusetts, USA). CFX Manager software (Bio-Rad Laboratories, Hercules, USA) was used for the calculation of resulting primer efficiencies and an online tool was used to calculate the amplification factor (https://www.thermofisher.com). Tested oligonucleotide combinations for *msfGFP* achieved a value of 1.97 and *rpoD* of 2.02.[146] Each cDNA sample was diluted 1:10 and analyzed as a technical duplicate. Volumes for each reaction are described above. As a negative control, the same amount of water was added to the reaction instead of cDNA. An examination from resulting Ct values was done with Microsoft Excel and a ΔCt method was applied.[147] To exclude genomic DNA in the samples, isolated RNA was used in separate reactions with oligonucleotides for *rpoD*.

2.5 Identification of suitable integration sites using RNA-Seq Data

The identification of suitable integration sites in *P. putida* KT2440[144] genome was done by analyzing two available sets of RNA-Seq data.[148] One set contains information about transcriptomic variations between different mutants and the wild type while cultivation with succinate and glucose as carbon source.[149] The second data set holds transcriptomic information about different carbon sources used for cultivations of *P. putida* KT2440 wild type.[150] Both data sets were merged into one data set and missing genes were identified and values deleted, hence the lack of information has no negative effect on further calculation like mean or standard deviation because otherwise a RPKM value of “0” would be used for calculation. Transcription values are given in reads per kilobase per million mapped reads (RPKM).

Furthermore, the start and end of the gene alike the orientation on the genome were added to the information.[151] Available gene names were given. Annotation of the merged data sets was done with a previously published annotation for *P. putida* KT2440 by Nelson *et al*.[151] All seven 16sRNA genes were removed from the data set due to their highly conserved sequences a false mapping of determined transcripts is possible.[152, 153] Indeed, integration into the genome without influences on basically expressed housekeeping genes is desired. Data sets were arranged next to each other in one Microsoft Excel file. For each gene, the mean and corresponding standard deviation were calculated. Dividing the standard deviation by the mean lead to the coefficient of variation (CV), which reflects the differential expression of one gene for all data sets. High CV values representing a high transcriptional change, low values showing genes expressed in the same magnitude. Since we are not interested in integrating our constructs into genes, we, therefore, analyzed the data for equally expressed clustered genes. For this purpose, the CV of five next to each other genes was determined and compared with the next CV of five genes. This framework covers each gene five times in comparison with neighboured genes (+/-2 genes from the candidate) by shifting the analysis window for one gene. In consequence, we have a high coverage for each gene and reliable predictions are possible. Starting with a degree of freedom of "zero" percent deviation, we found no neighboured genes with the same expression pattern. Increasing the degree of freedom to "ten" percent resulted in ten genes with the same expression pattern compared to the neighboured gene. A more increasing degree of freedom leads to a longer list of candidate genes and favored genes that are located next to each other on the genome. From this set of genes, we have chosen ten clustered genes and investigated the chromosomal arrangements. To avoid interruption of genomic clusters, we used the next free intergenic region for integration. The intergenic regions were identified using whole-genome information[151] and www.pseudomonas.com.[154] To prevent natural regulatory mechanisms from being disrupted, we analyzed possible integration sites for promoter sequences (BPROM[155], Neural Network Promoter Prediction[156]) and terminators (ARNold[157]).

2.6 Chemostat cultivation and FACS analysis

To obtain growth of all Tn5 bearing *P. putida* KT2440 clones, chemostat cultivation was performed to accomplish a steady-state of the whole population. Tn5 library was thawed and cells mixed. After centrifugation supernatant was discarded, to avoid growth on resting glycerol, a wash step with 1xKPi was inserted. For the cultivation a DASbox (Eppendorf AG, Germany) with small vessels and a total volume of 80 mL was used. As medium, minimal medium with 10 mM citric acid was used and pH was set to 7 inside the reactor with NaOH. Initial OD_{600} of the cultivation was adjusted to 0.5. After growth to OD_{600} of 1.0 and constant signal for dissolved oxygen pumps were activated. 8 mL of fresh medium was added to the vessels per hour, simultaneously a pipe in

the surface of the culture keeps volume constant. The temperature was set to 30°C and agitation to 500 rpm. pH-controlled with NaOH and HCl. Chemostat cultivation ended after 7.2 volume changes in 72 hours.

Cells from the cultivation analyzed for msfGFP fluorescence with a FACS device (BD, Becton, Dickinson Company, New Jersey, USA). Therefor cells were centrifuged for 5 minutes at 5,000 rpm. The resulting pellet was resolved in 1xPBS and diluted to a cell count of 3,000 per second by the FACS device. To avoid debris in the solution, a CellTrics 50 µm was used for filtering (Sysmex, Norderstedt, Germany).

2.7 Statistics

Each promoter construct was characterized in 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 6-9 biological replicates. For each construct, the mean and standard error of the mean was calculated from these combined biological replicates. The significance of the difference in the activity of constructs with different spacer lengths was analyzed by one-way ANOVA with Turkey´s post-hoc comparison. The coefficient of variation, determined by dividing the absolute difference of the predicted and experimental value by the experimental value, was used to compare the accuracy of prediction of stacked promoter activities.

2.8 Miller assay

The promoter strength of different pSEVA247y-lacZ plasmids was determined by applying the Miller Assay.[158] For the cultivation, we used minimal medium, as described above, with 20 mM glucose and 50 mg L^{-1} kanamycin. Overnight cultures were used as precultures and main cultures were inoculated with an initial optical density of 0.1. Cultivation was done in 24 well System Duetz plates (Enzyscreen, Heemstede, The Netherlands) at 30°C and 300 rpm. A total volume of 1.5 mL was used and cultivation performed for 16 hours. Optical density was measure at 600 nm. 20 µL of each culture was mixed with 980 µL KP_i-puffer (adjusted to pH 7.0). For cell lysis, 40 µL chloroform and 20 µL SDS (1 g L^{-1} solution) were added and mixed for 10 seconds. Initial incubation for 5 minutes at 37 °C and 200 rpm was performed in a thermomixer. With the addition of 200 µL ONPG (4 mg mL^{-1}) the reaction starts. The reaction was stopped by the addition of 0.5 mL of 1 M Na_2Co_3 solution and the time noted. The time of stopping the reaction depends on the yellow coloring of the solution. After a centrifugation step for 3 minutes at 13,000 rpm the supernatant was transferred to a cuvette. The solution was measured at 420 nm in spectrometer Cary 60 (Agilent, Santa Clara, California, USA). Resulting Miller Units were calculated with the following formula: $Miller\ Units = \frac{1000 * OD420}{t * V * OD600}$

Chapter 3

Results

Chapter 3.1
Characterization of context-dependent effects on synthetic promoters

Published as:
S. Köbbing, L. M. Blank, N. Wierckx. “Characterization of context-dependent effects on synthetic promoters”, *Front. Bioeng. Biotechnol*. (2020), 8(551), doi: 10.3389/fbioe.2020.00551

Contributions

Sebastian Köbbing generated plasmids and genomically modified strains and performed the experiments. Atscharah Panyot assisted in the generation of the single nucleotide polymorphism promoter library and characterized resulting strains for promoter activity. Atscharah Panyot and Sebastian Köbbing performed experiments separately and data were merged afterwards. Sebastian Köbbing prepared the figures and wrote the manuscript with the help of Nick Wierckx and Lars M. Blank. Sebastian Köbbing and Nick Wierckx analyzed the results. Nick Wierckx and Lars M. Blank conceived the project.

3 Results

3.1 Characterization of context-dependent effects on synthetic promoters

3.1.1 Abstract

Understanding the composability of genetic elements is central to synthetic biology. Even for seemingly long-time known elements such as a sigma 70 promoter the genetic context-dependent variability of promoter activity remains poorly understood. The lack of understanding of sequence to function results in highly limited de novo design of novel genetic elements combinations. To address this issue, we characterized in detail concatenated "stacked" synthetic promoters including varying spacer sequence lengths and compared the transcription strength to the output of the individual promoters. The proxy for promoter activity, the msfGFP synthesis from stacked promoters was consistently lower than expected from the sum of the activities of the single promoters. While the spacer sequence itself had no activity, it drastically affected promoter activities when placed up- or downstream of a promoter. Single promoter-spacer combinations revealed a bivalent effect on msfGFP synthesis. By systematic analysis of promoter and spacer combinations, a semi-empirical correlation was developed to determine the combined activity of stacked promoters.

3.1.2 Introduction

The Pseudomonads are promising group of bacteria for industrial applications.[1-3] A versatile metabolism enables them to grow on several carbon sources like glucose and glycerol, but also on a wide range of aliphatics and aromatics.[4-6] Different *Pseudomonas* strains have been engineered for the production of chemicals with industrial importance from different renewable carbon sources, like furandicarboxylic acid, rhamnolipids, and aromatics.[1, 12-16] *Pseudomonas* is highly tolerant to chemical stresses and can survive harmful conditions caused by oxidative stress.[1, 16, 19, 20] Some strains can thrive under a second phase of toxic hydrophobic solvents such as toluene or styrene.[21, 22] *P. putida* KT2440 is a non-pathogenic representative of this versatile group of bacteria.[151] The strain can produce and accumulate polyhydroxyalkanoates (PHA) as a storage polymer in granules under nitrogen depletion from different carbon sources like glycerol, glucose, ethylene glycol, 1,4-butanediol, or fatty acids.[9-11, 15]

Parallel to the increasing industrial interest in Pseudomonads, an ever-increasing set of synthetic biology tools is developed for this genus. These include genomic integration tools like transposon Tn5 [112, 114, 143] or Tn7 [159-161], as well as a suite of tools for targeted and marker-less integration.[26] To obtain gene replacements, counter-selection procedures were established, like *sacB* originating

from *B. subtilis*.[162] New tools are based on CRISPR/Cas9 showing high potential for whole-genome engineering approaches.[3, 163] These tools enable a deep genetic and metabolic re-factoring of Pseudomonads as exemplified in the engineering of streamlined chassis strains.[16, 31, 164]

Especially when such deep engineering entails the (over-) expression of many homologous or heterologous genes, balanced and reliable gene expression is required, which doesn't unnecessarily burden the cell. In this context, calibrated synthetic promoter libraries enable modulation of enzyme expression in metabolic pathways and protein production.[165, 166] Two major ways to generate a promoter library are prominently used. A low degeneracy approach, where only a few random nucleotides are introduced, reduces the number of possible generated promoter sequences and thus decreases the number of sequences, which have to be tested.[167] This allows a deeper insight into promoter sequence-activity relationships. On the other hand, high degeneracy promoter libraries based on a degenerated core promoter sequence lead to billions of different possibilities.[27, 30, 168] While the sequence space clearly outnumbers the experimental space possible to address, a high resolution of different promoter activities is possible. The use of calibrated and standardized synthetic promoters covering a range of activities are commonly used.[27] Constitutive synthetic promoters are generally based on sigma-70 (σ^{70}) factor core promoters.[126] The σ^{70} factor encoded by *rpoD* guides the RNA polymerase to many promoters active during growth including the expression of housekeeping genes.[125, 169] Varying the consensus sequences of the -10 and -35 elements, which are recognized by the holoenzyme as part of the core promotor [127, 170], leads to weaker expression strength in *E. coli* and *Pseudomonas aeruginosa*.[122] The σ^{70} factors consensus sequence of *P. putida* KT2440 and *P. aeruginosa* are identical.[27, 127]

Characterization of synthetic promoters has been performed using plasmid-based expression systems or genomically integrated probes.[27, 118, 120] However, varying plasmid copy numbers and high fitness costs for the host makes plasmid-based expression systems less suitable for promoter characterization in particular, and for metabolic engineering in general.[99, 100, 102, 171] Genomic integration of the probe is preferred for characterization procedures.[27] The major difference is the fact that many of the plasmids used are multicopy, which increases the variability of the reporter output by copy number variations. In addition, an often-overlooked disadvantage of using multicopy plasmids for synthetic promoter screening is that they favor the selection of relatively weak promoters, as the combined effect of a strong constitutive promoter at high copy number may pose a too high burden. Genomic integration abolishes these copy number effects, as well as clonal variations, which have also been observed for different *Pseudomonas* strains.[27, 101, 171] Nevertheless, the integration site in the genome must be chosen wisely and must be the same for all promoters. The expression activity differs not only for single genes, but also in larger regions on the genome ("hot" and "cold" spots). Therefore, we used a mini Tn7 transposon, which

integrates in a targeted manner downstream of the *glmS* gene in the *attTn7* site of a broad range of bacteria including *P. putida* KT2440, thereby enabling reliable and stable expression.[27, 115, 159] Synthetic promoter libraries are described for *P. putida* KT2440.[27, 30] However, the predictability and composability of these promoters in different genetic contexts is poorly understood. Li *et al*.[172] have shown that different numbers of promoters in tandem direction result in increased activities. Several other publications feature tandem promoters, but so far without a characterization of these promoter combinations that focusses on composability and predictability of the activity of these genetic elements.[173-175] The combination of promoters in different contexts is also a key element in logic gate construction in synthetic biology, and sensitivity to genetic context is considered a challenge there.[176]

In this work, we stacked (concatenated) promoters of a previously published promoter library from Zobel *et al*.[27] in series and analyzed the resulting activities as single genomically integrated probes by measuring *msfGFP* expression.[177] The obvious assumption that the combination of two promoters would yield their summed activity proved to be false. The reasons for this are investigated and a semi-empirical correlation was developed to reliably predict stacked synthetic promoter activities. This provided insights into the context-dependent activity of promoters that may foster a better predictability and composability of this key element of synthetic biology.

3.1.3 Results and discussion

3.1.3.1 Identification of the optimal distance between two promoters

For the characterization of stacked promoters, we used a mini Tn7 transposon, which integrates as single copy into the *attTn7* site downstream of the *glmS* gene in the genome of *P. putida* KT2440.[27, 115, 144] The transposon is designed to characterize promoters in a reliable and reproducible manner, featuring a BCD2 element to reduce GOI-based expression variability [167], a *msfGFP*[177] gene as reporter, and two flanking terminators to minimize genomic read-through (Figure 9).[27]

In order to determine the optimal distance between two promoters, we stacked the 14f and 14g promoters from a previously published synthetic promoter library[27] with spacer sequences with increasing length from 10 to 100 bp by extension at the 3′-end. The promoters are referred to solely by their SEVA code (Zobel *et al*.[27], 14a-g, with a being the weakest and g being the strongest) for ease of reference. The spacer was randomly generated and manually curated for unwanted restriction sites as well as putative ribosome binding sites, -35, and -10 like sequences, which could disturb the analysis due to intrinsic activity (http://www.faculty.ucr.edu/~mmaduro/random.htm).

The spacer was created with a GC content of 62%, similar to the genomic average of *P. putida* KT2440.[151]

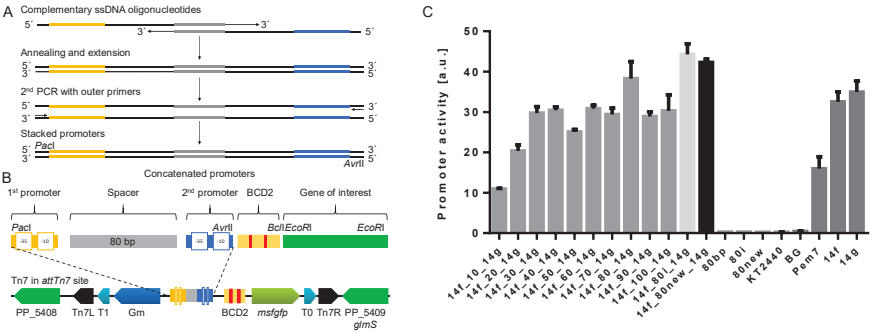


Figure 9: Identification of the optimal spacer length between the two promoters 14f and 14g using a mini Tn7 vector.[27] (A) Two PCR reactions were performed to generate stacked promoters with longer (>45 bp) spacer sequences and promoter combinations with the 80i spacer sequence. In a first PCR reaction, long single-stranded DNA oligonucleotides containing one promoter sequence (yellow or blue) were annealed via complementary sequences in the spacer (grey) and extended by Q5 polymerase to double-stranded DNA. The resulting dsDNA fragment was amplified in a second PCR. (B)Structural organization of stacked promoters and of the mini Tn7 used in this study after genomic integration. Stacked promoters consisting of promoter sequences at two positions separated by a spacer were inserted via restriction sites *Pac*I and *Avr*II. BCD2 element for translational coupling and *msfGFP* as reporter gene. Tn7 module contains a Gm^R marker for selection and two terminators (T0 and T1) for the insulation of the probe. Tn7R and Tn7L are recognized by a transposase. (C) Tested spacers contained between 10 and 100 bp. *P. putida* KT2440 *attTn7::*BG14f##g-msfGFP, where ## refers to the number of nucleotides in the spacer sequence (grey bars), were cultured in a BioLector in minimal medium with 20 mM glucose in a 96 well plate. The control strains BG13 with the P_{em7} promoter of average strength, the individual promoters 14f and 14g, and promoterless BG and wild type *P. putida* KT2440, as well as additional controls with two 80 bp spacers 80i and 80new are also shown. Identical strains from at least two from different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n>6).

A promoterless construct BG and wildtype *P. putida* KT2440 were used as negative controls. As positive controls, we used the single calibrated promoters described in Zobel *et al.*[27], including P_{em7}[33], which reaches half of the activity of promoter 14g. With short spacer sequences of <35 bp, the activity of the stacked promoter is lower than that of either of the single promoters (Figure 9). This result is likely caused by steric hindrance of the RNA polymerase holoenzyme since the combined sigma factor and RNA polymerase cover around 80 bp upstream of the promoter.[178] A spacer length of 40 - 70 and 90 -100 bp resulted in activities of comparable strength. Significantly higher activity was observed for 14f_80_14g with an 80 bp spacer compared to all other spacer lengths (one-way ANOVA with Turkey´s post-hoc comparison).

To exclude that this 80 bp spacer is an outlier due to possible activating sequences, the experiment was repeated with a reverse complement version of the spacer (80i) and a new, independently generated spacer sequence (80new, Table 5). All three 80 bp control spacers led to comparable activities, indicating that this distance between two promoters is promoting additive activity of the two promoters. It is interesting to note that the spacer length of 80 bp matches the sequence covered by the RNA polymerase holoenzyme[178], although this correlation should not be confused with

causation. Promoterless controls with only 80i and 80new also show no activity, excluding any intrinsic activity from the spacers themselves (Figure 9).

With the 80i spacer a cumulative effect occurred, with the total output of the stacked promoters being higher than the individual activities. However, the output was much lower than the sum of the two individual promoters for each tested spacer length. For further characterization, we used the 80i spacer since it enabled the highest promoter activity.

3.1.3.2 Characterization of context effects on stacked promoters

We hypothesized two possible ways how these stacked promoters are affected. The primary hypothesis is that of an effect of the spacer on the promoter. The alternative hypothesis is a mutual influence of one promoter on the other.[179, 180] To test these hypotheses, we constructed 14 different stacked promoter combinations and 12 controls to determine the influence of the 80i spacer and upstream sequences on single promoter activities. Following the rules provided by SEVA[33], the promoter is integrated between restriction sites *Pac*I and *Avr*II. The spacer is an additional sequence in the probe vector published by Zobel *et al*.[27] and is not replacing any sequences from the original construct. The constructs are named according to their composition, i.e., in construct 14f_80i promoter 14f is cloned upstream of the 80i spacer. After genomic integration of the Tn7 transposon into the genome all strains were characterized in a BioLector system (Figure 10).

The single promoter controls without spacer reached activities that are comparable to those initially described by Zobel *et al*.[27] In contrast, single promoter combined with the 80i spacer, either upstream or downstream, were strongly affected in their activity (Figure 10). With downstream placement of the spacer, all promoters were negatively affected, with decreases up to 70% for 14f_80i. In contrast, no clear trend could be discerned with upstream placement of the spacer, with most combinations having decreased activities up to 28%, but 80i_14c gained 12 % and 80i_14d even 50% activity.

These results show that the spacer has a drastic effect on all tested promoters. This is in spite of the fact that the spacer itself doesn't display any promoter activity (Figure 9), nor does it contain any discernible sequences that might affect promoter activity, such as AT-rich UP elements.[181] In addition, up- and downstream effects of the spacer are unpredictable. Most of the single promoters show a decreased activity when combined with the spacer, which could be explained by missing upstream activating elements potentially present in the original construct such as the AT-rich *Pac*I restriction site. While no consistent correlation between spacer position and promoter activity is discernible, the results do confirm the primary hypothesis that the activity of the promoters is affected by the spacer.

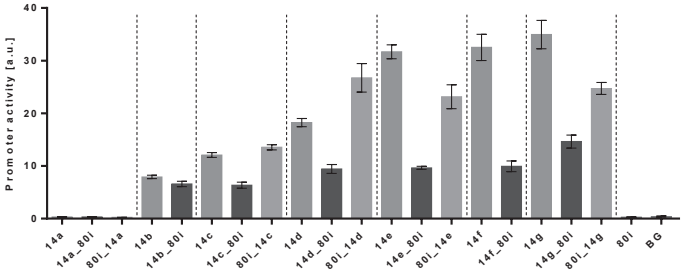


Figure 10: Characterization of context-depended promoter activities using the 80i spacer. Promoters with both up- and downstream spacer were tested. Shown are promoter activities for the original promoters (dark bars) from Zobel *et al.*[27] and promoter-spacer (14x_80i) as well as spacer-promoter (80i_14x) combinations (grey bars), where x stands for promoter 14a to 14g. All constructs were genomically integrated in *P. putida* KT2440. Strains were cultured in a BioLector in minimal medium with 20 mM glucose in a 96 well plate Identical strains from at least two different transformations were tested, with three biological replicates each. Vertical dotted lines are separating individual sets. Error bars indicate the standard error of the mean (n>6).

To further test if, besides the effect of the spacer, the stacked promoters also affect each other, all seven calibrated promoters were stacked with 14g at the second position. As an additional control, the reverse-order combination 14g_80i_14a was also included. As expected from the abovementioned results, all of these combinations led to much lower activities than expected from the sum of the individual promoters disregarding context-effects (Figure 11). Interestingly, the combinations 14a_80i_14g and 14g_80i_14a, which only differ in the order of the promoters, reached completely different activities. During the cloning of these stacks, a triple 'ffg' promoter consisting of two 14f and one 14g sequences separated by two spacers was accidentally created, in which the second 14f is shorter by two nucleotides between the -35 and -10 elements. Deriving sequence-function relationships from this promoter would be too complex, but the fact that it is around 45 % stronger than the strongest promoter combination 14g_80i_14g makes it useful in applications where very high expression is needed (Figure 11).[7, 17]

When comparing the activities of these stacks to the single promoter-spacer controls above, it becomes apparent that the immediate context of single promoters is the major determinant for the prediction of promoter activity. The sum of the single promoter activities greatly overestimates the activities of stacked promoters by as much as 140% for the 14g_80i_14a combination (Figure 11). In contrast, the sum of the context-specific controls provides a much more accurate prediction, i.e., 14g_80i + 80i_14a = 14g_80i_14a. In this case, the coefficient of variance between context-specific prediction and experimental values is lower than 15% for all tested combinations. This strongly suggests that, once the direct context of the individual promoters is sufficiently taken into account, the stacked promoters don't affect each other's activity.

Beyond having different promoter contexts, the abovementioned constructs also generate different 5'-terminal mRNA ends, which may cause differences in mRNA stability or translation initiation

rates. In order to minimize the effect of these differences, a bicistronic design was included in the reporter construct .[167] To verify whether the altered expression of context-affected constructs is caused by increased transcription, we performed quantitative real time PCR (qRT-PCR) of selected constructs. Determined transcript levels correlate well with promoter activities derived from fluorescence measurements ($r^2 = 0.95$, Figure 11), confirming that the spacer influences transcription rather than translation. Attempts to determine the relative contributions of the first or second promoter by qRT-PCR were inconclusive. In principle, stacked promoters generate two overlapping transcripts of different lengths, which might be distinguished with different primers pairs. However, longer transcripts show a shift in Ct value compared to shorter amplicons, and suitable primer pairs for similar lengths could not be found.[182]

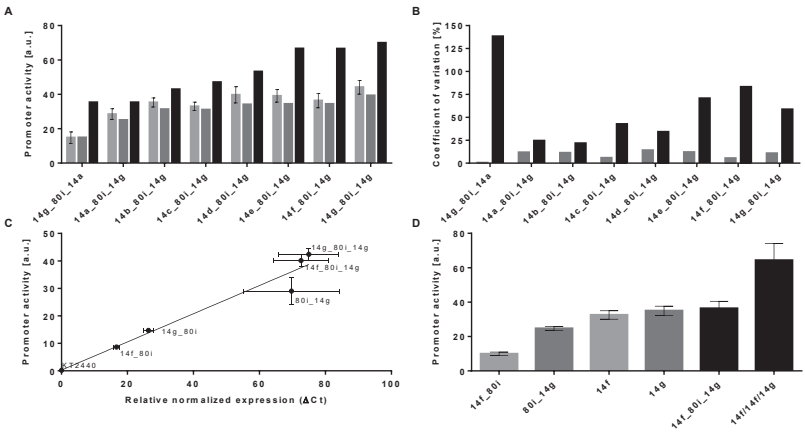


Figure 11: Comparison of experimental values, context-specific and context-unspecific prediction of promoter activities for stacked promoter with the 80i spacer separating promoters 14a to 14g. (A) Shown are determined promoter activities for stacked promoters (light grey bars), context-specific activities calculated with context-depended values (grey bars), and context-unspecific activities using original promoter activities (black bars). All constructs were genomically integrated into *P. putida* KT2440. Cultivation was done in a BioLector in minimal medium with 20 mM glucose in a 96 well plate. Identical strains from at least three different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n>9). (B) Coefficient of variation (CV) of context-specific (light grey bars) and context-unspecific (black bars) prediction of resulting promoter activities. (C) A plot of *msfGFP* transcription levels normalized to *rpoD* determined by quantitative real time PCR (qRT-PCR) and promoter activities from single promoter controls and stacked promoters. Wild type *P. putida* KT2440 was used as a negative control. (D) Characterized stacking promoters and controls genomically integrated in *P. putida* KT2440. Promoter 14f/14f/14g was accidentally generated during cloning procedures for combination 14f_80i_14g. Triple variance has an additional 14f_80i sequence inserted, whereat two additional nucleotides are integrated in the middle 14f promoter and one nucleotide is missing in the first 80 base pair spacer. All constructs are genomically integrated into the genome of *P. putida* KT2440. Cultivations to determine promoter activities were done in a BioLector in minimal medium with 20 mM glucose in a 96 well plate. Cultivation to determine transcription levels was done in 24 well System Duetz plates containing minimal medium with 20 mM glucose. Identical strains from at least two different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n>6).

3.1.3.3 Using an SNP promoter library for stacked promoters

A change in context greatly affects promoter activity, and there are large quantitative differences for each tested promoter-spacer combination. Given that the main variable between these constructs is the promoter sequence, this might be due to specific DNA-DNA interactions between promoter and spacer, which influence promoter activity, or RNA-RNA interactions, which affect RNA stability. To further investigate the sequence-activity relationship, we generated a single nucleotide polymorphism (SNP) library based on promoter 14g. Such a library yields promoters with very similar sequences, but large differences in activity. If the variability of the impact of the spacer on the activity is indeed caused by DNA-DNA or RNA-RNA interactions, using promoters with more similar sequences can be expected to reduce this variability. The library contained 90 different promoter sequences, which were generated in 30 PCR reactions with one degenerate nucleotide in the sequence. Changes of the core promoter sequence were inserted within the -35 element (position 1-6), in the interspaced region (position 7-23, 30), and the -10 element (position 24-29). For each position, four different promoters can occur, of which one will correspond to the original 14g sequence. Initial screening of mini libraries (14 clones each) of these low degeneracy promoters was performed in plasmid-bearing *E. coli* PIR2 strains by analyzing msfGFP fluorescence. Aberrations compared to original pBG14g-bearing *E. coli* PIR2 strains were recognized (Figure 12). For nearly each position clones were found with either a higher or lower fluorescence signal than the original pBG14g plasmid. Variations in the interspaced region generally had a lower effect on expression strength, while changing single nucleotides in the -35 and -10 consensus sequences yielded more clones with decreased promoter activity, as expected.[122, 170] We therefore focused further characterization on these elements in order to obtain a set of promoters with a range of activity that is comparable to the previously described calibrated promoter library from Zobel *et al*.[27] (Figure 12).

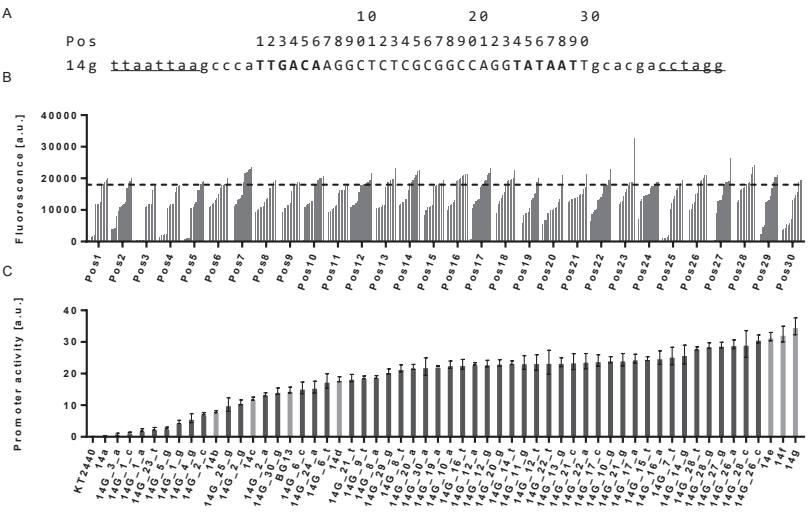


Figure 12: Comparison of screened and characterized promoter sequences based on P_{14g} with single nucleotide polymorphisms. (A) Original promoter sequence of 14g with highlighted -35 and -10 elements (in bold). Restriction sites *PacI* and *AvrII* are underlined and positions of the modified core promoter sequence are given with numbers above the sequence. (B) msfGFP fluorescence of *E. coli* PIR2 bearing plasmid pBG14g with degenerate bases at 30 positions along the core promoter sequence. Changed position is shown on the x-axis. Determined values are ranked by fluorescence intensity of 14 strains tested for each position. Strains were cultivated in 96 well System Duetz plates with LB medium supplemented with 50 mg L-1 kanamycin. Fluorescence and optical density were measured with a plate reader. The dotted line indicates the promoter activity of the 14g control. (C) Chosen SNP promoter constructs were genomically integrated into *P. putida* KT2440. Strains are named 14G_##n, whereas ## stands for the position in the promoter sequence and n for a nucleotide (A, C, G, or T). Cultivation was done in a BioLector in minimal medium with 20 mM glucose in 96 well plates. Identical strains from at least two different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n>6).

After an initial screening of positions in the SNP promoter sequences, we selected three such positions within the -35 or -10 elements for further characterization. Introducing a degenerate base in these three positions yields nine different promoters with a good spread of activity, and these promoters were combined with the 80i spacer in the downstream position (Figure 13). Changing position 26 in the -10 sequence resulted in a slightly decreased activity, while changes in positions 1 and 2 in the -35 sequence yielded larger decreases, which is in accordance with Lodge *et al*.[170] We have seen that small changes in the 14g promoter sequence can strategically affect activity in a mini-promoter-library (Table 9). In spite of the relative uniformity of the promoter sequences in this library, a combination of these promoters with the 80i spacer again strongly affected the activities with both in- and decreases up to 66 %. The reduced sequence variability did not reduce the quantitative variability of the spacer effect compared to the CalPro library from Zobel *et al*.[27] Both have a high coefficient of variation of 40% for the SNP library and 25% for the CalPro library (Figure 13). This strongly suggests that the promoter sequence *per se* does not cause the context-

dependent effect, suggesting that other factors such as the varying transcription-initiation rates are in play.

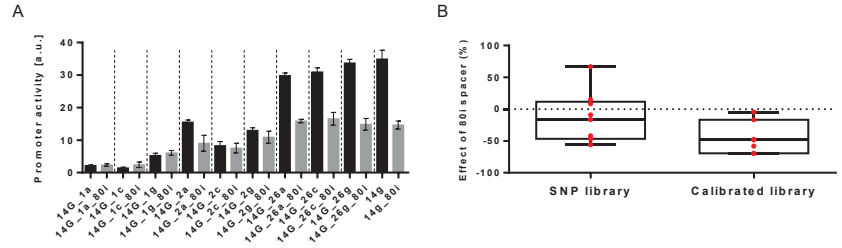


Figure 13: Characterization of the effect of the 80i spacer on a single nucleotide exchange promoter library based on promoter 14g genomically integrated in *P. putida* KT2440. (A) Promoter activities derived from GFP fluorescence analysis of SNP promoters with (grey bars) and without (black bars) downstream 80i spacer. Cultivation was done in a BioLector in minimal medium with 20 mM glucose in 96 well plates. Identical strains from at least two different transformations were tested, with three biological replicates each. Vertical dotted lines are separating individual sets. Error bars indicate the standard error of the mean (n>6). (B) Box and whiskers plot of the relative effect of the downstream 80i spacer on promoter activities of the SNP library and the original calibrated promoter library from Zobel *et al.*[27] The effect of the spacer is calculated as the % change of the promoter with downstream 80i sequence compared to the original promoter. Data points are indicated in red.

3.1.4 Conclusion

In this work, we aimed to increase the composability and predictability of synthetic promoters by investigating the effects of differing contexts on their activity. In the combination of two promoters, the length of the spacer region is crucial for reaching higher and cumulative effects, with 80 bp being optimal. Even though the spacer sequence has no intrinsic activity, it strongly and unpredictably influences the activity of promoters by up- and downstream effects. By accounting for this influence, the activity of two stacked promoters can be accurately predicted with coefficients of variance below 15%. A strong reduction of sequence variability was achieved using a SNP library, but this reduction did not reduce the quantitative variability of the spacer effect. This strongly indicates that nucleotide-nucleotide interactions between promoter and spacer do not play a prominent role. Clearly, context-specific effects of synthetic promoters are not yet fully understood in *Pseudomonas*. Although the semi-empirical approach for prediction of stacked promoter activities provides an accurate workaround to this, further work is needed to understand the fundamental interaction of genetic elements and their surroundings.

Chapter 3.2
Identification and characterization of suitable genomic integration sites in *P. putida* KT2440

Contributions

Sebastian Köbbing generated all plasmids and genomically modified strains and performed the experiments. FACS analysis was performed by Sebastian Köbbing and Volkan Besirlioglu from the Institute of Biotechnology at RWTH Aachen. Parts of the results will be published in a scientific paper. The manuscript is in preparation. Sebastian Köbbing and Nick Wierckx analyzed the results. Nick Wierckx and Lars M. Blank conceived the project.

3.2 Identification and characterization of suitable genomic integration sites in *P. putida*

3.2.1 Abstract

Using genomic integration tools like transposons or homologous recombination are state of the art to generate stable production hosts. However, with the increasing complexity of microbial chassis design, the integration of several genes and operons becomes needed to generate high performance cell factories. There is a limitation in the number of characterized integration sites in *P. putida* KT2440 for genomic integration events. To close this gap, we analyzed a set of RNA-Seq data from experiments while cultivated with different carbon sources. Focusing on genes with a low coefficient of variation under all tested conditions allowed us to identify potential genomic "landing sites" across the genome, which were further characterized by using different *msfGFP* expressing constructs and different carbon sources. A wide range of activities was observed, which shows that not all sites of the genome enable high expression and so-called "hot" and "cold" spots exist. Here we describe a set of genomic loci that enable the integration of heterologous constructs with low variability in expression while cultivated under changing conditions.

3.2.2 Introduction

Representatives of the genus *Pseudomonas* became common workhorses in biotechnology in recent years.[1-3] Thanks to their versatile metabolism and high tolerance towards organic compounds.[4, 6, 12] To increase the usability of *Pseudomonas putida* KT2440 tool development is in full progress[23, 24], enabling rapid and efficient metabolic engineering and genome editing.[3, 25-30] Besides frequently discussed CRISPR/Cas9 system for targeted genome modifications, also MAGE is available.[3, 28, 29] Both methods focus on genome modifications of the host by introducing deletions, insertions, and single nucleotide exchanges at multiple sites in parallel. But at the moment, still elaborate screening efforts are necessary to identify favored phenotypes, especially while using MAGE. State of the art is the use of homologous recombination-based gene integration and/or deletion.[16] Although this enables rapid and efficient insertion of heterologous constructs virtually anywhere in the genome, studies about characterization and description of suitable sites for integration are missing. Here we want to close this gap by systematically characterizing genomic integration sites for heterologous gene expression.

For more complex genetically modified microorganisms, for instance, for whole-cell catalysis, different genes and pathways have to be combined in a single cell.[183, 184] Here, plasmids reach their limit of usability. Plasmids are widely used in research and also play a major role in evolutionary processes in nature.[185] One of the most prominent examples for horizontal gene transfer of

plasmids or genes is the emergence of methicillin-resistant *Staphylococcus aureus* strains (MRSA) in hospitals in recent years. They are benefited by the false and unnecessary application of antibiotics.[186] While in the environment plasmid propagation is feasible, propagation of multiple plasmids in a single cell is hard to accomplish in a biotechnological context. A further challenge is posed by varying plasmid copy numbers leading to unreliable results.[99-101] Also, many plasmids cause decreased growth rates and high fitness costs for the host.[102] To maintain proper plasmid propagation and stability, the addition of antibiotics is necessary, which makes them unsuitable for large-scale fermentation.[187] Genomic integrations overcome the limitations of plasmids and are commonly used because they are more stable, less variable, and they don't require constant selective pressure for maintenance.

Transposons are often used for genomic integration. Tn5 integrates randomly into the genome and was the first discovered and characterized transposon.[113] However, targeted integration at a single site is our method of choice to reduce the variable influence of the unknown genomic context on gene expression. Here, transposon Tn7 is very popular across different species, because integration always occurs at the typically single *attTn7* site, thereby reducing noise from random genome context.[27, 115, 188] Zobel *et al*. developed a probe vector for the calibration of synthetic promoters, but marker-recycling was not considered.[27] For targeted, marker-less integration into the genome, as well as deletion and provoking single nucleotide exchanges, the pEMG system is currently used extensively.[26] This technique is based on homologous recombination followed by I-*Sce*I mediated double-strand breaks.[189] It is very popular due to its ease of use, allowing nearly all sites of the genome to be modified, including 16s rRNA genes of *P. putida* KT2440.[190-192] Using pEMG for genome modifications allows the user to delete large fragments and to generate genome reduced chassis strains with improved properties.[16, 23, 31, 193]

Many studies involve genomic integration of heterologous constructs. But often, the integration was done with transposon Tn7 or genes were directly replaced by applying the pEMG system.[7, 9, 16, 17] A systematic identification of different usable integrations sites along the full genome, which accounts for the effect of the insertion on the genomic context, and vice versa, was never done. Transposon Tn5 is often used for genomic integration in several gram-negative microorganisms and therefore, different transposable Tn5 vectors are available.[112, 114, 194] Here, random integration is used to drive heterologous gene expression from native promoters or inducible systems to investigate gene function or to generate production strains.[191, 195] However, these experiments often rely on genomic read through activity, based on native upstream promoters, which results in different activities depending on the integration site. Random transposon integration often disrupts native genes and can lead to lethal effects. Although this enables the investigation of gene-function relationships in microorganisms[196, 197], for biotechnological purposes, random integration requires

large screening efforts, especially when the product is not easily detectable. Characterization based on determinable fluorescence or color formation can help this screening but is not always feasible.[131, 191] To avoid these drawbacks in the dependency on native promoter activity for heterologous gene expression, the construct should contain its own promoter, isolated from its genomic context. In order to achieve this, we flanked our construct with transcriptional terminators, relying on an included constitutive promoter to drive expression of the gene of interest. However, even when excluding read-through from native promoters in the genomic context, also other factors can affect the expression of genomically integrated genes. Kuhlman *et al*. characterized four different integration sites in *E. coli*.[198] These sites were chosen after distribution along the genome in order to the distance to the origin of replication of the genome. A second publication from this group compared these four sites for the production of different compounds and has shown that the integration site affects product formation.[199] They hypothesized that the genomic context influences the expression of the integrated gene. On the other hand, DNA replication itself can affect heterologous expression at different sites.[200] During cell division, DNA replication is starting from multiple sites of the genome, in this step, so-called replication forks are formed. Depending on the distance to the origin of replication, the cell has variability in the copy number of integrated genes, which can lead to various expressed regions.[200] This leads to the assumption of the existence of "hot" and "cold" spots along the genome, which are differentially expressed during cultivation. "Cold" spots are less active due to, among other factors, local folding of DNA. "Hot" spots may more often be single-stranded, having better accessibility to RNA polymerase. On mRNA level 5′ untranslated region play a major role in post-translational regulation on translation.[135] Forming hairpins in 5′UTR leads to an inhibition of ribosome binding to the Shine Dalgarno (SD) sequence and consequently, no translation occurs. This can be caused by so-called standby sites, which is similar to a holding point and delays translation.[201] Relaxation of the single-stranded sequence (mRNA) again opens the SD sequence and protein biosynthesis can progress. The question is how to use the knowledge about "hot" and "cold" spots in the field of metabolic engineering and synthetic biology.[199, 202]

With the knowledge in hand that gene expression is affected by genomic context, we aimed to identify genomic integrations sites fostering efficient transcription in *P. putida* KT2440 by applying a transposon Tn5 library. This Tn5 library was sorted by FACS (similar to Kosuri *et al.*[131]) to correlate fluorescence intensities with integration sites (Figure 14). RNA-Seq analysis aimed on equally expressed genes across the genome and different data sets (Figure 17).[149, 150, 203] Identified sites were further characterized by targeted integration using pEMG.[26]

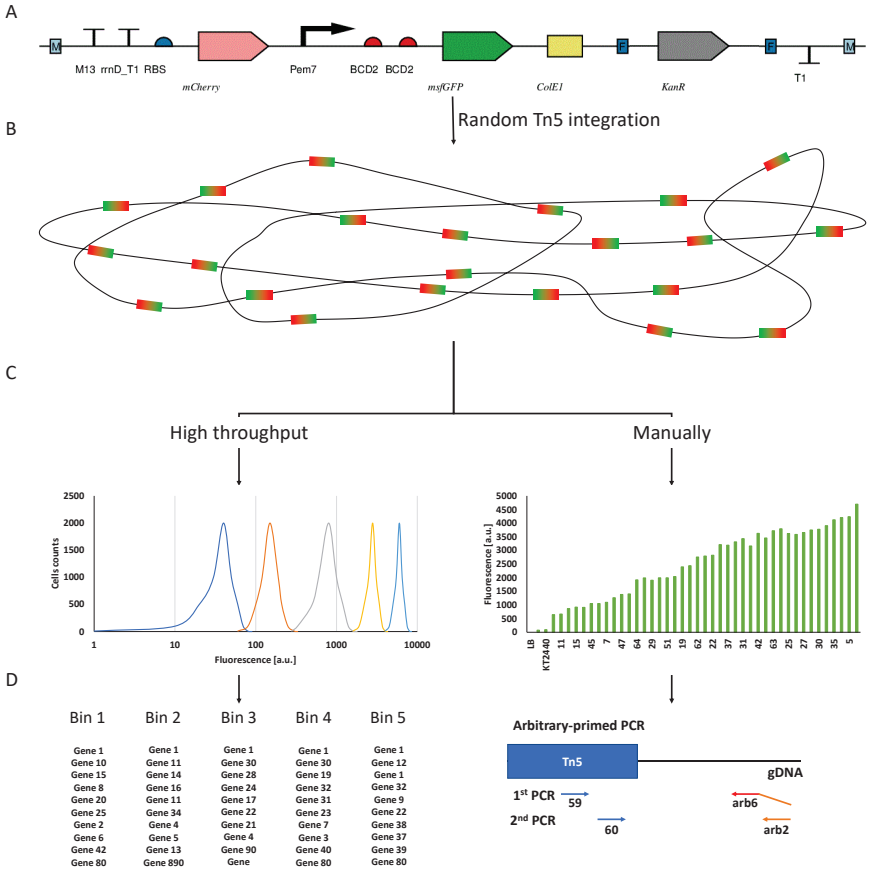


Figure 14:Workflow for the identification of possible integration sites using transposon Tn5_sensor as high throughput and manual approach. (A) Map of the generated probe sensor based on pBELK and integrated into a transposon.[114] The transposon is flanked by displayed mosaic ends (M). A double terminator was cloned in front of the construct, consisting of Phage M13 and *E. coli* K12 *rrnD1* terminator sequences, to isolate the sensor from upstream native promoter activity.[78] Promoterless *mCherry*[204] provides as a control for read-through activity. A combination of both allows the identification of "hot" and "cold" spots independent of read through promoter activity. Promoter P_{em7}[26] drives the expression of the reference gene *msfGFP*, BCD2[167] is a translational coupler. *ColE1* could be used for a plasmid rescue to identify the integration site of the Tn5_sensor. The kanamycin resistance gene is used for the selection of Tn5_sensor bearing strains; a marker recycling is possible by using flippase, which recognizes kanamycin flanking FRT sites. A second terminator T1 downstream of the kanamycin resistance gene insulates the sensor. (B) Schematic randomly integrated Tn5_sensor constructs in the *P. putida* KT2440 genome. Tn5 integrates as single copy into the genome and is only here shown at multiple sites of the genome, representing the cell population. Red part indicates the *mCherry* gene, green highlight *msfGFP* gene. (C) Measured msfGFP fluorescence signals of individual *P. putida* KT2440 Tn5_sensor strains. Coloured framed regions indication windows sizes for FACS sorting including 20 % of all measured cells. (D) Expected fluorescence distribution of remeasured bin containing sorted cells with different fluorescence intensities. Colours of (C) Comparison of excepted results for determined msfGFP fluorescence levels using high throughput FACS analysis (left) and manual measurements (right). (D) Sequencing approaches showing results for high throughput next generation sequencing of sorted cells (left) and manually identification of integration site by arbitrary-primed PCR.

3.2.3 Results and discussion

3.2.3.1 Hot and cold spots across the genome affect activity of randomly integrated Tn5 in *P. putida* KT2440

Identification of suitable integrations sites across the genome of *P. putida* KT2440 was based on determining the expression strength of randomly integrated transposon Tn5 containing a probe sensor (Figure 14).[114] The activity of our insulated probe sensor was determined by measuring msfGFP fluorescence during growth, which was driven by promoter P_{em7}.[26] We hypothesized that the resulting promoter activity is influenced by genomic "hot" and "cold" integration sites, even when read-through of upstream promoters is excluded. Depending on the activity suitable sites could be further characterized using pEMG for targeted and marker-less integration.[26]

Insulation of the probe sensor is ensured by two terminators (Figure 14 A). A fusion of phage M13 and *E. coli* K12 *rrnD1* terminators reported to block 99.8 % of transcription in *E. coli* [78] is placed on the 5'-end of the construct to block genomic read-through. A promoterless *mCherry*[204] is placed behind this terminator to determine if any read-through occurred. As an output for genomic expression, a *msfGFP* gene is driven by the constitutive promoter P_{em7}[26] linked with a BCD2[167] element for more stable and reliable expression. The transposon is based on pBELK published by Nikel *et al*.[114] We modified further parts of the transposon by addition of FRT recognition sites to the kanamycin resistance gene for later marker recycling, and the marker was inverted to avoid read through activity from this gene. An *oriColE1* origin of replication was integrated for a plasmid rescue to identify the integration site.

To assess the sensor construct, 64 randomly picked *P. putida* KT2440 Tn5_sensor bearing strains were tested. For comparison, Tn7 integration strains were used. *P. putida* KT2440 BG13 was used as a positive control for *msfGFP* expression. Strain *P. putida* KT2440 BG13_mCherry was generated as a positive control for *mCherry* expression.[27] Both strains contain the same promoter and BCD2 element as used in the sensor. For a direct comparison of sensor-bearing strains, a Tn7 was cloned containing the complete probe sensor resulting in strain *P. putida* KT2440 *attTn7*::sensor. Compared to *P. putida* KT2440 BG13, this *attTn7*::sensor control showed a 70% reduction of the fluorescence level, likely due to the presence of the double terminator (Figure 15).[27] The 64 randomly selected Tn5_sensor-bearing strains reveal a wide range of fluorescence. Although none of these showed higher GFP fluorescence than *P. putida* KT2440 BG13, there were 24 strains with stronger fluorescence than the more directly comparable control *P. putida* KT2440 *attTn7*::sensor. Measuring mCherry fluorescence reveals successful insulation of the sensor, with most strains showing only very low fluorescence signals caused by genomic read-through (Figure 15). Only *P. putida* KT2440 Tn5_sensor strains 4, 34 and 40 reached higher signals for mCherry.

These three strains also show high msfGFP fluorescence signals, which could be explained by high activity of the integration site.

Given that msfGFP is expressed from the same constitutive P_{em7} promoter in all 64 individual clones. That read-through from potential upstream promoters is excluded with the double terminator and promoterless mCherry, one would expect a rather homogenous GFP fluorescence distribution. In fact, exactly the opposite occurs, with a wide range in fluorescence signals measured for msfGFP. This confirms our hypothesis of "hot" and "cold" spots across the genome.

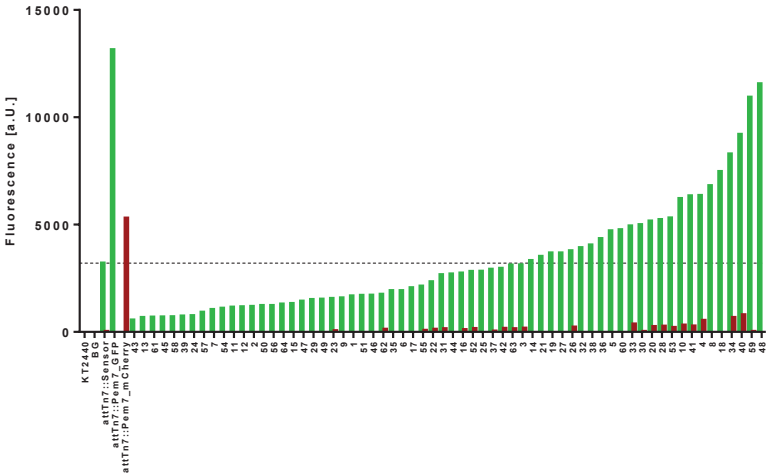


Figure 15: Characterization of 64 randomly picked clones with random Tn5_sensor integration by measuring msfGFP fluorescence. Strains were cultured in 96 well System Duetz plates with minimal medium[138] containing 20 mM glucose. Samples were measured for msfGFP fluorescence in black bottom 96 well plates at an excitation wavelength of 488 nm and emission wavelength was set to 520 nm. Absorption was measured at 600 nm in clear bottom 96 well plates. Green bars represent msfGFP fluorescence and red bars mCherry signals. *P. putida* KT2440 *attTn7*::sensor, *P. putida* KT2440 BG13_mCherry, and *P. putida* KT2440 BG13 were used as controls. All strains were tested in single reactions (n=1).

The integration sites in the different Tn5_sensor-bearing strains were identified by arbitrary-primed PCR and sequencing.[114] Arbitrary-primed PCR is based on a combination of targeted and randomized oligonucleotides. In a first-round, fragments with different lengths are generated randomly, which is only successful if the randomized oligonucleotide binds near the targeted oligonucleotide. In a second step, these fragments are amplified using two defined, nested oligonucleotides. Seven out of 64 individual clones were chosen for further characterization. These clones achieved different fluorescence signals with a wide range compared to the tested pool and are listed in Table 27. Integration sites for five clones were identified by sequencing and BLAST alignment against the *P. putida* KT2440 genome (Table 27).[151, 205] The transposon in clone 10 had integrated into a 16S rRNA gene, while in clones 3 and 12, integration had occurred in a 23S rRNA gene. The genome of *P. putida* KT2440 contains seven nearly identical *rrn* operons, with highly

conserved 16S and 23S sequences. Due to this, a clear assignment to one specific rRNA gene was not possible. Transposon integration of clone 1 was identified in *hutU*, (PP_5033) encoding an uroconate hydratase, which is part of the histidine metabolic pathway. Clone 2 seems to have an integration in cardiolipin synthase 1 (*clsA,* PP_5364). This protein is part of the lipid metabolism and generates cardiolipin from phosphatidylglycerol. Identification of clones 18 and 59 was not possible, because sequencing results and following BLAST search found a couple of sites with a short length of sequence accordance.

Table 27: Identified genomic integration sites of transposon Tn5 in chosen *P. putida* KT2440 Tn5_sensor strains. Flanking regions were amplified by arbitrary-primed PCR[114] and fragments sequenced. Given are the gene names and numbers, as well as the positions and orientations of the sensor compared to identified genomic site[151]. Corresponding fluorescence values are also given.

Clone	Gene name	Gene number	Position [bp]	Fluorescence [a.u.]	Orientation at integration site[a]
1	Uroconate hydratase *hutU*	PP_5033	5736423-5736644	1673	-
2	Cardiolipin synthase 1 *clsA*	PP_5364	6113184-6113031	1183	-
3	rRNA23S ribosomal RNA	n.d.	n.d.	3119	n.d.
10	rRNA16S ribosomal RNA	n.d.	n.d.	6216	n.d.
12	rRNA23S ribosomal RNA	n.d.	n.d.	1171	n.d.

a) Orientation is given with + and –, + means the same and - opposite direction of the hit gene

Arbitrary-primed PCR enabled the identification of Tn5 integration sites, but the approach is error-prone and time-consuming. Only after several attempted amplifications and sequencing reactions were usable results achieved.[114] Integration sites of clones with high fluorescence signals reveals 16S or 23S rRNA genes. These genes are known to be highly expressed and other studies have shown that these sites can be used for targeted integration of genes enabling the production of different metabolites.[191, 192] Identification of integrations sites of several hundreds of clones is not feasible with this method. Furthermore, highly expressed 16S rRNA genes can't be unambiguously assigned due to the high sequence similarity. For these hits, a separate PCR has to be performed for each of the seven operons.

3.2.3.2 High throughput FACS analysis of *P. putida* KT2440 Tn5_sensor

Analysis of thousands of individual clones would allow a detailed look on the influence of the genomic context on the integrated sensor, but this is not feasible by hand. To enable a more high-throughput approach, we generated a Tn5 library consisting of nearly 500,000 individual clones generated in multiple matings, which were merged in a single pool. This corresponds to a statistic coverage of one integration every 13 base pairs. To characterize as many clones as possible, we used a fluorescence-activated cell sorter (FACS, BD, Becton, Dickinson Company, New Jersey, USA) and tried to sort cells into bins with a distinct fluorescence range.

Random transposon integration can cause fitness defects, like lower growth rates and negative effects on gene expression. In batch cultivation, resulting slow growing mutants could be out-competed by fitter ones. To avoid this issue, we used a chemostat with a fixed slow growth rate so that all mutants have the same chance. Therefore, the library was cultivated in a mini bioreactor under chemostat conditions with a dilution rate of 0.1. To avoid surviving *E. coli* helper strains from the previous mating procedure, the cultivation was done for 72 hours (7.2 volume changes). During cultivation, the cell number was determined by dilution series to investigate the total number of cells. The number of colonies was higher on LB agar plates compared to cetrimide agar plates, indicating a significant presence of background *E. coli* or *P. putida* cells from the conjugation. Although these cells will provide noise in downstream analysis, they can be selected against by sorting and picking clones on selective plates.

An expected steady state of all cells in the chemostat cultivation is favored, due to similar cellular activity provided by an equalized growth rate. This ensures minimal variability in constitutive GFP expression but also protects slower-growing mutants from being out-competed. For comparison *P. putida* KT2440 BG13, *P. putida* KT2440 BG14g and *P. putida* KT2440 *attTn7*:: sensor strains were used.[27] PBS diluted cells were measured for msfGFP and mCherry fluorescence resulting in plots shown in Figure 16. msfGFP signal for *P. putida* KT2440 BG14g was very stable and located to the same area. Unfortunately, *P. putida* KT2440 BG13 shows a more bimodal distribution. *P. putida* KT2440 *attTn7*::sensor again revealed a more unimodal msfGFP signal. As expected, all three strains show low levels of mCherry fluorescence. Cultivated *P. putida* KT2440 Tn5_sensor library shows a wide range for msfGFP fluorescence, which reflects our previous results. At least five bins with sorted cells were focused on containing different fluorescence strength. Measured cells were plotted and five windows, each framing approximately 20 % of the cells were defined (Figure 16). The plot disclosed that msfGFP signals are not normally distributed as expected. Rather, two populations occurred during cultivation. Sorted cells were collected in reaction tubes for further cultivation and genomic DNA isolation. In addition, sorted cells were spotted directly on agar plates. From these plates, clones for each window were cultivated in minimal medium containing 20 mM glucose as carbon source and promoter activity was determined. Unfortunately, no difference in promoter activity was observed for different bins (Figure 16), indicating that the sorting was not successful due to technical difficulties. The size of single bacterial cells is much smaller than eukaryotic cells and therefore, the location inside the droplet, which is measured, is essential. Cells located more to the edge of the drop give a different fluorescence signal than the same cell located in the middle. This leads to lousy sorting and further analysis of the bins was not constructive.

We have shown that the integration of the transposon Tn5 containing the sensor construct is affected by the genomic context. This proves the presence of "hot" and "cold" spots in the genome of *P. putida* KT2440. The double terminator prevents genomic read through effectively but also reveals a negative influence on P_{em7} activity, as shown by the control construct integrated into the *attTn7* site. Therefore, characterized activities for *P. putida* KT2440 Tn5_sensor bearing strains could be higher by omitting the upstream terminator, which would likely increase the intrinsic activity of the sensor promoter, and also enable genomic read through to increase further. FACS analysis confirmed our first findings and disclosed a wide range of fluorescence intensities. But a FACS sorting and following next-generation sequencing (NGS) of genomic DNA originated from sorted bins was not realized due to technical difficulties in the sorting mechanism of the FACS. We have shown that identification of integration sites is possible by hand, but with an increasing number of interesting clones, the feasibility is decreasing. A next-generation sequencing (NGS) approach is more promising and resulting data can be compared. Kosuri *et al.*[131] demonstrated the power of NGS and FACS screening by analyzing thousands of plasmids with different promoter-RBS combinations in *E. coli* and created a predictable library. Van Opijnen *et al*.[206, 207] used a Tn-seq approach to identify pathways and regulatory relationships in *Streptococcus pneumoniae* and *Vibrio cholerae* by analyzing the genome-wide fitness of a transposon bearing library. At least lethal gene insertions can be identified because they will be missed in the resulting data set.[206] Van Opijnen *et al*.[206] used a Himar1 Mariner transposon, which has a large advantage compared to transposon Tn5, especially for library construction for NGS approaches. Comparing their results leads to large networks of genes which are playing key roles in regulation and pathways. In principle, the sensor construct and associated library described in this thesis are suitable for a high-throughput combination of FACS sorting and next-generation Tn-seq analysis. FACS sorted cells, with a distinct range of msfGFP fluorescence could be collected in different bins (Figure 14) and sequenced to identify integration sites. This would result in a high-resolution map of genomic "hot" and "cold" spots, which could be compared to RNA-Seq data to investigate parallels between the expression of native genes and context-dependent effects. Ideally, genomic integration sites could be chosen based on RPKM values with low coefficient of variation, and high fluorescence of the sensor, thereby enabling efficient expression under varying culture conditions.

Generated Tn5 library was designated to identify suitable integration sites by measuring msfGFP signals and identification of the integration site in the genome of *P. putida* KT2440. But Tn5_sensor bearing strains can also be used directly for integrating heterologous genes using pEMG vectors targeting the sensor.[26] Tn5 encoded kanamycin resistance gene is flanked by FRT sites. It can be removed out by the FLP flippase.[208] Kanamycin sensitive strains can be transformed with a novel pEMG vector including suitable target sequences for the sensor. As a result, on

molecular basis the sensor construct is replaced by the new sequence. This approach is also applicable when the exact integration of transposon Tn5 is unknown, but the Tn5_sensor-bearing clone is interesting. The pEMG vector allows using the Tn5_sensor sequence as a landing pad. We generated a test construct for heterologous expression and successfully integrated it in Tn5 of chosen strains.

Due to missing data from FACS analysis and following NGS sequencing we had to consider another strategy to accomplish our goal. We decided to use RNA-Seq data sets from different experiments to analyze them for suitable integration sites.

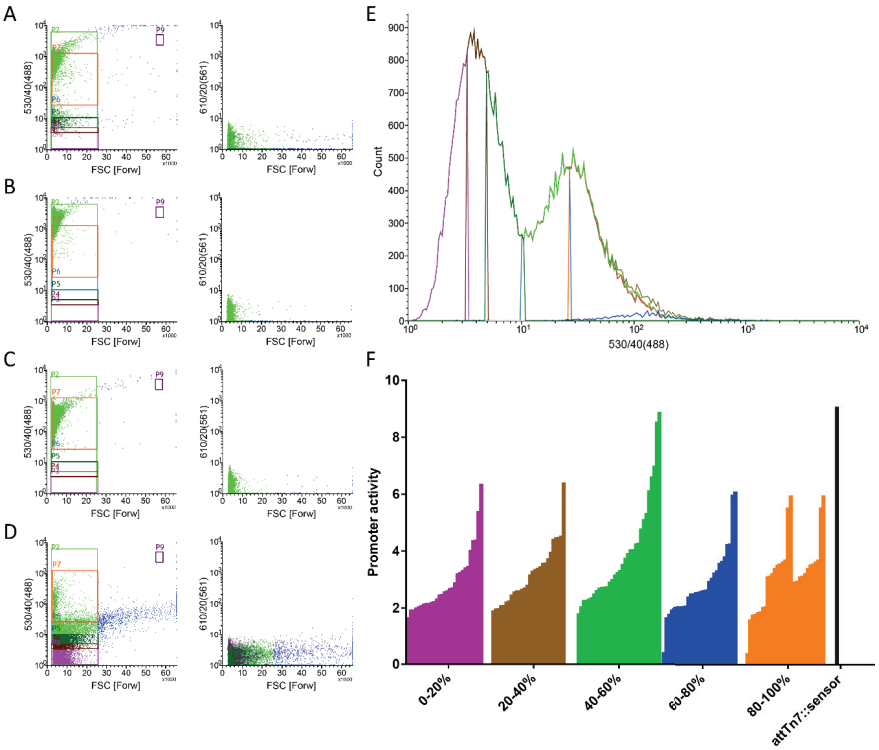


Figure 16: Plots for msfGFP (488/530 nm) and mCherry (561/610 nm) fluorescence over forward scattered of FACS measured *P. putida* KT2440 Tn5_sensor strains. Control strains (A) *P. putida* KT2440 BG13, (B) *P. putida* KT2440 BG14g, and (C) *P. putida* KT2440 *attTn7*::sensor were batch cultivated in shake flasks in minimal medium with 20 mM glucose.[138] Steady-state cultivation of (D) *P. putida* KT2440 Tn5_sensor in mini bioreactors was done with minimal medium[138] and 10 mM citric acid as a carbon source in chemostat cultivation with 7.2 volume changes with a dilution rate of 0.1. msfGFP was measured with an excitation of 488 nm and emission wavelength was set to 520 nm. mCherry was measured with an excitation wavelength of 561 nm and an emission of 610 nm. Cells were filtered to remove large particles and diluted in PBS.

3.2.3.3 Mining RNA-Seq data for equally expressed genes across experiments in *P. putida* KT2440

A second strategy for the identification of suitable integration sites across the genome was necessary. RNA-Seq data represent the transcriptomic state under defined cultivation conditions for an organism of interest.[149, 150, 203, 209] This data reflects differences in promoter activities of native genes. We hypothesized that it would be possible to use publicly available RNA-Seq data sets to identify suitable integration sites.

Publicly available RNA-Seq data from Frank *et al*.[203], Kim *et al*.[150], and Rosa *et al*.[149] (Table 28) were merged in a single Microsoft Excel file and missing information of genes, like gene names and orientation on the chromosome, were added. Values of missing genes for single data sets were deleted to avoid negative effects on calculated means and standard deviations. Also, 5S, 23S, and 16S rRNA genes were deleted from the data set. As described above, these sequences are highly conserved and a clear assignment of the transcripts to one of the seven rRNA operons is difficult. At least 5420 remaining open reading frames were analyzed for expression levels.

Table 28: RNA-Seq data sets used in this study. For each GEO accession number, strain with possible modifications, cultivation conditions, sequencing instruments, and author of the corresponding publication is given.

GEO accession number	Strain *P. putida* KT2440	Condition	System	Author
GSM594825	wild type	30°C, 15 mM succinate	Illumina Genome Analyzer II	Frank *et al*.[203]
GSM594826	wild type	10°C, 15 mM succinate	Illumina Genome Analyzer II	Frank *et al*.[203]
GSM1131198	wild type	30°C, 10 mM glucose	Illumina HiSeq 2000	Kim *et al*.[150]
GSM1131199	wild type	30°C, 10 mM fructose	Illumina HiSeq 2000	Kim *et al*.[150]
GSM1131200	wild type	30°C, 10 mM glycerol	Illumina HiSeq 2000	Kim *et al*.[150]
GSM1131201	wild type	30°C, 15 mM succinate	Illumina HiSeq 2000	Kim *et al*.[150]
GSM1561825	wild type	30°C, 20 mM glucose and 20 mM succinate	Illumina HiSeq 2000	Rosa *et al*.[149]
GSM1561826	wild type	30°C, 20 mM glucose and 20 mM succinate	Illumina HiSeq 2000	Rosa *et al*.[149]
GSM1561827	*crc*::tet allele[a]	30°C, 20 mM glucose and 20 mM succinate	Illumina HiSeq 2000	Rosa *et al*.[149]
GSM1561828	*crc*::tet allele[a]	30°C, 20 mM glucose and 20 mM succinate	Illumina HiSeq 2000	Rosa *et al*.[149]
GSM1561829	*crcZ*::tet and *crcY*::acc3 alleles[b]	30°C, 20 mM glucose and 20 mM succinate	Illumina HiSeq 2000	Rosa *et al*.[149]
GSM1561830	*crcZ*::tet and *crcY*::acc3 alleles[b]	30°C, 20 mM glucose and 20 mM succinate	Illumina HiSeq 2000	Rosa *et al*.[149]

a) inactivated *crc* allele[210]
b) deletion of the *crcZ* and *crcY* genes[211]

In a first step, we identified regions on the genome, which are not differentially expressed across experiments while cultivated with different carbon sources (Figure 17). The integration of a heterologous construct into a specific gene can lead to unpredictable impacts on the cell machinery and properties, resulting in reduced substrate spectrum or decreased growth rates. Also, the genomic context can significantly affect heterologous gene expression, as described in chapter 3.1. Therefore, instead of analyzing single genes, we used the mean of RPKM values of different

experiments to calculate the coefficient of variation (CV) for five genes located next to each other, focusing on the gene in the middle of the tested set (Figure 17). In this case, the CV reflects the variation of the average expression of neighboring genes, and thus a reliable genomic context for integration. This five gene-window CV was calculated for each gene in the data set (Figure 17 C), and standard deviations of individual genes across different experiments were subsequently plotted for selected clusters to select those with low variability in both dimensions: across experiments and genomic regions.

Starting with a very low cutoff CV for the selection of genes, only a few candidates met the specifications. Including higher CVs step by step, a longer list of suitable nearby genes develops and clustered genes can be identified (Figure 17 D).

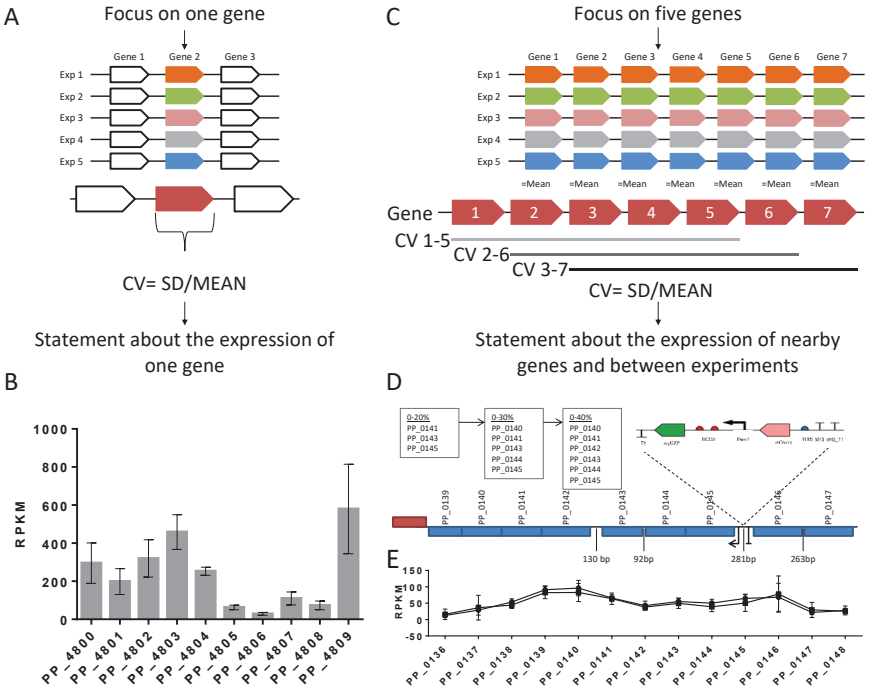


Figure 17: Workflow for the identification of suitable integration sites in the genome and the results for different approaches. (A) Focus on a single gene for equal expression across different conditions. CV is calculated for each gene across all used data sets. (B) Exemplary resulting cluster by using CV for one single gene. Here PP_4804 complies with the specifications, but nearby genes are differentially expressed. (C) Identification of genomic regions to identify equally expressed neighboring genes. For each gene, the mean RPKM from different RNA-Seq data sets is calculated and this value is further used to determine the CV of five consecutive genes. Different RNA-Seq data sets are pictured as clusters in different colors, whereas squares with top displaying genes. (D) Schematic workflow of the identification and investigation of integration sites using the example for PP_0145. Low CV reflecting small differences in expression, while relaxing specification by allowing higher CV (% variation of CV) for searching genomic regions is increasing the number of fitting genes; nearby genes can be seen and clustered to larger strings. Genome analysis reveals possible native genomic elements and the orientation of genes, which are used as a guide for manual selection of integration sites for heterologous constructs. (E) Expression of genes around PP_0145 given with mean RPKM and corresponding standard deviation across different RNA-Seq data sets. Data analysis was done in Microsoft Excel and data sets obtained from gene expression omnibus.[212]

Mapping RPKM values for each identified gene reveals clusters of neighboring genes with similar expression levels (Figure 18). Ten regions on the genome were identified, which show similar RPKM values across experiments. These regions were further characterized by using the genomic sequence and annotation from an online tool[154] to reveal gene names, orientation, and operon structures. Native regulatory sequences like promoters and terminators were also analyzed. Suitable integration sites are listed in Table 29. The first region spans genes PP_0136 to PP_0148. This region contains different transporters and membrane proteins. From PP_0340 to PP_0347 includes transferases and kinases primarily for saccharides. The third region is a long cluster of genes, starting from PP_0451 and ending at PP_0481. This region contains the 30S and 50S ribosomal proteins and is highly expressed. As RNA-Seq data often originated from cells at the exponential phase, the protein biosynthesis rate and associated ribosome expression is high.[150, 153] Ribosomal RNAs (rRNAs) can be found in ribosomes and represent around 80 % of total RNA in a cell.[131, 213] The region involved around 17,000 base pairs with high expression levels with low variability between genes, indicating that native promoters driving these genes have similar activity. Further regions are PP_1426 to PP_1434 and PP_1734 to PP_1743, containing sigma factors, elongation factors, and hypothetical proteins. Other identified regions are PP_3299 to PP_3309, PP_3756 to PP_3764, PP_4143 to PP_4153, PP_4823 to PP_4837 and PP_4936 to PP_4950. All these regions were characterized for regulatory elements and interspacing sequences, which are long enough to enable genomic integration of the sensor construct without affecting the expression of surrounding genes. Table 29 lists the selected integration sites, which are inside the regions mentioned above.

In addition to the rational approach based on low variability between genes and experiments, we also chose three sites with highly expressed genes, reflected by high mean RPKM values. Here we have chosen integration sites next to PP_2322, PP_3808, and PP_4709 (Figure 19). The highest mean RPKM values were found for 16s genes, but due to previously mentioned reasons, these sites were omitted.

Table 29: Chosen integration sites based on RNA-Seq data and *in silico* regulatory elements analysis. Given are 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 www.pseudomonas.com.[154]

Integration site	Gene names	Region	Orientation
PP_0145::PP_0146	*ygdQ* transport protein *yjbB* Na^+/Pi-cotransporter	PP_0136-PP_0148	-
PP_0340::PP_0341	*glnE* glutamate-ammonia-ligase adenylyltransferase *waaF* ADP-heptose:LPS heptosyltransferase II	PP_0340-PP_0347	+
PP_0480::PP_0481	*rplQ* 50S ribosomal protein L17 *katA* catalase	PP_0451-PP_0481	+
PP_1430::PP_1431	DegP-like serine endoprotease *lapA* elongation factor 4	PP_1426-PP_1434	+
PP_1738::PP_1739	hypothetical protein membrane protein	PP_1734-PP_1743	+

Integration site	Gene names	Region	Orientation
PP_3304::PP_3305	Bcr/CflA family multidrug resistance transporter TerC family membrane protein	PP_2399-PP_3309	+
PP_3761::PP_3762	two-component system sensor histidine kinase/response regulator two-component system response regulator	PP_3756-PP_3764	+
PP_4145::PP_4146	*mltD* membrane-bound lytic murein transglycosylase D peptide ABC transporter substrate-binding protein	PP_4143-PP_4153	+
PP_4832::PP_4833	*cobK* precorrin-6x reductase SURF1 domain-containing protein	PP_4823-PP_4837	+
PP_4944::PP_4945	*rlmJ* 23S rRNA (adenine(2030)-N(6))-methyltransferase carbamoyltransferase	PP_4936-PP_4950	-
PP_2322[a]	*oprI*, major outer membrane lipoprotein	PP_2320-PP_2322	-
PP_3808[a]	Antibiotic synthesis protein MbtH	PP_3807-PP_3808	-
PP_4709[a]	*rpsO*, 30S ribosomal protein S15	PP4708-PP_4709	-

a) gene chosen by high RPKM

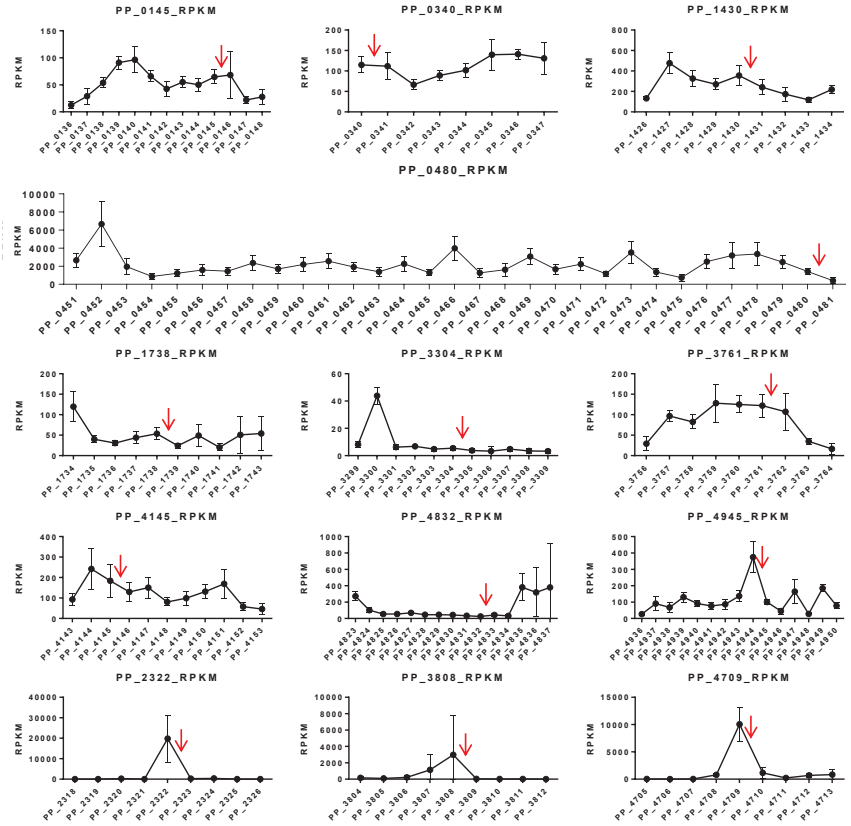


Figure 18: Ten identified regions for targeted genomic integration found by RNA-Seq data analysis and three genes chosen by high mean RPKM values. RPKM values of discovered genes are given with the mean of all experiments and the corresponding standard deviation. Integration site is highlighted with a red arrow.

We have mined RNA-Seq data to identify genomic regions with a low variance in expression levels or high average expression levels. The next task is to confirm our *in silico* findings *in vivo* by using genomically modified *P. putida* KT2440 strains.

3.2.3.4 Characterization of genomic integration sites in *P. putida* KT2440

Identified genomic integration sites described above were characterized *in vivo* using pEMG for targeted integration in *P. putida* KT2440.[26] Therefore, the before mentioned Tn5_sensor was adapted for pEMG-based integration. The part of the sensor starting at the double terminator, down to the T1 terminator, was used.[78] The insulated probe sensor further contained a promoterless *mCherry*, *msfGFP* with bicistronic design element BCD2[26, 167], driven by P_{em7}. This setup already allowed us to characterize randomly integrated transposon *P. putida* KT2440 Tn5_sensor strains for the identification of "hot" and "cold" spots by the exclusion of read-through activity caused by natural promoters as shown in section 3.2.3. In contrast to transposon Tn5 based experiments, here, genomic integration at targeted sites in the genome is investigated. Using the I-*SceI* mediated genomic integration tool pEMG allows precise integration at nearly every position in the genome. A modularly constructed cloning procedure was used to minimize the number of needed oligonucleotides to clone these ten integration vectors, containing target sequences and the sensor. We hypothesized that depending on the integration site in the genome, the activity of P_{em7} is affected by the genomic context, as described above for Tn5_sensor strains. Thus, a range in activity for these sites was expected. As described for *P. putida* KT2440 *attTn7*::sensor, the double terminator in front of the sensor, has a negative effect on the activity. We therefore also used a second, non-insulated sensor sequence without upstream double terminator or mCherry, containing only a promoter, BCD2 elements, *msfGFP*, and a terminator. This structure was already published by Zobel *et al.*[27] and is well established for the characterization of novel synthetic promoters.[192] Structural differences between the sensor and promoter constructs are expected to achieve different activities for all sites. In the case of the non-insulated construct, the upstream promoter read-through can affect the activity of the promoters in addition to other genomic context effects.

Figure 19 shows resulting promoter activities derived from msfGFP fluorescence driven by different promoter sequences after genomic integration of the abovementioned sensor variants in distinct sites across the genome (Table 29). *P. putida* KT2440 *attTn7*::sensor has an integration of the sensor by using transposon Tn7 in *attTn7* site next to *glmS* gene (PP_5409).[27, 115] This allows us to compare the activity of novel developed integrations sites with the commonly used *attTn7* site in *P. putida* KT2440. All strains contain the sensor as single genomic integration and most of the strains don't reach the activity determined for the reference site *attTn7*. Site PP_0340::sensor is the only one achieving a higher activity than *attTn7*. Integration next to PP_4945 reaches

approximately the same activity as *P. putida* KT2440 *attTn7*::sensor. Other integration sites reached less than 50 % of the activity of the control. Promoterless *msfGFP* as a control for genomic read through yielded 79-89% lower activity, indicating that for most constructs the contribution of read-through is relatively minor. *P. putida* KT2440 PP_0340::BG was the exception, being relatively more active than the other sites, possibly due to a faulty terminator analysis of the sequence. PP_1738::BG and PP_4945::BG also reached higher activities than the *attTn7* site, which was again used for comparison. For all tested constructs, integration sites PP_1738, PP_4945, and PP_0340 achieved the highest activities. The next most active site is PP_1430. Besides a wide range in promoter activity, the integration sites show very similar activities during different experiments with different clones. The integration into these sites is very stable and expression is reliable, one of the essential requirements in synthetic biology.

Zobel *et al*. generated and characterized a promoter library for *P. putida* KT2440 using a highly degenerate oligonucleotide. The strongest promoter sequence is P_{14g}, which reaches around twice the of the activity of P_{em7}.[27] As a negative control, a promoterless version (BG) was used. In previously described experiments, these three constructs were used to characterize the activity of different integration sites. Interestingly the behavior of these constructs is highly comparable to the *attTn7* site, not in the resulting activity, but the gradation of them (Table 30).

Integration sites chosen for their high RPKM mean values were tested in the same manner. This different rational approach leads to comparable observations as described before (Figure 19). While tested sensor constructs in site PP_2322 and PP_3808 reached equal activities like *P. putida* KT2440 *attTn7*::sensor, *P. putida* KT2440 PP_4709_sensor achieved a higher activity. Same for strain *P. putida* KT2440 PP_4709_BG, even when no promoter is present in this strain, the activity is comparable with *P. putida* KT2440 BG13. This result is likely caused by an incorrect determined natural terminator sequence. As a result, high read through activity increases the activity. For all other tested constructs at different positions, the gaining activity is comparable to previously shown integration sites (Table 30).

Table 30: Compared activities for different promoters and the sensor construct in percentage to P_{14g}. Promoter activity of P_{14g} was set to 100 % and other ones accordingly refer to it. BG is a promoterless construct, whereas P_{em7} and P_{14g} are synthetic promoters. Given are mean values of several experiments ± the deviation from the mean.

Integration site	**Activity in % compared to P_{14g} for each site**			
	Sensor [%][a]	BG [%]	P_{em7} [%]	P_{14g} [%]
attTn7 (Zobel *et al.*[27])	n.a.	0±5	45±10	100±6
attTn7	14±0.9	1±0.1	32±6	100±9.6
PP0145	24±1.4	6±0.1	23±0.8	100±5.7
PP0340	39±2.7	8±0.4	53±2.9	100±13.2
PP0480	26±3.1	8±0.8	63±3.9	100±6.1
PP1430	27±0.4	3±0.2	46±2.1	100±8.3
PP1738	28±2.2	11±0.6	60±2.4	100±4.0
PP3304	37±2.4	24±1.4	43±4.5	100±5.7
PP3761	33±2.4	7±0.8	36±1.5	100±4.2

Integration site	Activity in % compared to P_{14g} for each site			
	Sensor [%][a]	BG [%]	P_{em7} [%]	P_{14g} [%]
PP4145	40±3.4	21±3.6	72±12.9	100±17.5
PP4832	40±3.9	15±2.6	84±19	100±9.8
PP4945	39±4.6	10±0.5	80±14	100±14.7

a) the sensors contained an insulated P_{em7} with an upstream double terminator.

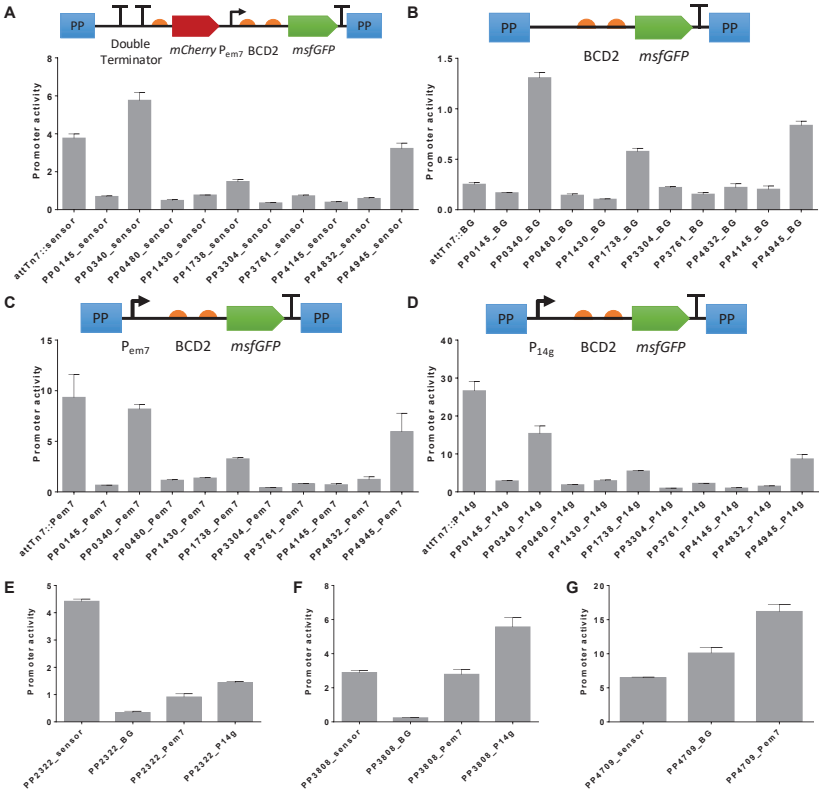


Figure 19: Characterized sites after genomic integration of the sensor and different promoters in *P. putida* KT2440. (A) Summary of genomically integrated sensor constructs in different sites. Promoterless control (B) BG, synthetic promoters (C) P_{em7}, and (D) P_{14g}were also used for characterization of all sites. All constructs were genomically integrated into *P. putida* KT2440 next to the indicated locus. Three integrations sites across the genome were chosen by high RPKM values, (E) PP_2322, (F) PP_3808, and (G) PP_4709. Constructs used for characterization are the sensor, promoterless (BG) control, and two synthetic promoters P_{em7} and P_{14g}. All constructs were genomically integrated into *P. putida* KT2440. Strains were cultured in a BioLector in minimal medium with 20mM 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).

Low variability across different experimental conditions was one of the fundamental selection criteria for suitable integration sites. In order to quantify this variability of the heterologously integrated constructs, their fluorescence was characterized during growth on different carbon sources. Therefore, we repeated the experiments with *P. putida* KT2440 strains containing the sensor in the *attTn7* site or identified integration sites described above. As carbon sources glucose,

citric acid, fructose, or succinate were used in minimal medium (Figure 20).[192] Otto *et al.* have shown that these different carbon sources have a high impact on growth rate.[192] In the case of integration of promoterless constructs into 16S rRNA genes, growth rate and fluorescence were directly linked to each other. Glucose with urea was also used as a stress condition. Most of the tested conditions reached lower activities compared to glucose, although in general, the variability is relatively small. The *attTn7* site, PP_0340, PP_1430, and PP_4945, accomplished the most stable phenotype across all three carbon sources. Activities determined during growth with fructose shows less decreased activity compared to glucose. The highest gain in activity was observed for PP_3761 and PP_0480, with 50 % and 34 %, respectively. Maximal measured fluorescence signals were comparable between glucose and fructose, even though the growth rate on fructose was much lower[192]. A longer lag phase of the cultures on fructose was observed, therefore also the fluorescence signal increased much slower. This influences the calculation of promoter activities, which is dependent on fluorescence signal and optical densities during early exponential phase. Higher growth rates were achieved while cultured with succinate, where the fluorescence signal was very similar to that on glucose. Most of the activities are higher than determined by glucose cultivation. The highest gain of activity was observed for PP_0340 with 64 %, PP_1738 with 68 %, and PP_4945 with 100 %. Growth of the strains with glucose and urea leads to much higher activities of up to 430 %. Otto *et al*. postulated that urea could influence the morphology of the cell.[192] Measuring biomass during cultivation with glucose and urea revealed higher optical densities than cultivation with glucose alone, which can be explained by morphological changes. Furthermore, measured optical densities decreased to expected values after entering stationary phase. Interestingly fluorescence signal is increased for all strains, which shows that protein biosynthesis is highly active. The addition of urea affects proteins stability and activity, which might be compensated for by higher expression.[192, 214] A direct comparison between cultured cells on glucose with and without urea is impossible, due to these high effects on cell growth. That said, promoter activity is determined by dividing fluorescence level by biomass over a certain time, and calculation with a higher biomass would thus reduce promoter activity. Since all activities are higher than on glucose alone this biomass effect does not explain the higher promoter activity with urea. A higher overall expression is more appropriate as an explanation.

Most of the integration sites show similar expression on different carbon sources compared to glucose. Leaving out determined promoter activities achieved while cultivated with glucose and urea, we see a low variability for some integration sites (Figure 20). Here again, reference *attTn7* site with integrated Tn7 constructs show low variability with a CV of 18 %. Weaker expression of sites PP_0145, PP_0480, PP_1430, PP_3304, PP_3761, PP_4145, and PP_4832 leads to CV from

6 % to 33 %. Higher expressed sites PP_0340 and PP_4945 show low variability for glucose, fructose, and citric acid with a CV of 30 % and 2 %, respectively. Including results from succinate, cultivation leads to an increased CV of 34 % to 45 % because twice as much activity was determined here. The same behavior was observed for PP_1738, which lead to an increase in a CV to 40 %. This 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. The influence of the genomic context is comparable to *attTn7* on promoter activities. Here we achieved a much lower variability of expression using different integration sites under the same conditions. However, cultivation under urea stress leads to much higher activity in 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 insertion site selection to counteract this effect.

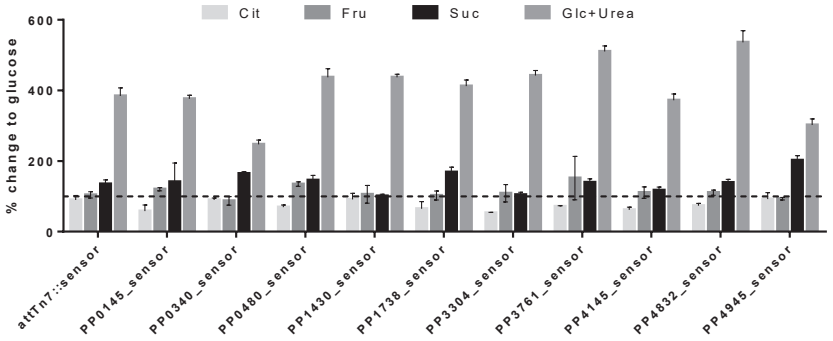


Figure 20: Characterization of genomically integrated sensor constructs at different integration sites in *P. putida* KT2440 while cultivated with different carbon sources. Promoter activity values are shown relative to glucose, which was set to 100% (dotted line). Strains were cultured in a BioLector in minimal medium in a 96 well plate. As carbon source 20 mM glucose, 20 mM fructose, 20 mM citric acid, 30 mM succinate, or 20 mM glucose with 0.8 M urea were added. Identical strains from at least two different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n>6).

3.2.3.5 Multiple integrations do not affect the expression of individual constructs

Often multiple heterologous genes have to be expressed simultaneously, or multiple copies of the same genes are needed to achieve very high expression levels. Here, the use of plasmid is advantageous, but this entails the drawbacks discussed above, such as plasmid instability and negative effects on growth.[99, 102] Having a set of genome integration vectors at hand, we integrated two of them into one strain to investigate possible interactive effects on promoter activity. For this purpose, we used previously described pEMG vectors for certain integration sites across the genome with different promoter constructs. Furthermore, the *attTn7* site was used for the integration of transposon Tn7 containing *mCherry* as a second reporter gene. Constructs were chosen with different activities to see differences more clearly. Single integration strains and

generated double integration strains were cultivated simultaneously and resulting activities determined (Figure 21). Single activities were summated to get a theoretical value for the expression of two constructs in different sites. A comparison of these theoretical activities and determined ones for double integration strains shows that the aberration is very low. Likely, the relatively large genomic distance between constructs prevents interference effects. A distance of expression cassettes reduces the competition for translational and transcriptional resources, as demonstrated by Qian *et al.* [71] We observed the same property for the integrated sensor in PP_0340 and *mCherry* driven by P_{em7} at *attTn7* site (Figure 21). Resulting activities of msfGFP and mCherry are comparable between strains with single and double integration.

We observed no synergistic effect between two integrated constructs. Furthermore, we also didn´t recognize a negative influence. This shows that double integrations in *P. putida* KT2440 are feasible without detectable effects on heterologous gene expression or growth rate of generated strains.

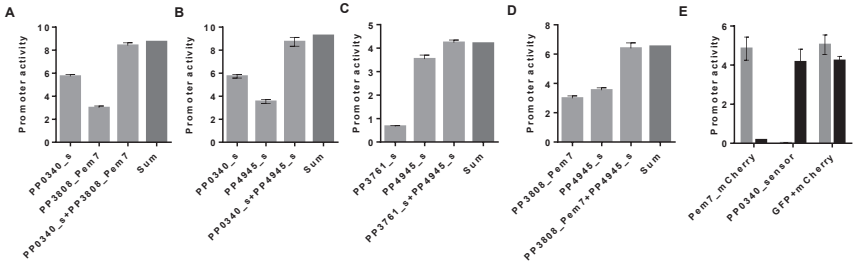


Figure 21: Characterization of genomic double integrations in different sites of the genome in *P. putida* KT2440. Simultaneously integration of different pEMG based vectors containing the sensor construct or *msfGFP* driven by P_{em7}. Furthermore, an integration in *attTn7* site was done with pBG13_mCherry for measuring different reporter genes. (A) Shows resulting promoter activities for single integrated PP_0340::sensor and PP_3808::BG13 constructs. The resulting activity of double integration for PP_0340::sensor+PP_3808::BG13 and the theoretical sum of the individual activities (dark grey bar). Results for integration of (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. Determined promoter activities for figures A to D are based on msfGFP fluorescence. The theoretical sum of the activities according to determined activities for double integration strains. (E) Shows integration of PP_0340::sensor and attTn7::BG13_mCherry, here next to msfGFP fluorescence (grey bars) also mCherry (dark bars) was measured and activity determined for each reporter gene. Strains were cultured in a BioLector in minimal medium with 20 mM glucose in a 96 well plate. Identical strains from at least two different transformations were tested, with three biological replicates each. Vertical dotted lines are separating individual sets. Error bars indicate the standard error of the mean (n>6)

3.2.4 Conclusion

RNA-Seq data was successfully used to identify several genomic integration sites. Table 31 summarizes the most important data for four selected sites, which enable high expression with relatively low variability. The expression of heterologous gene constructs in these sites doesn't rely on upstream promoters and is relatively constant across different cultivation conditions. The 2-dimensional assessment of genomic regions spanning five adjacent genes, as well as different RNA-Seq data sets, led to more reliable expression patterns. Integration of the same insulated

construct with its own constitutive promoter at different genomic sites caused drastic differences in reporter gene expression. This phenomenon was observed both with random Tn5-based integration and with targeted pEMG-based integration, proving the presence of genomic "hot" and "cold" spots, which affect expression based on the genomic locus, rather than promoter activity. Using uninsulated constructs with other promoters enabled a wide range in activity, enabling the tuning of gene expression. Characterization of double integration strains reveals no mutual influences or synergistic effects. Thus, this work provides a toolbox for the engineering of a multitude of traits by enabling multiple genomic integrations in a reliable and tunable fashion.

Table 31: Most promising integration sites found in this study. For each site name of the corresponding vector, RPKM values, promoter activities determined while cultivated with different carbon sources, and resulting CV are given.

Vector	RPKM[a]	Promoter activity					CV without urea[b]	iAMB strain collection
		glc[d]	fru[d]	cit[d]	suc[d]	glc + urea[d]		
pEMG_PP0145_sensor	54.6 69.6	1.2±0.0	1.4±0.1	0.7±0.2	1.7±0.6	4.5±0.1	0.33	6541
pEMG_PP0340_sensor	99.3 106.7	7.2±0.4	6.7±0.9	6.9±0.4	12.7±0.5	19.1±0.9	0.34	6545
pEMG_PP1738_sensor	51.7 21.5	2.1±0.1	2.2±0.3	1.4±0.4	3.6±0.3	8.7±0.4	0.40	6557
pEMG_PP4945_sensor	363.5 96.3	4.0±0.1	3.9±0.2	3.8±.08	8.4±0.6	12.6±0.7	0.45	6577
pBG_sensor (*attTn7* site)[c]	22.1 106.4	4.4±0.1	4.5±0.4	3.9±0.5	5.9±0.5	16.8±1.0	0.18	6539

a) RPKM values of the directly neighboring genes
b) Coefficient of variation calculated without cultivation with glucose and urea
c) *attTn7* site is located between PP_5408 and PP_5409
d) Carbon sources used glucose, fructose, citric acid, succinate, and glucose urea

Chapter 3.3
Development of molecular biology tools for *Pseudomonas*

Contributions

Sebastian Köbbing generated all plasmids and genomically modified strains and performed the experiments with the help of Justinas Valiulis and Raika Schäfer. Sebastian Köbbing and Nick Wierckx analyzed the results. Nick Wierckx and Lars M. Blank conceived the project.

3.3 Development of molecular biology tools for *Pseudomonas*

3.3.1 Abstract

The development of synthetic biology tools is often focused on gene deletion, gene expression, genomic integration systems, and multiplex manipulation technologies like MAGE and CRISPR. In this chapter, the existing genetic toolbox for *Pseudomonas* is extended with antibiotic marker recycling options for newly generated as well as pre-existing strains with genomic antibiotic markers. New Tn7 suicide vectors were constructed, which allow marker recycling by FLP-FRT recombination. Exchange and deletion of the marker have a low influence on the activity of downstream promoters. However, this effect does vary per promoter, and it can be dramatic in some cases, such as with the P_{em7} promoter. For pre-existing Tn7-bearing strains, a pEMG vector was developed to introduce a nonsense mutation in the gentamycin marker. Besides these genomic tools, plasmids with varying copy numbers were constructed using a *P. putida* KT2440 strain with genomic integration of the *rep* gene under the control of different synthetic promoters. The replication gene was removed from these plasmids, thus shifting replication and copy number control to the genome. In principle, the re-localization of the replication gene is feasible but resulted in varying fluorescence signals caused by clone to clone variation. These tools are delivered here to the community to generated markerless strains that lower translational and transcriptional burdens due to deletion or disturbance of genes.

3.3.2 Introduction

The Pseudomonads are interesting microbial candidates for biotechnological processes.[1-3] They can survive harmful conditions like oxidative stress and enables researchers and companies the production of chemicals, like furandicarboxylic acid, rhamnolipids, and aromatics, from renewable feedstocks.[1, 12-16] A wide range of utilizable substrate are known and still can be increased by metabolic engineering.[4-7, 9-11]

All these studies mentioned above have one key enabling factor in common, they used molecular genetics tools. Such tools, in general, include gene expression systems, genomic deletion and integration applications, and genome engineering.[3, 23, 27, 29, 115] They often rely on antibiotic selection, causing high costs, especially in large fermentation processes. This has led to a renunciation of plasmid-based expression systems.[187] Plasmid instability, varying plasmid copy numbers and rising fitness costs for the host makes the use of plasmids disadvantageous.[99, 101, 102] Transposon Tn5 can be used for genomic integration, but it integrates randomly, with all the screening and reproducibility issues that entail .[112, 114] Tn5 vectors generated by Nikel *et al*.[114], called pBELK and pBELG, include a kanamycin gene for selection which is flanked by flippase

recombination targets (FRT). Flp recombinase can recognize these sequences and recombine them.[215] With the right orientation of the FRT sites, the gene in between is lost through recombination, leaving a scar consisting of one FRT sequence. In contrast to Tn5, transposon Tn7 integrates into a targeted locus downstream of the *glmS* gene in the *attTn7* site across different bacterial species, including *Pseudomonas putida* KT2440.[115] This allows repeated transformations of a construct with reliable results, as also shown in chapter 3.1.[27] Unfortunately, this vector currently contains no system for marker recycling.

Commonly used promoters for heterologous gene expression are based on sigma-70 (σ^{70}) factors.[126] Sigma-70 is encoded by *rpoD* and the most prominent sigma factor in bacterial cells because it is the main factor for the expression of housekeeping genes.[125, 169] Sigma factors play a major role in each growth stage of bacteria. Still, they also function in responses to environmental changes and stress. While sigma factor *rpoS* (σ^{38}) is mainly active during stationary growth, it also is representing the central regulator of the stress response, like carbon starvation or acidification.[216-218] Both events occur by transitioning from exponential growth to the stationary phase. Other stressors, such as temperature shifts, activate *rpoE* (σ^{24}) and *rpoH* (σ^{32}).[219] As a result, heat shock protein expression is raised and the stability of these sigma factors is increased.[125] Next to carbon starvation, also nitrogen limitation has a huge effect on microbial behavior. Here *rpoN* (σ^{54}) is the dominant factor that initiates gene expression of a wide range of processes, including flagellar synthesis and virulence.[220, 221] The generation of synthetic promoters dependent on different sigma factors besides sigma-70 would enlarge the area of applications to each stage of microbial growth. The use of alternative sigma factors could be applied, for instance, in the biosynthesis of toxic products or proteins. Yu *et al*.[222] compared stationary and exponential expression by microarray analysis and identified several candidates which were further characterized to obtain late stage protein production in *B. subtilis*. Shimada *et al*.[223] analyzed upstream sequences from stationary phase related genes in *E. coli* and have shown that strong promoters can be identified which are RpoS and RpoD dependent. Comparison of the transcriptome of exponential and stationary phase reveals differences on the overall genetic activity and allows deep insight. D´Arrigo *et al*.[209] determined expression levels for both stages while cultivating *P. putida* KT2440 with different carbon sources. Here the opportunity of identifying “late expression” genes occurs by using these publicly available RNA-Seq data sets.

While previous chapters act on synthetic promoter characterization and genomic integration sites, this chapter focusses on additional tools like stationary phase promoter development and marker recycling. Here we show how to identify promoters depending on other sigma factors than sigma-70 and a resulting shift in initial activity compared to sigma-70 dependent promoters. Due to a missing marker recycling possibility, we developed a pEMG vector for targeted gentamycin

inactivation of previously generated Tn7 bearing strains. Furthermore, a set of new mini Tn7 vectors with the option to use FRT-FLP recombination for marker recycling is described.

3.3.3 Results and discussion

3.3.3.1 Marker recycling by homologous recombination based on pEMG vector system and flippase activity

Recycling of marker genes has two advantages; (I) genetic modification with the same antibiotic resistance marker is possible, and (II) application of marker-free engineered strains faces less regulatory hurdles. Here we show two ways of marker recycling using different approaches. The mini-Tn7 vector system developed by Zobel *et al.*[27] contains a gentamycin marker for the selection of successfully integrated transposon Tn7 (Figure 22 A). For selection and plasmid stability in *E. coli,* the backbone contains a kanamycin marker. We tried to equip gentamycin with FRT sites, which can be recognized by flippase and leads to a deletion of the gene by homologous recombination. Although recombination with this construct was successful, it proved unreliable and difficult to reproduce. To obtain a more useful system, we reassembled the mini-Tn7 plasmid by exchanging the gentamycin gene, inside transposon Tn7, against kanamycin from the backbone (Figure 22 B).

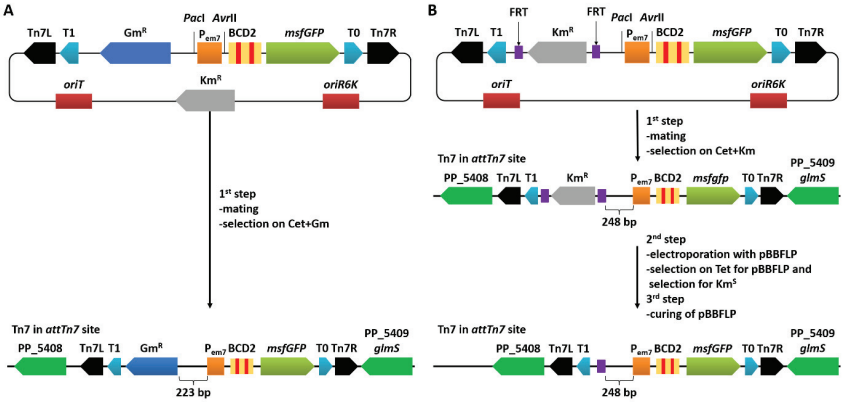


Figure 22: Structural organization of mini-Tn7 vectors from Zobel *et al.*[27] and FRT version developed in this study. (A) Original mini-Tn7 vector bearing a Km^R marker, an origin of replication *oriR6K*, and origin of transfer *oriT* on the backbone. Tn7 module contains a Gm^R marker for selection and two terminators (T0 and T1) for the insulation of the probe. Tn7R and Tn7L are recognized by a transposase. Promoters to be tested were integrated between *Pac*I and *Avr*II restriction sites. BCD2 element for translational coupling and *msfGFP* as a reporter gene. After genomic integration, the gentamycin marker is also integrated and remains in the genome. (B) Tn7 module contains a Km^R flanked with two FRT sites inside the Tn7 for selection and two terminators (T0 and T1) for the insulation of the probe. The backbone only contains the origin of replication *oriR6K* and origin of transfer *oriT*. Underneath the genomic organization of the inserted mini-Tn7 with Km^R before and after flippase activity and resulting recombination of the two FRT sites, which leads to the deletion of Km^R. In a first step, the mini Tn7 vector is integrated into the genome, including the Km^R gene. The second step consists of transformation with flippase encoding plasmid pBBFLP and selection for Tet^R clones. Identification for successfully recombined FRT sites was done on by selecting kanamycin sensitive clones on agar plates.

To test the effect of exchanging the resistance gene inside the Tn7 module on promoter activity, we generated a new mini-Tn7 library with promoters developed by Zobel *et al*.[27] (P_{14b} to P_{14g}) and P_{em7} from Martínez-García *et al*.[33] Here these promoters were tested with and without the kanamycin gene and were compared to the original construct containing the gentamycin marker (Figure 23).

Changing the antibiotic resistance gene shows only low effects on promoters developed by Zobel *et al*.[27] However, the activity of P_{em7} is decreased by 70 % compared to the original construct, independent of the presence of the kanamycin gene. The genomic structure between kanamycin- or gentamycin- marker bearing Tn7 modules is nearly the same. Distance between antibiotic resistance gene and promoter start site is 223 bp for the original gentamycin version and 248 bp for the new one. Apparently, even this far upstream of the promoter, the antibiotic resistance gene and/or the FRT sites affect P_{em7} activity. Except for this unexpected effect with P_{em7}, the generated marker-recycling vector enables reliable marker-free gene expression from Tn7-integrated genomic constructs.

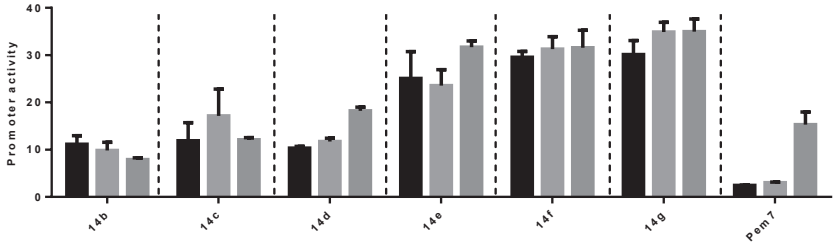


Figure 23: Characterization of genomically integrated mini-Tn7 vectors with different promoters and resistance genes in *P. putida* KT2440. Promoters P_{14b} to P_{14g} and P_{em7} were tested and activity determined. The original construct containing a gentamycin marker (grey bars) is compared with a newly developed mini-Tn7 construct containing kanamycin inside the Tn7 module (light grey bars) and deleted kanamycin gene after flippase activity (black bars). Vertical dotted lines are separating individual sets. All constructs were genomically integrated in *P. putida* KT2440. Strains were cultured in a BioLector in minimal medium with 20 mM glucose in a 96 well plate. Identical strains from at least three different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n=9).

Re-engineering of already constructed Tn7-bearing strains by using the above-mentioned marker recycling system would be highly laborious. Especially for deeply engineered strains, this would require the re-engineering of several genomic modifications. Therefore, we constructed a pEMG vector that enables marker recycling in already generated strains.[26] The first target sequence (TS) element contains a part from the gentamycin marker gene followed by an in-frame nonsense mutation, which causes a stop codon and thus translational abortion (Figure 24). The second TS element is homologous to PP_5408 and downstream sequences, which is located upstream of the *attTn7* site and next to *glmS* (PP_5409). The integration at the *attTn7* sites is guided by the recognition of a consensus sequence located at the end of the *glmS* gene by TnsD, one of four subunits from the transposase.[224, 225] The *glmS* encodes a L-glucosamine–fructose-6-phosphate

aminotransferase and is an essential gene in many prokaryotes and eukaryotes.[226] Therefore, the encoding sequence is highly conserved, including the consensus sequence needed for targeted integration.[227] TnsD acts as a target selector and recruits the rest of the transposase machinery to the *attTn7* site. Subunits TnsA and TnsB recognizes the ends of the transposon (Tn7R and Tn7L), which is located on the donor plasmid and excise the transposon Tn7 by introducing double strands breaks.[228] Also, the following insertion into the genome is carried out by TnsA and TnsB. Communication between transposon and target site indicated by TnsD is performed by TnsC.[229] Integration takes place 25 bp downstream of the TnsD binding site inside *glmS* and ensures that the open reading frame is unaffected.[227] Gringauz *et al*.[230] have shown that changing the sequence upstream of the TnsD binding site does not affect insertion frequencies of transposon Tn7 in the *attTn7* site in *E. coli*. This means that proper Tn7 integration is dependent on a TnsD binding site and not from downstream sequences.[226, 227]

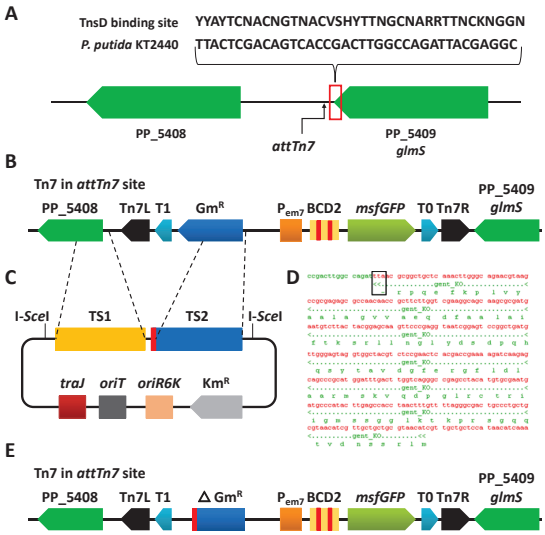


Figure 24: Consensus sequence of TnsD binding site and structural organization of genomically integrated transposon Tn7 targeted for pEMG based marker recycling. (A) TnsD-recognized consensus sequences and location in *P. putida* KT2440. The consensus sequence identified from Mitra *et al*.[227] transformed from a sequence motif applying IUPAC nomenclature for nucleotides to a single sequence. Alignment against *P. putida* KT2440 leads to the given sequence. The recognition site is located at the end of PP_5409 (*glmS*) and covers 37 nucleotides. (B) Genomically integrated Tn7 module contains a Gm^R marker for selection and two terminators (T0 and T1) for the insulation of the probe. Tn7R and Tn7L are recognized by a transposase. BCD2 element for translational coupling and *msfGFP* as a reporter gene. (C) Constructed pEMG vector for gentamycin marker gene recycling by the integration of a stop codon. The vector consists of two TS elements, flanked by I-*Sce*I sites inside a *lacZα* segment. Between TS elements, 80 bp of the surrounding sequence of *attTn7* and one additional nucleotide for in-frame stop codon integration in the gentamycin marker gene (red bar). Backbone contains a Km^R marker for selection, the origin of replication *oriR6K*, the origin of transfer *oriT*, and *traJ* for conjugative mobilization. (D) Integration of an in-frame stop codon in the gentamycin resistance gene. A newly generated stop codon abolishes translation and leads to sensitivity towards gentamycin. (E) Structural organization after co-integrated release and successful partial deletion of Tn7 module sequences. The remaining module consists of one terminator T0 and one Tn7 flank Tn7R. Promoter, BCD2 element, and *msfGFP* sequences are unattended. Part of gentamycin is deleted and instead a stop codon (red bar) integrated, which is named ΔGm^R.

We constructed three pEMG based vectors to knock out the gentamycin marker gene inside the transposon Tn7. Between TS elements, we cloned spacer sequences with 60, 80, and 100 bp, which include surrounding sequences from the *attTn7* site in both directions. These plasmids were named pEMG_GentKO_60, due to 60 bp inserted for new *attTn7* site, pEMG_GentKO_80 and pEMG_GentKO_100. We mated all plasmids into *P. putida* KT2440 BG42 and resulting clones contained the full-length integration inside the gentamycin resistance marker. The main bottleneck in this procedure was the release of the co-integrate. Usually, plasmid pSW-2 is used in this step to express the I-*Sce*I homing endonuclease. However, this plasmid also contains a gentamycin marker gene, making it unusable in combination with the genomic gentamycin marker. Also, unlike described[26] in the seminal paper, this plasmid does not require induction by the addition of 3-methyl benzoate.[16] Instead, we were forced to use pSW-1, an ampicillin resistance gene containing variant of pSW-2. Due to the high ampicillin resistance of *P. putida* KT2440, 500 mg L^{-1} ampicillin and 15 mM 3-methyl benzoate were required, but even under these conditions, I-*Sce*I was poorly induced. Resulting clones often contained the whole plasmid integrated and release of them while using pSW-1 was not feasible. Despite these difficulties, clones of *P. putida* KT2440 BG14g_△Gent_80 could be obtained, which were unable to grow on gentamycin and kanamycin containing agar plates. The successful co-integrate release was confirmed by sequencing reactions and revealed the correct in-frame stop codon formation in the gentamycin marker gene. Three individual clones of *P. putida* KT2440 BG14g_△Gent_80 were characterized and resulting activities were comparable to the original strain (Figure 25).

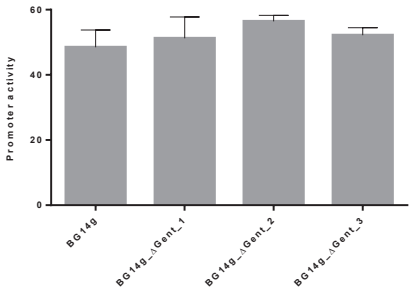


Figure 25: Determined promoter activities of gentamycin sensitive *P. putida* KT2440 BG14g_△Gent after targeted integration of a stop codon for marker recycling. As a control, *P. putida* KT2440 BG14g was used, which reflects the original strain. Strains were cultured in a BioLector in minimal medium with 20 mM glucose without antibiotics in a 96 well plate. Identical strains from at least three different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n>6).

Disruption of the gentamycin resistance gene is feasible but with low frequency, maybe due to low I-*Sce*I activity from pSW-1 after induction by 3-methyl benzoate on agar plates. Wirth *et al.*[231] generated new plasmids containing I-*Sce*I gene expressed from a synthetic promoter P_{em7}.[33] A

constitutive expression of the endonuclease could increase the co-integrate release and possibly yield more gentamycin sensitive clones arise. Alike liquid cultivation of *P. putida* KT2440 BG42 with the corresponding integration and plasmid pSW-1 as described by Martínez-García *et al*.[26] could be performed. The induction of the XylS/Pm expression system by 3MB is in liquid culture more effective. The generation of a second *attTn7* site for further Tn7 integration is unnecessary. Since the integration of transposon Tn7 is dependent on the presence of a TnsD binding site at the end of *glmS* and this sequence is untouched.[230] Therefore, a second use is thinkable, because the binding site is still present. The downstream sequence can vary without loss of integration efficiency in the *attTn7* site.[230] We tried to integrate a second transposon, but we focused on the wrong location. Our focus was on an upstream integration of the second transposon, but a downstream located integration is more likely. This could be tested by integrating a second transposon Tn7 plasmid into above mentioned Tn7 strains with a recycled kanamycin marker or gentamycin knock out strain describe here. Colony PCRs with new primer combinations would uncover what happened between Tn7 and *glmS* and different fluorescence genes can indicate whether simultaneous expression occurs from two Tn7 constructs.

Here we have developed a new mini-Tn7 vector system with marker recycling using FLP-FRT recombination. The effect of the marker recycling on promoter activity is low, except for P_{em7}, which is negatively affected. This promoter is different from the $P_{14a\text{-}g}$ library, being over 25 bp longer than promoters developed by Zobel *et al*.[27] Perhaps this additional sequence between -10 region and BCD2 start site is affected by nearby genes. But a similar effect was not seen by the characterization of different integrations sites across the genome (Chapter 3.2). Generation of a gentamycin marker knock-out caused by plasmid pEMG_GentKO_80 also shows no influence on the promoter activity of *P. putida* KT2440 BG14g_△Gm. Difficulties were encountered in the release of the co-integrate, which could be addressed by using plasmids with different antibiotic markers, expressing I-*Sce*I constitutively, such as pSEVA1213S.[231]

3.3.3.2 Tic Tac Lac – standard promoters

Promoters P_{lac}, P_{trp}, P_{lacUV5}, P_{tac}, P_{trc}, and P_{tic} are commonly used for gene expression and often included in standard vectors. Their widespread use is based mostly on historical reasons, due to their early availability. As such, they have become a gold standard for heterologous expression. These promoters are most often used in an inducible plasmid-based expression system, with previously discussed disadvantages. However, they are also often applied as constitutive promoters through the omission of the cognate regulator. They were best characterized in *E. coli*, although to our knowledge, no standardized comparison has been performed, neither in *E. coli* nor in *Pseudomonas.* The original P_{lac} is a natural promoter which initiates transcription of the lactose metabolic operon in *E. coli*.[232] Promoter P_{lacUV5} is a mutated version of P_{lac} with higher activity,

with two exchanged nucleotides in the -10 region (Figure 26).[233] The popular P_{tac} promoter was obtained by fusion of P_{lacUV5} with the P_{trp} promoter, with the -35 region originating from P_{trp} and -10 from P_{lacUV5}.[234] This promoter shows even higher activity than the two donor promoters. The P_{tac} derivatives P_{trc} and P_{tic} were obtained by increasing the length of interspacing sequences from 16 bp to 17 bp or 18 bp, respectively.[235]

```
Plac     TTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGA
PlacUV5  TTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGA
Ptrp     TTGACAATTAATCATCGAAC-TAGTTAACTAGTACGC
Ptac     TTGACAATTAATCATC--GGCTCGTATAATGTGTGGA
Ptrc     TTGACAATTAATCATCC-GGCTCGTATAATGTGTGGA
Ptic     TTGACAATTAATCATCCCGGCTCGTATAATGTGTGGA

P14a     TTGACA--------------TGGATATAATGTATGTA
P14b     TTGACATGCGTGATGTT-TAGAATTATAATTTGGGGA
P14c     TTGACATGTCAATTTTT-ATGTTGTATAATATAACTA
P14d     TTGACATCCGACATTCG-CGACTGTATAATAAGTTGA
P14e     TTGACAACACTCGAAAA-GCCGAGTATAATCAGATGA
P14f     TTGACATGACATGGTTT-TGAGGGTATAATGTGGCGA
P14g     TTGACAAGGCTCTCGCG-GCCAGGTATAATTGCACGA
```

Figure 26: Promoter core sequences commonly used for heterologous gene expression. P_{trp} and P_{lac} are naturally occurring promoters' sequences, where P_{lacUV5}, P_{tac}, P_{trc}, and P_{tic} are modifications. Promoter P_{14a} to P_{14g} originates from a promoter library and represents synthetic promoters.[27] Red letters highlight the -35 region and green letters the -10 region.

Most of these P_{lac} derivatives were generated and characterized decades ago, using reporter systems that were state of the art at that time. This mostly involved plasmid-based expression of *lacZ*, with all the disadvantages discussed above. Here, we systematically characterize these promoters in our single-copy genome integrated Tn7-based system, including a comparison with the P_{14} calibrated promoter library (Figure 22).[27, 234, 235]

Tested promoters show a wide range of activities. P_{lacUV5} and P_{trp} only reach low activities compared to P_{tac} (Figure 27), which is in accordance with the original description in *E. coli* .[234, 235] Also, P_{lacUV5} is weaker than P_{lac}, which is differently described in *E. coli*. Brosius *et al*.[235] generated P_{tic} and P_{trc} by extending the spacer and have reported decreased activities in *E. coli*. The same order of activities for P_{trc}, P_{tic}, and P_{tac} is described by Wang *et al*.[236] while testing these promoters in the Cyanobacterium *Synchocystis* sp. Here, in *P. putida*, both promoters have higher activity, with P_{trc} reaching the highest activity. A spacer length of 17 bp is optimal for *Pseudomonas*, at least for these promoters. This is also the length chosen by Zobel *et al*.[27] for the P_{14} promoter library, and indeed P_{14e}, P_{14f}, and P_{14g} achieve even higher activities than P_{trc}.

The basic structures of promoters are unchanged over decades. Developing synthetic promoters is still based on -35 sequence TTGACA and -10 TATAAT. Comparing natural and completely synthetic promoters, the trend of activity in relation to growth is not inherently different. Generating randomized sequences allows us to screen for defined activity levels, but at that time,

this technique was not available. Nevertheless, the promoters are still in use and probably more popular than any developed promoter library nowadays. This popularity makes these P_{lac} derivatives an important standard to put the activity of newly generated synthetic promoters into perspective.

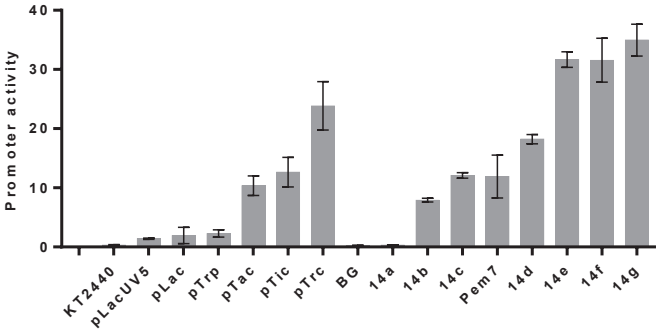


Figure 27: Characterization of standard promoters in *P. putida* KT2440 using transposon Tn7 for genomic integration. P_{lac}, P_{trp}, P_{lacUV5}, P_{tac}, P_{trc}, and P_{tic} were genomically integrated and characterized using *msfGFP* as a reporter gene in *P. putida* KT2440. Promoters BG and P_{14a} to P_{14g} were generated by Zobel *et al.*[27], P_{em7} originates from Martinez-Garcia *et al.*[33] All constructs were genomically integrated in *P. putida* KT2440. Strains were cultured in a BioLector in minimal medium with 20 mM glucose in a 96 well plate. Identical strains from at least three different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n=9).

3.3.3.3 Sigma-38 dependent promoters for expression outside the exponential phase

In essence, all constitutive sigma-70 core promoters have the same expression profile, only differing in their absolute activity, but not in the timing of this activity. Here we want to develop a promoter library with an altered expression profile, showing activity later in the growth curve of *Pseudomonas* without the need of inducible systems. This can be achieved, e.g., by using a sigma-38 dependent consensus sequence, which is unfortunately not available. But hints towards the importance of the -35 and -10 regions were found, alike the importance of upstream sequences that can repress premature expression.[237, 238] We analyzed an RNA-Seq data set from D`Arrigo *et al.*[209], who compared exponential and stationary phase expression of *P. putida* KT2440 while cultivated on glucose or citric acid. Their goal was to identify transcriptional start sites across the *P. putida* KT2440 genome. Still, this data set is also a useful starting point for identifying late-stage promoters, which become active towards the stationary phase of growth.

Data sets were sorted and filtered for exclusively expressed genes during stationary growth while cultivated with citric acid or glucose. The analysis was focused on genes only expressed during the stationary phase. Doing so for glucose 506 genes were identified, for citric acid 387 genes. Increasing the allowed RPKM level in the exponential phase from 0 to 0.5, the number of genes higher expressed in the stationary phase decreased to 151 genes for glucose and 118 genes for citric acid. A RPKM of 1 as a benchmark suggested only 90 genes for glucose and 71 genes for

citric acid with a higher expression level during the stationary phase. Sorting and filtering revealed many genes that are higher expressed during the stationary phase, most of them executively. Inverse, the number of genes exclusively expressed in the exponential phase was also determined and showed a similar trend (Figure 28). Focussing on generally higher expressed genes in stationary phase reveals 1096 gene for glucose and 869 genes while growing on citric acid.

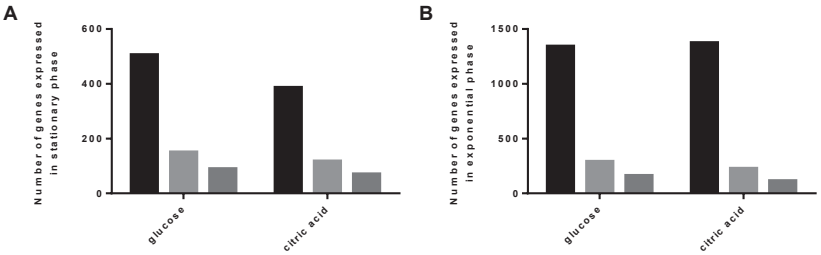


Figure 28: Presentation of exclusively expressed genes in stationary and exponential phase based on analyzing RNA-Seq data from D´Arrigo *et al.*[209] Number of genes exclusively expressed during (A) stationary phase and (B) exponential phase. RNA-Seq data were analyzed with Microsoft Excel. Genes identified by filtering for RPKM values of 0 (black bars), 0.5 (grey bars), and 1 (dark grey bars) for corresponding opposite data set. Stationary phase genes found by applying RPKM limits for exponential phase data set and sorting for genes higher expressed than in exponential phase in stationary phase data set. The evaluation was done with data sets originating from cultivations with glucose or citric acid.[209]

Genes with the highest RPKM from filtered data sets were compared to find overlapping genes in both sets (Table 32). Non-coding elements like rRNAs were omitted from this analysis. At least nine genes were present in both data sets with different functions in the cell. For orientation and to classify RPKM values from them, we add reference genes to Table 32. Sigma factors are responsible for transcription initiation. In the RNA-Seq data, we can observe that *rpoS* (σ38, PP_1623) is higher expressed during the stationary phase, while *rpoD* (σ70, PP_0387) is more prominent during the exponential phase. Both observations were expected due to their described task. *gyrB* (PP_0013) encodes a gyrase which belongs to topoisomerases.

Within the set of genes corresponding to our specifications, PP_1640 showed the highest RPKM change and was chosen for further characterization. This gene encodes *yohC* and is an inner membrane protein and seems to be part of an ABC-type transporter of unknown specificity, based on sequence identity to a similar *E. coli*.[239] The upstream sequence of PP_1640 was analyzed for natural promoters with online tools.[155, 156] However, both tools are designed to find sigma-70 dependent promoters and this bias may lead to the omission of promoters bound by other sigma factors. To further characterize which part of the sequence upstream of PP_1640 is responsible for the promoter activity, oligonucleotides for amplification of different lengths of the upstream sequence were used. Here four different constructs were characterized. pBG_PP_1640_Full contained 348 bp of the upstream sequence. Three other fragments were chosen based on the promoter sequences predicted by online tools inside the upstream sequence. Fragments were

cloned as usual in the transposon Tn7 vector following SEVA guidelines.[33] Constructed plasmids were called pBG_PP_1640_Full, pBG_PP_1640_P2, pBG_PP_1640_P4 and pBG_PP_1640_P6 (Figure 29). After genomic integration in *P. putida* KT2440, all constructs showing fluorescence were selected and promoter activities were determined (Figure 29). As controls, weak promoter strains like *P. putida* KT2440 BG14a and promoterless strain *P. putida* KT2440 BG were used. Promoter activities of tested upstream sequences from PP_1640 lead to very low values, comparable to 14a and promoterless version BG. However, since promoter activity is determined as a linear relation between fluorescence and growth, this calculation strongly favors exponential-phase promoters (Figure 29).

Table 32: Identified genes with high expression levels during the stationary phase. Given are gene names and numbers. For comparison, the change of RPKM values was calculated (stationary phase minus exponential phase) for glucose and citric acid. For orientation reference genes *rpoD* and *gyrB* are given, here negative values computed due to higher expression during the exponential phase.

Gene number	Gene name	RPKM glc[a)] exp[c)]	RPKM glc[a)] stat[c)]	Change[b)] glc[a)]	RPKM cit[a)] exp[c)]	RPKM cit[a)] stat[c)]	Change[b)] cit[a)]
PP_0103	cytochrome c oxidase subunit 2	0.0	32.5	32.5	0.5	38.0	37.5
PP_0536	hypothetical protein	0.5	65.0	64.5	0.5	41.5	41.0
PP_0576	lipoprotein	1.0	14.5	13.5	0.5	13.5	13.0
PP_1640	*yohC*	0.0	150.0	150	1.0	143.5	142.5
PP_2357	type I pili subunit CsuB	0.0	35.0	35.0	1.0	28.0	27.0
PP_2359	putative Type 1 pili subunit CsuA/B protein	0.5	44.5	44.0	0.0	41.5	41.5
PP_2360	type I pili subunit CsuA/B	0.0	30.0	30.0	0.0	29.5	29.5
PP_2942	two-component system response regulator	0.0	21.0	21.0	0.0	25.5	25.5
PP_3135	glycosyl transferase	0.5	19.5	19.0	0.5	16.5	16.0
PP_0387[d)]	*rpoD*	33.0	8.5	-24.5[d)]	23.5	9.5	-14.0[e)]
PP_0013[d)]	*gyrB*	6.0	0.0	-6.0[d)]	5.0	0.0	-5.0[e)]
PP_1623[d)]	*rpoS*	12.5	38.0	25.5	16.5	30.0	13.5

a) Carbon sources used for RNA-Seq glucose and citric acid.
b) Change of exponential and stationary phase based on RPKM. Stationary phase minus exponential phase RPKM value.
c) Mean RPKM values for exponential and stationary phase expression.
d) Reference genes for orientation.
e) Negative values are occurring from a higher expression in exponential phase than in stationary phase.

Analyzing RNA-Seq data from D`Arrigo *et al*.[209] allowed us to identify genes that are exclusively or higher expressed in the stationary phase than during exponential growth. From this set, we chose one candidate and characterized different parts of the upstream sequence and the complete regions for promoter activity. The tested upstream sequences from PP_1640 reached only low promoter activities, comparable to Zobel *et al*.[27] promoterless control BG and P_{14a}. Since promoter activity is determined at the beginning of the exponential phase, this criterion is unsuitable for making a statement about activity from the tested sequences.

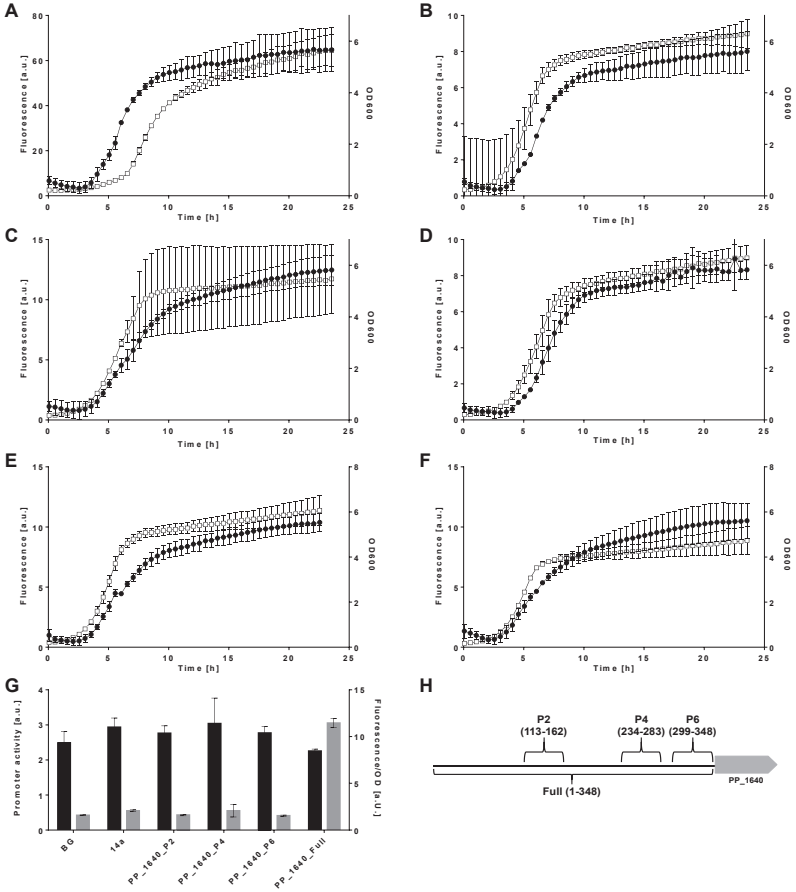


Figure 29: Characterization of different upstream sequences of PP_1640 for promoter activity. (A) PP_1640_Full, (B) PP_1640_P2, (C) PP_1640_P4, and (D) PP_1640_P6 containing upstream sequences or predicted promoters of PP_1640 were tested. As controls *P. putida* KT2440 BG14a (E) and *P. putida* KT2440 BG (F) were used. Black circles represent optical density (OD_{600}) and blank squares msfGFP fluorescence. All constructs were genomically integrated in *P. putida* KT2440. Strains were cultured in a BioLector in minimal medium with 20 mM glucose in a 96 well plate. Identical strains from at least two different transformations were tested, with three biological replicates each. Error bars indicate the standard error of the mean (n>6). (G) Determined promoter activities for different upstream sequences from PP_1640 and synthetic promoter P_{14a}, as well as promoterless control BG (black bars). As well as final fluorescence value over optical density (OD_{600}) for each tested construct (grey bars). The activity was calculated as usual at the beginning of the exponential phase with a corresponding fluorescence signal. (H) Origin and distance to PP_1640 (*yohC*) start site on the genome. P2, P4, and P6 are indicating tested sequences from the upstream region of PP_1640.

A lagged fluorescence formation was seen by comparing fluorescence signals over time for *P. putida* KT2440 BG14a and *P. putida* KT2440 PP_1640_P2, PP_1640_P4 and PP_1640_P6. The delay shows that our tested sequence induces *msfGFP* expression, later than sigma-70 dependent P_{14a}. PP_1640_Full reached the highest activity, which could be lead back to the presence of a regulatory sequence located upstream of the promoter sequence.[238] Franchini *et al*.[239]

have shown that *yohC* is downregulated in *E. coli* Δ*rpoS* strains, which could demonstrate the sigma-38 dependency of PP_1640 in *P. putida* KT2440. The native promoter of PP_1640 is still unknown, but we know that the full-length upstream sequence is gaining higher activities than shorter sections of that. Therefore, we will foster the sequence for possible promoter sequences with a walking approach.

We wanted to identify the minimal late-stage promoter sequence of PP_1640 since shorter parts of the upstream region lead to low activities. An alternative approach was applied, which is comparable to an arbitrary-primed PCR, as described for transposon Tn5.[240] A randomized oligonucleotide, which binds randomly to the DNA, was used to generate amplicons of different length (Figure 30), while using a specific reverse primer. The first PCR creates amplicons with varying lengths of the PP_1640 upstream sequence. In a second PCR reaction, a forward oligonucleotide is used which is complementary to the 5´ overhangs of the previously used one. Direct amplification of different length amplicons was aimed to generate a large library of possible promoter element-containing constructs. This method was also applied to generate a library focusing on the other site of the upstream sequence by using a randomized oligonucleotide for reverse amplification. To test all possibilities, also a combined PCR reaction was performed by using both randomized oligonucleotides in the first PCR. This procedure should generate a library with three different parts from the upstream sequence of PP_1640. Forward reaction focusses on the left part, while the reverse reaction is concentrated in the right part. Using both leads to a mix and centers on the middle of the sequence. These fragments can then be amplified in a 2nd reaction using specific primers that bind to the overhang of the random primers.

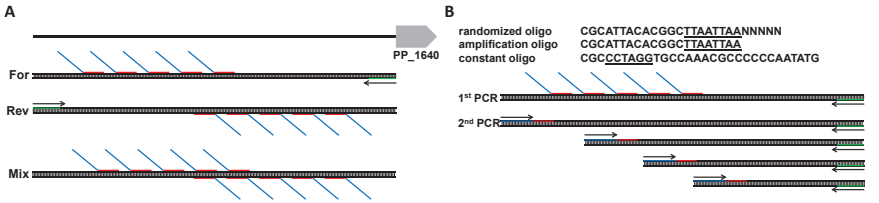


Figure 30: Schematic of the approach to generate amplicons with different length containing parts of upstream sequences of PP_1640. (A) Three PCR reactions were performed while applying a forward (For) or a reverse (Rev) randomized oligonucleotide for the generation of different length amplicons, a combination of both oligonucleotides is called mix (Mix). (B) Library generation was performed in two PCR reactions. The first PCR reaction includes a randomized oligonucleotide that binds randomly on the genomic DNA template to generate different length amplicons. 5´-end has a fixed sequence and is further used for targeted amplification during a second PCR reaction. The opposite oligonucleotide has a constant sequence; therefore, one part of the amplicon is known. Oligonucleotide sequences are given for the forward approach. Underlined nucleotides are highlighting *Pac*I and *Avr*II restrictions sites used for cloning into transposon Tn7.

In general, our genome walking method to isolate the minimal late-stage promoter was successful. Some of the tested and sequenced constructs of varying lengths contain different regions from the upstream sequence of PP_1640. But resulting fluorescence levels were either comparable to PP_1640_P2, PP_1640_P4, and PP_1640_P6 or equal to PP_1640_Full. This indicates that the

core promoter, critical for a higher expression, was not found. But we were successful in shortening the upstream sequence without decreasing the resulting activity. A larger set of varying upstream sequences could enable us to define the region where the core promoter is present.

Further, a longer upstream sequence shows the highest activity, which could be led back on regulatory sequences. The dominance of longer sequences can be explained by using a previously amplified upstream sequence as a template for the genome walking method. Since we were not able to predict the generated amplicon length, we cleaned up the complete PCR reaction, still including the template. Next to the optimization of the PCR procedures, a FACS analysis would improve the outcome since more clones can be analyzed in parallel and sorted depending on the fluorescence signal.

A successfully generated library contains hundreds of plasmid variants with different lengths of the upstream sequence after cloning procedures in *E. coli*. Since we want to test these constructs as single genomic integration in *P. putida* KT2440, we have to integrate them before characterization. Therefore, we used a droplet mating containing many different *E. coli* PIR2 cells bearing library derived plasmids. Arising *P. putida* KT2440 Tn7 integration clones from these matings contain different upstream sequences. From each approach, 12 clones were cultivated in minimal medium with glucose as carbon source, to measure resulting msfGFP fluorescence levels. Sequence clones with the corresponding part of the upstream sequence and fluorescence level are shown in Figure 31.

P. putida KT2440 clones bearing constructs from the forward approach only reached low fluorescence levels, comparable to tested short upstream sequences and promoterless control BG. Integrated Tn7 of chosen clones were sequenced and revealed that 82 to 103 bp of the upstream sequence were included. As planned, all sequences are starting at the same position of the upstream region. Since a low activity of these promoters is seen, it is feasible that the actual promoter was not included in the tested sequences. Clones derived from the reverse approach achieved higher fluorescence levels, comparable to *P. putida* KT2440 PP_1640_Full. The sequencing reaction shows that the integrated upstream sequence mostly contained 316 bp, nearly the same length used for PP_1640_Full. But *P. putida* KT2440 PP_1640_rev1 only contains 104 bp the upstream sequence, which is located further away from the transcriptional start site of PP_1640 but reached a comparable fluorescence level. Using the mix combinations revealed low and high fluorescence levels. Unfortunately, the sequencing of chosen clones shows that *P. putida* KT2440 PP_1640_mix4 and PP_1640_mix5 contained the same upstream sequence with 348 bp. However, fluorescence signals are very different.

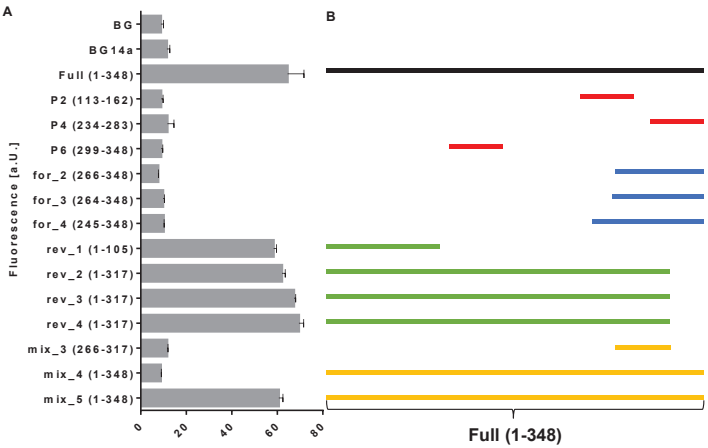


Figure 31: Determined fluorescence levels of sequenced clones from the genome walking method for the identification of the core promoter sequence of PP_1640. (A) Bars with naming for depending from the forward approach, rev corresponds to reverse and mix to the combined approach. Numbers in brackets are indicating the upstream sequence position from the full 348 bp sequence. As controls, promoterless *P. putida* KT2440 BG and weak promoter containing *P. putida* KT2440 BG14a were used. Alike the previously mentioned strain containing a defined length of the upstream sequence *P. putida* KT2440 BG_1640_P2, *P. putida* KT2440 BG_1640_P4, *P. putida* KT2440 BG_1640_P6 and *P. putida* KT2440 BG_1640. All constructs were genomically integrated in *P. putida* KT2440. Clones generated in three in parallel performed matings and each clone is tested as biological triplicate. Strains were cultured in a BioLector in minimal medium with 20 mM glucose in a 96 well plate. Error bars indicate the standard error of the mean (n=3). (B) Mapping of sequenced strain with the full 348 bp upstream sequence. Red lines are indicating the previously mentioned controls. Blue lines are sequences depending on the forward, green from the reverse, and yellow from the combined approach (mix).

3.3.3.4 Genomic control of plasmid copy number

For decades plasmids are used for molecular cloning and protein production in prokaryotes. Here, the expression efficiency is not only dependent on the promoter activity, but also on the number of copies in the cell and the used host. The plasmid copy number (PCN) is mostly fixed and controlled by proteins or RNAs, depending on the mechanisms of replication.[98, 241] The replication mechanisms are highly controlled and, therefore, not easily tunable in natural systems. However, manipulation of the replicon can lead to much higher PCNs.[97, 242] Plasmid copy number replication poses an interesting feedback loop problem. The *rep* gene is located on the plasmid, and therefore its expression increases with increasing PCN. This, in turn, can further increase PCN. On the other hand, a low copy number of a certain plasmid can cause heterogeneity[241] since the distribution of existing plasmids during cell division is unbalanced. Some low copy plasmids include a partitioning system to ensure proper propagation of the plasmid to daughter cells.[243] Jahn *et al.*[241] determined the PCN of seven different plasmids originated from the SEVA collection, in which only the replication system varies. They have shown that determined PCN is lower than reported for the corresponding replication systems. Misguided distribution of plasmids during cell growth can lead to clonal variability and promotes the formation of subpopulations.[99] Comparing plasmid-based and genomically integrated Tn7 reveals a large difference by using the same constructs for

testing (Figure 32). A low degeneracy promoter library was equipped with *lacZ* as a reporter gene and tested by applying the Miller assay.[158] The experiment was repeated three times on different days with new transformations. This led to high clonal variability in the plasmid-based strains. High clone to clone variation was the main driver for us to focus on genomic engineering with transposon Tn7. With this transposon, the variability was low between clones from different transformation approaches. The coefficient of variation was reduced from 32 % for plasmid-based down to 9.6 % after genomic integration. Large differences between plasmid-bearing clones are likely due to varying plasmid copy numbers, including subpopulations that have completely lost the plasmid.[241] However, not all plasmids are causing unreliable results. pRO1600 and ColE1 bearing pJNT and pBT using pBBR as *ori* are robust expression systems.

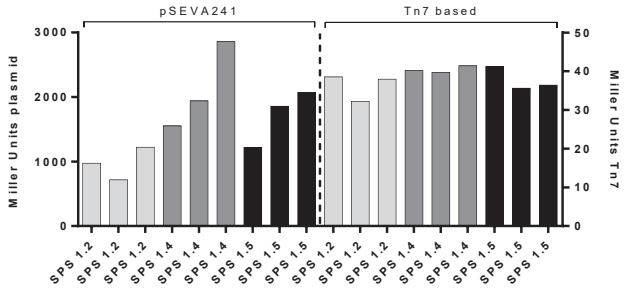


Figure 32: Comparison of pSEVA241 and genomically integrated Tn7 expressing the *lacZ* under the control of different promoters in *P. putida* KT2440. SPS1.2 (light grey bars), SPS1.4 (grey bars), and SPS1.5 (black bars) originating from a low degeneracy promoter library, which is based on P_{Trc}. Miller Units were determined for each single cultivation following experimental procedures. Transformation of plasmids and transposon Tn7 were done by mating at least three different ones on different days. The dotted vertical line separates results from plasmid-based (left y-axis) and Tn7 (right axis) experiments.

External control, stabilization, and reduced variability of PCN was achieved by genomically integrated *recA* and *pir* genes for replication of plasmids into the genome of *E. coli*.[244] Varying RBS sequences affect the expression of *recA* and *pir* and lead to different PCN in the cells with low variability. Such a system allows for tunable optimization of expressed genes, depending on the PCN, while also breaking the feedback loop caused by plasmid-encoded *rep* genes.

Jahn *et al*.[241] reported the use of plasmids derived from the SEVA collection could lead to unreproducible results, because of clone to clone variability caused by varying PCNs. To improve the reliability of these plasmids, we aimed to re-engineer pSEVA241 containing promoter P_{em7}, BCD2 element and *msfGFP* as a reporter gene (pSEVA241_P_{em7}_*msfGFP*) by genomic transplantation of the *rep* gene of the origin of replication pRO1600, which encodes the protein that initiates replication by binding to the origin of replication, similar to Kittleson *et al*.[98, 244] This *rep* gene was integrated into the genome of *P. putida* KT2440 controlled by different synthetic promoters with varying activities. The first question is whether a genomically controlled plasmid

replication and propagation is possible. The second question is how the external expression of the Rep protein, driven by promoters with different strengths, affects the PCN. Instead of determining the PCN itself, we use plasmid-based msfGFP expression as a proxy by measuring fluorescence signals, as we assume that a higher PCN would increase the amount of the reporter gene expression.

Origins of replication pRO1600 and ColE1 are present in pSEVA241, indicated by number 4 in the second position. We subcloned the full *rep* gene of pRO1600 into the Tn7 donor vector pBG containing promoter 14a, 14b, 14d, 14e, or 14g. As in previous experiments, the BCD2 element was used to minimize GOI-based variation. Tn7 transposons, bearing the replication gene, were integrated into the *attTn7* site of *P. putida* KT2440, to enable the strain to replicate the plasmid lacking the *rep* gene. Using PCR and self-ligation via restriction site *Aat*II of the resulting fragment, the *rep* gene was removed from the plasmid (Figure 33).

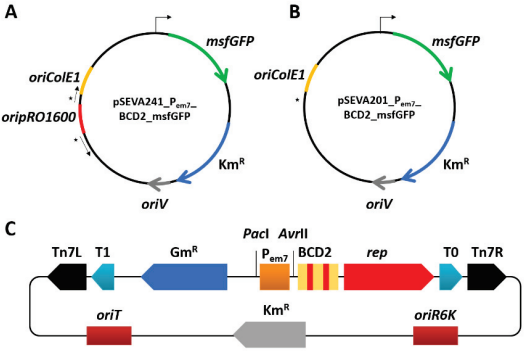


Figure 33: Relocation of pRO1600 replication gene from plasmid to transposon Tn7 vector. (A) Original plasmid pSEVA241_P_{em7}_BCD2_*msfGFP* including replication genes for origins of replication pRO1600 andColE1. (B) Novel pSEVA201_P_{em7}_BCD2_*msfGFP* without the replication gene of pRO1600. Red label highlights the origin of replication pRO1600, yellow origin of replication ColE1, grey the origin of transfer *oriV*, green *msfGFP* as reporter gene with promoter shown as a hooked arrow, and in blue resistance marker kanamycin. Replication gene was excluded by PCR (arrows beside the plasmid indicate oligonucleotides) with following self-ligation via *Aat*II restriction site, indicated by stars. (C) Transposon Tn7 with the replication gene of pRO1600 derived from pSEVA241. The mini-Tn7 vector is bearing a Km^R marker, an origin of replication *oriR6K,* and origin of transfer *oriT* on the backbone. Tn7 module contains a Gm^R marker for selection and two terminators (T0 and T1) for the insulation of the probe. Tn7R and Tn7L are recognized by a transposase. Different promoters to drive expression and a BCD2 element for translational coupling.

Screening for clones bearing a plasmid without the replication gene was done by colony PCR. Here often, two bands occurred. These colonies were re-streaked to obtain pure clones. However, even after repeated clone isolation, some clones yielded two PCR fragments with different lengths. From some of these clones, the plasmid was isolated and sequenced. All these clones share the first part of the alignment. Detailed analysis of the sequencing chromatograms shows two different sequences from the same point onwards (Figure 34). This is apparent from overlapping distinct peaks for each nucleotide starting from the point of *rep* deletion, leading to the assumption that two different plasmids are present in one culture, stabilized by the same resistance marker. This

was not expected due to the general opinion that plasmids using the same mechanism for propagation are not able to co-exist and after some generations, only one plasmid will persist.[245] Alternatively, the cultures consist of mixed clones, although this is unlikely considering the careful clone isolation and repeated re-streaking. Manual analysis of the chromatograms reveals that the *Aat*II restriction site is included in all tested sequences. One restriction site was introduced by PCR and the second is present in pSEVA241 upstream of the *rep* gene. Ultimately, more intensive screening and clone isolation yielded three clones, which only showed one PCR band corresponding to the *rep* gene deletion. Sequencing of these clones revealed extensive unexpected mutations. Instead of a clear deletion of the replication gene, additional sequences were present. These additional sequences are part of the original pSEVA241. The three plasmids were named pSEVA201_P_{em7}_*msfGFP*_g3, pSEVA201_P_{em7}_*msfGFP*_h7 and pSEVA201_P_{em7}_*msfGFP*_i6, to shorten these names further only called pSEVA201_g3, pSEVA201_h7 and pSEVA201_i6. The plasmid pSEVA201_g3 is missing 225 nucleotides and contains 34 nucleotides between the two *Aat*II restriction sites. The origin of this artifact is unclear. The other two plasmids share the same sequence but are missing 76 nucleotides.

To ensure that pSEVA201_P_{em7}_BCD2_*msfGFP* is not able to replicate in *P. putida* KT2440, several matings were performed. As expected, no colonies grow on kanamycin-containing agar plates. Transformation of pSEVA201_P_{em7}_BCD2_*msfGFP* into *P. putida* KT2440 strains, which contain the replication gene encoded on transposon Tn7, did lead to growth of colonies on kanamycin-containing agar plates. This confirms that genomically controlled plasmid replication by re-engineering pSEVA201 is possible.

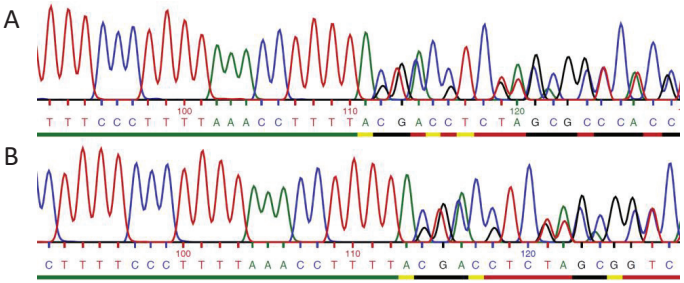


Figure 34: Sequencing chromatograms of two isolated plasmids from different *E. coli* clones. Chromatograms indicate the presence of two plasmids. Onwards of a distinct nucleotide double peaks occur. The manual analysis shows that introduced *Aat*II restriction site is present in one of the plasmids.

P. putida KT2440 strains bearing transposon Tn7 with the replication gene were verified by sequencing, to ensure the correctness of integrated sequences. Plasmids missing an own replication gene pSEVA201_g3, pSEVA201_h7 and pSEVA201_i6 were transferred to *P. putida* KT2440 clones bearing the replication gene. Promoter activity was determined and led again to high clone

to clone variation (Figure 35). *P. putida* KT2440 pSEVA241_P_{em7}_BCD2_*msfGFP* reached an activity of 3.1. For comparison, *P putida* KT2440 BG13, containing the promoter P_{em7} with BCD2 element and *msfGFP* integrated single-copy into the genome, reached an activity of 23. Sequencing reactions revealed that the plasmid is correct. In theory, a plasmid with multiple copies in a cell would lead to a higher expression level of *msfGFP*. However, here this is not observed. Possibly, the plasmid was not equally distributed during cell division. This issue was previously reported by Jahn *et al*.[241] They have shown that cultivation with a pSEVA241 derivative plasmid ends up with two subpopulations.[100, 241] Since we only measured the overall fluorescence signal, this issue was not noticeable to us. But the emergence of a subpopulation without plasmid, with varying PCN, would reduce the average fluorescence signal. Using different promoters to drive expression of the replication gene might solve this issue by tuning the *rep* expression and, thereby, the plasmid copy number to a more stable level. However, no clear trend can be discerned from this. Plasmid pSEVA201_g3 achieved comparable activity independent of the promoter controlling replication gene expression, with the same high clone-to-cone variation occurring as seen before. However, also no effect on growth in the presence of kanamycin can be recognized, indicating that plasmid loss does not play a significant role since this would likely reduce the growth rate due to subpopulation marker loss. *rep* gene expression is driven by P_{14a}, $P_{14b,}$ and P_{14d} combined with plasmid pSEVA201_h7 did show the expected increasing trend of promoter activity depending on replication gene expression. But the trend is non-linear with expression decreased by using P_{14e} and P_{14g}. The plasmid pSEVA201_i6 reached for all promoters comparable activities with pSEVA241_P_{em7}_BCD2_*msfGFP*. Only 14e leads to lower and 14g to higher ones, but here again with a very high standard deviation.

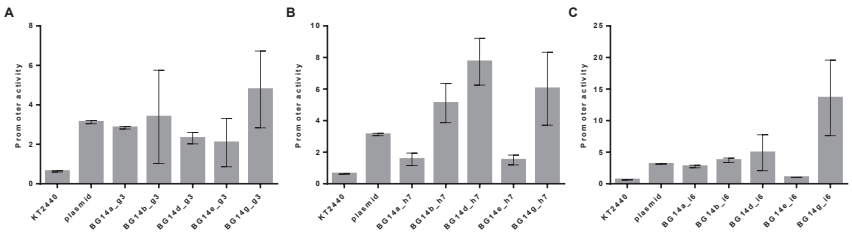


Figure 35: Characterization of genomically controlled plasmid replication in *P. putida* KT2440 Tn7 strains. The replication gene was relocated from plasmid pSEVA241 into the genome in the *attTn7* site. Gene expression is driven by promoter 14a, 14b, 14d, 14e, or 14g.[27] Replication gene of pRO1600 was deleted from the plasmid resulting in (A) pSEVA201_g3, (B) pSEVA201_h7 and (C) pSEVA201_i6. All plasmids have different aberrations around the restriction site used for self-ligation. As control plasmid *P. putida* KT2440 wild type and *P. putida* KT2440 pSEVA241_P_{em7}_BCD2_*msfGFP* was used. All replication gene expressing constructs were genomically integrated in *P. putida* KT2440, and plasmids pSEVA201_g3, pSEVA201_h7 or pSEVA201_i6 were introduced episomally. Strains were cultured in a BioLector in minimal medium with 20 mM glucose and 50 mg L^{-1} kanamycin in a 96 well plate. Identical strains from at least two different transformations were tested, with three biological replicates each. Some clones did not grow during cultivation. Error bars indicate the standard error of the mean (n>4).

All these constructs rely on the genomic expression of the replication gene. Still, a clear correlation between the strength of *rep* gene expression and resulting plasmid-based reporter gene expression is not recognizable. The issue of clone to clone variation is also not solved or reduced, which can be seen by varying determined promoter activities. Also, resulting activities are much weaker than the same construct genomically integrated as a single copy in *P. putida* KT2440 BG13. This indicates that the issue of plasmid stability and plasmid copy number is not solved by genomic *rep* gene expression and that plasmid-based factors for maintenance and expression are strongly interlinked.

3.3.4 Conclusion

Expansion of available tools is an important field for the Pseudomonads community and in synthetic biology in particular. Herein antibiotic-free expression system plays an interesting role. Avoidance of expensive addition of antibiotics, as needed for proper plasmid stability, can be circumvented by using tools for genomic integration. Zobel *et al*.[27] generated a mini Tn7 vector for single targeted integration to increase the reliability of heterologous gene expression. We modified this mini Tn7 plasmid by introducing a FLP-FRT recombination system for kanamycin marker recycling. Here, no or a small effect on promoter activity was seen by using synthetic promoters P_{14b} to P_{14g}. However, promoter P_{em7} shows a dramatically decreased activity by 70 %, independent from the presence or absence of the kanamycin marker gen. This again shows a connection between the upstream element and promoter sequence, which spans farther than thus far anticipated. To achieve marker recycling in already constructed Tn7 legacy strains, a pEMG vector was constructed targeting common sequences present in all *P. putida* KT2440 strains bearing a Tn7 integration. Although this system still suffers from poor I-*Sce*I expression from helper plasmid pSW-1,[26] alternative vectors are available for this purpose, making the construct widely applicable.[231] Nonetheless, this tool could be a promising application for old, still marker containing Tn7 strains.

Besides the expansion of the genetic toolbox of *Pseudomonas*, also some well-known legacy promoters were characterized in a systematic comparison to modern synthetic promoters, providing important context and a sense of scale to constitutive heterologous expression. The toolbox was further expanded with a late-stage promoter, which provides new options for the expression of genes that inhibit growth or genes related to PHA production, which is generally only activated in the stationary phase. Genomic control of plasmid replication by transplantation of the *rep* gene proved possible but did not solve the main issues associated with the seminal plasmid, namely clonal variation and instability

Chapter 4
General discussion

Contributions

This chapter was written by Sebastian Köbbing and reviewed by Nick Wierckx and Lars M. Blank.

4 General discussion

4.1 *Pseudomonas putida* as a suitable host for the bioeconomy

P. putida KT2440 is a promising host for biotechnological applications.[3] Its ability to overcome stressful condition makes this microorganism a favored host for the biosynthesis of challenging chemicals, like aromatic and glycolipids.[1, 20] The availability of tools allows genome editing leading to new properties of the resulting strains including the use of novel substrates and the efficient synthesis of products of choice.[26-28, 34, 114]

Glucose is a preferred carbon source for biotechnological approaches and utilized by most of the microorganisms. Due to its omnipresence, glucose is a cheap source of energy that can be produced from cereals or sugar cane. Carbon substrates that compete with food production will be challenged in the future, for sure, due to a growing world population. The fuel vs. food debate should not be repeated when changing the chemical industry from fossil to renewable carbon sources. The expansion of utilizable carbon sources is required to circumvent this issue.[246] The pentose xylose is one of the most promising alternative carbon sources and after glucose, the second most common monosaccharide on earth. Bator *et al.*[7] modified *P. putida* KT2440 to grow on xylose and to produce rhamnolipids simultaneously. The rhamnolipid synthesis genes were under the control of a strong synthetic triple promoter generated in this study ($P_{14f\_80i\_14f\_80i\_14g}$). Rhamnolipids can replace chemically derived surfactants in many everyday goods.[247-249] Using renewable resources, thereby being CO_2 neutral, is a major aim of the envisaged bioeconomy. Plant-derived triacylglycerides are used to form biofuels, and as a side product, glycerol is produced.[250] Glycerol can be utilized by *P. putida* KT2440 and many other microorganisms as a source of carbon and energy. Genetically modified strains from genus *Pseudomonas* can produce 4-hydroxybenzoate, trans-cinnamate, or phenol from glycerol.[16-18] All mentioned products are usually produced from fossil feedstocks and are essential chemical building blocks. To achieve a straight forward or reverse engineering of *P. putida* KT2440, reliable tools are needed. The choice of a suitable promoter is crucial. Here the strength can decide about the efficiency of novel pathways to be heterologously expressed in a cell. Alike the used expression system is essential, many plasmids, transposons, and integration tools are available.[26, 33, 114, 160] The major difference is the number of gene copies provided for protein synthesis in a single cell and thereby impacts the efficiency of novel pathways. Copy number and the used promoter sequence has a significant impact on the success of metabolic engineered strains like *P. putida* KT2440, by modulating heterologous gene expression.

4.2 Characterized tools promote the power of synthetic biology

Synthetic biology aims to create an integrated toolbox containing different modules that can be combined and behave in the same manner in different microorganisms.[58, 66, 76] Biologically diversity, in contrast, prohibits this dream for decades since the dynamics of biological systems lead to unpredictable effects in different hosts.[59] Calibrated and characterized tools are therefore necessary to make one step towards predictable genetic engineering. For the well-studied strain, *P. putida* KT2440, many tools are available, but defined modules are in general rare.[251] Some promoters were described after genomic integration and used for heterologous gene expression.[17, 27, 192] More extensive studies on tool characterization and calibration are missing.

This study is closing a small gap in *Pseudomonas* synthetic biology by characterization of different available molecular tools. The development of stacked promoters highlights the importance of promoter position. By only changing a short sequence upstream or downstream of the promoter leads to dramatic effects on transcriptional activity. If Zobel *et al*.[27] had used another upstream sequence while characterization and calibration of the synthetic promoter library, we nowadays would work with entirely different promoter sequences derived from that study. This example indicates how little we know about sequence-sequence interaction and the influence of promoter location on activity. An impact of upstream sequences is vaguely reported by introducing a longer upstream adenine repetitive sequence. Still, the study reported a small set of constructs.[181, 252] Keeping the surrounding context of a promoter constant indeed allows the generation of reliable promoter libraries. Varying the DNA sequence is the fundamental basis on how to create promoter libraries resulting in a wide range in promoter activity.[120, 121] Here, the influence on promoter activity varies depending on the position of changes in the sequence. The consensus sequences -35 and -10 are susceptible to nucleotides exchanges resulting in reduced activity.[120, 170] Which is a result of an inefficient recognition of the promoter by the sigma factor. Subsequently, a reduced transcription rate occurs and fewer RNA molecules for translation by ribosomes are present. Nonetheless, promoter libraries are a powerful tool in synthetic biology and metabolic engineering. They enable fine-tuning of novel pathways and lead to increased protein production.[253, 254] Depending on the approach used to generate the library, the number of promoter variants can reach up to billions; the number of promoters has a significant impact on the characterization procedures. Only a small number of promoters were characterized in detail. By reducing the number of different promoters, a detailed characterization of the sequence-activity correlation is possible, as shown with a single nucleotide polymorphism (SNP) library in this study (Chapter 3.1.3).

Fine-tuning of gene expression can be ensured by using a promoter with a certain activity and a specific sequence. The deciphering of the genotype-phenotype relationship of promoters will be challenging, maybe benefitting from advanced analyses like machine learning. We characterized

different promoters sequences with single nucleotide exchanges, which allows generating a model-based machine learning. Liebal *et al*. (in progress at iAMB, RWTH Aachen University) are using promoter sequences and activities derived from the SNP library (Chapter 3.1.3.3). Ideally, machine learning not only predicts well the relationship of promoter sequence and resulting activity but also allows the design of a promoter of a certain transcriptional activity. Here artificial intelligence meets synthetic biology to support the aim of being as predictable as possible. Integration of artificial intelligence, deep and machine learning is broadening possibilities for research and targeted molecular modifications to improve the outcome of genetic interventions. These technologies, in combination with -omics data, will increase researcher potentials in metabolic engineering and streamline chassis design. Today reliable protein structure predication is feasible by applying deep learning from previously determined proteins.[74] The combination of different technologies empowers protein engineering in a new way with promising impacts on further research. However, for many applications of machine learning, the training sets are small compared to the enormous sequence space opened by the genetic code. The need for high-quality data is also very prominent here.

The overall goal is to generate a whole-cell catalyst that combines multiple properties is one cell, accomplished by genetic and metabolic engineering.[255] Chassis allows the efficient microbial synthesis of fine and bulk chemicals from renewable feedstocks.[16] The increasing amount of reliable tools for genome editing and metabolic engineering promotes the generation of whole-cell catalysts. Since pathway construction involves heterologous gene expression, different systems are in use to facilitate proper transcription and translation. Due to plasmid instability and other disadvantages, the integration of the recombinant DNA into the genome is favored.[99, 102, 256] The transposon Tn7 integrates into many microorganisms downstream of *glmS* and allows reliable experiments due to a single copy.[115, 227] Since only one *attTn7* site is present in the genome, this tool is limited to one integration.[115] Thereby, Tn5 can integrate randomly across the genome as a single copy. Still, the location has to be laboriously identified by arbitrary PCR, and only one by one is doable.[240] Tn5 bearing pBELK developed by Nikel *et al*.[114] allows a resistance marker recycling after random integration by acting flippase, which recombines flanking FRT sequences. Due to random integration, the resulting strains are showing different expression patterns caused by the integration site. The genomic context influences transcription and translation of integrated transposons bearing heterologous genes. Activity is varying caused by the availability of resources needed to protein biosynthesis, mainly RNA polymerases, and ribosomes. Further read through activity deriving from native promoters can positively affect gene expression. These differences lead to the assumption of “hot” and “cold” spots across the genome. Marker recycling is a common approach to decrease the metabolic burden for the cell. Therefore, we transferred the kanamycin

gene with flanking FRT sites from pBELK into Tn7 vectors generated by Zobel *et al*.[27] to allow marker recycling. Strains bearing a new Tn7 with a deleted marker gene enable the transfer of a plasmid with a kanamycin marker. As mentioned above, the use of plasmids is unpopular nowadays and unsuitable for the generation of whole-cell catalysts. Genomic integration by homologous recombination is state of the art in *Pseudomonas* species by using the pEMG system developed by Martínez-García *et al*.[26] to integrate and or delete genes. However, no study is presently describing suitable integration sites across the genome. For *E. coli* and *B. subtilis* are some sites characterized. Often chosen sites depend on their distance to the replication of origin of the genome.[199, 202] Here, we used RNA-Seq data to identify equally expressed genes under a range of different growth conditions and characterized these sites by the integration of a sensor construct. Two terminators insulated the sensor and revealed to us by using a Tn5 mutant library that the activity is different across the genome. No read through activity from native promoters was observed. Identified sites for homologous based integration were tested while cultivated with varying carbon sources, showing low activity variability for some sites. These findings lead to the conclusion that our approach to identifying suitable integrations sites is applicable. Different promoters and promoterless constructs were characterized and have shown that the range of activity of no promoter, P_{em7}, and P_{14g} is the same as integrated into *attTn7* site (Table 30). Resulting promoter activity is mainly affected by the genomic context. Different to stacked promoters surrounding sequences are varying depending on the integration site in the genome.

Taking these findings together, we generated a large set of tools for gene expression and targeted markerless genomic integration in *P. putida* KT2440. This enlarges the available toolbox by well-characterized modules on two levels. To understand the connection between genotype and phenotype is an essential goal in all domains of life science.[257] High throughput sequencing enables genome sequencing on lower prices than in the past and turns attention to understanding genes and their function.[92, 258] Physiology is depending on the presence or absence of certain metabolic pathways allowing to perform a specific task in a cell, resulting in a phenotype.[259] This correlation can describe and define bacterial species based on different physiological properties. Genomics is used to get an overview of genetic information of a cell to understand the relationship of genotype to phenotype and vice versa. But results from the sequencing of microbial communities from environmental samples show that the presence of certain genes is not correlating to a specific phenotype.[260] One prominent example is the presence of resistance genes found in environmental bacterial communities. Many microorganisms are owning genes conferred a possible antibiotic resistance, but do not show the corresponding phenotype. This shows that a genotype does not always result in an expected phenotype. But the environmental context can modify these genes by the creation of mutations, which can lead to the predicted phenotype.

Prominent examples of mutation causing resistance are *rpsL* and *rpoB* in *E. coli*.[261, 262] A single point mutation in gene *rpsL* leads to resistance against streptomycin. Ricaurte *et al*.[29] used this condition for the adaptation of MAGE in *P. putida* KT2440.[25] *rpsL* encode a subunit from the RNA polymerase, which is typically attacked by streptomycin. Still, a single amino acid change leads to resistance. Introducing a single point mutation can change the genotype-phenotype relationship towards the expected one, but naturally, the resistance is not present.[260] Unfortunately, the availability of high throughput sequencing is also not able to clarify the connection of genotype-phenotype in cells. Uncovering the underlying relationships will need more research or will remain a mystery of nature.

4.3 Applications of engineered *P. putida* in the not so far future: from plastic upcycling to environmental engineering

Environmental engineering is a new field to enlarge biotechnological approaches into nature.[263] Bioremediation, much investigated in the last two decades of the past century, is a major topic and focuses on the repletion of contaminants derived from ecological pollution and industry.[264] The idea is to increase the existing microbial community in certain areas to allow the remediation of toxic compounds. In the last years, several strains with beneficial properties were found in contaminated areas. A most prominent example is *Ideonella sakaiensis*, which was found at a PET-bottle recycling site in Sakai city, Japan.[265] This strain can grow on PET-debris and degrades this polymer into accessible monomers. Two identified proteins enable this process performed by the cell. Adaptation of these properties towards other microorganisms and even other contaminants is the goal of environmental engineering.

For decades researchers are working on microorganisms to tackle pollution in air, ground, and water by so-called bioremediation.[264] Microbial communities inhabited in contaminated areas are often not able to use these compounds for growth caused by inefficient pathways or toxic intermediates accumulate leading to dead ends. Transfer of new or foreign genes and operons into other microbial species is a matter of genetically modified organisms (GMO) with beneficial properties towards the removal of pollutants, like polyaromatic hydrocarbons or PET.[266] But a release of GMO into nature is a very sensitive topic in the public discussion about biosafety and risk assessment.[267-269] Legal regulations varying between countries, and thus, high burdens have to overcome to get permission to test GMOs in the field. Therefore, only a small set of field studies using GMOs for bioremediation are present in the literature.[268] Such ideas are not followed in Europe, as the release of a GMO is strictly prohibited. Each microorganism has an ecological niche in an ecosystem, which was supported by evolution, and every organism is adapted accordingly.[270] Microorganisms generated in the laboratory, equipped with new pathways optimized for

bioremediation of certain pollutants, can perform perfectly under controlled conditions. But in nature, these organisms have to overcome biotic and abiotic stresses, alike competition for resources with other inhabitants of the area.[264, 269] At the same reduced fitness, due to higher metabolic burdens caused by integrated genes, weakens the cell and complicate growth. Next to the biotic competition, several issues can occur while bringing GMOs into nature. A missing induction of catabolic operons, stress caused by environmental conditions, toxic intermediates, and thermodynamic balances in the field has significant impacts on the chance of survival and assertiveness of the GMO.[263, 269] The field of environmental engineering has to struggle with the same issues as described for synthetic biology, the knowledge about molecular networks, physicochemical landscapes, and performance of the GMO under complex environmental parameters are rarely known.[65, 185] Here, not only the predictability of modules used in different organisms has to be characterized, but also the influence of competitors and abiotic parameters have to be investigated for new strains.

4.4 Conclusion and Outlook

This study focussed on the development and characterization of existing and new genetic tools for *P. putida* KT2440. Genomic integration was preferred to circumvent the need for antibiotic stabilization, which is necessary while using plasmids. We generated different promoters, with varying nucleotide composition of the core promoter sequence or placed them next to each other, to observe their behavior. Here a standardized integration site *attTn7* was used to make sure that the genomic context under certain conditions has no changing effects on promoter activity. While exchanging single nucleotides in the core sequence of promoter P_{14g}, we have seen that no stronger version arose, which means P_{14g} includes the most optimal nucleotide sequence for sigma-70 factor-based gene expression in *P. putida* KT2440. Nonetheless, this library increased the availability of synthetic promoters for the community and offered promoters with nuances in different activities for fine tuning gene expression. Using stacked promoters, which were separated by an 80 bp spacer sequence, taught us that upstream sequences and the distance to the translational start site have a significant impact on promoter activity. Characterization of context-dependent promoter controls reveals the differences and allows us to generate a semi-empirical correlation framework to predict the outcome of stacked promoters. Since upstream effects are unpredictable for each promoter, a separate context control is needed to determine the influence on promoter activity. Heterologous based integration allows us to delete or integrate genes at nearly every position in the genome of *P. putida* KT2440. We analyzed RNA-Seq data sets to obtain regions which are equally expressed under different cultivation conditions, to characterize these sites further. In total, 14 different integration sites were tested with different constructs and cultivated

with different carbon sources. Resulting in low variability of promoter activity for some genomic integration sites shows that the use of RNA-Seq data is feasible. This set of integrable vectors enlarges the toolbox for several approaches for *P. putida* KT2440. Furthermore, no study is available, describing and comparing different integration sites across the genome, and this also applies to other microorganisms.

msfGFP is a small protein, and the metabolic burden is manageable for the cell. Expression of larger operons under the control of SNP derived promoters could be the next step further to investigate the correlation of promoter activity and product concentration. In the best case, a correlation between promoter activity and production can be seen. This could allow us to predict the outcome of a production process in *P. putida* KT2440. Identified integration sites could be checked in the same manner with the focus on the correlation between sites for promoter activity and product formed by the cell.

Characterized genetic tools developed in this study will contribute to the field of synthetic biology in general since predictability and repeatability are overall goals of this discipline. The *Pseudomonas* community will benefit from these tools. Reliable gene expression is a crucial step in metabolic engineering and is ensured by well-characterized promoters. Subsequent genomic integration is forwarded by providing a rational chosen and characterized integration site across the genome of *P. putida* KT2440. We will, for sure, promote the use of these valuables tools for future-oriented applications.

References

5 References

[1] Wierckx, N. J. P., Ballerstedt, H., De Bont, J. A. M., and Wery, J. (2005) Engineering of solvent-tolerant *Pseudomonas putida* S12 for bioproduction of phenol from glucose, *Appl. Environ. Microb. 71*, 8221-8227.

[2] Tiso, T., Wierckx, N., Blank, L., and Grunwald, P. (2014) *Non-pathogenic Pseudomonas as platform for industrial biocatalysis*, 1st ed., Pan Stanford Publishing, Singapore.

[3] Aparicio, T., de Lorenzo, V., and Martinez-Garcia, E. (2018) CRISPR/Cas9-based counterselection boosts recombineering efficiency in *Pseudomonas putida*, *Biotechnol. J. 13*, e1700161.

[4] Nikel, P. I., Kim, J., and de Lorenzo, V. (2014) Metabolic and regulatory rearrangements underlying glycerol metabolism in *Pseudomonas putida* KT2440, *Environ. Microbiol. 16*, 239-254.

[5] Köhler, K. A. K., Blank, L. M., Frick, O., and Schmid, A. (2015) D-Xylose assimilation via the Weimberg pathway by solvent tolerant *Pseudomonas taiwanensis* VLB120, *Environ. Microbiol. 17*, 156.

[6] Jiménez, J. I., Miñambres, B., García, J. L., and Díaz, E. (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440, *Environ. Microbiol. 4*, 824.

[7] Bator, I., Wittgens, A., Rosenau, F., Tiso, T., and Blank, L. M. (2020) Comparison of three xylose pathways in *Pseudomonas putida* KT2440 for the synthesis of valuable products, *Front. Bioeng. Biotechnol. 7*.

[8] Dragosits, M., and Mattanovich, D. (2013) Adaptive laboratory evolution – principles and applications for biotechnology, *Microb. Cell Fact. 12*, 64.

[9] Li, W.-J., Jayakody, L. N., Franden, M. A., Wehrmann, M., Daun, T., Hauer, B., Blank, L. M., Beckham, G. T., Klebensberger, J., and Wierckx, N. (2019) Laboratory evolution reveals the metabolic and regulatory basis of ethylene glycol metabolism by *Pseudomonas putida* KT2440, *Environ. Microbiol. 21*, 3669-3682.

[10] Li, W.-J., Narancic, T., Kenny, S. T., Niehoff, P.-J., O'Connor, K., Blank, L. M., and Wierckx, N. (2020) Unraveling 1,4-butanediol metabolism in *Pseudomonas putida* KT2440, *Front. Microbiol. 11*.

[11] Franden, M. A., Jayakody, L. N., Li, W. J., Wagner, N. J., Cleveland, N. S., Michener, W. E., Hauer, B., Blank, L. M., Wierckx, N., Klebensberger, J., and Beckham, G. T. (2018) Engineering *Pseudomonas putida* KT2440 for efficient ethylene glycol utilization, *Metab. Eng. 48*, 197-207.

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

[13] Blank, L. M., Ionidis, G., Ebert, B. E., Bühler, B., and Schmid, A. (2008) Metabolic response of *Pseudomonas putida* during redox biocatalysis in the presence of a second octanol phase, *FEBS. J. 275*, 5173.

[14] Koopman, F., Wierckx, N., de Winde, J. H., and Ruijssenaars, H. J. (2010) Efficient whole-cell biotransformation of 5-(hydroxymethyl)furfural into FDCA, 2,5-furandicarboxylic acid, *Bioresour. Technol. 101*, 6291.

[15] Sun, Z., Ramsay, J. A., Guay, M., and Ramsay, B. A. (2007) Carbon-limited fed-batch production of medium-chain-length polyhydroxyalkanoates from nonanoic acid by *Pseudomonas putida* KT2440, *Appl. Microbiol. Biot. 74*, 69.

[16] Wynands, B., Lenzen, C., Otto, M., Koch, F., Blank, L. M., and Wierckx, N. (2018) Metabolic engineering of *Pseudomonas taiwanensis* VLB120 with minimal genomic modifications for high-yield phenol production, *Metab. Eng. 47*, 121-133.

[17] Lenzen, C., Wynands, B., Otto, M., Bolzenius, J., Mennicken, P., Blank, L. M., and Wierckx, N. (2019) High-yield production of 4-hydroxybenzoate from glucose or glycerol by an engineered *Pseudomonas taiwanensis* VLB120, *Front. Bioeng. Biotechnol. 7*.
[18] Otto, M., Wynands, B., Lenzen, C., Filbig, M., Blank, L. M., and Wierckx, N. (2019) Rational engineering of phenylalanine accumulation in *Pseudomonas taiwanensis* to enable high-yield production of t*rans*-cinnamate, *Front. Bioeng. Biotechnol. 7*.
[19] Isken, S., and de Bont, J. A. (1998) Bacteria tolerant to organic solvents, *Extremophiles : life under extreme conditions 2*, 229-238.
[20] Ramos, J. L., Duque, E., Gallegos, M. T., Godoy, P., Ramos-Gonzalez, M. I., Rojas, A., Teran, W., and Segura, A. (2002) Mechanisms of solvent tolerance in gram-negative bacteria, *Annu. Rev. Microbiol. 56*, 743.
[21] Heipieper, H. J., Neumann, G., Cornelissen, S., and Meinhardt, F. (2007) Solvent-tolerant bacteria for biotransformations in two-phase fermentation systems, *Appl. Microbiol. Biot. 74*, 961-973.
[22] Kusumawardhani, H., Hosseini, R., and de Winde, J. H. (2018) Solvent tolerance in bacteria: Fulfilling the promise of the biotech era?, *Trends Biotechnol. 36*, 1025-1039.
[23] Nikel, P. I., and de Lorenzo, V. (2018) *Pseudomonas putida* as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism, *Metab. Eng. 50*, 142-155.
[24] Calero, P., and Nikel, P. I. (2019) Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms, *Microb. Biotechnol. 12*, 98-124.
[25] Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009) Programming cells by multiplex genome engineering and accelerated evolution, *Nature 460*, 894-898.
[26] Martínez-García, E., and de Lorenzo, V. (2011) Engineering multiple genomic deletions in gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440, *Environ. Microbiol. 13*, 2702-2716.
[27] Zobel, S., Benedetti, I., Eisenbach, L., de Lorenzo, V., Wierckx, N., and Blank, L. M. (2015) Tn7-Based device for calibrated heterologous gene expression in *Pseudomonas putida*, *ACS Synth. Biol. 4*, 1341-1351.
[28] Aparicio, T., Jensen, S. I., Nielsen, A. T., de Lorenzo, V., and Martinez-Garcia, E. (2016) The Ssr protein (T1E_1405) from *Pseudomonas putida* DOT-T1E enables oligonucleotide-based recombineering in platform strain *P. putida* EM42, *Biotechnol. J. 11*, 1309-1319.
[29] Ricaurte, D. E., Martinez-Garcia, E., Nyerges, A., Pal, C., de Lorenzo, V., and Aparicio, T. (2018) A standardized workflow for surveying recombinases expands bacterial genome-editing capabilities, *Microb. Biotechnol. 11*, 176-188.
[30] Elmore, J. R., Furches, A., Wolff, G. N., Gorday, K., and Guss, A. M. (2017) Development of a high efficiency integration system and promoter library for rapid modification of *Pseudomonas putida* KT2440, *Metab. Eng. Commun. 5*, 1-8.
[31] Shen, X., Wang, Z., Huang, X., Hu, H., Wang, W., and Zhang, X. (2017) Developing genome-reduced *Pseudomonas chlororaphis* strains for the production of secondary metabolites, *BMC Genomics. 18*, 715.
[32] Martínez-García, E., Nikel, P. I., Aparicio, T., and de Lorenzo, V. (2014) Pseudomonas 2.0: genetic upgrading of *P. putida* KT2440 as an enhanced host for heterologous gene expression, *Microb. Cell Fact. 13*, 159-159.
[33] Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S., and de Lorenzo, V. (2015) SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities, *Nucleic Acids Res. 43*, 1183.
[34] Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de Las Heras, A., Páez-Espino, A. D., Durante-Rodríguez, G., Kim, J., Nikel, P. I., Platero, R., and de Lorenzo, V. (2013) The Standard European Vector Architecture (SEVA): a coherent

platform for the analysis and deployment of complex prokaryotic phenotypes, *Nucleic Acids Res. 41*, D666-D675.

[35] Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., and Law, K. L. (2015) Plastic waste inputs from land into the ocean, *Science 347*, 768-771.

[36] Mani, T., Hauk, A., Walter, U., and Burkhardt-Holm, P. (2015) Microplastics profile along the Rhine River, *Sci. Rep. 5*, 17988.

[37] Gourmelon, G. (2015) Global plastic production rises, recycling lags, *New Worldwatch Institute analysis explores trends in global plastic consumption and recycling. Recuperado de http://www.worldwatch.org*

[38] Geyer, R., Jambeck, J. R., and Law, K. L. (2017) Production, use, and fate of all plastics ever made, *Sci. Adv. 3*, e1700782.

[39] Wierckx, N., Prieto, M. A., Pomposiello, P., de Lorenzo, V., O'Connor, K., and Blank, L. M. (2015) Plastic waste as a novel substrate for industrial biotechnology, *Microb. Biotechnol. 8*, 900-903.

[40] Eriksen, M., Lebreton, L. C., Carson, H. S., Thiel, M., Moore, C. J., Borerro, J. C., Galgani, F., Ryan, P. G., and Reisser, J. (2014) Plastic pollution in the world's oceans: more than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea, *PLoS One 9*, e111913.

[41] Yaradoddi, J. S., Hugar, S., Banapurmath, N. R., Hunashyal, A. M., Sulochana, M. B., Shettar, A. S., and Ganachari, S. V. (2019) Alternative and renewable bio-based and biodegradable plastics, In *Handbook of Ecomaterials* (Martínez, L. M. T., Kharissova, O. V., and Kharisov, B. I., Eds.), pp 2935-2954, Springer International Publishing, Cham.

[42] Philp, J. C., Bartsev, A., Ritchie, R. J., Baucher, M.-A., and Guy, K. (2013) Bioplastics science from a policy vantage point, *N. Biotechnol. 30*, 635-646.

[43] Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Jr., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R., and Tschaplinski, T. (2006) The path forward for biofuels and biomaterials, *Science 311*, 484-489.

[44] Wang, Q., and Nomura, C. T. (2010) Monitoring differences in gene expression levels and polyhydroxyalkanoate (PHA) production in *Pseudomonas putida* KT2440 grown on different carbon sources, *J. Biosci. Bioeng. 110*, 653-659.

[45] Eerhart, A. J. J. E., Faaij, A. P. C., and Patel, M. K. (2012) Replacing fossil based PET with biobased PEF; process analysis, energy and GHG balance, *Energy Environ. Sci. 5*, 6407-6422.

[46] Spiekermann, P., Rehm, B. H., Kalscheuer, R., Baumeister, D., and Steinbuchel, A. (1999) A sensitive, viable-colony staining method using Nile red for direct screening of bacteria that accumulate polyhydroxyalkanoic acids and other lipid storage compounds, *Arch. Microbiol. 171*.

[47] Lu, J., Tappel, R. C., and Nomura, C. T. (2009) Mini-review: biosynthesis of poly(hydroxyalkanoates), *Polym. Rev. 49*, 226-248.

[48] Liu, Y., Yang, S., and Jia, X. (2020) Construction of a “nutrition supply–detoxification” coculture consortium for medium-chain-length polyhydroxyalkanoate production with a glucose–xylose mixture, *J. Ind. Microbiol. Biotechnol. 47*, 343-354.

[49] Tan, G.-Y., Chen, C.-L., Li, L., Ge, L., Wang, L., Razaad, I. M. N., Li, Y., Zhao, L., Mo, Y., and Wang, J.-Y. (2015) Start a research on biopolymer polyhydroxyalkanoate (PHA): a review, *Polymers 6*, 706-754.

[50] Nikel, P. I., de Almeida, A., Melillo, E. C., Galvagno, M. A., and Pettinari, M. J. (2006) New recombinant *Escherichia coli* strain tailored for the production of poly(3-hydroxybutyrate) from agroindustrial by-products, *Appl. Environ. Microb. 72*, 3949-3954.

[51] Steinbüchel, A. (2001) Perspectives for biotechnological production and utilization of biopolymers: metabolic engineering of polyhydroxyalkanoate biosynthesis pathways as a successful example, *Macromol. Biosci. 1*, 1-24.

[52] Doğan, B., and Erol, D. (2019) The future of fossil and alternative fuels used in automotive industry, In *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp 1-8.
[53] Demirbas, A. (2007) Importance of biodiesel as transportation fuel, *Energy Policy 35*, 4661-4670.
[54] Roser, M. (2020) Future Population Growth, *OurWorldInData.org*.
[55] PlasticsEurope. (2015) Plastics – the facts 2014/2015: an analysis of European plastics production, demand and waste data.
[56] Tournier, V., Topham, C. M., Gilles, A., David, B., Folgoas, C., Moya-Leclair, E., Kamionka, E., Desrousseaux, M. L., Texier, H., Gavalda, S., Cot, M., Guémard, E., Dalibey, M., Nomme, J., Cioci, G., Barbe, S., Chateau, M., André, I., Duquesne, S., and Marty, A. (2020) An engineered PET depolymerase to break down and recycle plastic bottles, *Nature 580*, 216-219.
[57] Kenny, S. T., Runic, J. N., Kaminsky, W., Woods, T., Babu, R. P., and O'Connor, K. E. (2012) Development of a bioprocess to convert PET derived terephthalic acid and biodiesel derived glycerol to medium chain length polyhydroxyalkanoate, *Appl. Microbiol. Biot. 95*, 623-633.
[58] Pleiss, J. (2006) The promise of synthetic biology, *Appl. Environ. Microb. 73*, 735-739.
[59] Szybalski, W. (1978) Nobel prizes and restriction enzymes, *Gene 4*, 181-182.
[60] Chao, R., Yuan, Y., and Zhao, H. (2015) Recent advances in DNA assembly technologies, *FEMS Yeast Res 15*, 1-9.
[61] Serrano, L. (2007) Synthetic biology: promises and challenges, *Mol. Syst. Biol. 3*, 158.
[62] Research, E. C. D. G. f. (2005) *Synthetic biology: applying engineering to biology : report of a NEST high-level expert group*, Office for Official Publications of the European Communities.
[63] Sprinzak, D., and Elowitz, M. B. (2005) Reconstruction of genetic circuits, *Nature 438*, 443-448.
[64] Chan, L. Y., Kosuri, S., and Endy, D. (2005) Refactoring bacteriophage T7, *Mol. Syst. Biol. 1*, 2005.0018.
[65] Endy, D. (2005) Foundations for engineering biology, *Nature 438*, 449-453.
[66] Lei, J., and Sun, X. (2013) Synthetic biology, predictability and reliability, In *Encyclopedia of Systems Biology* (Dubitzky, W., Wolkenhauer, O., Cho, K.-H., and Yokota, H., Eds.), pp 2042-2046, Springer New York, New York, NY.
[67] Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., and Causton, H. C. (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data, *Nat. Genet. 29*, 365.
[68] Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Le Novère, N., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J., and Forum:, a. t. r. o. t. S. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, *Bioinformatics 19*, 524-531.
[69] Mackenzie, A., Waterton, C., Ellis, R., Frow, E. K., McNally, R., Busch, L., and Wynne, B. (2013) Classifying, constructing, and identifying life: standards as transformations of "The Biological", *Science, Technology, & Human Values 38*, 701-722.
[70] Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., and Church, G. (2004) Accurate multiplex gene synthesis from programmable DNA microchips, *Nature 432*, 1050-1054.

[71] Qian, Y., Huang, H.-H., Jiménez, J. I., and Del Vecchio, D. (2017) Resource competition shapes the response of genetic circuits, *ACS Synth. Biol. 6*, 1263-1272.
[72] Dwyer, M. A., Looger, L. L., and Hellinga, H. W. (2004) Computational design of a biologically active enzyme, *Science 304*, 1967-1971.
[73] Rocha, E. P. C. (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization, *Genome Res. 14*, 2279-2286.
[74] Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Zidek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., and Hassabis, D. (2020) Improved protein structure prediction using potentials from deep learning, *Nature 577*, 706-710.
[75] Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis, *Nature 181*, 662-666.
[76] Benner, S. A., and Sismour, A. M. (2005) Synthetic biology, *Nat. Rev. Genet. 6*, 533-543.
[77] Alvarez, M., Rhodes, S. J., and Bidwell, J. P. (2003) Context-dependent transcription: all politics is local, *Gene 313*, 43-57.
[78] Cambray, G., Guimaraes, J. C., Mutalik, V. K., Lam, C., Mai, Q.-A., Thimmaiah, T., Carothers, J. M., Arkin, A. P., and Endy, D. (2013) Measurement and modeling of intrinsic transcription terminators, *Nucleic Acids Res. 41*, 5139-5148.
[79] Buchholz, K., and Collins, J. (2013) The roots—a short history of industrial microbiology and biotechnology, *Appl. Microbiol. Biot. 97*, 3747-3762.
[80] Samuel, D. (1989) Their staff of life: initial investigations of ancient Egyptian bread baking, In *Armana Reports V*. (Kemp, B. J., Ed.), pp 253-290, London: Egypt Exploration Society.
[81] Fleming, A. (1929) On the antibacterial action of cultures of a Penicillium, with special reference to their use in the Isolation of *B. influenzae*, *Brit. J. Exp. Pathol. 10*, 226-236.
[82] Weaver, W. (1970) Molecular Biology: Origin of the Term, *Science 170*, 581.
[83] Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid, *Nature 171*, 737-738.
[84] Cohen, S. N., Chang, A. C., Boyer, H. W., and Helling, R. B. (1973) Construction of biologically functional bacterial plasmids in vitro, *Proc. Natl. Acad. Sci. U. S. A. 70*, 3240-3244.
[85] Baeshen, N. A., Baeshen, M. N., Sheikh, A., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., Saini, K. S., and Redwan, E. M. (2014) Cell factories for insulin production, *Microb. Cell Fact. 13*, 141-141.
[86] Erlich, H. A. (1989) Polymerase chain reaction, *J. Clin. Immunol. 9*, 437-447.
[87] Sanger, F., and Coulson, A. R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, *J. Mol. Biol. 94*, 441-448.
[88] Bentley, D. R., and Balasubramanian, S., and Swerdlow, H. P., and Smith, G. P., and Milton, J., and Brown, C. G., and Hall, K. P., and Evers, D. J., and Barnes, C. L., and Bignell, H. R., and Boutell, J. M., and Bryant, J., and Carter, R. J., and Keira Cheetham, R., and Cox, A. J., and Ellis, D. J., and Flatbush, M. R., and Gormley, N. A., and Humphray, S. J., and Irving, L. J., and Karbelashvili, M. S., and Kirk, S. M., and Li, H., and Liu, X., and Maisinger, K. S., and Murray, L. J., and Obradovic, B., and Ost, T., and Parkinson, M. L., and Pratt, M. R., and Rasolonjatovo, I. M. J., and Reed, M. T., and Rigatti, R., and Rodighiero, C., and Ross, M. T., and Sabot, A., and Sankar, S. V., and Scally, A., and Schroth, G. P., and Smith, M. E., and Smith, V. P., and Spiridou, A., and Torrance, P. E., and Tzonev, S. S., and Vermaas, E. H., and Walter, K., and Wu, X., and Zhang, L., and Alam, M. D., and Anastasi, C., and Aniebo, I. C., and Bailey, D. M. D., and Bancarz, I. R., and Banerjee, S., and Barbour, S. G., and Baybayan, P. A., and Benoit, V. A., and Benson, K. F., and Bevis, C., and Black, P. J., and Boodhun, A., and Brennan, J. S., and Bridgham, J. A., and Brown, R. C., and Brown, A. A., and Buermann, D. H., and Bundu, A. A., and

Burrows, J. C., and Carter, N. P., and Castillo, N., and Chiara E Catenazzi, M., and Chang, S., and Neil Cooley, R., and Crake, N. R., and Dada, O. O., and Diakoumakos, K. D., and Dominguez-Fernandez, B., and Earnshaw, D. J., and Egbujor, U. C., and Elmore, D. W., and Etchin, S. S., and Ewan, M. R., and Fedurco, M., and Fraser, L. J., and Fuentes Fajardo, K. V., and Scott Furey, W., and George, D., and Gietzen, K. J., and Goddard, C. P., and Golda, G. S., and Granieri, P. A., and Green, D. E., and Gustafson, D. L., and Hansen, N. F., and Harnish, K., and Haudenschild, C. D., and Heyer, N. I., and Hims, M. M., and Ho, J. T., and Horgan, A. M., and Hoschler, K., and Hurwitz, S., and Ivanov, D. V., and Johnson, M. Q., and James, T., and Huw Jones, T. A., and Kang, G.-D., and Kerelska, T. H., and Kersey, A. D., and Khrebtukova, I., and Kindwall, A. P., and Kingsbury, Z., and Kokko-Gonzales, P. I., and Kumar, A., and Laurent, M. A., and Lawley, C. T., and Lee, S. E., and Lee, X., and Liao, A. K., and Loch, J. A., and Lok, M., and Luo, S., and Mammen, R. M., and Martin, J. W., and McCauley, P. G., and McNitt, P., and Mehta, P., and Moon, K. W., and Mullens, J. W., and Newington, T., and Ning, Z., and Ling Ng, B., and Novo, S. M., and O'Neill, M. J., and Osborne, M. A., and Osnowski, A., and Ostadan, O., and Paraschos, L. L., and Pickering, L., and Pike, A. C., and Pike, A. C., and Chris Pinkard, D., and Pliskin, D. P., and Podhasky, J., and Quijano, V. J., and Raczy, C., and Rae, V. H., and Rawlings, S. R., and Chiva Rodriguez, A., and Roe, P. M., and Rogers, J., and Rogert Bacigalupo, M. C., and Romanov, N., and Romieu, A., and Roth, R. K., and Rourke, N. J., and Ruediger, S. T., and Rusman, E., and Sanches-Kuiper, R. M., and Schenker, M. R., and Seoane, J. M., and Shaw, R. J., and Shiver, M. K., and Short, S. W., and Sizto, N. L., and Sluis, J. P., and Smith, M. A., and Ernest Sohna Sohna, J., and Spence, E. J., and Stevens, K., and Sutton, N., and Szajkowski, L., and Tregidgo, C. L., and Turcatti, G., and Vandevondele, S., and Verhovsky, Y., and Virk, S. M., and Wakelin, S., and Walcott, G. C., and Wang, J., and Worsley, G. J., and Yan, J., and Yau, L., and Zuerlein, M., and Rogers, J., and Mullikin, J. C., and Hurles, M. E., and McCooke, N. J., and West, J. S., and Oaks, F. L., and Lundberg, P. L., and Klenerman, D., and Durbin, R., and Smith, A. J. (2008) Accurate whole human genome sequencing using reversible terminator chemistry, *Nature 456*, 53-59.

[89] Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J. A., Costa, G., McKernan, K., Sidow, A., Fire, A., and Johnson, S. M. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning, *Genome Res. 18*, 1051-1063.

[90] Rusk, N. (2011) Torrents of sequence, *Nat. Methods 8*, 44-44.

[91] Nyren, P., Pettersson, B., and Uhlen, M. (1993) Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay, *Anal. Biochem. 208*, 171-175.

[92] Wetterstrand, K. A. (2019) DNA sequencing costs: data from the NHGRI genome sequencing program (GSP)

[93] Shalf, J. (2020) The future of computing beyond Moore's Law, *Philos. Trans. A Math. Phys. Eng. Sci. 378*, 20190061.

[94] Crick, F. H. (1958) On protein synthesis, *Symp. Soc. Exp. Biol. 12*, 138-163.

[95] Crick, F. (1970) Central dogma of molecular biology, *Nature 227*, 561-563.

[96] Novick, R. P. (1987) Plasmid incompatibility, *Microbiol. Reviews 51*, 381-395.

[97] 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.

[98] Khan, N. T. (2017) Mechanisms of plasmid replication, *J. Proteomics Bioinform. 10*.

[99] Lindmeyer, M., Jahn, M., Vorpahl, C., Müller, S., Schmid, A., and Bühler, B. (2015) Variability in subpopulation formation propagates into biocatalytic variability of engineered *Pseudomonas putida* strains, *Front. Microbiol. 6*, 1042.

[100] Jahn, M., Vorpahl, C., Turkowsky, D., Lindmeyer, M., Buhler, B., Harms, H., and Muller, S. (2014) Accurate determination of plasmid copy number of flow-sorted cells using droplet digital PCR, *Anal. Chem. 86*, 5969-5976.
[101] Friehs, K. (2004) Plasmid copy number and plasmid stability, *Adv. Biochem. Eng. Biot. 86*.
[102] San Millan, A., and MacLean, R. C. (2017) Fitness costs of plasmids: a limit to plasmid transmission, *Microbiol. Spectr. 5*, 1-12.
[103] Projan, S. J., Carleton, S., and Novick, R. P. (1983) Determination of plasmid copy number by fluorescence densitometry, *Plasmid 9*, 182-190.
[104] Summers, D. K., and Sherratt, D. J. (1984) Multimerization of high copy number plasmids causes instability: Cole 1 encodes a determinant essential for plasmid monomerization and stability, *Cell 36*, 1097-1103.
[105] Jones, K. L., Kim, S.-W., and Keasling, J. D. (2000) Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria, *Metab. Eng. 2*, 328-338.
[106] Langley, K. E., Villarejo, M. R., Fowler, A. V., Zamenhof, P. J., and Zabin, I. (1975) Molecular basis of beta-galactosidase alpha-complementation, *Proc. Natl. Acad. Sci. U. S. A. 72*, 1254.
[107] Ullmann, A., Jacob, F., and Monod, J. (1967) Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*, *J. Mol. Biol. 24*, 339-343.
[108] Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[109] Wein, T., Hülter, N. F., Mizrahi, I., and Dagan, T. (2019) Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance, *Nat. Commun. 10*, 2595-2595.
[110] McGrayne, S. B. (1998) *Nobel prize women in science: their lives, struggles, and momentous discoveries*, Carol Publishing Group.
[111] Freeling, M. (1992) The dynamic genome: Barbara McClintock's ideas in the century of genetics, *Trends in Genetics 8*, 463.
[112] de Lorenzo, V., Herrero, M., Jakubzik, U., and Timmis, K. N. (1990) Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative bacteria, *J. Bacteriol. 172*, 6568-6572.
[113] Reznikoff, W. S. (2008) Transposon Tn 5, *Annual. Rev. Genet. 42*, 269-286.
[114] Nikel, P. I., and de Lorenzo, V. (2013) Implantation of unmarked regulatory and metabolic modules in gram-negative bacteria with specialised mini-transposon delivery vectors, *J. Biotechnol. 163*, 143-154.
[115] Choi, K. H., Gaynor, J. B., White, K. G., Lopez, C., Bosio, C. M., Karkhoff-Schweizer, R. R., and Schweizer, H. P. (2005) A Tn7-based broad-range bacterial cloning and expression system, *Nat. Methods 2*, 443-448.
[116] Avery, O. T., Macleod, C. M., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolate from *Pneumococcus* TYPE III, *J. Exp. Med 79*, 137-158.
[117] Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation, *Nat. Rev. Microbiol. 2*, 241-249.
[118] Jensen, P. R., and Hammer, K. (1998) Artificial promoters for metabolic optimization, *Biotechnology and bioengineering 58*, 191-195.
[119] Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pavé, A. (1980) Codon catalog usage and the genome hypothesis, *Nucleic Acids Res. 8*, r49-r62.
[120] Hammer, K., Mijakovic, I., and Jensen, P. R. (2006) Synthetic promoter libraries-tuning of gene expression, *Trends Biotechnol. 24*, 53-55.

[121] Jensen, P. R., and Hammer, K. (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters, *Appl. Environ. Microb. 64*, 82.
[122] McLean, B. W., Wiseman, S. L., and Kropinski, A. M. (1997) Functional analysis of sigma-70 consensus promoters in *Pseudomonas aeruginosa* and *Escherichia coli*, *Can. J. Microbiol. 43*, 981-985.
[123] Ishihama, A. (2000) Functional Modulation of *Escherichia coli* RNA Polymerase, *Annu. Rev. Microbiol. 54*, 499-518.
[124] Szafranski, P., Smith, C. L., and Cantor, C. R. (1997) Principal transcription sigma factors of *Pseudomonas putida* strains mt-2 and G1 are significantly different, *Gene 204*, 133-138.
[125] Potvin, E., Sanschagrin, F., and Levesque, R. C. (2008) Sigma factors in *Pseudomonas aeruginosa*, *FEMS Microbiol. Rev. 32*, 38-55.
[126] Gruber, T. M., and Gross, C. A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space, *Annu. Rev. Microbiol. 57*, 441-466.
[127] McLean, B. (1995) Sequence and functional analysis of sigma70-type promoters in *Pseudomonas aeruginosa*: The influence of the intervening spacer region, *Can. J. Microbiol. 43*, 981-985.
[128] Solem, C., and Jensen, P. R. (2002) Modulation of gene expression made easy, *Appl. Environ. Microb. 68*, 2397-2403.
[129] Shine, J., and Dalgarno, L. (1975) Determinant of cistron specificity in bacterial ribosomes, *Nature 254*, 34-38.
[130] Berwal, S. K., Sreejith, R. K., and Pal, J. K. (2010) Distance between RBS and AUG plays an important role in overexpression of recombinant proteins, *Anal. Biochem. 405*, 275-277.
[131] Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., Endy, D., and Church, G. M. (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A. 110*, 14024-14029.
[132] Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering, *Proc. Natl. Acad. Sci. U. S. A. 102*, 12678-12683.
[133] Dekel, E., and Alon, U. (2005) Optimality and evolutionary tuning of the expression level of a protein, *Nature 436*, 588-592.
[134] Nevoigt, E., Kohnke, J., Fischer, C. R., Alper, H., Stahl, U., and Stephanopoulos, G. (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*, *Appl. Environ. Microb. 72*, 5266-5273.
[135] Lou, C., Stanton, B., Chen, Y.-J., Munsky, B., and Voigt, C. A. (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context, *Nat. Biotechnol. 30*, 1137-1142.
[136] Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator, *Genome Res. 14*, 1188.
[137] Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences, *Nucleic Acids Res. 18*, 6097.
[138] Hartmans, S., Smits, J. P., van der Werf, M. J., Volkering, F., and de Bont, J. A. M. (1989) Metabolism of styrene oxide and 2-phenylethanol in the styrene-degrading *Xanthobacter* strain 124X, *Appl. Environ. Microb. 55*, 2850.
[139] Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids, *J. Mol. Biol. 166*, 557-580.
[140] Boyer, H. W., and Roulland-Dussoix, D. (1969) A complementation analysis of the restriction and modification of DNA in *Escherichia coli*, *J. Mol. Biol. 41*, 459.
[141] Keen, N. T., Tamaki, S., Kobayashi, D., and Trollinger, D. (1988) Improved broad-host-range plasmids for DNA cloning in gram-negative bacteria, *Gene 70*, 191.
[142] de Lorenzo, V., and Timmis, K. N. (1994) Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons, *Methods Enzymol. 235*, 386-405.

[143] Herrero, M., de Lorenzo, V., and Timmis, K. N. (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria, *J. Bacteriol. 172*, 6557.

[144] Bagdasarian, M., Lurz, R., Ruckert, B., Franklin, F., Bagdasarian, M., Frey, J., and Timmis, K. (1981) 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 16*, 237-247.

[145] de las Heras, A., Carreno, C. A., and de Lorenzo, V. (2008) Stable implantation of orthogonal sensor circuits in Gram-negative bacteria for environmental release, *Environ. Microbiol. 10*, 3305-3316.

[146] Udvardi, M. K., Czechowski, T., and Scheible, W.-R. (2008) Eleven golden rules of quantitative RT-PCR, *Plant Cell. 20*, 1736-1737.

[147] Pfaffl, M. W. (2001) A new mathematical model for relative quantification in real-time RT–PCR, *Nucleic Acids Res. 29*, e45-e45.

[148] Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., and Soboleva, A. (2012) NCBI GEO: archive for functional genomics data sets—update, *Nucleic Acids Res. 41*, D991-D995.

[149] La Rosa, R., Nogales, J., and Rojo, F. (2015) The Crc/CrcZ-CrcY global regulatory system helps the integration of gluconeogenic and glycolytic metabolism in *Pseudomonas putida*, *Environ. Microbiol. 17*, 3362-3378.

[150] Kim, J., Oliveros, J. C., Nikel, P. I., Lorenzo, V., and Silva-Rocha, R. (2013) Transcriptomic fingerprinting of *Pseudomonas putida* under alternative physiological regimes, *Environ. Microbiol. Rep. 5*, 883-891.

[151] Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A., 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., and Fraser, C. M. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440, *Environ. Microbiol. 4*, 799-808.

[152] Chu, Y., and Corey, D. R. (2012) RNA sequencing: platform selection, experimental design, and data interpretation, *Nucleic Acid Ther. 22*, 271-274.

[153] Kukurba, K. R., and Montgomery, S. B. (2015) RNA sequencing and analysis, *Cold Spring Harb Protoc 2015*, 951-969.

[154] Winsor, G. L., Griffiths, E. J., Lo, R., Dhillon, B. K., Shay, J. A., and Brinkman, F. S. (2016) Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database, *Nucleic Acids Res. 44*, D646-653.

[155] Solovyev, V., and Salamov, A. (2011) Automatic annotation of microbial genomes and metagenomic sequences. In metagenomics and its applications in agriculture, biomedicine and environmental studies, (Li, R. W., Ed.), pp 61-78, Nova Science Publishers.

[156] Reese, M. G. (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome, *Comput. Chem. 26*, 51-56.

[157] Gautheret, D., and Lambert, A. (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles, *J. Mol. Biol. 313*, 1003-1011.

[158] Low, K. B. (1974) Experiments in Molecular Genetics. Jeffrey H. Miller, *9+ 49*, 151-151.

[159] Lambertsen, L., Sternberg, C., and Molin, S. (2004) Mini-Tn7 transposons for site-specific tagging of bacteria with fluorescent proteins, *Environ. Microbiol. 6*, 726-732.

[160] Damron, F. H., McKenney, E. S., Barbier, M., Liechti, G. W., Schweizer, H. P., and Goldberg, J. B. (2013) Construction of mobilizable mini-Tn7 vectors for bioluminescent detection of gram-negative bacteria and single-copy promoter lux reporter analysis, *Appl. Environ. Microb. 79*, 4149.
[161] Silva-Rocha, R., and de Lorenzo, V. (2014) Chromosomal integration of transcriptional fusions, *Methods. Mol. Biol. 1149*, 479-489.
[162] Schweizer, H. P. (1992) Allelic exchange in *Pseudomonas aeruginosa* using novel ColE1-type vectors and a family of cassettes containing a portable oriT and the counter-selectable *Bacillus subtilis* sacB marker, *Mol. Microbiol. 6*, 1195-1204.
[163] Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems, *Nat. Biotechnol. 31*, 233.
[164] Sánchez-Pascuala, A., Fernández-Cabezón, L., de Lorenzo, V., and Nikel, P. I. (2019) Functional implementation of a linear glycolysis for sugar catabolism in *Pseudomonas putida*, *Metab. Eng. 54*, 200-211.
[165] Solem, C., Koebmann, B., Yang, F., and Jensen, P. R. (2007) The las enzymes control pyruvate metabolism in *Lactococcus lactis* during growth on maltose, *J. Bacteriol. 189*, 6727.
[166] Rud, I., Jensen, P. R., Naterstad, K., and Axelsson, L. (2006) A synthetic promoter library for constitutive gene expression in *Lactobacillus plantarum*, *Microbiology 152*, 1011.
[167] 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., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements, *Nat. Methods 10*, 354-360.
[168] Gilman, J., and Love, J. (2016) Synthetic promoter design for new microbial chassis, *Biochem. Soc. T. 44*, 731-737.
[169] Kang, J.-G., Hahn, M.-Y., Roe, J.-H., and Ishihama, A. (1997) Identification of sigma factors for growth phase-related promoter selectivity of RNA polymerases from *Streptomyces coelicolor* A3(2), *Nucleic Acids Res. 25*, 2566-2573.
[170] Lodge, J., Williams, R., Bell, A., Chan, B., and Busby, S. (1990) Comparison of promoter activities in *Escherichia coli* and *Pseudomonas aeruginosa*: use of a new broad-host-range promoter-probe plasmid, *FEMS Microbiol. Lett. 67*, 221-225.
[171] Gao, Y., Liu, C., Ding, Y., Sun, C., Zhang, R., Xian, M., and Zhao, G. (2014) Development of genetically stable *Escherichia coli* strains for poly(3-hydroxypropionate) production, *PLoS One 9*, e97845.
[172] Li, M., Wang, J., Geng, Y., Li, Y., Wang, Q., Liang, Q., and Qi, Q. (2012) A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*, *Microb. Cell Fact. 11*, 19.
[173] Dixon, R. (1984) Tandem promoters determine regulation of the *Klebsiella pneumonia* glutamine synthetase (gln) gene, *Nucleic Acids Res. 12*, 7811-7830.
[174] Martens, J. A., Laprade, L., and Winston, F. (2004) Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene, *Nature 429*, 571-574.
[175] Tamsir, A., Tabor, J. J., and Voigt, C. A. (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires', *Nature 469*, 212-215.
[176] Stanton, B. C., Nielsen, A. A. K., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. A. (2013) Genomic mining of prokaryotic repressors for orthogonal logic gates, *Nat. Chem. Biol. 10*, 99.
[177] Landgraf, D. (2012) Quantifying localizations and dynamics in single bacterial cells, Harvard University, Cambridge, Massachusetts, USA.
[178] Schmitz, A., and Galas, D. J. (1979) The interaction of RNA polymerase and lac repressor with the lac control region, *Nucleic Acids Res. 6*, 111-137.
[179] Shearwin, K. E., Callen, B. P., and Egan, J. B. (2005) Transcriptional interference-a crash course, *Trends Genet. 21*, 339-345.

[180] Callen, B. P., Shearwin, K. E., and Egan, J. B. (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter, *Mol. Cell. 14*.
[181] Estrem, S. T., Gaal, T., Ross, W., and Gourse, R. L. (1998) Identification of an UP element consensus sequence for bacterial promoters, *Proc. Natl. Acad. Sci. U. S. A. 95*, 9761-9766.
[182] Debode, F., Marien, A., Janssen, É., Bragard, C., and Berben, G. (2017) Influence of the amplicon length on real-time PCR results, *Biotechnol. Agron. Soc. 21*, 3-11.
[183] Martinez-Garcia, E., and de Lorenzo, V. (2017) Molecular tools and emerging strategies for deep genetic/genomic refactoring of *Pseudomonas*, *Current opinion in biotechnology 47*, 120-132.
[184] Panke, S., de Lorenzo, V., Kaiser, A., Witholt, B., and Wubbolts, M. G. (1999) Engineering of a stable whole-cell biocatalyst capable of (S)-styrene oxide formation for continuous two-liquid-phase applications, *Appl. Environ. Microb. 65*, 5619-5623.
[185] Davison, J. (1999) Genetic exchange between bacteria in the environment, *Plasmid 42*, 73-91.
[186] Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H., and Spratt, B. G. (2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA), *Proc. Natl. Acad. Sci. U. S. A. 99*, 7687-7692.
[187] Berenjian, A. (2019) *Essentials in Fermentation Technology*, Springer.
[188] Craig, N. L. (1996) Transposon Tn7, *Curr. Top. Microbiol. Immunol. 204*, 27-48.
[189] Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., de Arruda, M., Burland, V., Harcum, S. W., and Blattner, F. R. (2006) Emergent properties of reduced-genome *Escherichia coli*, *Science 312*, 1044-1046.
[190] Domröse, A., Hage-Hülsmann, J., Thies, S., Weihmann, R., Kruse, L., Otto, M., Wierckx, N., Jaeger, K.-E., Drepper, T., and Loeschcke, A. (2019) *Pseudomonas putida* rDNA is a favored site for the expression of biosynthetic genes, *Sci. Rep. 9*, 7028.
[191] Domröse, A., Klein, A. S., Hage-Hülsmann, J., Thies, S., Svensson, V., Classen, T., Pietruszka, J., Jaeger, K. E., Drepper, T., and Loeschcke, A. (2015) Efficient recombinant production of prodigiosin in *Pseudomonas putida*, *Front. Microbiol. 6*, 972.
[192] Otto, M., Wynands, B., Drepper, T., Jaeger, K.-E., Thies, S., Loeschcke, A., Blank, L. M., and Wierckx, N. (2019) Targeting 16S rDNA for stable recombinant gene expression in *Pseudomonas*, *ACS Synth. Biol. 8*, 1901-1912.
[193] Nikel, P. I., Chavarría, M., Danchin, A., and de Lorenzo, V. (2016) From dirt to industrial applications: *Pseudomonas putida* as a Synthetic Biology chassis for hosting harsh biochemical reactions, *Curr. Opin. Chem. Biol. 34*, 20-29.
[194] de Bruijn, F. J., and Lupski, J. R. (1984) The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids — a review, *Gene 27*, 131-149.
[195] Martínez-García, E., Aparicio, T., de Lorenzo, V., and Nikel, P. I. (2014) New transposon tools tailored for metabolic engineering of gram-negative microbial cell factories, *Front. Bioeng. Biotechnol. 2*, 46-46.
[196] Miller, J. H., Calos, M. P., Galas, D., Hofer, M., Büchel, D. E., and Benno, M.-H. (1980) Genetic analysis of transpositions in the *lac* region of *Escherichia coli*, *J. Mol. Biol. 144*, 1-18.
[197] Shanabruch, W. G., Behlau, I., and Walker, G. C. (1981) Spontaneous mutators of *Salmonella typhimurium* LT2 generated by insertion of transposable elements, *J. Bacteriol. 147*, 827-835.
[198] Kuhlman, T. E., and Cox, E. C. (2010) Site-specific chromosomal integration of large synthetic constructs, *Nucleic Acids Res. 38*.
[199] Englaender, J. A., Jones, J. A., Cress, B. F., Kuhlman, T. E., Linhardt, R. J., and Koffas, M. A. G. (2017) Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*, *ACS Synth. Biol. 6*, 710-720.

[200] Cooper, S., and Helmstetter, C. E. (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r, *J. Mol. Biol. 31*, 519-540.
[201] Espah Borujeni, A., Channarasappa, A. S., and Salis, H. M. (2013) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites, *Nucleic Acids Res. 42*, 2646-2659.
[202] Sauer, C., Syvertsson, S., Bohorquez, L. C., Cruz, R., Harwood, C. R., van Rij, T., and Hamoen, L. W. (2016) Effect of genome position on heterologous gene expression in *Bacillus subtilis*: an unbiased analysis, *ACS Synth. Biol. 5*, 942-947.
[203] Frank, S., Klockgether, J., Hagendorf, P., Geffers, R., Schöck, U., Pohl, T., Davenport, C. F., and Tümmler, B. (2011) *Pseudomonas putida* KT2440 genome update by cDNA sequencing and microarray transcriptomics, *Environ. Microbiol. 13*, 1309-1326.
[204] Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein, *Nat. Biotechnol. 22*, 1567-1572.
[205] Belda, E., van Heck, R. G., Jose Lopez-Sanchez, M., Cruveiller, S., Barbe, V., Fraser, C., Klenk, H. P., Petersen, J., Morgat, A., Nikel, P. I., Vallenet, D., Rouy, Z., Sekowska, A., Martins Dos Santos, V. A., de Lorenzo, V., Danchin, A., and Medigue, C. (2016) The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis, *Environ. Microbiol. 18*, 3403-3424.
[206] van Opijnen, T., Bodi, K. L., and Camilli, A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms, *Nat. Methods 6*, 767-772.
[207] van Opijnen, T., Lazinski, D. W., and Camilli, A. (2014) Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms, *Curr. Protoc. Mol. Biol. 106*, 7 16 11-24.
[208] Leprince, A., de Lorenzo, V., Voller, P., van Passel, M. W., and Martins dos Santos, V. A. (2012) Random and cyclical deletion of large DNA segments in the genome of *Pseudomonas putida*, *Environ. Microbiol. 14*, 1444-1453.
[209] D'Arrigo, I., Bojanovic, K., Yang, X., Holm Rau, M., and Long, K. S. (2016) Genome-wide mapping of transcription start sites yields novel insights into the primary transcriptome of *Pseudomonas putida*, *Environ. Microbiol. 18*, 3466-3481.
[210] Hernandez-Arranz, S., Moreno, R., and Rojo, F. (2013) The translational repressor Crc controls the *Pseudomonas putida* benzoate and alkane catabolic pathways using a multi-tier regulation strategy, *Environ. Microbiol. 15*, 227-241.
[211] Moreno, R., Fonseca, P., and Rojo, F. (2012) Two small RNAs, CrcY and CrcZ, act in concert to sequester the Crc global regulator in *Pseudomonas putida*, modulating catabolite repression, *Mol. Microbiol. 83*, 24-40.
[212] Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, *Nucleic Acids Res. 30*, 207-210.
[213] Christodoulou, D. C., Gorham, J. M., Herman, D. S., and Seidman, J. G. (2011) Construction of normalized RNA-seq libraries for next-generation sequencing using the crab duplex-specific nuclease, *Curr. Protoc. Mol. Biol. 94*, 4.12.11-14.12.11.
[214] Auton, M., Holthauzen, L. M., and Bolen, D. W. (2007) Anatomy of energetic changes accompanying urea-induced protein denaturation, *Proc. Natl. Acad. Sci. U. S. A. 104*, 15317-15322.
[215] Cox, M. M. (1983) The FLP protein of the yeast 2-microns plasmid: expression of a eukaryotic genetic recombination system in *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A. 80*, 4223-4227.
[216] Lange, R., and Hengge-Aronis, R. (1991) Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*, *Mol. Microbiol. 5*, 49-59.

[217] Hengge-Aronis, R. (2002) Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase, *Microbiol. Mol. Biol. Rev. 66*, 373-395.

[218] Schuster, M., Hawkins, A. C., Harwood, C. S., and Greenberg, E. P. (2004) The *Pseudomonas aeruginosa* RpoS regulon and its relationship to quorum sensing, *Mol. Microbiol. 51*, 973-985.

[219] Rouvière, P. E., De Las Peñas, A., Mecsas, J., Lu, C. Z., Rudd, K. E., and Gross, C. A. (1995) rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*, *EMBO J. 14*, 1032-1042.

[220] Tsang, J., and Hoover, T. R. (2014) Themes and variations: Regulation of RpoN-dependent flagellar genes across diverse bacterial species, *Scientifica (Cairo) 2014*, 681754-681754.

[221] Hoffmann, N., and Rehm, B. H. A. (2004) Regulation of polyhydroxyalkanoate biosynthesis in *Pseudomonas putida* and *Pseudomonas aeruginosa*, *FEMS Microbiol. Lett. 237*, 1-7.

[222] Yu, X., Xu, J., Liu, X., Chu, X., Wang, P., Tian, J., Wu, N., and Fan, Y. (2015) Identification of a highly efficient stationary phase promoter in *Bacillus subtilis*, *Sci. Rep. 5*, 18405.

[223] Shimada, T., Makinoshima, H., Ogawa, Y., Miki, T., Maeda, M., and Ishihama, A. (2004) Classification and strength measurement of stationary-phase promoters by use of a newly developed promoter cloning vector, *J. Bacteriol. 186*, 7112-7122.

[224] Peters, J. E., and Craig, N. L. (2001) Tn7: smarter than we thought, *Nat. Rev. Mol. Cell Biol. 2*, 806-814.

[225] Waddell, C. S., and Craig, N. L. (1988) Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes, *Genes Dev. 2*, 137-149.

[226] Kuduvalli, P. N., Mitra, R., and Craig, N. L. (2005) Site-specific Tn7 transposition into the human genome, *Nucleic Acids Res. 33*, 857-863.

[227] Mitra, R., McKenzie, G. J., Yi, L., Lee, C. A., and Craig, N. L. (2010) Characterization of the TnsD-attTn7 complex that promotes site-specific insertion of Tn7, *Mob. DNA. 1*, 18.

[228] Sarnovsky, R. J., May, E. W., and Craig, N. L. (1996) The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products, *EMBO J. 15*, 6348-6361.

[229] Bainton, R. J., Kubo, K. M., Feng, J.-n., and Craig, N. L. (1993) Tn7 transposition: Target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system, *Cell 72*, 931-943.

[230] Gringauz, E., Orle, K. A., Waddell, C. S., and Craig, N. L. (1988) Recognition of *Escherichia coli attTn7* by transposon Tn7: lack of specific sequence requirements at the point of Tn7 insertion, *J. Bacteriol. 170*, 2832-2840.

[231] Wirth, N. T., Kozaeva, E., and Nikel, P. I. (2020) Accelerated genome engineering of *Pseudomonas putida* by I-*Sce*I—mediated recombination and CRISPR-Cas9 counterselection, *Microb. Biotechnol. 13*, 233-249.

[232] Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., VanCleemput, M., and Wu, A. M. (1981) The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*, *Nucleic Acids Res. 9*, 6647-6668.

[233] Silverstone, A. E., Arditti, R. R., and Magasanik, B. (1970) Catabolite-Insensitive Revertants of *Lac* Promoter Mutants, *Proc. Natl. Acad. Sci. U. S. A. 66*, 773-779.

[234] de Boer, H. A., Comstock, L. J., and Vasser, M. (1983) The tac promoter: a functional hybrid derived from the trp and lac promoters, *Proc. Natl. Acad. Sci. U. S. A. 80*, 21-25.

[235] Brosius, J., Erfle, M., and Storella, J. (1985) Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity, *J. Biol. Chem. 260*, 3539-3541.

[236] Wang, B., Eckert, C., Maness, P.-C., and Yu, J. (2018) A genetic toolbox for modulating the expression of heterologous genes in the cyanobacterium *Synechocystis* sp. PCC 6803, *ACS Synth. Biol. 7*, 276-286.

[237] Wise, A., Brems, R., Ramakrishnan, V., and Villarejo, M. (1996) Sequences in the -35 region of *Escherichia coli rpoS*-dependent genes promote transcription by E sigma S, *J. Bacteriol. 178*, 2785-2793.
[238] Lim, H.-J., Kim, K., Shin, M., Jeong, J.-H., Ryu, P. Y., and Choy, H. E. (2015) Effect of promoter-upstream sequence on σ38-dependent stationary phase gene transcription, *J. Microbiol. 53*, 250-255.
[239] Franchini, A. G., Ihssen, J., and Egli, T. (2015) Effect of global regulators RpoS and cyclic-AMP/CRP on the catabolome and transcriptome of *Escherichia coli* K12 during carbon- and energy-imited growth, *PLoS One 10*, e0133793.
[240] Das, S., Noe, J. C., Paik, S., and Kitten, T. (2005) An improved arbitrary primed PCR method for rapid characterization of transposon insertion sites, *J. Microbiol. Methods 63*, 89-94.
[241] Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H., and Müller, S. (2016) Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR, *Microb. Cell Fact. 15*, 211.
[242] Lin-Chao, S., Chen, W. T., and Wong, T. T. (1992) High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II, *Mol. Microbiol. 6*, 3385-3393.
[243] Dmowski, M., and Jagura-Burdzy, G. (2013) Active stable maintenance functions in low copy-number plasmids of Gram-positive bacteria I. Partition systems, *Pol. J. Microbiol. 62*, 3-16.
[244] Kittleson, J. T., Cheung, S., and Anderson, J. C. (2011) Rapid optimization of gene dosage in *E. coli* using DIAL strains, *J. Biol. Eng. 5*, 10.
[245] del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., and Díaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids, *Microbiol. Mol. Biol. Rev. 62*, 434-464.
[246] Prasad, S., and Ingle, A. P. (2019) Impacts of sustainable biofuels production from biomass, In *Sustainable Bioenergy* (Rai, M., and Ingle, A. P., Eds.), pp 327-346, Elsevier.
[247] Lang, S., and Wullbrandt, D. (1999) Rhamnose lipids – biosynthesis, microbial production and application potential, *Appl. Microbiol. Biot. 51*, 22-32.
[248] Loeschcke, A., and Thies, S. (2015) *Pseudomonas putida*—a versatile host for the production of natural products, *Appl. Environ. Microb. 99*, 6197-6214.
[249] Tiso, T., Zauter, R., Tulke, H., Leuchtle, B., Li, W. J., Behrens, B., Wittgens, A., Rosenau, F., Hayen, H., and Blank, L. M. (2017) Designer rhamnolipids by reduction of congener diversity: production and characterization, *Microb. Cell Fact. 16*, 225.
[250] Ma, F., and Hanna, M. A. (1999) Biodiesel production: a review, *Bioresour. Technol. 70*, 1-15.
[251] Sleight, S. C., Bartley, B. A., Lieviant, J. A., and Sauro, H. M. (2010) Designing and engineering evolutionary robust genetic circuits, *J. Biol. Eng. 4*, 12.
[252] Aiyar, S. E., Gourse, R. L., and Ross, W. (1998) Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase alpha subunit, *Proc. Natl. Acad. Sci. U. S. A. 95*, 14652-14657.
[253] De Mey, M., Maertens, J., Lequeux, G. J., Soetaert, W. K., and Vandamme, E. J. (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering, *BMC Biotechnol. 7*, 34.
[254] Hartner, F. S., Ruth, C., Langenegger, D., Johnson, S. N., Hyka, P., Lin-Cereghino, G. P., Lin-Cereghino, J., Kovar, K., Cregg, J. M., and Glieder, A. (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*, *Nucleic Acids Res. 36*, e76.
[255] Lin, B., and Tao, Y. (2017) Whole-cell biocatalysts by design, *Microb. Cell Fact. 16*, 106.
[256] Lin-Chao, S., and Bremer, H. (1986) Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*, *Mol. Gen. Genet. 203*.
[257] van Opijnen, T., and Camilli, A. (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen, *Genome Res. 22*, 2541-2551.

[258] Pommerenke, C., Müsken, M., Becker, T., Dötsch, A., Klawonn, F., and Häussler, S. (2010) Global genotype-phenotype correlations in *Pseudomonas aeruginosa*, *PLoS Pathog. 6*, e1001074-e1001074.
[259] Balows, A. (1975) Bergey's Manual of Determinative Bacteriology. Eighth Edition, *Am. J. Public Health. 65*, 315-315.
[260] Hughes, D., and Andersson, D. I. (2017) Environmental and genetic modulation of the phenotypic expression of antibiotic resistance, *FEMS Microbiol. Rev. 41*, 374-391.
[261] Funatsu, G., and Wittmann, H. G. (1972) Ribosomal proteins: XXXIII. Location of amino-acid replacements in protein S12 isolated from *Escherichia coli* mutants resistant to streptomycin, *J. Mol. Biol. 68*, 547-550.
[262] Jin, D. J., and Gross, C. A. (1988) Mapping and sequencing of mutations in the *Escherichia coli rpoB* gene that lead to rifampicin resistance, *J. Mol. Biol. 202*, 45-58.
[263] de Lorenzo, V. (1992) Genetic engineering strategies for environmental applications, *Curr. Opin. Biotechnol. 3*, 227-231.
[264] Timmis, K. N., and Pieper, D. H. (1999) Bacteria designed for bioremediation, *Trends Biotechnol. 17*, 201-204.
[265] Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., and Oda, K. (2016) A bacterium that degrades and assimilates poly(ethylene terephthalate), *Science 351*, 1196.
[266] King, J. M., Digrazia, P. M., Applegate, B., Burlage, R., Sanseverino, J., Dunbar, P., Larimer, F., and Sayler, G. S. (1990) Rapid, sensitive bioluminescent reporter technology for naphthalene exposure and biodegradation, *Science 249*, 778-781.
[267] de Lorenzo, V. (2010) Environmental biosafety in the age of Synthetic Biology: Do we really need a radical new approach?, *Bioessays 32*, 926-931.
[268] Sayler, G. S., and Ripp, S. (2000) Field applications of genetically engineered microorganisms for bioremediation processes, *Curr. Opin. Biotechnol. 11*, 286-289.
[269] Cases, I., and de Lorenzo, V. (2005) Genetically modified organisms for the environment: stories of success and failure and what we have learned from them, *Int. Microbiol. 8*, 213-222.
[270] Pocheville, A. (2015) The ecological niche: history and recent controversies, In *Handbook of Evolutionary Thinking in the Sciences* (Heams, T., Huneman, P., Lecointre, G., and Silberstein, M., Eds.), pp 547-586, Springer Netherlands, Dordrecht.

Curriculum vitae

Curriculum vitae

Personal data

Last Name	Köbbing
First Name	Sebastian
Date of birth	January 2nd, 1986
Place of birth	Münster, Germany
Nationality	German
Email	sebastian.koebbing@rwth-aachen.de

Education

2015-2020	Ph.D. student at the Institute of Applied Microbiology, RWTH Aachen University, Aachen, Germany
2016	3rd International Synthetic & Systems Biology Summer School – SSBSS2016, Volterra (Pisa), Italy
2012-2015	Master of Science in Biology at RWTH Aachen University, Aachen, Germany
2009-2012	Bachelor of Science in Biosciences at Westfälische Wilhems University, Münster, Germany
2006-2009	Abitur at the Overberg-Kolleg, Münster, Germany
2003-2006	Training as a biological laboratory technician in the field of animal disease investigation at the Chemischen Landes- und Staatlichen Veterinäruntersuchsamt, Münster (CVUA), Germany

Work Experience

2015-2020	Scientific employee and doctoral candidate at the Institute of Applied Microbiology, RWTH Aachen University, Germany
2003-2006	Training at the Chemischen Landes- und Staatlichen Veterinäruntersuchsamt (CVUA), Münster, Germany

Publications

S. Köbbing, L. M. Blank, N. Wierckx. “Characterization of context-dependent effects on synthetic promoters”, *Front. Bioeng. Biotechnol*. (2020), 8(551), doi: 10.3389/fbioe.2020.00551

M. Wehrmann, E. M. Elsayed, S. Köbbing, L. Bendz, A. Lepak, J. Schwabe, N. Wierckx, G. Bange, J. Klebensberger. “An engineered PQQ-dependent alcohol dehydrogenase for the oxidation of 5-(hydroxymethyl)furfural and 5-(hydroxymethyl)furoic acid”, *ACS Catal*. (accepted), doi: 10.1021/acscatal.0c01789

R. Hosseini, J. Kuepper, S. Köbbing, L. M. Blank, N. Wierckx, J. H. de Winde. “Regulation of solvent tolerance in *Pseudomonas putida* S12 mediated by mobile elements”, *Microb. Biotechnol*. (2017), 10:1558-1568, doi: 10.1111/1751-7915.12495

D. Bittner, S. Köbbing, D. Neves, S. Nies, N. Wierckx, B. E. Ebert, L.M. Blank, F. Eiden. “Genetic engineering of a recombinant pathway for superior acetone production by *Escherichia coli*”, *J. Biosci. Bioeng*. (in preparation)

Oral presentations

S. Köbbing, L. M. Blank, N. Wierckx. “Development of synthetic biology tools for *Pseudomonas putida*” at P4SB project meetings in Lisbon (Portugal, 24.11.-27.11.2015), Madrid (Spain, 28.09.-29.09.2016), Strasbourg (France, 20.09.-21.09.2017), and Surrey (England, 20.09.-21.09.2018)

S. Köbbing, L. M. Blank, N. Wierckx. “P4SB - From Plastic waste to Plastic value using *Pseudomonas putida* Synthetic Biology” at Seminartag Bioplastik, Münster, Germany, 11.06.2016

S. Köbbing, L. M. Blank, N. Wierckx. “Synthetic promoter stacking: When two plus two equals three” at 4th Applied Synthetic Biology in Europe, Toulouse, France, 24.10.-26.10.2018

S. Köbbing, L. M. Blank, N. Wierckx. “Development and application of molecular tools for *Pseudomonas*” at 1st Pseudomonas Grassroots Meeting, Frankfurt am Main, Germany, 08.11.-09.11.2018

S. Köbbing, H. Ballerstedt, L. M. Blank, N. Wierckx. “From Plastic waste to Plastic value” at GreenWin Conference – Green Chemistry – White Biotechnology, Brussels, Belgium, 08.05.2019

Poster presentations

S. Köbbing, B. Wynands, L. M. Blank, N. Wierckx. “Synthetic promoter libraries, a powerful tool for tuning gene expression” 3rd Congress on Applied Synthetic Biology in Europe, Lisbon, Portugal, 22.02.-24.02.2016

B. Wynands, S. Zobel, S. Köbbing, I. Benedetti, V. de Lorenzo, L. M. Blank, N. Wierckx. “Low degeneracy synthetic promoter libraries” at Gordon Research Conference - Synthetic Biology 2015, Newry, USA, 28.06.-03.07.2016

S. Köbbing, L. M. Blank, N. Wierckx. “Synthetic promoter stacking: when two plus two equals three” at GASB German Conference on Synthetic Biology, Aachen, Germany, 12.09.-13.09.2019

S. Köbbing, L. M. Blank, N. Wierckx. “Identification of reliable genomic landing sites for heterologous constructs in *Pseudomonas putida* KT2440” at 2nd Pseudomonas Grassroots Meeting, Leiden, Netherlands, 07.11.-08.11.2019

Projects

2015-2019	P4SB - “From Plastic waste to Plastic value using *Pseudomonas putida* KT2440 Synthetic Biology”
2017/2018	Contract research for Corbion N.V., Netherlands
Since 2019	Wrote an application in cooperation with WZL (RWTH Aachen University) for an ERS grant from RWTH Aachen University for the construction of an automated genome editing device for MAGE (Multiplex Automated Genomic Engineering). The project was approved and is still running

Tools published in other publications

I. Bator, A. Wittgens, F. Rosenau, T. Tiso, L. M. Blank. “Comparison of three xylose pathways in *Pseudomonas putida* KT2440 for the synthesis of valuable products”, *Front. Bioeng. Biotechnol.* (2020), 7:480, doi: 10.3389/fbioe.2019.00480

C. Lenzen, B. Wynands, M. Otto, J. Bolzenius, P. Mennicken, L. M. Blank, N. Wierckx. “High-yield production of 4-hydroxybenzoate from glucose or glycerol by an engineered *Pseudomonas taiwanensis* VLB120”, *Front. Bioeng. Biotechnol.* (2019), 7:130, doi: 10.3389/fbioe.2019.00130

Tools shared with the community

P. Demling from Institute of Applied Microbiology, RWTH Aachen University, Germany

M. Filbig from Institute of Applied Microbiology, RWTH Aachen University, Germany

I. Bator from Institute of Applied Microbiology, RWTH Aachen University, Germany

C. Lenzen from Institute of Applied Microbiology, RTWH Aachen University, Germany

U. Liebal from Institute of Applied Microbiology, RTWH Aachen University, Germany

C. Alvarez Ortega from BacMine, Madrid, Spain

R. Kniewel from CSIC, Madrid, Spain

J. Bohrlinghaus from Institute of Plant Physiology, RWTH Aachen University, Germany

Panke Group at ETH Zürich, Basel, Switzerland

M. Salvadore de Lara from University of Surrey, England