Technische Universität München

TUM School of Life Sciences

Novel components of acquisition of competence for natural transformation in *Actinobacteria*

Enzo Joaquín Torasso Kasem

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (*Dr. rer. nat.*)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Kay Heinrich Schneitz

Prüfer der Dissertation: 1. Prof. Dr. Wolfgang Liebl

2. Prof. Dr. Johan Philipp Benz

Die Dissertation wurde am 25.02.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 05.07.2022 angenommen.

Table of Contents

Figures Index

Figures Index

Tables Index

Acknowledgements

I wanted to dedicate this work and specially thank…

To the DAAD (German Academic Exchange Service) for giving me the possibility of fulfilling my dream of doing my Doctorate thesis abroad, and for the opportunity of having one of the most rewarding experiences of my life.

To Professor Dr. Wolfgang Liebl, for his unconditional help and for always offering me everything within his reach to solve any situation that could have arisen during my years in Germany and after. He was all a Ph.D. tutor was supposed to be and I am really thankful for that.

To the second examiner of this thesis, Prof. Dr. Johan Philipp Benz, and the chair of the examination committee, Prof. Dr. Kay Heinrich Schneitz, for their time and dedication.

To Dr. Angel Angelov for his patience, leadership and precise advice that allowed me to learn and progress a lot during my work.

To Anthony Lichev, for his help and support during the years we worked side by side. Thank you for your disinterested friendship and guidance.

To María Überlacker, for her permanent good predisposition and her kindness, for being always aware of our needs at the lab and ready to help.

To all my collegues at the lab 002: Dr. María Cecilia Rasuk, Dr. Sonja Vanderhaeghen, Dr. Matthias Mörch, Dr. Maximilian Surger, Dr. Trang Pham, Yajing Liu and Luis Serrano.

To the bachelor and master students that went through the lab during my years of work: Constanza Oszwald, Pablo Valiente, Elisa Werner, Helena Ortiz-Cañavate, Alberto Salas Ruiz and Lidia Menéndez.

To the teams and employees at the Microbiology chair, to Dr. Zverlov, Dr. Ehrenreich, Melanie Baudrexl, Regina Rettenmaier, Nills Thieme, Martin Bimmer, Kornelia Garus, and Sibylle Schadhauser.

To my dear friends Jimena, Betiana, Joel, Pamela, Carlita and Gabriel who were always there for me when I needed them.

To all my family, but specially to my parents and siblings, that have always accompanied me in my dreams, loved me and never had doubts about what I could achieve.

Abbreviations & Acronyms

Table 0.1. General abbreviations

Abbreviation	Meaning
°C	Degrees Celsius
µ-	Micro (10^{-6})
A	Ampere, base unit of electric current.
AA	Aminoacids
Bp	Base pairs
CFU	Colony forming unit
CSP/F	Competence stimulating peptide/factor
CTAB	Cetyltrimethylammonium bromide
Da	Dalton (unified atomic mass unit)
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
ds-	Double stranded
dTTP	Deoxythymidine triphosphate
DU	DNA uptake
EtOH	Ethanol
F	Farad, unit of electrical capacitance
g	Gram
GMM	Glutamate minimal medium
GTAs	Gene transfer agents
HR	Homologous recombination
HGT/LGT	Horizontal/lateral gene transfer
HTH	Helix-turn-Helix motif
ICEs	Integrative conjugative elements
K-	Kilo (10^3)
Kan	Kanamycin resistance cassette
KS	K-state

Abbreviations & Acronyms

L	Liter
LB	Lysogeny broth
m	Meter
M	Molar (mol L^{-1})
m-	Mili (10^{-3})
MW	Molecular weight
n	Number
n-	Nano (10^{-9})
NGT	Natural genetic transformation
NC	Natural competence
O/N	Overnight
OD	Optical density
ORF	Open reading frame
-P	Phosphorylated
PAIs	Pathogenicity Islands
PCR	Polymerase chain reaction
pH	Negative decimal logarithm of protons concentration
QS	Quorum sensing
RNA	Ribonucleic acid
rpm	Revolutions per minute
RR	Response regulator
s	Second
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sp.	Species
ss-	Single stranded
ssp.	Subspecies
T	Time
T°	Temperature
TCS	Two component system
tDNA	Transforming DNA
Tfp	Transformation pilus
V	Volt (unit for electric potential)
x	Times/ multiple of

Abbreviations & Acronyms

x g	Times 9.8 m/s^2 (the acceleration due to gravity)
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
comEA/EC	*comEA-comEC* transcription unit
Ω	Ohm (unit for electrical resistance)

Summary

Natural transformation, a process by which a DNA molecule is internalized into a recipient cell and integrated into its genome, is one of the most common ways of horizontal gene transfer. Some organisms, among them several human pathogens, can perform this naturally under specific circumstances, depending on a specific physiological state called competence. Most of the existent knowledge about natural transformation has been unveiled in many Gram-negative and a few Gram-positive bacteria. Curiously, not much is known in any member of the *Actinobacteria* phylum despite the fact that this phylum contains numerous human pathogens as well as biotechnologically important representatives. This study focuses on the use of the ubiquitous bacterium *Micrococcus luteus* as an actinobacterial model organism to study this process in greater detail. It is hypothesized that the transformation apparatus and also the regulatory system for induction of competence are different in this organism and its phylogenetically close relatives compared with other well-studied model bacteria. Among the described genes participating in natural transformation in previously studied bacteria, only the marker genes *comEA* and *comEC* are present in *M. luteus*. The major goal of this work was to find previously unreported genes or proteins taking part in the regulation of *comEA/EC*. For this, random chemical mutagenesis was carried out in a *comEA/EC-lacZ* reporter strain, with the subsequent screening for dysregulated mutants. Two mutant groups were isolated, the first with an upregulated *com* promoter under competence-repressive conditions and a second with repressed *com* promoter in a competence-inducing environment. Following an intensive phenotype and genotype screening, the genomes of selected mutant strains were fully sequenced to identify point mutations that could be responsible for the observed mutant phenotypes. Interesting putative causative mutations were cleanly re-inserted into the genome of the *M. luteus comEA/EC-lacZ* reporter strain. In this way, it was possible to verify which of them indeed generated the evaluated phenotypes. This approach allowed the identification of new loci potentially involved in *comEA/EC* regulation. Transformation frequencies were altered by several orders of magnitude when genes at these genetic loci were altered, showing their importance for the process of competence development. Certain genes at investigated loci were found to be similar to others belonging to the already described *tad* cluster of *M. luteus* and other naturally transformable bacteria.

Zusammenfassung

Natürliche Transformation, ein Prozess, bei dem DNA in eine Empfängerzelle aufgenommen und dort in das Genom integriert wird, ist eine der wichtigsten Möglichkeiten des horizontalen Gentransfers. Einige Organismen, darunter mehrere Krankheitserreger des Menschen, können natürliche Transformation unter bestimmten Umständen auf natürliche Weise ausführen, wenn sie sich in einem bestimmten, als Kompetenz bezeichneten physiologischen Zustand befinden. Das heutige Wissen über natürliche Transformation wurde Großteils bei diversen Gram-negativen und einigen Gram-positiven Bakterien enthüllt. Seltsamerweise ist bei keinem Mitglied des Phylums *Actinobacteria* darüber viel bekannt, obwohl diese Entwicklungslinie zahlreiche humane Krankheitserreger sowie biotechnologisch wichtige Vertreter enthält. Die vorliegende Studie konzentriert sich auf die Nutzung des ubiquitär verbreiteten Bakteriums *Micrococcus luteus* als Modellorganismus der *Actinobacteria*, um diesen Prozess genauer zu untersuchen. Es wird vermutet, dass der Transformationsapparat und auch das Regulationssystem zur Kompetenzinduktion bei diesem Organismus und seinen phylogenetisch nahen Verwandten sich unterscheidet von anderen gut untersuchten Modellbakterien. Von den beschriebenen Genen, die in anderen Bakterien an natürlicher Transformation beteiligt sind, sind in *M. luteus* nur die Markergene *comEA* und *comEC* vorhanden. Das Hauptziel dieser Arbeit war es, bisher unbekannte Gene bzw. Proteine zu finden, die an der Regulation von *comEA/EC* beteiligt sind. Dazu wurde in einem *comEA/EC-lacZ*-Reporterstamm eine zufällige chemische Mutagenese mit anschließendem Screening auf dysregulierte Mutanten durchgeführt. Dabei konnten zwei Gruppen von Mutanten isoliert werden, die erste mit einem hochregulierten *com*-Promotor unter kompetenzunterdrückenden Bedingungen und eine zweite mit reprimiertem *com*-Promotor in einer kompetenzinduzierenden Umgebung. Nach einem intensiven Phänotyp- und Genotyp-Screening wurden die Genome ausgewählter mutierter Stämme vollständig sequenziert, um Punktmutationen zu identifizieren, die für die beobachteten mutierten Phänotypen verantwortlich sein könnten. Interessante, mutmaßlich für die Phänotypen ursächliche Mutationen wurden als Einzelmutationen rückstandslos in das Genom des *M. luteus* comEA/EC-lacZ-Reporterstamms eingebracht. Auf diese Weise konnte überprüft werden, welche von ihnen tatsächlich zu den beobachteten Phänotypen geführt haben. Dieser Ansatz ermöglichte die Identifizierung neuer Loci, die möglicherweise an der *comEA/EC*-Regulation beteiligt sind. Die Transformationshäufigkeiten waren um mehrere Größenordnungen verändert, wenn Gene an diesen Genorten mutiert wurden, was ihre Bedeutung für den Prozess der Kompetenzentwicklung aufzeigt.

In den untersuchten Loci wurden Gene gefunden, die anderen ähnlich sind, die zu dem bereits beschriebenen tad-Cluster von *M. luteus* und anderen natürlich transformierbaren Bakterien gehören.

Publications

- **Torasso Kasem, E. J., Angelov, A., Werner, E., Lichev, A., Vanderhaeghen, S., & Liebl, W.** (2021). Identification of New Chromosomal Loci Involved in com Genes Expression and Natural Transformation in the Actinobacterial Model Organism *Micrococcus luteus*. *Genes*, *12*(9):1307. **https://doi.org/10.3390/genes12091307**

- **Torasso Kasem, E. J., Angelov, A., Werner, E., Lichev, A., & Liebl, W.** (2020). First hints of main regulatory elements of natural competence in *Actinobacteria*. International Conference on "Microbial Stress" hosted by the European Federation of Biotechnology. Held online due to Covid pandemic, 16-18.11.2020.

1 Introduction

1.1 Horizontal gene transfer

Among several processes through which organisms increase genetic diversity, horizontal or lateral gene transfer (HGT) has been found to play a significant role. This group of mechanisms refers to the ways of unidirectional flow of genetic information, regardless of the usual mating barriers, between more or less distantly related organisms and, thus, stands in distinction to the standard vertical transmission of genes from parents to offspring (Keeling & Palmer, 2008). Beginning in the 1950s, the evidence showing the impact of HGT on the evolution of prokaryotes and eukaryotes has increased decade after decade (Davies, 1995). The common hypothesis was that the influence in multicellular complex eukaryotes was limited due to the barrier of the germline in animals and apical meristem in plants, but the data gathered were in contradiction to this (Yue, Hu, Sun, Yang, & Huang, 2012). HGT seems to have exerted an influence over processes as crucial as cellular differentiation or defense mechanisms against both biotic and abiotic agents (Alcázar *et al*., 2010). In the case of *Bacteria* and *Archaea*, some solid data suggests that the role of HGT in genetic diversity was even of greater importance than the one of single point mutations accumulation. It is difficult to account for the ability of prokaryotes to exploit new environments by the accumulation of point mutations alone (Ochman *et al*., 2000; Treangen and Rocha, 2011). Prokaryotes and other unicellular organisms are normally under different selective pressures regarding DNA exchange than most eukaryotes. They experience recombination much less frequently than they reproduce, and the quantity of the DNA exchanged is small. Under these circumstances, there is little selective advantage in preventing such rare interspecific exchange. If a novel gene arises that provides a selective advantage, this new trait can be shared between unrelated organisms through HGT (Gogarten and Townsend, 2005).

A representative human health threatening example of a harmless bacteria becoming a serious concern through HGT, was the case of the Shiga-toxin–producing enterohaemorrhagic *E. coli* O104:H4. This strain caused severe gastroenteritis during a German outbreak in 2011. This strain had integrated the virulence factors of two different pathogens in its genome, i.e. Shiga-toxin-producing *E. coli* and enteroaggregative *E. coli* (Bielaszewska *et al*., 2011). Other studies on the topic found out that there is substantial evidence that HGT contributes to the dissemination of antibiotic resistance, virulence traits, and vaccine evasion mechanisms (Ochman *et al.*, 2000; Croucher *et al.*, 2016). Virulence genes acquired through HGT are integrated and maintained as pathogenicity islands (PAIs) in bacterial genomes (Seitz & Blokesch, 2013). In light of all the findings from the last decades, our view of

evolution development has changed drastically. From a branched tree of life to a complex and not yet fully understood net/web of life that connects phylogenetically distant organisms of all three domains.

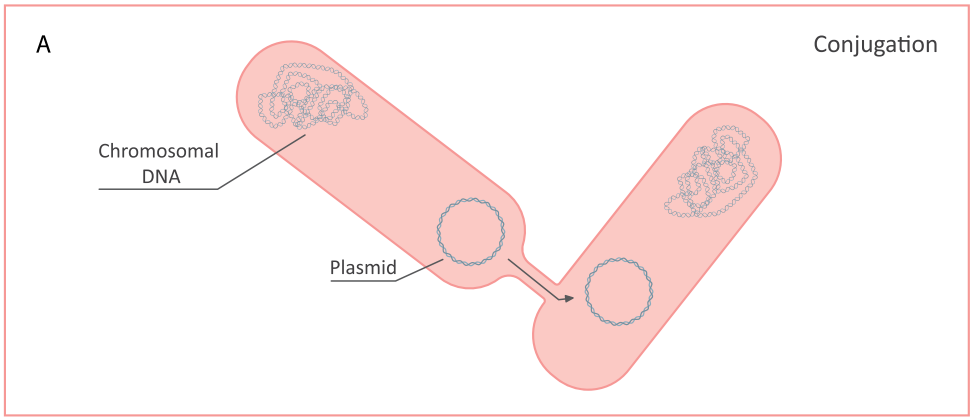
1.2 Horizontal gene transfer in *Bacteria*

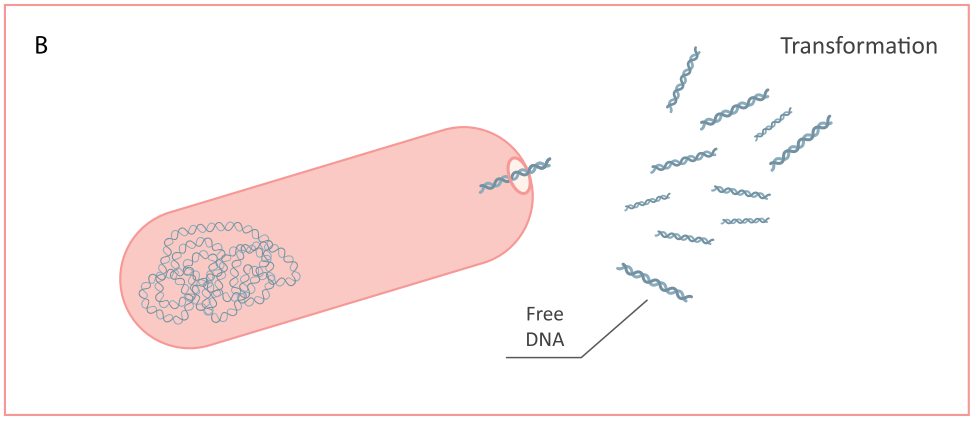
1.2.1 Conjugation

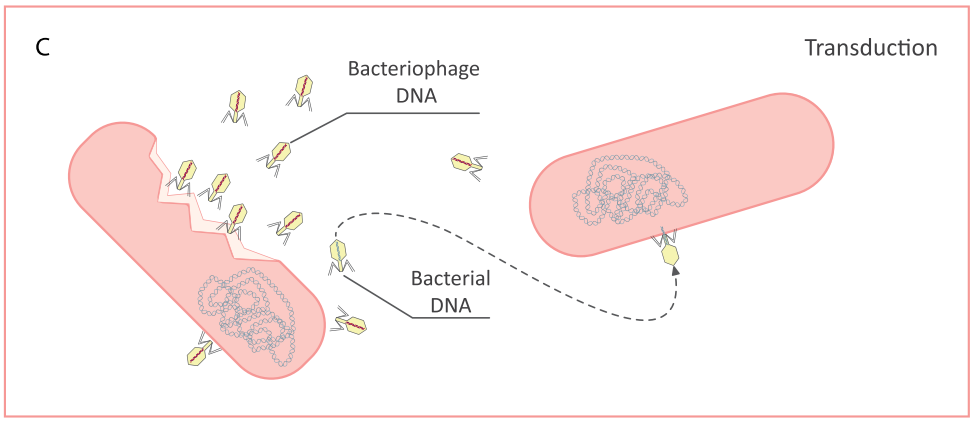
One of the most broadly studied methods of HGT is bacterial conjugation (see Figure 1.1 A), which not only can happen between two unicellular prokaryotes from the same or different species, but also between one of them and a multicellular eukaryote. Conjugation needs a direct physical link between the donor and the recipient cell. The donor cell generates a macromolecular proteinaceous apparatus that builds a tunnel-like pilus that connects with the recipient cell. The genetic material to be transmitted is processed to become ssDNA and is later on reconverted to dsDNA by the replication machinery of the recipient cell. The conjugative machinery is encoded either by genes on autonomously replicating plasmids or by integrative conjugative elements (ICEs) in the chromosome (Cabezón, Ripoll-Rozada, Peña, de la Cruz, & Arechaga, 2015; Wozniak & Waldor, 2010).

1.2.2 Transduction

Among all organisms in the microbiome of any environment, bacteriophages play a significant role in the distribution of many genetic traits. They can serve as shuttles for genes that are advantageous to their hosts, thus promoting their survival and dissemination. The process of transfer of one host's genetic material to another one through a viral particle is known as transduction (see Figure 1.1 C). This process represents one of the most common ways of transfer of antibiotic resistance genes or PAIs (Fortier & Sekulovic, 2013; Von Wintersdorff *et al*., 2016). During a viral "life" cycle, some new particles acquire DNA sequences from the host instead of, or in addition to, the phage genome. Two sorts of transductions have been described in great detail: the specialized and the generalized transduction. In the first one, one virion obtains a specific region of the bacterial chromosome in its capsid, usually one of the flanking sequences near the prophage insertion position. For instance, phage λ occasionally packages galactose or biotin metabolism genes located right next to its insertion site in *E. coli* chromosome. When generalized transduction occurs, some phages like P1 of *E. coli* nonspecifically package random sections of the bacterial DNA, even without any sequence of the viral genome along with it. These particles transfer the DNA to the next host and according to the particular sequence there is a chance to be integrated into the chromosome (Lang *et al.*, 2013).


A
Conjugation
Chromosomal DNA
Plasmid


B
Transformation
Free DNA


C
Transduction
Bacteriophage DNA
Bacterial DNA

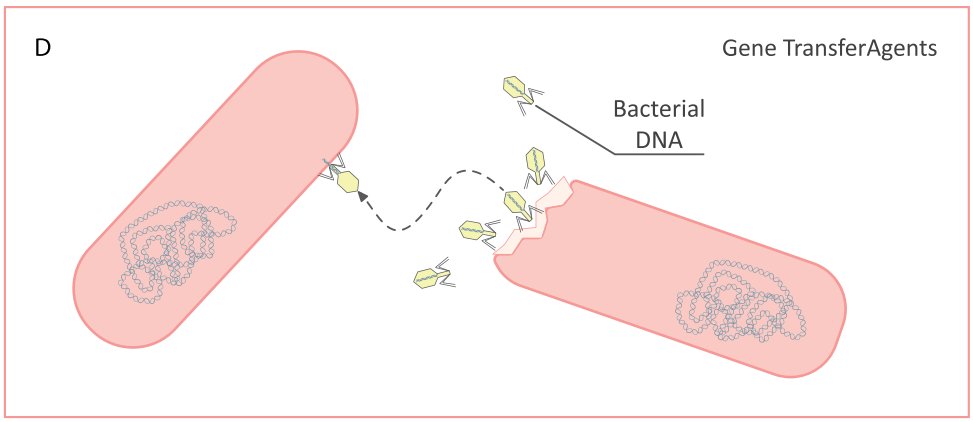

D
Gene TransferAgents
Bacterial DNA

Figure 1.1 Mechanisms of horizontal gene transfer in bacteria.
Each box shows one distinct process of horizontal gene transfer. **(A)** Conjugation is a process requiring cell to cell contact via cell surface pili or adhesins, through which DNA is transferred from the donor cell to the recipient cell. **(B)** Transformation is the uptake, integration, and functional expression of naked fragments of extracellular DNA. **(C)** Through specialized or generalized transduction, bacteriophages may transfer bacterial DNA from a previously infected donor cell to the recipient cell. During generalized transduction, bacterial DNA may be accidentally loaded into the phage head (shown as a phage with a blue DNA strand). During specialized transduction, genomic DNA neighboring the prophage DNA is co-excised, loaded into a new phage and integrated in the genome of the next infected bacterium. **(D)** Gene transfer agents (GTAs) are bacteriophage-like particles that carry random pieces of the producing cell's genome. GTA particles may be released through cell lysis and spread to a recipient cell.

1.2.3 Gene transfer agents (GTAs)

First discovered in *Rhodobacter capsulatus* in 1974 (Marrs, 1974), GTAs are gene delivery systems harnessed by host cells that contain and sometimes regulate the genetic elements to build up these shuttle particles (see Figure 1.1 D) They have been found in both bacteria and archaea and are very much like bacteriophage particles that are believed to have evolved from one ancestor that lost its capability of targeting its DNA for packaging. Thus they carry small random sections of the host genome to nearby new hosts. GTAs are not able to transfer all the genes that encode themselves to a new host, which represents an important difference compared with transduction mechanisms and opening the still unanswered question of how these genes remain under selection for function. Transfer through GTAs has been documented between members of different phyla and some studies show that not all cells, including those that contain GTA encoding genes, are "competent" to receive the donated genes (Lang *et al.*, 2013; Soucy *et al.*, 2015). One study suggests that the cell recipient capability requires the expression of essential competence genes known to be involved in transformation (see next section for details), thus meaning that GTAs may combine main features of transduction and transformation, being the homologous recombination mediated by DprA/RecA (Brimacombe *et al.*, 2014; Brimacombe *et al.*, 2015). Another characteristic of GTAs is that even though they are released through cell lysis, bacterial cultures do not display observable cell disruption. This is because only a subpopulation of cells (~3%) generates ~95% of the GTAs in a population (Von Wintersdorff *et al*., 2016).

1.2.4 Natural transformation

Natural transformation is a parasexual process through which a cell acquires DNA from the surrounding environment and integrates it into its genome through homologous recombination (see Figure 1.1 B). This is different from, and should not be confused with the more widely known artificial transformation techniques for bacteria, carried out often in most molecular biology laboratories i.e. heat shock transformation or electroporation. Natural competence for transformation is a specific

physiological state that is genetically encoded by the *com* regulon (Blokesch, 2016). Since its discovery by Griffith in 1928 (F. Griffith, 1928), many bacterial species from several taxonomic groups have shown to bear this trait, but the question of what is the primary purpose of this process is still a matter of discussion among experts. The first hypothesis states that the main reason for DNA absorption from the environment serves a nutritional requirement. This is supported by many studies that show how different species reach competence once they encounter nutrient-limiting conditions in high cell density cultures. The acquisition of ready-to-use nucleotides can alleviate the metabolic stress in such a context. However, DNA-uptake of specific recognized sequences, as it occurs in certain organisms, as well as the competence-induced protection of internalized ssDNA do not fully support this idea. As in other HGT processes, internalized DNA could be also harnessed as a source of novel genetic information, that upon recombination can lead to the acquisition of new genes or operons. A third hypothesis would be that the internalized DNA molecules can be used as templates for chromosomal DNA repair of bacteria. Most probably, there is not only one reason to keep this trait among species, and they likely vary from one organism to another and are dependent on the population context.

So far, most of the studies aiming to unveil the molecular mechanism behind natural competence and transformation were carried out on pathogens and model organisms. Generally, transformation occurs in a more or less conserved fashion. In Gram-negative bacteria, the DNA molecule binds to the cell and enters into the periplasm, whereas in Gram-positive bacteria, the DNA uptake requires transversal translocation through the thick cell wall. In almost all transformable bacteria documented, DNA uptake uses type IV pili (T4P) or a pilus-like structures as a binding anchor, with one notable exception described in the bacteria *Helicobacter pylori* which uses a type IV secretion system for DNA uptake. In Gram-negative bacteria, the PilQ secretin channel enables the pilus to cross the outer membrane. The DNA uptake pilus retracts and introduces the transforming DNA (tDNA) to the DNA-binding competence protein ComEA. This protein has different names or topological features depending on the organism. Subsequently, the tDNA finds the transmembrane channel protein ComEC, through which it enters the cytosol. According to what has been studied so far, all transformable species harbor the transmembrane channel ComEC, and a DNA binding protein such as the already mentioned ComEA (Johnsborg *et al.*, 2007). While being driven through the cytoplasmic membrane, one strand of the tDNA is degraded and the other is internalized in a single-stranded form in a 3'–5' orientation. ssDNA is bound by the transformation-dedicated DprA (DNA processing A) protein, which loads the recombinase RecA. RecA polymerizes on the ssDNA and promotes a homology search in the host chromosome. When a homologous sequence is found, RecA-driven homologous recombination occurs, which involves the pairing of the ssDNA with a complementary

strand of homologous chromosomal DNA to form a transformation heteroduplex. These duplexes can be fully homologous or can contain heterologous DNA (such as PAIs), which remain in single-stranded form and are flanked by the homologous regions. The donor and recipient strands are resolved by replication or in some cases by mismatch repair. The nature of the tDNA limits the usefulness to the acquiring cell because homologous recombination is necessary for a successful integration event. Therefore, the more closely related the tDNA is to the genome of the competent cell, the more likely a productive recombination event will occur (Shanker & Federle, 2017). Lastly, the competent state is reversed, a necessary step in bacteria in which the state of competence imposes growth arrest.

The upcoming pages are focused on the details regarding the regulatory pathways that trigger the previously mentioned genes that carry on DNA uptake. For a better insight on the molecular mechanisms that drive DNA uptake there are several comprehensive reviews, being the most recent Blokesch and Dubnau, 2019; Muschiol *et al.*, 2019; Johnston *et al.*, 2014; and Claverys *et al.*, 2009. Even though most of the studied organisms share the basic DNA uptake machinery, it is noteworthy that their competence-triggering mechanisms differ significantly according to their environmental contexts and particular niches.

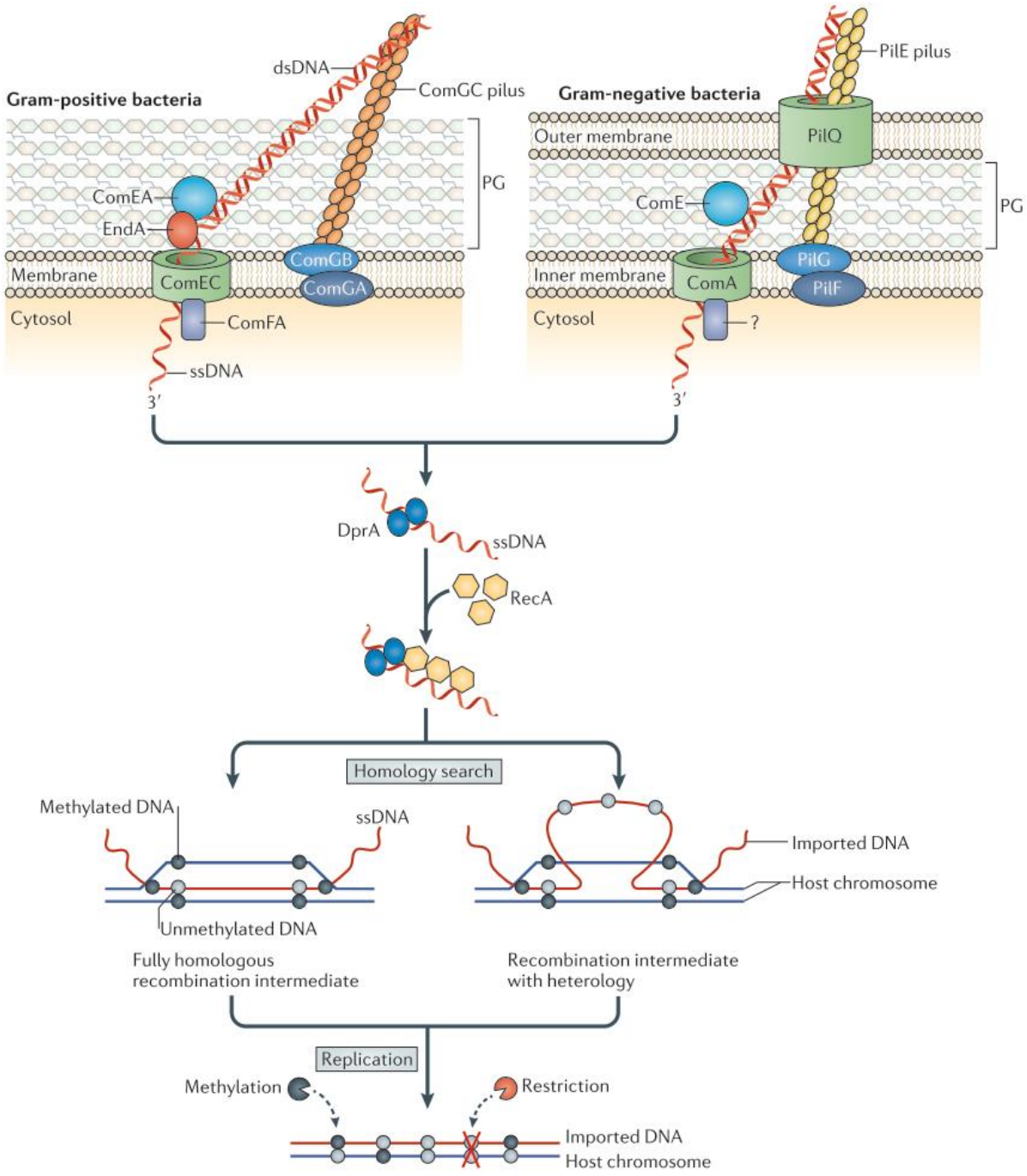


Figure 1.2 An overview of the natural transformation process

Key aspects of transformation in Gram-positive and Gram-negative species are represented. The DNA-uptake machinery generally comprises a transformation pilus (Tfp) that captures exogenous double-stranded DNA (dsDNA), the DNA receptor ComEA and the transmembrane pore ComEC. In *Streptococcus pneumoniae*, the EndA nuclease receives DNA from the DNA receptor ComEA and degrades one DNA strand, while unidentified nucleases (or strand-separating proteins) generate single-stranded DNA (ssDNA) for uptake in other species. In Firmicutes, ssDNA internalization through ComEC is presumably driven by the ATP-dependent translocase ComFA. In Gram-negative bacteria, such as *Neisseria gonorrhoeae*, the PilQ secretin channel enables the pilus (which is mainly composed of PilE subunits) to cross the outer membrane and dsDNA is transported across the outer membrane through PilQ. In both Gram-positive and Gram-negative cells, additional proteins are required for DNA uptake (for example, the ComGA and ComGB proteins of Firmicutes). Internalized ssDNA is presumably bound by DprA (DNA processing protein A), which recruits the recombinase RecA. RecA polymerizes on ssDNA

and promotes a homology search along chromosomal DNA, followed by strand exchange. The transformation heteroduplex that forms can be a fully homologous double-stranded recombination intermediate, or if the imported DNA contains heterologous sequences (such as a pathogenicity island) flanked by homology, a recombination intermediate with a single-stranded loop is formed. If heterologous donor DNA is unmethylated (light grey circles), this DNA remains fully unmethylated in the recipient chromosome after replication. The methylation and restriction activities of the restriction–modification (R–M) system compete (dashed arrows) for access to this sensitive DNA, and restriction can kill transformants and limit heterologous transformation. PG, peptidoglycan (Johnston *et al.*, 2014).

1.3 Natural transformation in Gram-negative bacteria

Most of the knowledge about the regulation of competence in Gram-negative organisms was obtained from human pathogens. This is because most research on natural genetic transformation has been influenced by medically important questions such as how virulence factors are acquired or how antibiotic resistance genes are spread between species. Another reason is the availability of whole-genome sequences from many different pathogenic isolates, which is not always the case for harmless environmental bacteria. Having different genome sequences from one species facilitates the identification of horizontally transferred regions and the regulatory paths behind these processes.

In some Gram-negative bacteria, natural competence is a transient and tightly regulated process, although it occurs constitutively in others. After almost 100 years of research on the topic, a correlation between the diversity encountered in an organism´s niche and the complexity of the regulatory network driving competence induction was drawn. The hypothesis states that the greater the fluctuations in the environmental conditions, the more complex and carefully regulated is competence acquisition. Examples of this are, on one side the pathogen *H. pylori* whose only reported reservoir is the epithelium of the human stomach. This niche does not normally present sudden and/or dramatic changes in pH or nutrient availability, but in the case of long-term infection of the host, natural transformation can help beneficial mutations to spread and provide a means for adaptation. On the other side, a good example is the aquatic pathogen *Vibrio cholerae*. This pathogen´s environmental niche is constantly fluctuating, and once the bacterium enters a host, conditions are also changed dramatically. *Vibrio sp.* can be found in a planktonic state or as a sessile biofilm community. Key environmental conditions such as carbon amount, source, and temperature vary frequently. Thus, the adaptation between famine and feast and between growth and growth arrest including the ability to become competent must be fine-tuned (Seitz & Blokesch, 2013).

As mentioned, *Helicobacter pylori* is an important human pathogen. This bacterium can be detected in the gastric epithelia of more than half of the world´s population and is believed to have been there since humans migrated out of Africa around 60 000 years ago. One interesting feature of this species is that it represents one of the most diverse bacterial species isolated up to date. High

frequencies of recombination events and flow of genetic material were crucial to reach this outstanding diversity, and these processes are directly linked with its ability to carry on natural transformation. However, in comparison to most other Gram-negative transformable bacteria, *H. pylori* differs in two important aspects: the DNA-uptake mechanism and its constitutive competence state. Most competent bacteria rely on type IV pilus-related DNA-uptake machinery. In *H. pylori*, the components forming the DNA translocation apparatus resemble a type IV secretion system (T4SS), called ComB system. After entry into the periplasm, tDNA binds to a unique DNA receptor protein, ComH, which is not an orthologue of the conserved periplasmic protein ComEA present in other organisms. It´s believed that ComH delivers the tDNA to the transmembrane channel ComEC (Damke *et al*., 2019; Seitz & Blokesch, 2013). Regarding the question about when *H. pylori* cells become competent, the organism is often described as being constitutively in such a state. However, it is important to mention that this conclusion was drawn after a study in which peaks of high transformability were observed at different growth phases, but in between these, there were periods of lower transformability too. This suggests that some sort of regulation must be occurring at a certain level in this species. The only proven external stimulus that seems to trigger an increase in natural transformation is DNA damage. There is still further work needed before making final conclusions about how competence acquisition happens in this organism. The other human pathogens that have been reported to have a constitutive state of competence are *Neisseria gonorrhoeae* and *N. meningitidis*, whose transformability was already reported more than 50 years ago. Both mentioned species have shown to bear similar pathways of competence regulation and, thus, many findings derived from one organism can readily be applied to the other one. As for *H. pylori*, for *N. gonorrhoeae* no other environmental niche is known other than a mucosal tissue, in this case of the human urogenital tract (Johnston, Martin, Fichant, Polard, & Claverys, 2014).

One of the latest Gram-negative human pathogens to be proven naturally competent is *Vibrio cholerae* (Meibom, Blokesch, Dolganov, Wu, & Schoolnik, 2005). In nature, this bacterium can be found as a planktonic organism in the water, in the mucilaginous sheaths of blue-green algae, and on the chitinous exoskeletons and moths of copepods. It attaches to and degrades chitin (a β-1,4-linked N-acetylglucosamine (GlcNAc) polymer), one of the most abundant biopolymers in nature. Besides being a source of nutrients, chitin also induces natural competence for transformation in *V. cholerae* and other *Vibrio* species. Chitin leads to the up-regulation of *tfoX* which is translated to the main competence activator TfoX, a transcriptional co-regulator essential for activation of the *com* regulon. The mechanism through which $GlcNAc_2$ activates *tfoX* transcription is yet not clear. However, a 102 bp sRNA -TfoR- is known to modulate TfoX translation by direct binding to its mRNA, enabling the

ribosome to find the RBS which is otherwise sequestered by a secondary structure (Yamamoto *et al*., 2011, 2014). Besides this, other pathways like quorum sensing (QS) signals and carbon catabolite repression (CCR) are also necessary to trigger the competence genes of *V. cholerae*. Two autoinducers are known to be produced and secreted in this organism: the intra-species cholera autoinducer 1 (CAI-1) and the universal autoinducer 2 (AI-2). When high cell densities are reached, the concentration of autoinducers is enough to promote the production of HapR, the master regulator of QS known to regulate virulence, biofilm formation, and natural competence. HapR regulates natural transformation by repressing *dns* and concomitantly activating *qstR*. Dns is a nuclease that upon DNA degradation in the surrounding environment prevents its uptake, and QstR is a transcription factor that directly activates some genes of the *com* regulon. Recent findings indicate that there seem to be one or more still undetected intermediates for the activation of certain essential genes like *comEA* (Jask, Stutzmann, Stoudmann, & Blokesch, 2018). In addition to these two independent pathways of natural transformation promotion, a third one occurs. CCR is a mechanism by which, in the presence of a preferred carbon source such as glucose, the expression of the genes necessary for metabolizing other carbon sources is repressed. Four main components take part in this pathway: the phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS), adenylate cyclase (CyaA), the metabolite 3',5'-cyclic adenosine monophosphate (cAMP), and the CRP protein. Under low PTS sugars concentrations, the unsaturated PTS transporters enhance the synthesis of cAMP by CyaA. High concentrations of cAMP trigger the binding with CRP to form the cAMP-CRP complex, which binds to and activates the promoters of target genes like the ones involved in the transport and usage of alternative carbon sources. Regarding natural transformation, the presence of PTS sugars significantly decreases the transformability of *V. cholerae*, to the point that knockout strains of *crp* or *cyaA* are not transformable. CRP forms dimers that bind two cAMP molecules and recognize the CRP sites (22 bp length symmetrical sequences) at the target promoters. Based on work done in *Haemophilus influenzae*, which has a very similar set of regulatory mechanisms, it was hypothesized and finally proven that the CRP-cAMP complex binds to CRP-S sites at the target promoters along with TfoX (Scrudato *et al*., 2014). This is an example of an intricated regulatory network, that requires more than one originally independent signal to act together to finely tune all genes required for competence and natural transformation (Johnston *et al*., 2014).

Nutrient starvation is a common trigger of competence development, although which nutrients specifically do that varies from species to species. *H. influenzae* has been shown to increase transformability in response to nucleotide pools, more specifically purines. A depletion of adenine and/or guanine has been found to activate translation of *sxy* mRNA, which encodes the central

competence regulator Sxy, an orthologue of TfoX. Regulation of this protein's translation is dependent on a stem structure in the mRNA (Cameron *et al.*, 2008; Sinha *et al.*, 2013). This is not that clear in *V. cholerae,* in which the absence of glucose induces competence through CCR but the presence of chitin does it too.

1.4 Natural transformation in Gram-positive bacteria

The field of competence development for natural transformation and its regulation was born around 90 years ago after studies were carried out on the Gram-positive human pathogen *Streptococcus pneumoniae* (F. Griffith, 1928). After Dawson and Sia (1931) managed to transform the "pneumococcus", the details of the transformation mechanisms and the regulation of their genes have slowly been unraveled in different bacteria. Up to date, most of the documented naturally competent bacteria are Gram-negative, however, the best understood organisms at the transformation machinery and regulatory levels are the low G+C Gram-positive Bacteria which comprise the phylum *Firmicutes*: *B. subtilis* and members of *Streptococcus* genus like *S. pneumoniae*, *S. mitis*, and *S. mutants* (Johnsborg *et al.*, 2007; Shanker and Federle, 2017; Underhill *et al.*, 2019). Studies in *B. subtilis* and *S. pneumoniae* showed a growth-stage dependence in both organisms. Competence seems to be inhibited in the stationary phase in *S. pneumoniae* while activated in the exponential phase in nearly all cells of the population; it is maintained for up to 40 min and then it decays almost as fast as it arose. In *B. subtilis* competence develops at the onset of the stationary phase and typically is observed in only part of the population (Claverys *et al.*, 2006; Salvadori *et al.*, 2019).

Competence regulatory pathways of the above-mentioned organisms have significant similarities. Both species depend on a small peptide pheromone for competence development. A dedicated two-competent regulatory system detects this pheromone and transmits a signal downstream that induces transcriptional activation of the *com* regulon. Although, when looking in detail, it is clear that despite these similarities the effectors of competence in these bacteria have not evolved from a common ancestor (Claverys and Martin, 2003; Claverys *et al.*, 2006).

1.4.1 Firmicutes

1.4.1.1 Streptococcus sp.

The competence factor that stimulates *S. pneumoniae*, CSP, is a 17-residue peptide that gets exported by ComAB, an ABC exporter/protease that cuts off a leader peptide after a Gly-Gly motif in the 45 residue pre-CSP encoded by *comC* (Håvarstein *et al.*, 1995). No other modifications to CSP are known to occur in *S. pneumoniae*, thereby distinguishing CSP from the analogous mutacin-inducing

peptide (MIP) in *S. mutans*, which is further processed extracellularly by the protease SepM. Extracellular CSP then binds to and activates the polytopic membrane-embedded cognate histidine kinase receptor ComD of the ComCDE two-component signal-transduction pathway. Activation of ComD results in autophosphorylation and the subsequent transfer of a phosphoryl group (~P) to the cytosolic response regulator ComE. Phosphorylated ComE (ComE~P) binds to a conserved promoter sequence, referred to as a ComE-binding site (Ceb), found upstream of the alternative sigma factor gene σ^X/ComX, for which there are two identical genomic copies within the *S. pneumoniae* genome. ComX activity is essential for the expression of many genes required for competence. This conserved Ceb element is also found near the *comAB* operon as well as the *comCDE* operon. ComE~P binds at the Ceb in the promoter region of these operons among others, to activate gene expression by recruiting RNA polymerase, resulting in the amplification of the core components of the CSP signal transduction pathway (Johnston *et al*., 2014; Shanker & Federle, 2017).

The induction of competence in *S. pneumoniae* is divided into two temporally distinct phases, the early and the late, regulated by ComE~P or ComX, respectively. Besides activating transcription of *comX*, *comAB*, and *comCDE*, ComE~P also activates transcription of at least 17 other genes, several of which are required for the development of the competent state. The early gene (defined by the maximal expression between 7.5 and 10 min post-CSP treatment) *comW*, is important in natural transformation as it is required for stabilizing σ^X protein levels by preventing its degradation by the ClpP protease, a negative regulator of competence induction (Johnsborg & Håvarstein, 2009). Some early genes that contribute to transformation have shown to be dispensable, e.g. *comM*, a membrane protein that imparts cellular immunity towards lytic enzymes, or *blpAB* which encodes a bacteriocin transporter that shares homology with the CSP exporter, *comAB*. The expression levels of ComX-induced late genes peak approximately 12.5 to 15 min post-CSP induction. Although the late genes comprise over 80 genes, only 14 have been identified as essential for transformation. These required genes include those encoding the DNA-uptake machinery, like *comEA-comEC*, and the recombination machinery (also, referred to as the transformasome) (Peterson *et al.*, 2004; Claverys *et al.*, 2009; Shanker and Federle, 2017). In *streptococci*, the competence state depends upon the single master regulator σ^X, but not all species share the ComDE two-competent regulatory system. There is an alternative circuit that involves the transcriptional activation system ComRS. This system is induced by the expression of the ComS peptide, which is processed by a protease to produce the *comX*-inducing peptide (XIP). XIP initially accumulates extracellularly, but at a critical concentration is imported into the cell by the oligopeptide permease Opp and directly interacts with ComR, a Rgg transcriptional regulator. The comR-XIP complex activates transcription of *comX,* which promotes the

development of competence. The groups pyogenic, salivarius, bovis, and mutans harness this pathway. Except for of *S. mutans*, which utilizes both cascades, all species employ either ComDE or ComRS pathways to induce competence (Underhill *et al.*, 2018; Salvadori *et al.*, 2019; Underhill *et al.*, 2019).

1.4.1.2 Bacillus subtilis

Most genes involved in natural transformation were originally identified in *B. subtilis* by the work of Dubnau *et al.* (1987). 28 genes from the *com* regulon were detected after a systematic transposon screening, lately characterized by sequencing and subsequently identified in other bacteria. Gene and protein nomenclature varies between species, but *B. subtilis* nomenclature is the one used for general comprehensive reviews of the topic. ComEA in Gram-positives is a membrane-bound DNA binding protein localized in the periplasm that mediates between the DNA uptake pilus and the transmembrane channel ComEC (Johnston *et al*., 2014). Additionally, all seven genes within the *comG* operon are necessary for DNA binding, and ComG proteins allow tDNA to find the ComEA receptor (Chung & Dubnau, 1998; Provvedi & Dubnau, 1999). ComG proteins share a degree of similarity with those involved in the formation of T4P and T2SS from other Gram-positive and Gram-negative organisms. ComGA is an ATPase that is located on the inner face of the cell membrane and is required for assembly of the DNA uptake pilus; ComGB is a conserved protein with several membrane-spanning domains, believed to be involved in the DNA uptake pilus retraction process; ComGF is predicted to be an integral membrane protein; ComGC, -GD, -GE, and -GG are similar to type 4 pilins. Such pilin-like proteins are also required for type 2 secretion, but because they were not known to form a pilus, they have been called “pseudopilins”, a term adopted also for ComGC, -GD, -GE, and -GG. ComGC is the most abundant of the mature pseudopilins, therefore is referred to as the major pilin, and the other ones as the minor pilins. All of them are made as precursors and undergo proteolytic cleavage by the prepilin peptidase ComC, also required for translocation across the membrane. Finally, the proteins thiol-disulfide oxidoreductases BdbD/C are needed for the introduction of disulfide bonds in ComEC and between the major pilin at the membrane (David Dubnau & Blokesch, 2019). At least in *B. subtilis*, protrusion of the transformation pseudopilus seems to be limited, considering that a typical cell holds 40-100 monomers and, assuming that each monomer is around 1 nm in length, this would result in a structure with an overall length of 40-100 nm (compared with the average 55 nm thickness of the periplasm and cell wall). (Chen *et al.*, 2006; Johnston *et al.*, 2014). Nevertheless, up until now, no direct observation of this DNA uptake pilus has been published in the literature, in contrast to the already shown structures in *V. cholerae* and *S. pneumoniae*.

Besides the iconic ComEA and ComEC, other proteins also have shown to be important or indispensable for optimal performance of the transformasome in the low-GC Gram-positive bacteria. The ComFA protein, a DEAD-box helicase, and its co-expressed protein ComFC, whose exact function remains unknown, seem to have more than one task in transformation. ComFC holds a phosphoribosyl-transferase domain in the C-terminal portion and a C_2-C_2 zinc finger domain in the N-terminal region. There is evidence that both ComF proteins are implicated in ssDNA translocation across ComEC, and simultaneously serve as a connection to the recombination protein DrpA and the next step in the tDNA journey to the chromosome (Chilton, Falbel, Hromada, & Burton, 2017; Diallo *et al*., 2017).

All genes from *comE*, *comF*, and *comG* operons belong to the late competence genes and are regulated by the transcriptional activator ComK. In *B. subtilis*, under nutrient limitation, competence is induced in the early stationary phase when ComK binds to its promoter and creates an autostimulatory positive loop essential for late competence genes activation. Only around 10-15% of the population becomes competent, due to stochastic fluctuations in the cellular levels of ComK.

Because ComK also activates the expression of several dozen genes not needed for transformation (Berka *et al.*, 2002; Hamoen *et al.*, 2003), the ComK ON physiological conditions in this subpopulation is referred to as the K-state (KS). Besides competence, all cells in the KS are also known for being arrested in growth and cell division, giving them the capacity to resist exposure to cell wall-targeting β-lactam antibiotic penicillin, kanamycin, and the quinolone oxolinic acid (Jeanette Hahn, Tanner, Carabetta, Cristea, & Dubnau, 2015). One of the main proteins involved in growth arrest in the KS is the already mentioned ComGA. In addition to its function -in DNA uptake pilus elongation, ComGA blocks replisome assembly through a still unknown mechanism, it also binds to RelA, thus inhibiting hydrolysis of the alarmone (p)ppGpp. As a result of these increased alarmone concentrations, rRNA synthesis is inhibited and cells repress growth (David Dubnau & Blokesch, 2019; Haijema, Hahn, Haynes, & Dubnau, 2001).

In contrast to CSP of *S. pneumoniae*, in *B. subtilis* a 10 amino acid competence pheromone called ComX is produced which is post-transcriptionally modified (Magnuson *et al.*, 1994). Before being exported, the inactive 55-residues precursor of ComX is cleaved and then isoprenylated on a tryptophan residue by the transmembrane ComQ protein. The ComX extracellular concentration increases and it stimulates the histidine kinase ComP, which donates a phosphate to its cognate response regulator ComA (D. Dubnau, 1991). ComA belongs to the LuxA-like response regulator family, whose members have an helix-turn-helix (HTH) structure at the C terminus, a feature not shared with

S. pneumoniae ComE, which contains a distinctive, non-HTH DNA-binding domain. ComA~P activates transcription of *comS*, a small ORF embedded in the surfactin-production operon *srfA*. ComS then allows the accumulation of ComK by preventing its proteolysis, thus representing an example of posttranslational control of competence acquisition (Turgay and Hahn, 1998; Claverys *et al.*, 2006).

The KS regulation through ComK levels is controlled by a two-pronged system. On one side, there is a transcriptional control module of *comK* and, in parallel, a stability control module of its protein. There are seven proteins known to act at the transcriptional level: the RRs, ComA, and DegU; the transcriptional activators, ComK itself and PerR; and three transcriptional repressors, AbrB, CodY, and Rok. The stability control module, on the other hand, consists of the already mentioned effector ComS and the proteolytic complex of ClpP and ClpC that is targeted to ComK by direct binding to the adapter protein MecA. ComS plays a key role in the stabilization of ComK. This happens by displacement of ComK from the ternary complex it forms with MecA and ClpC., thus, preventing its ClpP-mediated degradation. Transcription of *srfA* is modulated negatively by CodY and positively by PerR. Once stabilized, ComK contributes directly, together with DegU, to its accumulation. Other effectors of competence regulation include the proteins and small molecules that convey or modulate cell-cell signals: ComX-ComQ-ComP and Rap/Phr (Claverys *et al.*, 2006).

The Rap/Phr QS system encompasses three operons, all of them code for one or more Rap proteins and a *rap*-associated *phr* gene right after them. This system, together with ComA (ComX QS pathway), plays an important role in the interlinking of adaptive processes in *B. subtilis*. Rap proteins antagonize response regulators involved in the control of sporulation, degradative enzyme production (DEP), and competence. RapA and RapE specifically dephosphorylate the sporulation response regulator Spo0F∼P; RapC and RapF inhibit DNA binding of competence response regulator ComA∼P; and finally, RapG acts similarly with DEP/motility response regulator DegU∼P. Rap proteins are inhibited by specific Phr pentapeptides that result from maturation and export of *phr*-encoded prepro-products. They act within the cell after import through the Opp oligopeptide permease. Expression of *phrCEFG* genes is under control of σ^H that binds to the internal promoters within the *rap-phr* operons (Omer Bendori, Pollak, Hizi, & Eldar, 2015). PhrC, also known as competence stimulating factor (CSF), and PhrF inhibit RapC and RapF respectively, relieving the antagonist effect of them on ComA∼P. ComA∼P itself activates the expression of *rapACEF*, thereby forming a complex regulatory network between various QS pathways (Lazazzera, Kurtser, Mcquade, & Grossman, 1999). This further illustrates the interdependence of the regulation of competence and other adaptive processes, as does the involvement in the regulation of the K-state of global regulators such as CodY, a GTP-binding protein that senses the intracellular GTP concentration as an indicator of nutritional conditions and

regulates the transcription of early stationary phase and sporulation genes; AbrB, a transition state regulator that represses many stationary-phase processes; PerR, an oxidative stress-response transcription activator; or Rok, which represses genes for cell surface and extracellular functions.

1.4.2 Actinobacteria

Some members of the high G+C Gram-positive *Actinobacteria* were among the first organisms studied to have the capacity to naturally take up DNA. However, little progress has been achieved so far apart from reporting a couple of species as being able to aquire natural competence (Gelbart & Juhasz, 1970; Kloos & Schultes, 1969; Norgard & Imaeda, 1978). Until now, only *Mycobacterium smegmatis* and *Micrococcus luteus* were described as naturally competent and proven to acquire tDNA through natural transformation (Angelov *et al*., 2015). Besides this, there have been reports of radioactive DNA being taken up in *Streptomyces* species, but no genetic transformation was observed (Roelants, Konvalinkova, Mergeay, Lurquin, & Ledoux, 1976).

Some of the competence marker genes, like *comEA* and *comEC*, are broadly present in the transformable actinobacterial species genomes, but almost none of the other competence genes known from other organisms were found (Angelov *et al*., 2015). This fact suggests that other novel mechanisms could be in control of transformation, especially at the regulatory level where greater diversity exists among transformable organisms. Studying bacteria that belong to this phylum-level lineage will most probably lead to the discovery of new transformable species, which may have completely novel pathways of competence regulation.

Curiously, the data published about natural transformation in actinobacterial species seems to have been forgotten in the scientific literature of the last decades. *M. luteus* was reported as naturally competent in several studies, however, it cannot be found listed as such in recent comprehensive reviews about the topic (Johnsborg *et al.*, 2007).

1.4.2.1 Micrococcus luteus

Previously called *M. lysodeikticus*, *M. luteus,* the type species of the genus *Micrococcus,* family *Micrococcaceae,* order *Actinomycetales* (Skerman *et al.*, 1980), is a spherically shaped, nonmotile, and non-spore-forming obligated aerobe (see Figure 1.3). Frequently found forming tetrads, this organism is ubiquitous, having been isolated from many and diverse environments, ranging from human skin to high altitude wetlands to marine sponges and 120 million years old fossils (Dib, Wagenknecht, Hill, Farías, & Meinhardt, 2010; Greenblatt *et al*., 2004; Kloos & Musselwhite, 1975; Santos *et al*., 2019).

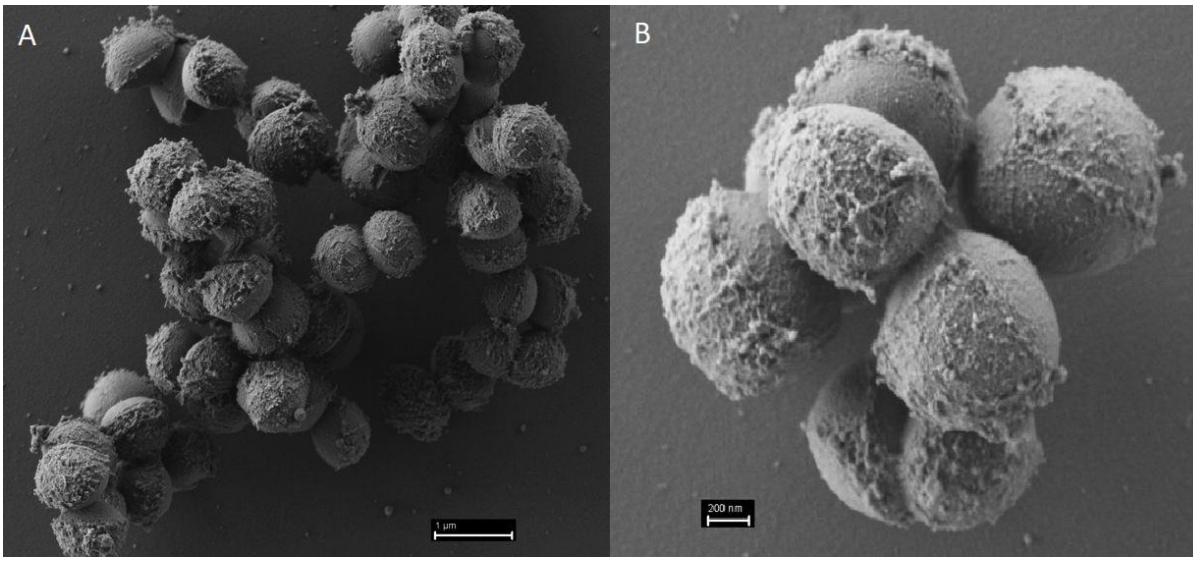


Figure 1.3 Electron micrographs (EM) of aggregated *Micrococcus luteus* cells.
(A) Scanning EM showing bacterial clusters. **(B)** Scanning EM showing single tetrad. This images were kindly provided by Prof. Dr. Wolfgang Liebl.

This strain is of historical importance since Fleming used it to demonstrate bacteriolytic activity -due to lysozyme- in a variety of body tissues and secretions (Fleming, 1922). But it has also been used in a plethora of research contexts: as a source of bacterial cell membranes and protoplast for investigations in bioenergetics, to investigate the relationship between codon usage and tRNA representation in bacterial genomes, to study dormancy and resuscitation processes; and as biotechnological tools, for their capacity of synthesizing long-chain alkenes and concentrate heavy metals (Ohama *et al.*, 1989; Young *et al.*, 2010; Surger *et al.*, 2018). In recent years, *M. luteus* has been proposed as a well-suited actinobacterial model organism to study natural transformation (Angelov *et al*., 2015).

Like in other transformable organisms, in *M. luteus,* competence for DNA acquisition is an inducible and transient physiological state. The peak of competence, where the higher transformation frequencies can be measured, is during the mid to late exponential growth phase, with a maximum when the cell density reaches an OD of 1 (Angelov *et al*., 2015). Different experiments showed that competence was also triggered by nutritional stress, given that cells grown in complex media (LB) displayed ca. 100-fold lower transformation frequencies than those in minimal medium (GMM). Due to the diversity of niches, this organism has in nature, the regulatory mechanisms that control these traits are most likely intricate. Apart from ComA, no homologs of ComX, ComP, ComS, ComQ, ComK, CodY or PhrC were found in its genome. This could also support the hypothesis that the regulatory pathways are different from the already better-studied organisms. The systematic search of the *M.*

luteus genome for ORFs with sequence similarity to known competence genes from prototypical species did lead to the identification of orthologs of the *B. subtilis* ComEA (Mlut_12640; Gene ID: 937827), ComEC (Mlut_12650; Gene ID: 937839), and DprA (Mlut_09260; Gene ID: 940135) proteins. The deletion of these genes led to the loss of transformability (Angelov *et al*., 2015; Angelov, Li, Geissler, Leis, & Liebl, 2013), meaning as one would expect that they are essential for the process.

Considering that for most model transformable organisms there is a link between transformation and the involvement of a macromolecular transport machinery (T4P or T2SS), a search of these sorts of elements was conducted (Angelov *et al*., 2015). Two separate gene clusters were found in the *M. luteus* genome, *tad-1* (Mlut_07500 to Mlut_07560) and *tad-2* (Mlut_18150 to Mlut_18190), both containing ORFs with sequence similarities to prototypical Flp pili biogenesis proteins, Flp pilin and Flp-like proteins of the reference organism for these structures, *Aggregatibacter actinomycetemcomitans* (*E*-values below 10^{-5}) (Kachlany *et al*., 2000). All genes in these *M. luteus tad* gene clusters were reported to be essential for transformation except for *tadB* (Mlut_07510), which only generated a >100-fold reduction in transformation frequencies (Angelov *et al*., 2015). This extracellular Flp pili structure seems to be present in only 3% of the population in the mid to late exponential phase when grown in a minimal medium. Given that after deletion of the ORFs of the *tad* clusters there was no obvious difference in adherence, cell aggregation, or microcolony morphology with the wild-type strain (trpE16), it is believed that the main physiological role of the Flp pili in *M. luteus* is transformation.

In a more recent publication (Lichev, Angelov, Cucurull, & Liebl, 2019), the regulation of competence in *M. luteus* was studied further by using a β-galactosidase reporter strain for the essential *comEA/EC* genes (see Figure 1.4). After observing that competence is a transient regulated trait that is affected by the growth phase and the nutritional state of the organism, the effects of different amino acids on the transcriptional levels of *comEA/EC* and the overall transformability were assayed. Considering the absence of any data suggesting a quorum sensing system, the role of the stringent response as a consequence of amino acids deprivation was evaluated.

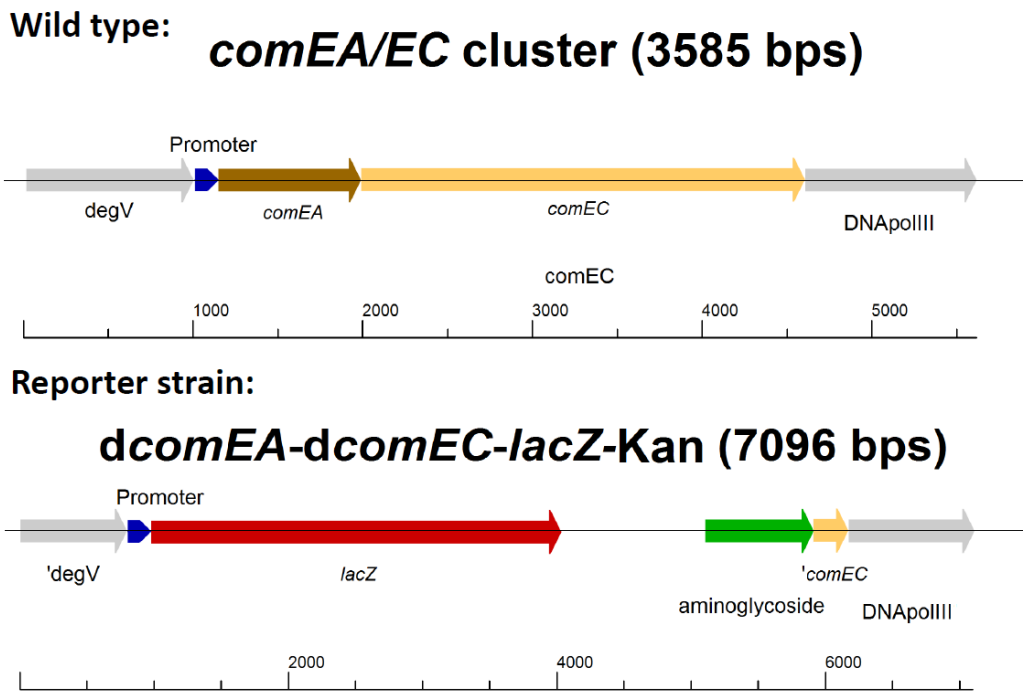


Figure 1.4 Δ*comEA/EC*:*lacZ-Kan* construct.
LacZ reporter construct scheme in contrast to the wild-type disposition of the *comEA/EC* transcriptional unit. The 3.5 kbp ORF of *comEA-EC* of *M. luteus* trpE16 was replaced by a 5 kbp construct including the full length of the *E. coli lacZ* gene followed by the resistance cassette for the aminoglycoside kanamycin.

The stringent response is a broadly distributed physiological response of bacteria to nutrient starvation, heat shock, or other stress factors. A massive switch in the transcription profile occurs after a change in the concentration of the alarmone (p)ppGpp that triggers this response system. This molecule also acts by direct binding to specific enzymes or by direct modulation of cytosol GTP levels, affecting for example master regulators as CodY in the case of *B. subtilis* (Inaoka & Ochi, 2002). In *M. luteus*, on the transcriptional level, the expression of the late competence-related genes *comEA/EC* was greater when the amino acids concentration was low. Regarding transformability, a similar correlation was observed. The regulatory model proposed by Lichev *et al*. (2019) states that the intracellular concentrations of (p)ppGpp would regulate the transformability of the cells in such a manner that the concentration-response curve would have a bell-shaped form. During the transition from full medium to minimal medium, the alarmone would accumulate in the cells due to the lack of amino acids. Because of this, the bacteria would increase their amino acid biosynthesis which in turn would cause oscillations in the concentrations of (p)ppGpp. After equilibrium sets in, the cells would reach their maximum transformability in the mid-late exponential phase.

Two distinct loci which code for proteins with typical features of stringent response enzymes were localized. *Mlut_12840* and *Mlut_22200*, which encode a putative RelA-type protein with both hydrolase and synthetase functions (RSH), and a GTP pyrophosphokinase with a single RelA_SpoT

synthesis domain, respectively. After deletion of each gene and both together, none of the single mutants showed a significant change in transformability in contrast to the wild-type reference. Although, with the double mutant a nearly 100-fold decrease in the transformation frequency was measured. These genes are presumably involved in (p)ppGpp synthesis and might be required for induction of competence in *M. luteus* during growth in minimal medium. Interestingly, none of the mutants exhibited more than 2-fold change in the relative expression of the investigated genes when compared to the reference strain trpE16. Thus it seems that cells can have wild-type levels of transformability even when they exhibit much stronger repression of the competence-related genes. This apparent contradiction between the observed gene expression profiles and the transformability of the mutants could mean that the stringent response and natural transformation are linked not only on a transcriptional but also on a post-transcriptional level (Lichev *et al*., 2019).

All the data published and the information gathered so far are just the tip of the iceberg considering all that remains unknown regarding natural competence and transformation in *M. luteus* and other transformable members of *Actinobacteria*. To further unravel how competence-related genes are regulated, the reporter strain *M. luteus lacZ* Δ*comEA*/*EC*:*lacZ-Kan* (see Figure 1.4) was utilized in this work.

2 Goal and starting hypothesis

The main goal of the research in our group, and of this work specifically, is to recognize and characterize previously unknown elements (genes, proteins, or any biomolecule) involved in the regulation of essential genes for natural competence and/or transformation in *Actinobacteria*. For this purpose, we have proposed an actinobacterial model organism, *Micrococcus luteus*. Considering that the best-studied genes that play a role in this process are the natural transformation marker co-transcribed genes *comEA*-*comEC*, it was decided to start further investigations on the basis of these genes. The focus of the work is to find out how the promoter of this operon is regulated. For this, the introduction of multiple and random changes in the genome sequence of a *comEA/EC* promoter-*lacZ*-reporter strain was planned. Subsequently, a phenotype screening of dysregulated mutants should be done to find the best candidates for whole-genome sequencing. It was anticipated that it should be possible to identify the genetic loci causing the phenotypes by sequencing and in-depth bioinformatic analysis of the aligned genome sequences. In order to confirm their role in elicitating the observed dysregulated *comEA/EC*-*lacZ*-reporter phenotypes, it was planned to reintroduce the most promising individual point mutations identified in this way into the *comEA/EC*-*lacZ*-reporter strain. This strategy would allow a clean evaluation of the impact of specific single-nucleotide genome editions on competence gene expression and natural transformation, and would be the foundation for further studies on gene expression rates and protein-gene interactions.

The general working hypothesis of this study states that, for *M. luteus*, natural competence is a tightly regulated and transient physiological state, governed by an intricate regulatory network sensitive to the environment, and distinct to what has been seen in other naturally transformable bacteria so far.

3 Materials & Methods

3.1 Chemicals

Table 3.1. List of used substances

Name	Producer
Agarose LE	Biozym Scientific GmbH (Hessich Oldenburg, Germany)
Bradford Reagent	SERVA Electrophoresis GmbH (Heidelberg, Germany)
$CaCl_2$	J.T. Baker, Inc. (Phillipsburg, NJ, USA)
$CuSO_4$	Sigma-Aldrich Corp. (St. Louis, MO, USA)
D-Maltose	Sigma-Aldrich Corp. (St. Louis, MO, USA)
DMF	AppliChem GmbH (Darmstadt, Germany)
DMSO	AppliChem GmbH (Darmstadt, Germany)
D-Saccharose	Merck KGaA (Darmstadt, Germany)
EDTA	AppliChem GmbH (Darmstadt, Germany)
Ethidium Bromide	AppliChem GmbH (Darmstadt, Germany)
Acetic Acid	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)
$FeCl_3$	Merck KGaA (Darmstadt, Germany)
$FeSO_4.7H_2O$	Merck KGaA (Darmstadt, Germany)
Glucose	Sigma-Aldrich Corp. (St. Louis, MO, USA)
Meat extract	VWR International (Radnor PA, USA)
H_3BO_3	AppliChem GmbH (Darmstadt, Germany)
Yeast extract	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)
KCl	J.T. Baker, Inc. (Phillipsburg, NJ, USA)
KH_2PO_4	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)
K_2HPO_4	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)
L-Tryptophan	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)
$MgCl_2$	VWR International (Radnor PA, USA)
$MgSO_4.7H_2O$	Merck KGaA (Darmstadt, Germany)
$MnCl_2$	Merck KGaA (Darmstadt, Germany)
$MnCl_2.4H_2O$	Merck KGaA (Darmstadt, Germany)
NaCl	AppliChem GmbH (Darmstadt, Germany)

NaH_2PO_4	Merck KGaA (Darmstadt, Germany)
Na-Glutamate	Sigma-Aldrich Corp. (St. Louis, MO, USA)
NH_4Cl	Merck KGaA (Darmstadt, Germany)
Peptone	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)
p-Nitrophenyl β-D-galactopyranoside	Carbosynth Ltd. (Compton, England)
Tris-HCl	AppliChem GmbH (Darmstadt, Germany)
Tryptone	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)
X-Gal	Carbosynth Ltd. (Compton, England)
$ZnCl_2$	Merck KGaA (Darmstadt, Germany)

3.2 Growth media

All media described below were, unless otherwise indicated, sterilized by autoclaving at 121 °C for 15 min (Systec VX-150) or 20 min (VARIOKLAV® 25T).

Table 3.2. Composition of utilized full and minimal media.

Name	Components	Amount
Lysogeny Broth medium (LB) pH 7.2	Tryptone	10.0 g
	Yeast Extract	5.0 g
	NaCl	5.0 g
	H_2O	add 1000 mL
Glutamate Minimal Medium (GMM/Naylor medium) pH 7.5	K_2HPO_4	2.0 g
	NH_4Cl	1.0 g
	L-Glutamate (Monosodium)	10 g
	$MgSO_4 \cdot 7H_2O$	0.1 g
	$FeSO_4 \cdot 7H_2O$	0.004 g
	$MnCl_2 \cdot 4H_2O$	0.002 g
	H2O	add 940 mL
	Added after autoclaving:	
	(filter-sterilized)	0.1 g
	L-Tryptophan	7.0 g
	Glucose	add 60 mL
	H_2O	
CAH- pH 7.2	K_2HPO_4	2.0 g
	NH_4Cl	1.0 g
	L-Glutamate (Monosodium)	5.0 g
	Casein (acid) Hydrolysate	5.0 g
	$MgSO_4 \cdot 7H_2O$	0.05 mg
	$FeSO_4 \cdot 7H_2O$	0.0016 mg
	$MnCl_2 \cdot 4H_2O$	0.0004 mg
	H_2O	add 970 mL

	Added after autoclaving: (filter-sterilized)	
	L-Tryptophan	7.0 g
	Glucose	0.1g
	H_2O	30 mL
SOB medium pH 7.0	Tryptone	20 g
	Yeast Extract	5.0 g
	NaCl	0.5 g
	H_2O	add 1000 mL
	Added after autoclaving: (filter-sterilized)	
	KCl	2.5 mM
	$MgCl_2$	10.0 mM
	$MgSO_4$	10.0 mM

3.3 Buffer and Solutions

Table 3.3. Composition of buffers and solutions

Name	Components	Producer
Anode Buffer pH 8.9	Tris-HCl	0.04 M
Cathode Buffer	Tris-HCl	0.1 M
	Tricine	0.1 M
	SDS	0.1 %
Column Buffer pH 7.4	Tris-HCl	20 mM
	NaCl	200 mM
	EDTA	1 mM
	β-mercaptoethanol	10 mM
CTAB pH 8.0	Tris	100 mM
	NaCl	1.4 M
	EDTA	20 mM
	Polyvinylpyrrolidone	1%
	CTAB	2%
	H_2O	
Phosphate Buffer (PBS) pH 7.4	Na_2HPO_4	60 mM
	NaH_2PO_4	40 mM
	Citrate	200 mM
SDS-PAGE running buffer	Tris	0.25 M
	SDS	0.035 M
	Glycin	1.924 M
Separating/resolving gel buffer pH 8.8	Tris	182 g/L
	SDS	4 g/L
Stacking gel buffer pH 8.8	Tris	61 g/L
	SDS	4 g/L

TAE-Buffer **pH 7.5**	Tris-HCl EDTA Acetic acid H_2O	40 mM 1 mM 20 mM Add 1000 mL
TB-Solution **pH 6.7**	PIPES $CaCl_2$ KCl Adjust pH with HCl or KOH Filter-sterilized: $MgCl_2$	10 mM 15 mM 250 mM 55 mM
Transformation Buffer 1 **pH 7.0**	$CaCl_2$ Tris Adjust pH with HCl	10 mM 50 mM
Transformation Buffer 2 **pH 7.0**	$CaCl_2$ Tris Adjust pH with HCl	190 mM 50 mM
Tris-Borate binding Buffer **pH 8.1**	Tris-HCl H_3BO_3 DTT BSA Filter-sterilized: NaCl K-Acetate $MgCl_2$ $MnCl_2$ $ZnCl_2$	45 mM 45 mM 2 mM 0.25 mg/mL 100 mM 100 mM 5 mM 5 mM 5 mM
Tris-Borate Buffer **pH 8.1**	Tris-HCl H_3BO_3 Adjust pH with HCl	45 mM 45 mM

3.4 Kits, Commercial Solution & Reagents

Table 3.4. List of commercial kits used in this work.

Name	Producer
DNA Clean & Concentrator Kit	Zymo Research Corp. (Irvine CA, USA)
Monarch® Plasmid Miniprep Kit	New England Biolabs, Inc. (Ipswich MA, USA)
ZR Fungal/Bacterial RNA Miniprep Kit	Zymo Research Corp. (Irvine CA, USA)
TURBO DNA-*free*™ Kit	Thermo Fisher Scientific, Inc. (Waltham MA, USA)
iScript™ cDNA synthesis kit	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
5´/3´RACE Kit 2nd Generation	Sigma-Aldrich Corp. (St. Louis, MO, USA)
High Pure PCR product purification Kit	Sigma-Aldrich Corp. (St. Louis, MO, USA)
MBP-Spin Protein Miniprep Kit	Zymo Research Corp. (Irvine CA, USA)

Table 3.5. List of commercial solutions and reagents used in this work.

Name	Producer
Bradford Reagent (5x)	SERVA Electrophoresis GmbH (Heidelberg, Germany)
Ethidium Bromide	AppliChem GmbH (Darmstadt, Germany)
Factor Xa protease	New England Biolabs, Inc. (Ipswich MA, USA)
GelRed® Nucleic Acid Gel Stain	Biotium, Inc. (Fremont CA, USA)
GeneRuler™ 1kb DNA Ladder PLUS, ready-to-use (0.1 µg/µL)	Thermo Fisher Scientific, Inc. (Waltham MA, USA)
Lysozyme	Thermo Fisher Scientific, Inc. (Waltham MA, USA)
NEBuilder HiFi DNA Assembly Master Mix	New England Biolabs, Inc. (Ipswich MA, USA)
PageRuler™ Plus Prestained Protein Ladder, 10 – 250 KDa	Thermo Fisher Scientific, Inc. (Waltham MA, USA)
PageRuler™ Unstained Low Range Protein Ladder 3.4 – 100 KDa	Thermo Fisher Scientific, Inc. (Waltham MA, USA)
Proteinase K	Biozym Scientific GmbH (Oldendorf, Germany)
RNA loading dye	
RNAase solid	AppliChem GmbH (Darmstadt, Germany)
Tissue and cell lysis buffer	Lucigen Corp. (Middelton WI, USA)

3.5 Technical equipment

Table 3.6. List of technical equipment used in this work.

Name	Producer
ÄKTATM explorer system	Amersham Pharmacia Biotech, Uppsala, Sweden
Alphaimager MINI	ProteinSimple (San Jose CA, USA)
Analytical balance Precisa® 125A SCS	Precisa Gravimetrics AG (Dietlikon, Switzerland)
Axio Scope.A1 (Microscope) along with AxioCam ICc 1 (Microscope camera)	Carl Zeiss Microimaging GmbH (Göttingen, Germany)
Biometra TAdvance, thermocycler	Analytic Jena AG (Jena, Germany)
Bosch GSP 30A21 (Freezer)	Robert Bosch GmbH (Stuttgart, Germany)
Cabinet dryer UM	Memmert GmbH & Co. KG (Schwabach, Germany)
CHEF-DR® III Variable Angle System 962BR, Cooling Module 816BR and Variable Speed Pump 260BR (PFGE-System)	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
Gene Pulser IITM Electroporation System	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
Glass measuring cylinder, SILBERBRAND, borosilicate (50 mL, 100 mL, 250 mL, 500 mL, 1.000 mL)	BRAND GmbH & Co. KG (Wertheim, Germany)
Haake DC1 (Circulating Bath)	Thermo Haake GmbH (Karlsruhe, Germany)
HERAsafe HSP 18 (Safety Workstation)	Heraeus Holding GmbH (Hanau, Germany)
HTL Discovery Comfort (Pipette) (2 µL, 20 µL, 200 µL, 1 mL)	PZ HTL SA (Warschau, Poland)
Integra Voyager II Multichannel pipette (125 µL, 300 µL)	INTEGRA Biosciences AG (Zizers, Switzerland)
Measuring glass, Polypropylene (2 L, 3 L, 5 L)	VITLAB GmbH (Grossostheim, Germany)
Multipette® plus (Multidispense pipette)	Eppendorf AG (Hamburg, Germany)
MyCycler™ Thermal Cycler	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
NanoDrop™ ND-1000	NanoDrop Technologies LLC (Wilmington DE, USA)
New Brunswick™ Innova® 42/42R (Incubator shaker)	New Brunswick Scientific Co., Inc. (Edison NJ, USA)
pH-Meter 902	Berrytec GmbH (Grünwald, Germany)

Precision balance EW 600-2M	KERN & SOHN GmbH (Balingen-Frommern, Germany)
PX1 PCR plate sealer	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
Qubit™ 4 Fluorometer	Thermo Fisher Scientific, Inc. (Waltham MA, USA)
QX200™ Droplet Generator	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
QX200™ Droplet Reader	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
Reagent bottle, borosilicate (50 mL, 100 mL, 250 mL, 500 mL, 1.000 mL)	Schott AG (Mainz, Germany)
Refrigerated Centrifuge Universal 320R	Hettich AG (Bäch, Switzerland)
SLM Aminco FA-078 (French pressure cell press)	American Instruments Company Inc. (Silver Spring MD, USA)
Small centrifuge Mikro 22R	Hettich AG (Bäch, Swizerland)
Sorvall™ LYNX™ Superspeed-Zentrifuge mit Fiberlite™ F9-6×1000 LEX Festwinkelrotor / Fiberlite™ F14-6×250y LEX Festwinkelrotor / A27-8×50 Festwinkelrotor	Thermo Fisher Scientific, Inc. (Waltham MA, USA)
Sorvall™ RC6+ Centrigufe with Fiberlite™ F12S-6×500 LEX Fixed angle rotor	Thermo Fisher Scientific, Inc. (Waltham MA, USA)
Systec V-150, vertical standing autoclave 150 L	Systec GmbH (Linden, Germany)
Thermomixer comfort	Eppendorf AG (Hamburg, Germany)
TKA® MicroPure (Ultrapure water source)	TKA Wasserau□ereitungssysteme GmbH (Niederelbert, Germany)
Ultrasound processor UP200S	Hielscher Ultrasonics GmbH (Teltow, Deutschland)
Ultrospec 2100 pro (UV/VIS-Spectrophotometer)	GE Healthcare Bio Sciences AB (Uppsala, Schweden)
VARIOKLAV® 25 T, steam sterilizer, 25 L	H+P Labortechnik GmbH (Oberschleißheim, Deutschland)
VARIOKLAV® steam sterilizer	H+P Labortechnik GmbH (Oberschleißheim, Deutschland)
VARIOMAG® Mobil 60 (magnetic beads)	H+P Labortechnik GmbH (Oberschleißheim, Germany)
Vortext-Genie 2 (vortex mixer)	Scientific Industries, Inc. (Bohemia NY, USA)
WTC Binder BD 240 (Incubator)	BINDER GmbH (Tuttlingen, Deutschland)
WTW Inolab pH 720 (pH-Meter)	Xylem Analytics Germany Sales GmbH & Co. KG (Weilheim, Deutschland)

3.6 Dispensable Materials

Table 3.7. List of dispensable material used in this work.

Name	Producer
Cover glass	VWR International (Radnor PA, USA)
Electroporation cuvette, 2 mm gap (VWR Collection)	VWR International (Radnor PA, USA)
Eppendorf Combitips advanced® multipipette tips (2,5 mL)	Eppendorf AG (Hamburg, Germany)
Glass beads (2,85-3,45 mm)	Sarstedt AG & Co (Nümbrecht, Germany)
HSW SOFT-JECT® dispensable syringe (sterile) (10 mL, 20 mL, 50 mL)	Henke-Sass, Wolf GmbH (Tuttlingen, Germany)
Low-Profile PCR Tubes 8-tube strip, white	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
Microcentrifuge tube with lid, polypropylene (2 mL)	Sarstedt AG & Co (Nümbrecht, Germany)
Microtiter plate 96 wells, flat bottom (sterile)	Sarstedt AG & Co (Nümbrecht, Germany)
Ministart® NML syringe filter, 0.2 µm SFCA sterile	Sartorius AG (Göttingen, Germany)
NATURE Star® BIO Toothpick, 8 cm	Franz Mensch GmbH (Buchloe, Germany)
Glass slide, VWR Collection	VWR International (Radnor PA, USA)
Optical flat 8-Cap Strips (for 0.2 mL reaction tubes/plates)	Bio-Rad Laboratories, Inc. (Hercules CA, USA)
PCR SoftTubes, DNS-, DNase-, RNase free (0.2 mL)	Biozym Scientific GmbH (Hessisch Oldenburg, Germany)
Petri dishes, polystyrene 92 mm (sterile with vents)	Greiner Bio-One International GmbH (Kremsmünster, Austria)
Polypropylene tubes (15 mL, 50 mL)	Sarstedt AG & Co (Nümbrecht, Germany)
Polypropylene tubes (1.5 mL)	Zefa-Laborservice GmbH (Harthausen, Deutschland)
Polypropylene tubes (1.5 mL)	Greiner Bio-One International GmbH (Kremsmünster, Austria)
TipOne® pipette tips (2 µL, 200 µL, 1.250 µL)	STARLAB International GmbH (Hamburg, Germany)
Vivaspin Turbo 15 RC, 100,000 MWCO, 12/PK	Sartorius AG (Göttingen, Germany)

3.7 Computer Software

Table 3.8. List of Software used in this work and their respective version number**.**
All programs were installed and used in Microsoft® Windows 10.

Program	Version number
Artemis Software	V18.0.0
Bio-Rad CFX Manager 3.1	3.1.1517.0823
Clone Manager Professional 9	9.51
GraphPad PRISM	7.0
ImageJ Image Processing and Analysis	1.46r
Microsoft® Excel® 2016	16.0.9001.2080
R	3.51
R Studio	1.1.463

3.8 Microorganisms

Table 3.9. List of microorganisms used during this study.

Strain	Genotype and relevant phenotype	Source
Micrococcus luteus **ATCC 27141**	"*Micrococcus lysodeikticus*" ISU, *trp*$^{+}$	(Kloos, 1969)
Micrococcus luteus **trpE16**	trpE16, mutagenesis derivative of ATCC 27141, *trp*$^{-}$, RefSeq(NCBI): ASM87779v1	(Kloos and Rose, 1970)
Micrococcus luteus ***ΔcomEA/EC*:*lacZ***	trpE16 Δ*comEA*/*EC*:*lacZ-Kan*;Kan^{R}, LacZ expression	(Lichev *et al*, 2019)
Escherichia coli **XL-1 Blue**	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacI*^{q}*ZΔM15 Tn10 (Tet*^{r}*)]*	Stratagene Corp. (La Jolla, CA, USA)
Escherichia coli **BL21**	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs (Ipswich, USA)

3.8.1 Culturing of Microorganisms

When solid media was required, 13 g/L of agar were added to the corresponding liquid media before autoclaving. Then 25 mL of the medium with dissolved agar were poured on a 92 mm polystyrene petri dish and left 10 min at room temperature for solidification.

For culturing in liquid medium, 100 mL capacity Erlenmeyer flasks were used. After sealing with a cotton stopper, cultivation was carried out at 180 rpm and constant temperature in a shaking

incubator (New Brunswick™ Innova® 42). Test tubes with aluminum caps were used for smaller volumes (5 mL).

The standard cultivation conditions of each microorganism used in this study are listed below. The composition of the media can be found in Table 3.2.

Table 3.10. Culture media used for each microorganism and their culturing temperature.

Strain	Medium	Cultivation Temperature
Micrococcus luteus	LB, GMM	30°C
Escherichia coli	LB	37°C

In case of different incubation conditions, they will be described for the particular method.

3.8.2 Antibiotics

In the upcoming Table the antibiotics used during this work can be found, along with their working concentration and due solvent.

Table 3.11. Used antibiotics and their respective working concentrations.

Name	Abbreviation	Working concentration	Solvent
Ampicillin	Amp	100 µg mL^{-1}	H_2O
Kanamycin	Kan	60 µg mL^{-1}	H_2O
Tetracycline	Tet	12.5 µg mL^{-1}	EtOH
Rifampicin	Rif	1 µg mL^{-1}	MeOH

When used with solid medium, the antibiotic was added only when the medium had a temperature below 55°C in order to prevent degradation.

3.9 Optical cellular density determination

The determination of the optical density (OD) of a bacterial culture is a simple and quick method for indirect quantification of the cell number. In this way, growth and physiological phases can be estimated and different cultures can be normalized. A measured volume of bacterial culture is placed in a cuvette and the determination is carried out using a spectrophotometer at a wavelength of 600 nm. For cultures with OD_{600nm} above 1.5, a 1:10 dilution was prepared using a suitable, sterile buffer or medium.

3.10 Genetic material isolation

3.10.1 Genomic DNA preparation for *M. luteus*

After O/N incubation in full medium, 4 mL of culture were centrifuged at 14000 x g for 5 min. The pellet was resuspended in 25 µL of PBS and mixed with 25 µL of 0.1 mg/mL of Lysozyme. 1 µL of Proteinase K were mixed with 300 µL of TC buffer, and the mixture was added to the cells. After mixing well, the cells were incubated 15 min at 65°C. The suspension was mixed every 5 min. After cooling the sample to 37°C, 1 µL of 5 µg/µL of RNAase A was added and the extract was incubated at the same temperature for 30 min. 150 µL of CTAB were pipetted in the sample after having cooled it on ice. The mixture was vortexed and centrifuged at 4°C and 14000 x g for 15 min. The supernatant was transferred into a clean 1.5 mL tube and mixed by repeated inversion with 500 µL of absolute isopropanol. The suspended DNA was pelleted and washed twice with ethanol 70%. After removing all the ethanol, the tube was left to dry. The DNA pellet was then resuspended in 50 µL of ultrapure H_2O.

3.10.2 Plasmids

Prior to the plasmid extraction and purification, the corresponding plasmid-carrier strain was cultured overnight in a full medium under antibiotic selective pressure. Enough plasmid was prepared by using the Monarch® Plasmid Miniprep Kit (New England Biolabs, Inc.) following the manufacturer´s instructions. Plasmid concentration and quality was determined by using a NanoDrop™ ND-1000 and running the samples on a 0.8% Agarose gel (see Section 3.16 for more details).

Table 3.12. List of plasmids used in this work.

Name	Features	Source
pMAL-c2x (6645 bps)	Expression and purification system. *lacIq* promoter, *lac* repressor; *tac* promoter, *malE* *AmpR* promoter, *AmpR*	New England Biolabs, Inc. (Ipswich MA, USA)
pKOS6b (6689 bps)	Cloning vector. *KanR, codBA* integrated, FC^S	Kostner *et al.,*2013
pWLTK6 (3101 bps)	*KanR*	-
pJET (2974 bps)	Cloning vector. *AmpR, trpE*	Thermo Fisher Scientific, Inc. (Waltham MA, USA)

3.10.3 Total RNA isolation

For total RNA preparation, 4mL of overnight cell culture were centrifuged at 14000 x g for 5 min and washed with PBS. After resuspending the cells in 1mL of buffer, 0.1 mg/mL of Lysozyme (Thermo Fisher Scientific, Inc. Waltham MA, USA) were added and the suspensions incubated 10 min at 37°C. Following cell lysis, ZR Fungal/Bacterial RNA Miniprep Kit (Zymo Research, USA) was used under the manufacturer´s specifications. An additional rigorous DNase treatment was performed with the TURBO DNA-free Kit (Thermo Fisher Scientific, USA) to minimize the amount of genomic DNA left in the samples.

3.11 Polymerase chain reaction (PCR)

To easily and effectively generate copies of specific DNA sequences *in vitro*, the Polymerase chain reaction (PCR) technique was harnessed using a Q5 polymerase (New England Biolabs). All steps of the reaction, *denaturation*, *annealing* and *extension* (see Tables 17 and 19 for detailed information) were carried out by using MyCycler™ Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules CA, USA). The specific oligonucleotide primers were ordered to Eurofins Genomics GmbH (Ebersberg, Germany). A list and description of all primers used in this work can be seen in Table 3.17. Either a genomic DNA solution (see Table 3.13) or a cell suspension (see Table 3.15) were used as template source. When the second option is used, the technique is called Colony-PCR, which enables faster screening of high numbers of bacterial colonies without having to grow them on cultures and prepare genomic DNA solutions.

Table 3.13. Components of a Q5-PCR reaction with a final volume of 25 µL.

Component	Amount
Reaction Buffer Q5 (5x)	5 µL
dNTPs (10 mM)	0.5 µL
High GC Enhancer Q5 (5x)	5 µL
***Forward*-Primer (10 µM)**	1.25 µL
***Reverse*-Primer(10 µM)**	1.25 µL
DNA-Polymerase Q5	0.25 µL
DNA Template	1 µL*
H_2O ultrapure	10.75 µL

* 1:100 or 1:1000 dilutions of the DNA suspensions were usually used.

* This temperature worked for most primers used in this study. Although, when no amplification or amplification of unspecific products happened, a temperature gradient from 55 to 72°C was carried out to find the optimal temperature.
** The elongation time was set up according to the expected amplicon length. The polymerization rate of the Q5-polymerase was used as calculation factor (2 kbps min^{-1})

Table 3.14. Q5-PCR cycle steps and temperatures.

Step	Temperature	Time
Initial Denaturation	98°C	30 s
30 Cycles:		
Denaturation	98°C	10 s
Annealing	64°C*	20 s
Extension	72°C	Variable**
Final extension	72°C	120 s

* This temperature worked for most primers used in this study. Although, when no amplification or amplification of unspecific products happened, a temperature gradient from 55 to 72°C was carried out to find the optimal temperature.
** The elongation time was set up according to the expected amplicon length. The polymerization rate of the Q5-polymerase was used as calculation factor (2 kbps min^{-1})

Table 3.15. Components of a Colony Q5-PCR reaction with a final volume of 25 µL.

Component	Amount
Reaction Buffer Q5 (5x)	5 µL
dNTPs (10 mM)	0.5 µL
High GC Enhancer Q5 (5x)	5 µL
***Forward*-Primer (10 µM)**	1.25 µL
***Reverse*-Primer(10 µM)**	1.25 µL
DNA-Polymerase Q5	0.25 µL
DNA Template (cell suspension)	2.5 µL*
H_2O ultrapure	9.25 µL

* 1 *M. luteus* average size colony (2-3 mm) resuspended in 30 µL of ultrapure H2O was used as template

Table 3.16. Q5-PCR cycle steps and temperatures.

Step	Temperature	Time
Initial Denaturation	98°C	300 s
30 Cycles:		
Denaturation	98°C	10 s
Annealing	64°C*	20 s
Extension	72°C	Variable**
Final extension	72°C	120 s

Unless the amplicons were for analytical purposes, all were purified after the reaction (DNA Clean & Concentrator Kit) and then run on an 0.8% Agarose gel to check their size.

Table 3.17. List and sequences of oligonucleotide primers used in this study

Name	Sequence	Template	Transformant/ Vector	Used for
341241F	GAGCTCGGTACCCGGGGATCC TGAAGGCCAGCAGGTTCTTCG	*M. luteus* Δ*comEA/EC*:*lacZ* M43	*M. luteus* trpE16/pKOS6b	Single SNP insertion
341241R	GTAAAACGACGGCCAGTGCCA GCGGATGGGTGGAACTCAAG			
1061676F	GAGCTCGGTACCCGGGGATCC TTCGCCATGATGCTCGACCAG	*M. luteus* Δ*comEA/EC*:*lacZ* M89	*M. luteus* trpE16/pKOS6b	Single SNP insertion
1061676R	GTAAAACGACGGCCAGTGCC ATGCGCGGTGAACCACTTGTC			
1063999F	GAGCTCGGTACCCGGGGATCC TTCGGCAAGTCGAACATCCTC	*M. luteus* Δ*comEA/EC*:*lacZ* M93 - M28B	*M. luteus* trpE16/pKOS6b	Single SNP insertion
1063999R	GTAAAACGACGGCCAGTGCCA CCGCGAGCGACATGATCTTG			
1606810F	GAGCTCGGTACCCGGGGATCC TTGACCTTCGCGTCCACGATG	*M. luteus* Δ*comEA/EC*:*lacZ* M27 - M89 - M1090 - M93 - R4	*M. luteus* trpE16/pKOS6b	Single SNP insertion
1606810R	GTAAAACGACGGCCAGTGCC AATGACCGATGTCCGGCACTC			
226684F	GAGCTCGGTACCCGGGGATCC TGCATGGCATGACAGGGTCTC	*M. luteus* Δ*comEA/EC*:*lacZ* M33	*M. luteus* trpE16/pKOS6b	Single SNP insertion
226684R	GTAAAACGACGGCCAGTGCC AGGCGATCAGCTCCGATTTGG			
810889F	GAGCTCGGTACCCGGGGATCC TGGTTGGTCTCCAGGCTCATC	*M. luteus* Δ*comEA/EC*:*lacZ* R33	*M. luteus* trpE16/pKOS6b	Single SNP insertion
810889R	GTAAAACGACGGCCAGTGCC ACTGCTGTTCGGTGTGCTGTC			
1541663F	GAGCTCGGTACCCGGGGATCC TTGGACTTCCTGCACCAACTG	*M. luteus* Δ*comEA/EC*:*lacZ* R34	*M. luteus* trpE16/pKOS6b	Single SNP insertion
1541663R	GTAAAACGACGGCCAGTGCCA GCTGGACCATGGAGCACATC			
KanF2	AGGTTGGGAAGCCCTGCAAAG	pWLTK6	*M. luteus* trpE16/None	Kan cassette amplification
KanR3	GAGTCCCGCTCAGAAGAACTC			
14650AF	GAGCTCGGTACCCGGGGATCC TGACAGCTCCTCGTCCTGTTC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
14650AR	GAACATCCGGTGTCCACGACC GTGGACGCGAAGGTCAAGG			
14650BF	CCTTGACCTTCGCGTCCACGG TCGTGGACACCGGATGTTC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
14650BR	GTAAAACGACGGCCAGTGCCA ATGACCGATGTCCGGCACTC			
14660AF	GAGCTCGGTACCCGGGGATC CTTCCTTGACCTTCGCGTCCAC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
14660AR	CACTTCCTTCTGACGCTCCAG TCGGTCAGGGTGAGAAAGC			

14660BF	GCTTTCTCACCCTGACCGAC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
14660BR	GTAAAACGACGGCCAGTGCCA CCAGAAGCAGGACGGTGATG			
14660AF	GAGCTCGGTACCCGGGGATCC TTCCTTGACCTTCGCGTCCAC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
14660AR2	CTCAGTGGGCCTTCGCGTCCG TCGGTCAGGGTGAGAAAGC			
14660BF2	GCTTTCTCACCCTGACCGACG GACGCGAAGGCCCACTGAG	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
14660BR2	GTAAAACGACGGCCAGTGCCA CGACGGGTCAACTTGTCCAG			
02150AF	GAGCTCGGTACCCGGGGATCC TCTCTCCGTGAGGAATGAACC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
02150AR	GACTTGACCCAGACACGATG GAAGCCGGACCTGTTCATGC			
02150BF	GCATGAACAGGTCCGGCTTC CATCGTGTCTGGGTCAAGTC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
02150BR	GCATGAACAGGTCCGGCTTC CATCGTGTCTGGGTCAAGTC			
03135AF	GAGCTCGGTACCCGGGGATCC TTGGATGTCGAAGACGTTCTC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
03135AR	GACCAGATTTCTCGCCACGA GTATTTCTGGCTCTCCTGAC			
03135BF	GTCAGGAGAGCCAGAAATACT CGTGGCGAGAAATCTGGTC	*M. luteus* trpE16	*M. luteus* trpE16/pKOS6b	Markerless deletion
03135BR	GTAAAACGACGGCCAGTGCCA TCGAGATCTACCGCGACCAC			
KO14660AF	TCCTTGACCTTCGCGTCCAC	*M. luteus* trpE16	*M. luteus* trpE16/none	Gene deletion through Kan cassette exchange
KO14660AR	CTTTGCAGGGCTTCCCAACCTG TCGGTCAGGGTGAGAAAGC			
KO14660BF	AGTTCTTCTGAGCGGGACTCT GGAGCGTCAGAAGGAAGTG	*M. luteus* trpE16	*M. luteus* trpE16/none	Gene deletion through Kan cassette exchange
Seq14650	AGCACGGTGATGTCCGTCTG			
SP1-14660	GCGTCCGCCTCGGTCCACTC	*M. luteus* trpE16 cDNA	-	5´RACE Mlut_14660
14660PromR	CCTTCTGACGCTCCACGTAG			
SP3-14660	ACCTGGATGGCCTGCAGCTC			
1606810F	(detailed above)	*M. luteus* trpE16 cDNA	-	5´RACE Mlut_14650
SP2-14650	ATGGCCACCCAGACGTCCAC			
SP3-14650	AGCACGAGACCGACGAGGAG			
ComEA-F	CGAGGATCATGGCCACGTTG	*M. luteus* trpE16 cDNA	-	ddPCR
ComEA-R	TCACTCCGCAGAGGACATGG			
14660F	GCTGAACATCTCCGGTCAGG	*M. luteus* trpE16 cDNA	-	ddPCR
SP3-14660	(detailed above)			
comPr-F	GTGGGCGATGATCTTCTCCG	*M. luteus* trpE16	-	EMSA

comPr2-R	TGATCTGCCCGCTGGAATCG			
60ORF-pMAL-F	AACCTCGGGATCGAGGGAAG GGTGCAGCGCTTTCTCACCCT	*M. luteus* trpE16	*E.coli* BL21/pMAL-c2x	Mlut_14660 protein expression
60ORF-pMAL-R	TAAAACGACGGCCAGTGCCA TCAGTGGGCCTTCGCGTCCG			

3.12 Droplet digital PCR

In order to quantify specific mRNAs of interest, a droplet digital PCR was used. After total RNA preparation and DNAase treatment, a routine Q5 PCR was carried out to check for significant gDNA traces left. cDNA synthesis followed by using iScript™ cDNA synthesis kit from Biorad. The final amount of cDNA was measured using Qubit Fluorometer with dsDNA High Sensitivity Assay Kit. Two standards were prepared with 190 µL of Qubit working solution and 10 µL of each Qubit standard. For the cDNA samples, 2 µL were used along with 198 µL of Qubit working solution. All mixtures were vortexed and incubated for two minutes at room temperature before measurement.

Samples were carefully thawed on ice, and the reaction mix was pipetted according to the scheme in Table 21. Each sample was diluted to a concentration of approximately 20,000 copies/20 µL, calculated by the formula (1). A final volume of 22.5 µL was pipetted so that exactly 20 µL of them could be mixed with QX200™ Droplet Generation Oil for EvaGreen (Biorad) as indicated in the manufacture´s manual. Between 16000 and 22000 droplets were generated per sample with a QX200™ Droplet Generator (Biorad) and transferred to 96-well plate. Before starting the PCR, the plate was sealed with a pierceable PCR Plate Heat Seal Foil (Biorad) using a PX1 PCR plate sealer (Biorad).

(1) $660\frac{g}{mol}\,x\ average\ target\ cDNA\ size(bps)$/ 6,022x$10^{23}$ mol $^{-1}$ x 10^{9} x $\frac{20000}{20}$ µL = concentration (ng/ µL)

Table 3.18. ddPCR reaction set-up.

Component	Volume µL/sample
2xQX200 ddPCR EvaGreen Supermix	11.25
10 µM primer *comEA-F*	0.1125
10 µM primer *comEA-R*	0.1125
10 µM primer *14660-F*	0.1125
10 µM primer *SP3-14660*	0.1125
cDNA sample	2.25
Ultrapure H_2O	8.55

Final Volume	22.5

Table 3.19. Program for ddPCR. The PCR program included a ramp of 2°C/s in each step.

Temperature	Time	
95°C	5 min	
95°C	30 s	
60°C	1 min	30x
72°C	3 min	
4°C	∞	

The PCR products in each droplet was analyzed using the QX200™ Droplet Reader (Bio-Rad) in combination with the QuantaSoft™ Software (Bio-Rad).

3.13 DNA Quantification

To measure nucleic acid concentration and purity, 1 µL of purified DNA or RNA sample was pipetted into NanoDrop™ ND-1000. The used wavelength for nucleic acids concentration measurement is 260 nm. To assess the purity of the samples, the ratios between the absorbance at 260 nm and the absorption wavelength (280 nm) of phenolic compounds, proteins, and other contaminants were observed. The same for absorbance at 230 nm, typical of EDTA, phenolic molecules, and/or certain carbohydrates.

An alternative method for DNA concentration measurement is based on Agarose gel electrophoresis. In parallel to 1 µL of DNA sample, 3 different measured volumes of GeneRuler™ 1 Kb DNA ladder PLUS were loaded on the same gel. After running and staining, the band intensities can be linked to the ladder known DNA amounts/concentrations. This serves as a reference to link our sample intensity to a specific DNA concentration. This method doesn´t allow purity evaluation. Bands intensities are determined by using ImageJ Image Processing and Analysis program.

3.14 Restriction digestion

DNA restriction digestion was always carried out according to the enzyme manufacturer´s instructions. For analytical purposes, 30 min incubations were the choice, and for preparative digestions, between 4 and 16 h. The enzymes were always inactivated by heat treatment and the samples were cleaned and concentrated as required. In cases of double restriction, only enzymes that have compatible buffers were selected.

3.15 Gibson DNA molecules assembly

The Gibson assembly reaction is a method that allows the ligation of two or more DNA molecules that share a common sequence at their ends. In only one step it is possible to assemble several DNA molecules in a particular order and orientation.

The first and most important step for a successful Gibson assembly is the primer design, in which the different fragments that will form the final molecule are determined. The introduction of 20 bp long overhang sequences at each site of the primers determines the molecules with which the ends will fuse. When, for example, the goal is to bind 3 DNA fragments (A, B, and C) the primers that generate fragment B should carry in their 5´-end the overhangs that match and hybridize with the 3´-end of the corresponding sequences (A and C). After all the fragments were amplified by Q5-PCR, thus, they have a 20 bp specific overlap with adjacent sections. The assembly reaction is then carried out by an enzyme mix consisting of a 5'-exonuclease, a DNA polymerase, and a DNA ligase. The 5'-exonuclease activity exposes the compatible sections of the fragments, which are then able to hybridize. Gaps at the fragment are filled by the existing polymerase and the DNA strands are finally linked by the ligase.

The assembly was carried out using the NEBuilder® HiFi DNA Assembly enzyme mix (New England Biolabs, Inc.) according to the manufacturer's instructions.

3.16 Agarose gel electrophoresis

To evaluate the length of DNA fragments in a given sample (for example, PCR amplicons or a plasmid after restriction digestion) an Agarose based gel electrophoresis was used. A solution of 0.8% Agarose in TAE-Buffer (Table 3.3) was heated up until total dissolution. The solution was stored at 60°C to prevent unwanted solidification. Between 20 and 25 mL of the liquid solution were poured on a horizontal rectangular tray and a 16 or 25-well comb was inserted to create several identical pockets. After the solidification of the gel, TAE-buffer was poured into the tray until 75% of its volume capacity was reached. The samples were loaded on the pockets after being mixed with Loading Dye (6x DNA Loading Dye, Thermo Fisher Scientific, Inc.) and a potential difference of 180 V was generated for 20 min. The GeneRuler™ 1 kb (Thermo Fisher Scientific, Inc.) was used as a molecular size reference.

After due separation of the fragments in each sample, the gels were submerged for 10 min in an Ethidium-bromide solution (1 mg mL^{-1}) and again shortly in water. Finally, the gels were exposed to UV radiation to see the stained DNA bands, and pictures were taken.

3.17 DNA sequencing methods

The DNA Sanger amplicon sequencing was performed at the GATC Biotech AG (LightRun) using LightRun barcodes. For sample preparation, 5 µL of 80-100 ng/µL of purified template DNA were mixed with 5 µL of the sequencing primer at a final concentration of 5 µM. The sequencing results were analyzed with Clone Manager Professional 9.

Whole-genome sequencing was performed on a MiSeq platform (Illumina) with standard manufacturer protocols (TruSeq LT DNA sample preparation kit, 2 × 150 bp paired-end reads) at the NGS core facility (WZW, TUM), the genomic DNA of *M. luteus* was isolated according to the CTAB extraction protocol (Section 3.10.1). Quality controls, meaning verifying the number of reads per base pair, were performed using the Artemis software. Sequence reads assembly, alignment, and SNP calling were performed using the Snippy script (https://github.com/tseemann/snippy).

To organize the data in clear plots, R Studio was utilized to generate a script that takes the mutations data as input.

3.18 Chemical random mutagenesis

For the generation of enough random mutations, cells from an O/N culture of the transcriptional reporter *M. luteus* trpE16 Δ*comEA*/*EC*:*lacZ-Kan* (Lichev *et al*., 2019) were used to inoculate LB medium with and without 5 mM of ethyl-methanesulfonate (EMS) and the antibiotic. With a starting OD_{600nm}= 0.2, the cultures were incubated O/N. The treated cells were washed twice with PBS after centrifugation. For cell recovery and SNPs stabilization, samples were inoculated in 5 mL of fresh LB^{Kan} and incubated O/N. Some aliquots with 500 µL of the O/N culture were stored in 25% glycerol at -20°C and others were plated on LB for SNP stabilization and qualitative phenotype screening (see section 3.20.1).

3.18.1 Mutagenesis efficiency analysis

Since rifampicin resistance (Rif^R) is an easily acquirable trait, that is generated by single nucleotide mutations in different positions of the *rpoB* gene (Baltz, 2014), it´s commonly used to

assess the effectiveness of mutagenic methods. Samples from the EMS mutagenesis assay were plated on LB^{Rif} 1 µg/mL to evaluate the number of Rif^{R} CFU/mL. Untreated cells were used as controls.

3.19 Genetic linkage analysis

As a form of genotype screening, a genetic linkage analysis was performed. To check each mutant isolate for linkage between its causative SNP and the 7 Kb reporter construct (ΔcomEA/EC:lacZ-Kan) genomic DNA from each mutant was prepared and used to transform naturally competent *M. luteus* trpE16. Following the transformation step, 100 µL aliquots were plated on the corresponding media along with X-Gal and Kan. Negative controls lacking DNA were always performed with the same cell suspensions. Additional control reactions were prepared, including transformations with genomic DNA from the trpE16 strain as well as from the not mutagenized reporter strain. The linkage degree (%) was calculated by dividing the number of colonies with the mutant phenotype by the ones with the reporter strain phenotype.

3.20 β-galactosidase activity screening

A few different methods were used to detect and determine the enzymatic activity of the β-galactosidase.

3.20.1 Qualitative screening

For immediate and rapid qualitative screening of β-galactosidase activity of bacterial colonies, cell suspensions were firstly diluted in subsequent decimal dilutions. Either LB or GMM agar plates were prepared with the chromogenic substrate X-Gal and Kan. To determine the best X-Gal concentration to use, a range from 10 to 200 µg/mL was assayed. One hundred microliters of each dilution were plated to assess the most convenient one, so the cells would grow in separate single colonies. After having set the variables, screenings for 2 phenotypes were carried out - one on *comEA-comEC* promoter repressing conditions i.e. LB^{Kan} + 45 µg/mL of X-Gal and the second one on normally *comEA-comEC* promoter inducing medium, GMM^{Kan} + X-Gal 80 µg/mL. Thus, two specific mutant phenotypes were looked for, one showing enhanced promoter activity on LB and the other one with no promoter activity under inducing context (GMM)

3.20.2 Quantitative screening methods

For a more specific determination of enzymatic activity, two different assays were used.

3.20.2.1 p-Nitrophenol substrate-based activity assay

Each strain was cultured in 5 mL LB O/N. After measuring the optical density, all cultures were normalized to the same OD_{600nm} =1.5 and pelleted by centrifugation. The pellets were resuspended in 2 mL of PBS and treated at 37°C for 10 min with 0.1 mg/mL of lysozyme. The now translucent suspensions were pelleted at 14000 x g for 10 min and the crude extracts were divided into two fractions of 1 mL. 50 µL of 10 µg/µL of the 4-Nitrophenyl-β-D-galactopyranoside substrate were added to each fraction. The tubes were incubated for 30 min at 30°C and all enzymatic reactions were detained simultaneously by adding 200 µL of 1M Na_2CO_3. Negative controls with *M. luteus* trpE16 wild-type and the reporter strain were prepared. Two hundred microliters aliquots of each fraction were pipetted in a transparent flat-bottom 96-well microtiter plate, and the ODs were measured at 420 and 550 nm with a microtiter plate reader FLUOstar Omega (BMG LABTECH, Germany). The time of the measurement was taken into consideration for calculating the activity units by using Miller´s equation:

(2) Activity unit = $1000 \frac{OD420-(1.75*OD550)}{Time\ [min]*Volumen\ [mL]*OD600}$

OD_{420} is the absorbance of the yellow p-nitrophenol,

OD_{550} is the scatter from cell debris, which, when multiplied by 1.75 approximates the scatter observed at 420nm.

3.20.2.2 4-methylumbelliferyl- β-D-galactopyranoside substrate-based activity assay

Precultures of all the evaluated strains containing the transcriptional reporter Δ*comEA/EC*:*lacZ* were grown overnight in LB. Adequate aliquots of the cell suspensions were harvested and washed with PBS by centrifugation at 14000 x g for 10 min. All samples were normalized so that they would have the same OD_{600nm} at 1.5. All suspensions were treated with 0.1 mg/mL of Lysozyme for 10 min at 37°C. Later, 200 µL aliquots were transferred to a transparent flat-bottom 96-well microtiter plate. For the measurement of LacZ activity, β-galactosidase substrate 4-methylumbelliferyl β-D-galactopyranoside (MUG) (Sigma-Aldrich, USA) was added to each well at a final concentration of 250 µg/mL. Kinetic measurements were performed by incubating the cells in a FLUOstar Omega (BMG LABTECH, Germany) microtiter plate reader at 30 °C for 10 hours and measuring MUG fluorescence every 10 minutes at 460 nm. The wild-type strain of *M. luteus* was incubated with and without a substrate to establish the background fluorescence. There were at least 3 biological replicates and 3

technical ones for every strain assayed. Controls without cells and substrate were made to evaluate background noise. To determine the relative activity of the promoter of *comEA/EC*, a four-parameter logistic function was fitted to the measured kinetic curves. The linear range of the kinetic curves before substrate depletion was used for evaluation. This was done by calculating the steepest slope within the linear range of the curves which corresponds to the maximal velocity of the enzymatic reaction and is directly proportional to the enzyme concentration i.e. to the promoter activity.

3.21 Single Nucleotide Polymorphism (SNP) calling and analysis

All the mutations that were not in the reference genome (from GeneBank) but were present also in *M. luteus* trpE16 strain were filtered out. Every SNP not present in the reference was highlighted and the number of SNPs per mutant was counted. A visual clustering analysis was done, by highlighting groups of mutations within the same region or mutations present in more than one mutant. Also, mutations affecting interesting genes, like ones coding for transcriptional regulators or proteases, were marked. After the identification of all the potentially interesting mutations, the mutation type (missense, nonsense, or synonym), the original and new aminoacids, the affected genes annotations, and the gene orientation were checked.

An R script was written by Dr. Angel Angelov (see Supplementary material) taking the mutations table as input. Thus, a scattered plot, which has all the 2.5 million positions of the *M. luteus* genome on the X-axis and each mutant on the Y-axis with all its SNPs in the same line. Simultaneously, a programmable "sliding window" function with changeable width (e.g. 1000 bp) was built. After it is set, it moves through the genome detecting SNPs and the distance between them. It calculates a clustering score for each position in the genome, according to the amount of SNPs in the covered area and the distance between them. This script enables the recognition of regions with a higher density of mutations.

3.22 *E. coli* transformation

3.22.1 Preparation of chemically competent *E. coli*

To obtain competent cells from *E. coli,* 10 to 12 isolated colonies grown O/N were inoculated in 250 mL of SOB medium in a 1 L Erlenmeyer flask. The incubation occurred at 19°C with 120 rpm shaking, until the culture reached an OD_{600} = 0.5 (around 36 hours). The flask with the cells was placed on ice for 10 min, and then the cooled cells were pelleted at 4°C by centrifugation at 4000 x g. The pellets were gently resuspended in 80 mL of ice-cold TB-solution and stored for 10 min on ice. Another

10 min centrifugation step was carried out at 4°C and 4000 x g. Afterward, the cells were resuspended gently in 20 mL ice-cold TB-Solution with 1.4 mL DMSO. The cells suspension was fractioned in aliquots of 50 µL and stored at -80°C.

3.22.2 Heat shock transformation

Previously prepared competent cells were taken from storage at -80°C and placed 10 min on ice. After that time, between 0.1 and 100 ng of plasmid were pipetted into the cells suspension. The mixture was incubated for 30 min on ice and then placed for exactly 1 min in a 42°C water bath. Immediately after the heat shock, they were placed again on ice for 5 min. LB-medium (950 µL) was added to the cells, and they were incubated for 1 h at 37°C and 180 rpm. Finally, 50 µL of the suspension were plated on LB agar with the corresponding antibiotics for selective pressure. The remaining culture was distributed on a second plate in case the overall efficiency has not been high. The plates were incubated O/N at 37°C.

3.22.3 Electroporation

Alternatively, *E. coli* was also transformed by electroporation. For this, following the growth of *E. coli* XL1 on LB agar, 1 or 2 isolated colonies were inoculated in a 15 mL tube containing 1 mL of LB medium. The tube was placed on a shaker at 180 rpm and 37°C O/N. The next day, a tube with 1.4 mL of fresh liquid LB was inoculated with 30 µL of the O/N culture and incubated for 3 h at 100 rpm and 37°C. Afterward, the cells were centrifuged for 30 s at 11000 x g in a cooled benchtop centrifuge at 2°C. The supernatant was discarded and the pellet placed on ice. 1 mL of ddH20 was used to resuspend the cells and the following centrifugation and resuspension steps were repeated. Twenty to 30 µL of supernatant were left in the tube along with the pellet. The tube was vortexed on kept on ice. One microliter of DNA was added to the cells and mixed by pipetting. The mixture was transferred to a chilled and clean electroporation cuvette with an electrode spacing of 1 mm (PeqLab Biotechnology GmbH, Erlangen), and then placed in the Bio-Rad Gene Pulser IITM Electroporation System. The electroporation parameters used were 2.5 kVmm, 1.25 µFarad, and 200 ohms with time constants around 5 ms. Immediately thereafter, 1 mL of fresh LB without antibiotics was added to the cuvette, and the whole mixture was poured in a 1.5 mL microfuge tube. Lastly, a recovery incubation was carried out at 37°C for 70 min and 1000 rpm. In the end, cells were streaked out on LB with the corresponding antibiotic and incubated O/N at 37°C.

3.23 Natural competence induction of *M. luteus*

Overnight precultures grown in LB medium were washed with PBS and used to inoculate 25 mL of fresh GMM medium with a starting OD_{600nm}=0.2. By the time the culture OD reached 1 (around 20 h), 2 mL were taken in small tubes and pelleted at 10000 x g and 4°C for 8 min. The cell pellets were resuspended in 500 µL of transformation buffer I. In parallel 500 µL of transformation buffer II were mixed with the DNA to be uptaken. The DNA solution was added to the cell suspension and the mixture was incubated for 30 min at 30°C and 180 rpm shaking. The transformation was stopped by placing the tubes in ice. For better quality of CFUs, the suspension was firstly sonicated for 1 min with 30% amplitude and 0.25 duty cycle. This disrupts cell aggregates and allows better plating of the individual cells. Dilutions were made when required and the suspension was plated in a suitable solid medium (CAH^-). Negative controls without DNA were always performed in parallel.

3.24 Natural transformation frequency analysis of *M. luteus*

For transformation frequency assays, cells from an overnight culture grown in full medium were transferred to 25 mL GMM (with the appropriate antibiotic supplementation if needed) to give an initial OD_{600nm}= 0.2. After 20 hours of incubation at 30 °C in a shaking incubator (180 rpm), the OD_{600nm} was determined to serve as a marker for the physiological state of the cells. 2 mL of the GMM cultures were pelleted by centrifugation at 10000 × g for 8 min at 4 °C. The supernatant was discarded and the cells were resuspended in 1 mL transformation buffer (100 mM $CaCl_2$, 50 mM Tris-HCl, pH 7.2). Next, 300 ng of a plasmid carrying the wild-type *trpE* allele from wild-type *M. luteus* ATCC 27141 (pJET-*trpE*, constructed by cloning of the wild-type *trpE* gene in pJET in *E. coli*) was added to the cell suspension. If internalized by natural transformation, this DNA is suited to allow conversion of trpE16 tryptophan auxotrophic cells back to prototrophy. Following incubation for 30 min at 30 °C in a shaking incubator, the transformation reaction was stopped by placing the cells on ice. The cell suspensions were sonicated for 1 min with 30% amplitude and 0.25 duty cycle to gently disrupt cell aggregates before plating on CAH^- agar plates to score transformants and on LB plates to determine the total viable cell counts. Control reactions performed without the addition of DNA delivered the rate of spontaneous reversion to prototrophy which determined the detection limit of the assay.

3.25 Gene knockout by kanamycin cassette exchange through HR

A relatively easy and direct way of evaluating the effect of the loss of a certain gene function is to make an exchange of most of the gene of interest by a Kan^R cassette (911 bp) through homologous recombination (HR). The Kan^R cassette was obtained by PCR amplification with the primers KanF2 and

KanR3 (see Table 3.17) and the plasmid pWLTK6 as DNA template. In parallel, and according to the specific sequence that should be deleted, 2 pairs of primers must be designed so that they would yield fragments identical to the immediately flanking sequences. After amplification of both flanking regions and the Kan^R cassette, they were ligated through Gibson assembly (see section 2.15). The final linear fragment holding the Kan^R cassette in the middle of the flanking sequences was used as DNA template (2-5 µL of assembly product) on the naturally competent *M. luteus* (see section 2.22). The transformants were plated on LB supplemented with the antibiotic as a selective marker.

3.26 *codAB*-based markerless modification system

By using the *codAB*-based markerless modification system (Kostner, Peters, Mientus, Liebl, & Ehrenreich, 2013) it is possible to introduce clean gene deletions and single changes in specific sequences of interest (see Figure 3.1). Each insert was built by amplifying either the mutated region from a certain mutant or the upstream and downstream flanking sequences of a gene. All amplicons inserts had sizes between 1 – 2 kbp and were cloned in the linearized pKOS6b vector (6661bp) through in vitro assembly of equimolar amounts in a 20 µL Gibson Assembly reaction. Competent cells of *E. coli* XL1 were transformed with the Gibson assembly through heat shock and transformants were selected for on LB^{Amp}. After verification of the plasmid by restriction digestion or Sanger sequencing, it was prepared and used to directly transform *M. luteus* competent cells. The transformation with the plasmids resulted in the generation of kanamycin-resistant and 5-fluorocytocine (5-FC) sensitive mutants. 5-FC is a non-toxic pyrimidine analog that is converted by the cytosine deaminase (encoded by *codA* gene) to a toxic compound 5-fluorouracil (5-FU) (Neuhard 1983, Martinussen and Hammer 1994). The gene *codB* encodes for a cytosine permease that enhances the penetration of 5-FC inside the cell. Transformants were selected for on LB^{Kan} plates. For the counter selection, 15 colonies were harvested and cultured at 30°C for 4 h in 1 mL LB media. Culturing transformants under a non-selective environment facilitates the events of a second homologous recombination to excise the integrated plasmid. The plasmid excision will result in the generation of 5-FC resistant strains that can easily be selected on a growth medium containing the analog. Thus, appropriate dilutions of the cell suspension were plated on LB supplemented with 500 µg/mL 5-FC and incubated at 30°C for 72h. Plasmid excision can generate two types of mutants, having either the wild-type copy of chromosome or carrying the desired insert within the plasmid. The distribution of both types is a random process and requires additional tests for confirmation. The deletion of desired genes was verified by locus amplification and amplicon length comparison to the wild-type sequence. In the case of SNP insertions, the regions encompassing them were amplified and sent for sequencing.

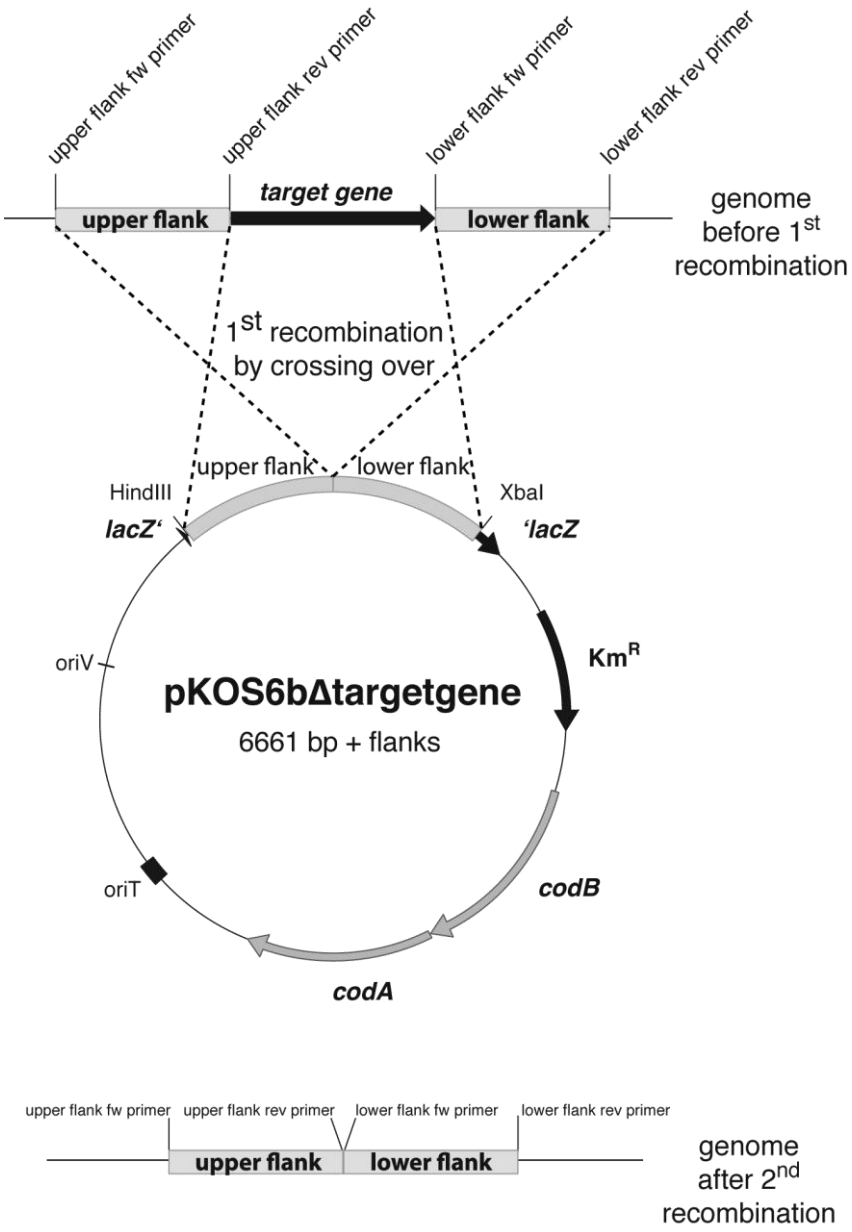


Figure 3.1. Vector map of pKOS6bΔtarget gene used for clean deletions and genome editing via double-crossing over.

Homologous recombination occurs via ~1 kbp regions flanking the target gene or the genomic position to be changed. After a second recombination event, there are 50% chances to obtain the new sequence in the genome, either the deleted locus or the new nucleotide. Annotated in the exemplary vector are the origin of replication (oriV), origin of transformation (oriT), cytosine deaminase gene (*codA*), cytosine permease gene (*codB*), both restriction sites (XbaI, HindIII) found in the multiple cloning site within *lacZ* gene and used for plasmid digestion, two representative flanks (upper, lower), and the kanamycin resistance cassette (Km^R). This Figure was extracted from the master thesis titled "Application of *codBA*-based counterselection for genome modification in high-GC Gram-positive bacteria" carried out under the supervision of this doctoral thesis author.

3.27 Databases and online tools

Table 3.20. List of online bioinformatic tools used in this work.

Program	Function	Link
Blast	Basic local alignment search tool.	https://blast.ncbi.nlm.nih.gov/Blast.cgi
HMM logo	Quick overview of the properties of an HMM in a graphical form.	http://pfam.xfam.org/family/PF12728#tabview=tab4
RBS calculator (DeNovo DNA)	Predict a mRNAs Translation Initiation Rate in Bacteria	https://salislab.net/software/predict_rbs_calculator
RNA secondary structure prediction	Building of highly probable, probability annotated group of secondary structures.	https://rna.urmc.rochester.edu/RNAstructureWeb/
Microbial genomic context viewer	Interactive gene-context visualization tool for genome small scale analysis.	https://mgcv.cmbi.umcn.nl/

3.28 Rapid amplification of cDNA ends (5´RACE)

Total RNA samples were used as a template for reverse transcription into cDNA. This was achieved by using the iScript™ cDNA synthesis kit (Bio-Rad Laboratories, USA) in a 20 µL reaction containing 1 µg RNA. To determine the length and sequence of specific mRNAs, 5´/3´RACE Kit 2nd Generation (Sigma-Aldrich, USA) protocol was followed. High Pure PCR product purification kit (Sigma-Aldrich, USA) was used when required and the final products were sent for Sanger amplicon sequencing.

3.29 pMAL™ protein fusion and purification system

An MBP fusion-based pMAL-vector system (New England Biolabs, Ipswich, MA) was harnessed. The 226 bp PCR-amplified ORF of Mlut_14660 was introduced into the double digested (HindIII, XmnI) pMAL-c2x (6607 bp) by mixing equimolar amounts of both insert and vector in a Gibson assembly reaction. The Gibson product was used to electroporate *E. coli* XL1. Ampicillin-resistant colonies were cultured and used for high concentration plasmid preparation (Monarch Plasmid Miniprep Kit, New England Biolabs). After due restriction digestion verification, the plasmid pMAL-c2x+Mlut_14660ORF was electroporated into *E. coli* BL21.

The pMAL-c2x protein expression was done according to the manufacturer´s instruction (pMAL Protein Fusion and Purification System Manual, New England BioLabs). After IPTG induction, the bacterial culture was incubated for 4 hours at 37°C and 120 rpm. Cells were pelleted at 4500 x g and

4°C for 20 min, 1.5 g were resuspended in 8 mL of column buffer and sonicated in a water-ice bath for 7 min with 50% amplitude and 0.4 duty cycle. After the protein concentration stopped raising the suspension was centrifuged for 5min at 14000 x g. The supernatant was purified using the MBP-Spin Protein Miniprep Kit (Zymo Research). Parallel to the fusion protein-encoding vector, *E. coli* with the empty pMAL-c2x without Mlut_14660 was processed in the same way. Each P1-Spin-column (Zymo Research) was packed with 300 µL of amylose resin (NEB). The columns were washed with 500 µL of column buffer, followed by 800 µL of the crude extract. After loading, the crude extract was incubated for 3 min and then washed with 500 µL of column buffer. The final elution was carried out with 500 µL of elution buffer. Between each step, there was a centrifugation of 1 min at 4°C and 13000 xg. Protease Factor Xa was used (10 µg/mL) for cutting the fused proteins for 5 hours at room temperature. The final protein concentration was measured using the Bradford reagent.

3.30 Protein concentration determination by Bradford

The Bradford assay can be utilized to establish the protein concentration from a sample in the mg/mL range with a photometer. The Coomassie brilliant blue dye forms complexes with the nonpolar and cationic residues of the proteins and thereby shifts the absorption maximum from 479 nm to 595 nm. The color change depends on the amount of protein, so a calibration curve created with the standard protein bovine serum albumin (BSA), was necessary for a correct determination. Five microliters of the sample were placed in a 1 mL cuvette and it was filled up to 1 mL with Bradford reagent. After mixing, the samples were incubated for 5 minutes (protected from light) and then the OD_{595} was measured. One microliter of pure Bradford reagent was used as a reference. To obtain the protein concentration in µg/µL, the following was then divided by the sample volume used:

Protein concentration (µg/cuvette) = OD_{595} * 164

(3) Protein concentration (µg/µL) = $$\frac{\text{Protein concentration}\left(\frac{\mu g}{\text{cuvette}}\right)}{\text{Sample volumen } (\mu L/\text{cuvette})}$$

3.31 Polyacrylamide Gel Electrophoresis (PAGE)

For the separation of proteins between 10 and 150 KDa, Tris-glycerol-SDS-polyacrylamide gels were prepared according to the pipetting scheme in the table below. The gels were poured using glass chambers from Biorad in such a way that the collecting gel was approx. 3 cm long. The solidified gels were clamped in the Mini-PROTEAN® Tetra System gel devices. The protein samples were mixed with 4x SDS loading buffer and denatured at 96°C for 5 minutes. After briefly cooling, they were then pipetted into the pockets of the SDS gel. PageRuler Plus™ Prestained Protein Ladder from Thermo

Scientific was used as the size standard. Gel electrophoresis was started at 80 V until the bands at the bottom of the collection gel formed a straight line. It was then increased to 120 V until the lowest band of the marker had reached the lower edge of the gel. The gel was then carefully removed from the chamber and stained O/N at room temperature in a Coomassie staining solution (25% isopropanol, 10% acetic acid, 0.5% Coomassie Blue) and then decolorized with 10% acetic acid until clear bands were visible.

Table 3.21. Composition of 2 polyacrylamide gels for middle-range size proteins.

Component	Resolving gel (10%)	Stacking gel (5%)
Separating gel buffer	2.5 mL	-
Stacking gel buffer	-	1.2 mL
Acrylamide/Bisacrylamide (29:1)	2.5 mL	0.625 mL
H_20	5 mL	3.175 mL
APS 10%	60 µL	40 µL
TEMED	30 µL	20 µL

3.31.1 Low range PAGE (Tris-Tricine)

For the separation of proteins between 3 and 50 KDa, Tris-Tricine-polyacrylamide gels were prepared according to the pipetting scheme in the table below. Apart from the gel composition, all steps described in section-3.31 were the same. The PageRuler™ Unstained Low Range Protein Ladder was used as a molecular size standard.

Table 3.22. Composition of 2 polyacrylamide gels for low range size proteins.

Component	Resolving gel (18%)	Stacking gel (5%)
Ethylene glycol	1.3 mL	-
3 M Tris-HCl buffer pH 8.4	4 mL	1.2 mL
Acrylamide/Bisacrylamide	(19:1) 5.1 mL	(29:1) 0.625 mL
H_20	1.6 mL	3.175 mL
APS 40%	20 µL	10 µL
TEMED	54 µL	20 µL

3.31.2 Silver staining

Silver staining is a process for making proteins, carbohydrates, and nucleic acids visible in a polyacrylamide gel. With this method, a much higher sensitivity (detection limit: 0.1 ng - 1 ng/band) is achieved compared to Coomassie-stained gels or other protein stains. The silver ions in the silver nitrate solution attach to the negatively charged residues of the proteins and are reduced to elemental silver by the subsequent addition of formaldehyde. This colors the proteins in the gel black.

After the polyacrylamide gel has been run, the proteins were fixed by incubation of the gel slab in a fixation solution (50:5:45 v/v/v methanol: acetic acid: water) for 30 min. After having rinsed the gel with water (2 changes, 2 minutes per change) it was left further in water for 1 h on a shaking platform. A solution with 0.02% sodium thiosulfate was used to sensitize the gel for 2 min. The solution was then discarded and the gel slab quickly rinsed with two changes of water (10 s each). After this, the gel was immersed in chilled 0.1% $AgNO_3$ for 30 min. After two changes of water, the gel was developed with a developing solution composed of 0.04% formaldehyde in 2% $NaCO_3$ until a sufficient degree of staining was achieved. Then, the developing solution was exchanged with 1% acetic acid to quench the staining. The gel was stored at 4°C.

3.32 Chromatography

After overexpression of the fusion protein Mlut_14660-MBP, the MBP was cut with the Factor Xa protein according to the manual. For the purification of the Mlut_14660 encoded protein, chromatographic separation was attempted. The target protein was purified using the ÄKTA™ explorer system (Amersham Pharmacia Biotech, Uppsala, Sweden) with a P-920 pump, a UPC-900 monitor, an INV-907 distributor, and a Frac-900 fraction collector. The Äkta could be operated via the UNICORN software and the running conditions were set with the WIZARD program. Interfering particles were removed from samples, solutions, and buffers using a filter (Ministart 0.2 µm pore size, Sartorius-Stedim, Göttingen). Pipes were connected through constant liquid contact, also known as “drop-to-drop”, in order to prevent air bubbles from entering. Between the buffer change from ethanol to the running buffer, the system was always rinsed with filtered MilliQ water. After use, the column should be rinsed with 0.5 mM NaOH for regeneration and with 20% ethanol (v/v) for storage.

A Hi Load 16/60 Superdex 200 column (GE Healthcare) with buffer A (50 mM Tris-HCl; 150 mM NaCl; pH 7.5) was used for protein purification by size exclusion. The flow rate of the column was set to a constant 1 mL/min and first, a column volume (approx. 120 mL) was equilibrated with filtered MilliQ water and then with buffer A. The sample was pumped into the column through a 1 mL loop

and the eluates were collected in 2 mL fractions by the fraction collector. The proteins with a larger molecular weight are eluted first because they can flow more easily through the column material, while smaller proteins penetrate the pores of the column material and it takes longer for them to travel through the column.

3.33 Electrophoretic mobility shift assay (EMSA)

For evaluating protein-DNA interaction and comparing affinities, an EMSA assay was set up for our system. Different amplicons were generated by colony PCR using *M. luteus* trpE16 as a source of template. They were purified using Clean & Concentrator Kit (Zymo Research) and then adjusted to a final concentration of 4 ng/µL. Tris-borate buffer was used as a solvent for agarose solution, as running buffer and, as the basic component of the binding buffer. Agarose gel solutions (2%) were prepared. To ensure uniform depth of the viscous high-percentage gels, 7x10 cm gel trays were pre-incubated at 60°C while the gel solution was prepared. From the solution, 22 mL were poured on the tray, a 25-well comb was inserted. The tray was left for cooling until the gel was formed. To estimate the dissociation constant (Kd), binding reactions were assembled on ice in 20 µL according to what is detailed in TTT, adding each component in the order listed from top to bottom. After mixing all components, 10 µL of each mixture were pipetted in the gel.

Table 3.23. EMSA binding reactions set-up 1 (broad range).

Components	1	2	3	4	5	6	7	8	9	10	11	12	13	14
DNA [4ng/µL]	1	0	1	1	1	1	1	1	1	1	1	1	1	1
5x Binding buffer	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Glycerol [50 %]	2	2	2	2	2	2	2	2	2	2	2	2	2	2
MBP-Mlut_14660 [2.5 mg/mL]	0	4	0	4	3.5	3	2.5	2	1.5	1	0.5	0	0	0
MBP- Mlut_14660 [0.25 mg/mL]	0	0	0	0	0	0	0	0	0	0	0	2.5	2	1.5
MBP [2.5 mg/mL]	0	0	4	0	0	0	0	0	0	0	0	0	0	0
Column buffer	13	10	9	9	9.5	10	10.5	11	11.5	12	12.5	10.5	11	11.5
Final Protein conc [µM]	0	9.8	9.8	9.8	8.6	7.4	6.1	4.9	3.7	2.5	1.2	0.6	0.4	0.3

Table 3.24. EMSA binding reaction set-up 2 (low range).

Components	1	2	3	4	5	6	7	8	9	10	11	12	13	14
DNA [4ng/µL]	1	0	1	1	1	1	1	1	1	1	1	1	1	1
5x Binding buffer	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Glycerol [50 %]	2	2	2	2	2	2	2	2	2	2	2	2	2	2
MBP-Mlut_14660 [0.509 mg/mL]	0	4	0	4	3.75	3.5	3.25	3	2.75	2.5	2.25	2	1.75	1.5
MBP [0.59 mg/mL]	0	0	4	0	0	0	0	0	0	0	0	0	0	0

Column buffer	13	10	9	9	9.25	9.75	10	10.3	10.5	10.8	11	11.3	11.5	11.8
Final Protein conc [μM]	0	2	2	2	1.88	1.75	1.63	1.5	1.38	1.25	1.13	1	0.88	0.75

4 Results

4.1 Chemical random mutagenesis with ethyl methanesulfonate of the *M. luteus comEA/EC-lacZ* transcriptional reporter strain.

To expand our knowledge about how the *comEA*/*EC* promoter is regulated, a LacZ reporter strain was constructed. The β-galactosidase gene was placed under the control of the promoter of interest (see Figure 1.4.) in the parental strain *M. luteus* trpE16, yielding a strain which can be used to analyze with an assay with colorimetric readout under which conditions and when the *comEA*/*EC* promoter is more or less active.

To localize one or more elements involved in the regulation of the *comEA/EC* Promoter, random chemical mutagenesis was carried out on cells of the transcriptional reporter strain *M. luteus* Δ*comEA*/*EC*:*lacZ-Kan* (see Methods section 3.18). To evaluate if the mutagenesis had been successful, a mutagenesis efficiency analysis was performed (see Methods 3.18.1) by detecting the numbers of rifampicin-resistant colonies for the treated and untreated cultures. EMS-treated samples showed 2.8 x 10^3 CFU/mL capable of growing on 1 µg/mL of rifampicin, while the EMS-untreated controls with the reporter strain and the wild-type yielded on average less than 3 CFU/mL. In all assayed cultures, the total CFU/mL was about 10^8 when grown without any antibiotics. The $\log^{10}$ of each CFU/mL was calculated and plotted in Figure 4.1. It can be seen that there is a more than 100-fold difference of the rifampicin resistant cell concentrations between the EMS treated *lacZ* reporter strain sample and the untreated one, while there was no significant difference between both untreated reporter strain and wild-type. By calculating the relation between the number of rifampicin resistant CFU/mL and the total number of CFU/mL for each assayed strain, it was possible to obtain the frequency of the rifampicin resistance trait appearance. The $\log^{10}$ transformed values were plotted in Figure 4.2. The frequency of resistance appearance when cells were exposed to EMS was around 3 $\log^{10}$ scale units higher than when cells were not.

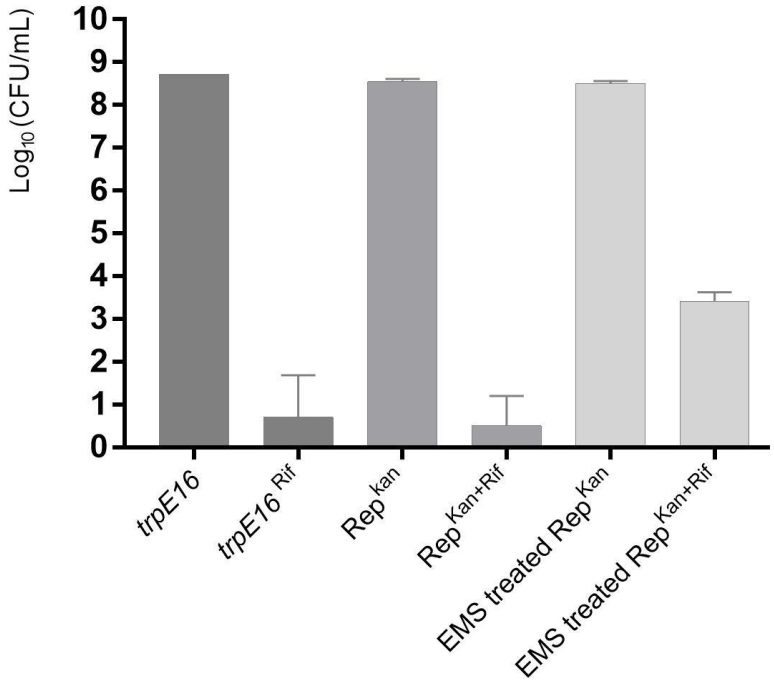


Figure 4.1 Log10 transformed numbers of rifampicin-resistant cell concentrations obtained with and without EMS treatment.
Appropriate dilutions of samples of both the wild-type (trpE16) and the *lacZ* reporter strain were plated on LB along with the EMS treated *lacZ* reporter. The medium with the lacZ reporter strain additionally contained kanamycin. The error bars show the standard deviation of three biological replicates. An unpaired, two-tailed Student´s test showed a significant difference between the mean values of the treated and untreated cultures ($p<0.05$, $n>3$), meaning a higher mutation rate for the EMS-treated cells.

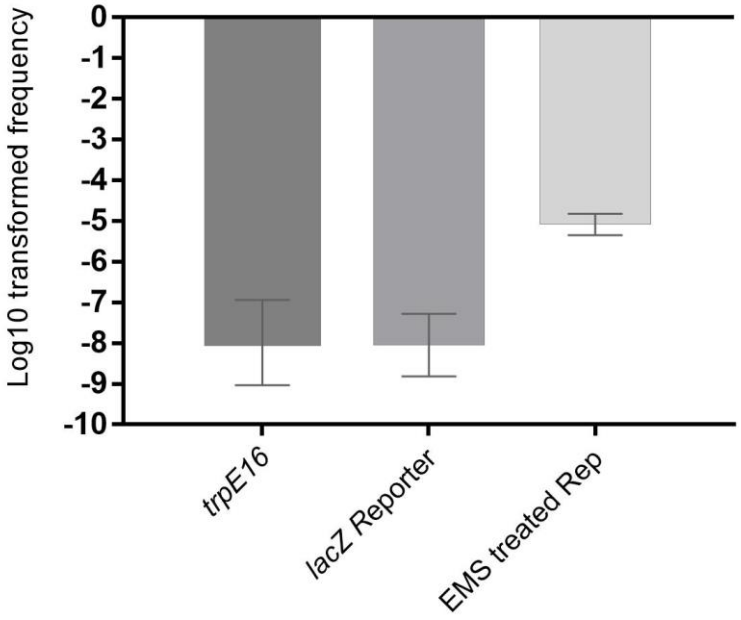


Figure 4.2 Frequency of rifampicin resistance appearance.

Frequency of Rif^R appearance of the EMS untreated trpE16 and the treated and untreated *lacZ* reporter strain; There is around 3 log^{10} scale difference between the frequency of the treated and none treated cultures. An unpaired, two-tailed Student´s test showed a significant difference between the mean values of the treated and untreated cultures ($p<0.05$, $n>3$).

To adjust an optimal X-Gal concentration for mutant screening, different amounts of the chromogenic LacZ substrate were mixed with the screening media, LB, and GMM. Two different types of mutants were of interest, the so-called "M-mutants" which showed enhanced activity on a normally competence-repressive full medium, and the "R-mutants", showing no activity when grown on a normally competence-inducing minimal medium (see Figure 4.3). For spotting the M-mutants, a concentration of X-Gal was needed at which only a relatively high β-galactosidase activity would yield blue color. On the other hand, for the R-mutants screening, a higher X-Gal concentration was required, at which only mutants with a significantly repressed *lacZ* wouldn´t yield blue color. Around 50 µg/mL of X-Gal in LB showed a very faint blue color when the *lacZ* reporter strain was streaked out on it. So, to increase the contrast of the screening, 45 µg/mL were chosen. For the same X-Gal concentration on GMM, there was a strong release of the chromogenic molecule. Therefore, an even higher but still not growth inhibitory concentration was chosen (80 µg/mL). The LB and GMM plates with X-Gal 10 µg/mL, 50 µg/mL, and 200 µg/mL streaked out with the *lacZ* reporter strain can be observed in Figure 4.4.

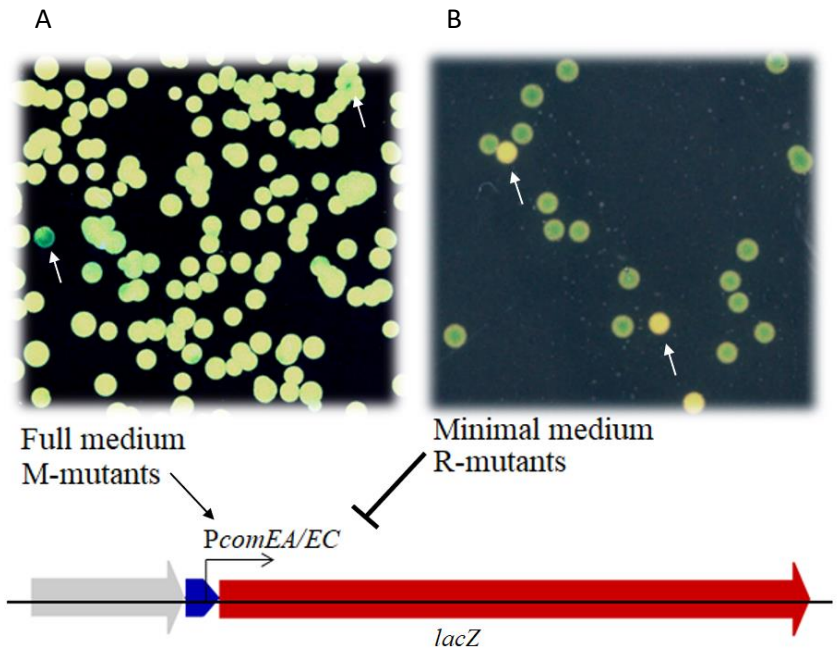


Figure 4.3 Screening of mutants with comEA/EC deregulated promoter.
On the left panel (A) some mutants showing higher enzymatic activity, meaning an increase in the promoter activity under usually repressive conditions (LB full medium). The mutants with this phenotype were called M-mutants. On the right panel (B), plated on inductive minimal media (GMM), some mutants with no enzymatic activity are present. These were named R-mutants (Torasso Kasem *et al*., 2021).

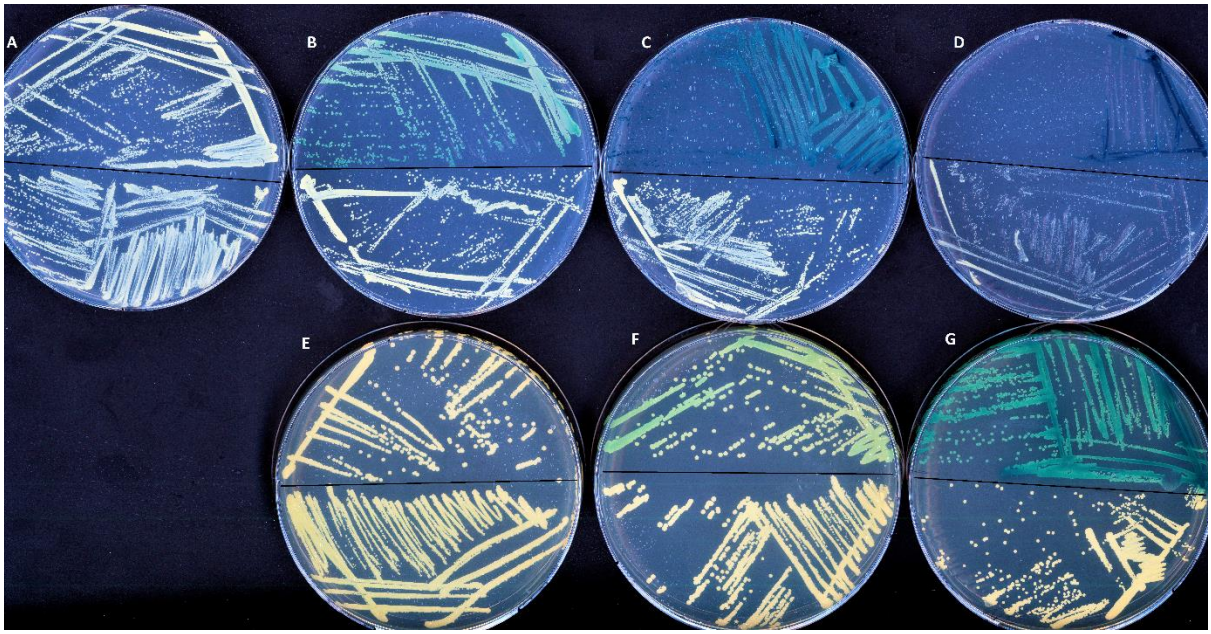


Figure 4.4 X-Gal concentration adjustment for screening for *comEA/EC* promoter dysregulated mutants. The upper half of each plate shows the *M. luteus* Δ*comEA*/*EC*:*lacZ-Kan* reporter strain, and the lower half of the plates the trpE16 wild-type strain as no enzymatic activity control. (A) Negative control with no X-Gal in GMM. (B) 10 µg/mL X-Gal in GMM. (C) 50 µg/mL X-Gal in GMM. (D) 200 µg/mL X-Gal in GMM. (E) 10 µg/mL X-Gal in LB. (F) 50 µg/mL X-Gal in LB. (G) 200 µg/mL X-Gal in LB.

Decimal serial dilutions of the chemically mutagenized *M. luteus* Δ*comEA*/*EC*:*lacZ-Kan* cultures were prepared so that they would yield isolated colonies when plated. More than 100 colonies showing enhanced β-galactosidase activity under competence-repressing conditions and around 80 colonies with no β-galactosidase activity under competence-inducing conditions were isolated and repeatedly re-streaked to obtain stable phenotypes (see Figure 4.5). The mutants were labeled either M or R depending on their phenotype, followed by the isolate number e.g. R25. After obtaining stable and pure isolates, two different tests were carried out on them.

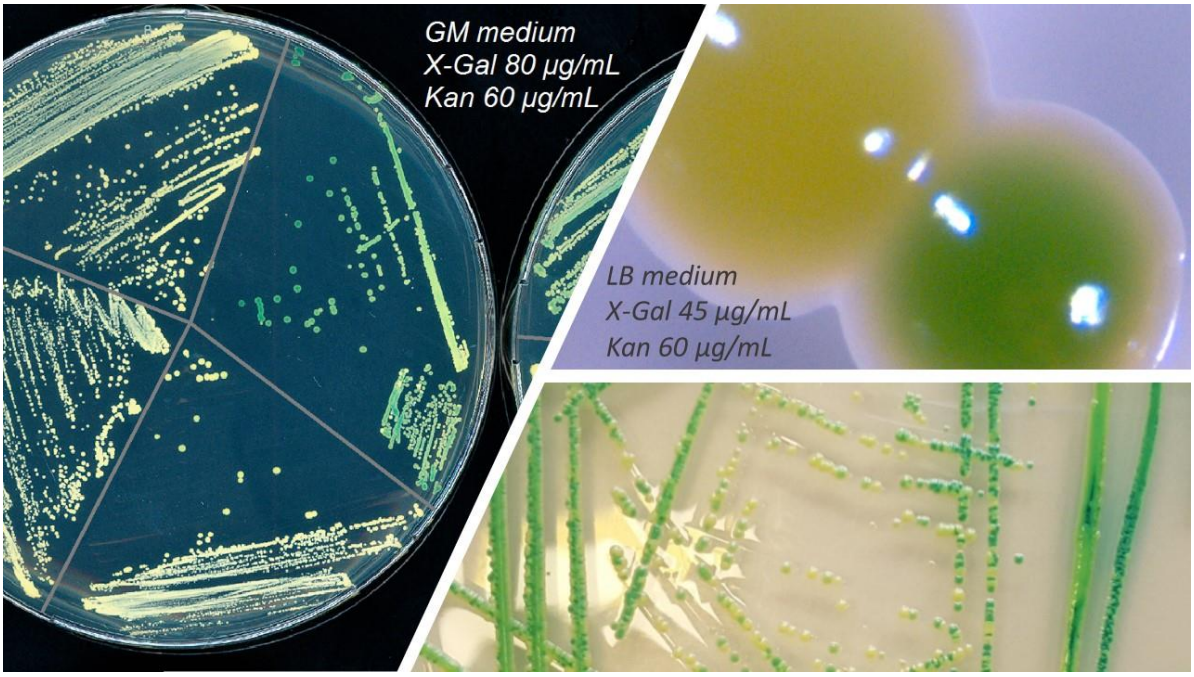


Figure 4.5 *M. luteus* reporter strain R and M-mutants.
On the left image, some R-mutants showing a range from none to low β-galactosidase activity. On the right images, M-mutants (green) not completely isolated from strains with the usual β-galactosidase activity levels.

4.2 Mutant screening

4.2.1 M-mutant screening

First, stable mutants with unusually high expression levels of the *comEA*/*EC* promoter were selected for further work. These M-mutants were subjected to a following assays in order to select the best candidates for whole-genome sequencing: first, a β-galactosidase activity assay based on the chromogenic substrate 4-nitrophenyl-β-D-galactopyranoside was carried out, and thereafter, a genetic linkage analysis was performed. In this way, it was possible to select for the strongest phenotypes, in particular also for those whose mutations causing the phenotype were not present within the 7 kbps Δ*comEA*/*EC*:*lacZ-Kan* region.

The activity values were plotted as arbitrary units calculated according to the methods described in section 3.20.2.1 (see Figure 4.6). In parallel, the genetic linkage analysis was carried out for all the mutants according to the protocol in methods section 3.19. Every isolate with more than 3% linkage was left aside (see Table 4.1). Of all 52 of the analyzed M-mutants, 7 showed a linkage percentage higher than 98% and only one a linkage percentage of 36%. Among the mutants with less than 3% of linkage, the 22 with the highest activities were selected for further work.

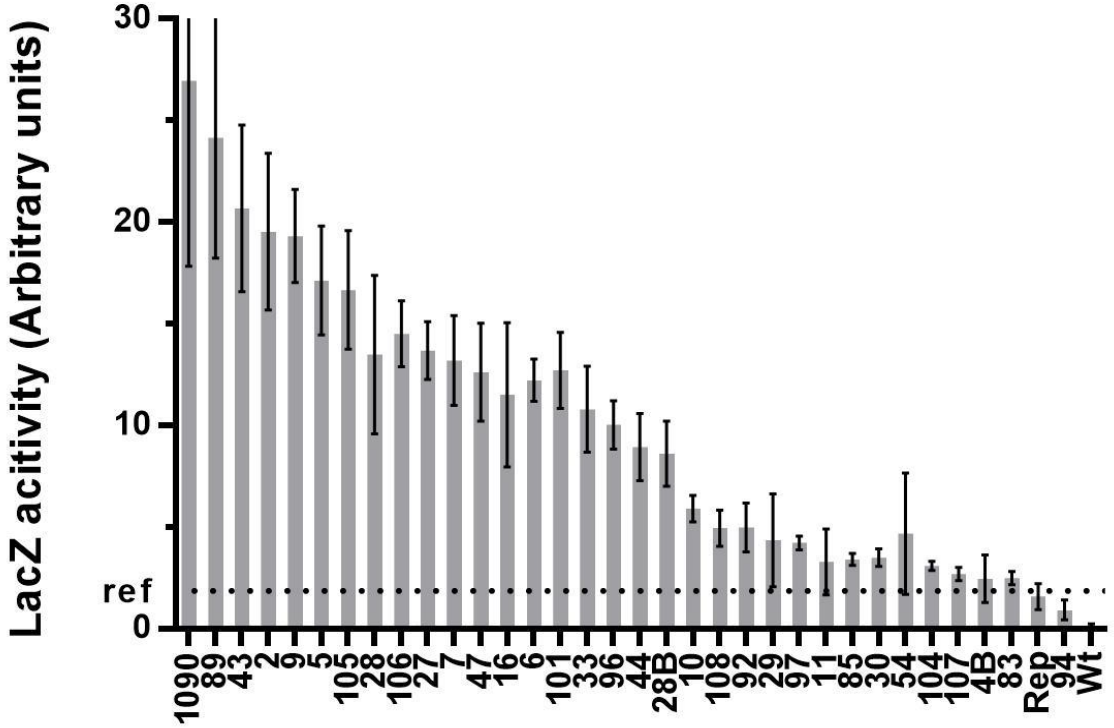


Figure 4.6 β-Galactosidase activity assay of the M-mutants based on the chromogenic substrate 4-nitrophenyl-β-D-galactopyranoside.

The *M. luteus* trpE16 wild-type, the Δ*comEA/EC:lacZ* reporter strain, and all the M-mutants were grown and normalized by OD. After centrifugation and lysozyme treatment, the crude extracts were mixed with 0.5 µg/mL of the chromogenic substrate 4-nitrophenyl-β-D-galactopyranoside. The reactions were incubated for 30 min at 30°C before the addition of Na_2CO_3. Aliquots of each fraction were pipetted in a 96-well microtiter plate and the ODs at 420 nm and 550 nm were measured. The activity units were calculated based on the Miller assay´s equation. All mean values and standard deviations were plotted. The wild-type and the reporter strain were used as negative control and common activity reference respectively. Rep: *M. luteus* Δ*comEA/EC:lacZ* reporter strain; Wt: wild-type.

Table 4.1. M-mutants linkage assay results.

Mut N°	Total transformants	White CFU	Blue CFU	% linkage
M43	41	41	0	0.00%
M6	17	17	0	0.00%
M16	23	23	0	0.00%
M92	80	80	0	0.00%
LacZ rep	8000	8000	0	0.00%
M94	3124	3124	0	0.00%
M4B	25	25	0	0.00%
M87	672	672	0	0.00%
M33	5088	5088	0	0.00%
M108	20	20	0	0.00%
M54	3121	3120	1	0.03%
M11	2515	2512	3	0.12%

M10	1605	1600	5	0.31%
M107	27021	26870	151	0.56%
M101	4829	4800	29	0.60%
M89	2690	2672	18	0.67%
M97	3544	3520	24	0.68%
M97	3544	3520	24	0.68%
M28	558	554	4	0.72%
M2	1976	1960	16	0.81%
M44	5414	5360	54	1.00%
M7	1707	1688	19	1.11%
M47	6052	5984	68	1.12%
M105	3876	3832	44	1.14%
M85	4680	4624	56	1.20%
M27	5835	5760	75	1.29%
M28B	3890	3840	50	1.29%
M1090	10925	10784	140	1.29%
M96	6278	6192	86	1.37%
M29	5877	5760	117	1.99%
M106	6581	6400	181	2.75%
M104	47	30	17	36.17%
M5	2305	45	2260	98.05%
C(-)	0	0	0	-

4.2.2 R-Mutant screening

In the case of the mutants with repressed *comEA/EC* promoters under normally inducing conditions, a β-galactosidase activity assay was also performed. This time the assay was based on the fluorescent LacZ substrate 4-methylumbelliferyl β-D-galactopyranoside. Only the mutants that had less than half of the *lacZ* reporter activity were included in further work (see Figure 4.7). From the 20 mutants with the lowest activities, 6 showed more than 99.96% linkage between the phenotype-causing mutation and the Δ*comEA/EC:lacZ* reporter construct (see Table 4.2).

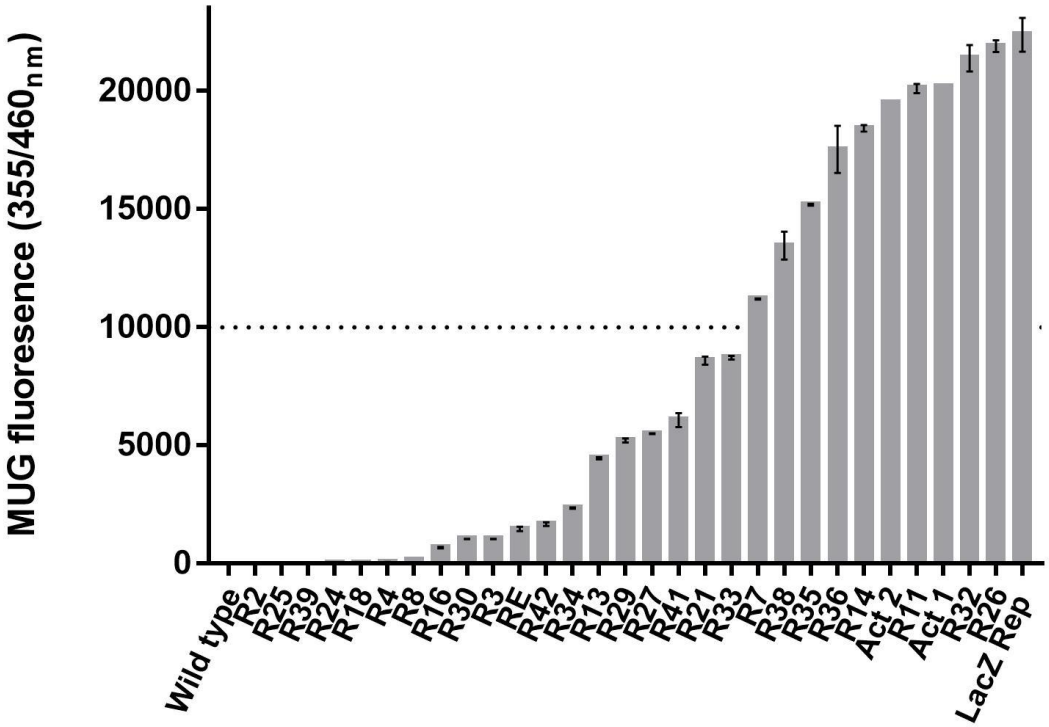


Figure 4.7 β-Galactosidase activity assay of R-mutants based on the fluorescent substrate 4-methylumbelliferyl β-D-galactopyranoside.
The assayed strains were cultured O/N on LB and they were all normalized to an OD_{600} = 1.5. After lysozyme treatment, 200 µL of crude extract from each sample were mixed with the substrate to a final concentration of 250 µg/mL. A 10 min end-point measurement was made at 460 nm absorbance. Only those R-mutants which had less than half of the Δ*comEA/EC*:*lacZ* reporter strain activity were taken for further analysis. The trpE16 wild-type strain was used as negative control and de *lacZ* transcriptional reporter as maximum activity reference. Act1 and Act2 are two isolates with a known high β-galactosidase activity used as positive controls. All mean values and standard deviations were represented and derive from at least 3 independent repetitions.

Table 4.2. R-mutants linkage assay results.

Mut N°	Total transf.	Blue	White	% linkage
C(-)	0	0	0	-
LacZ	5001	5000	1	0.02%
R21	4002	4000	2	0.05%
R27	5005	5000	5	0.10%
R3	1906	1904	2	0.10%
R42	4005	4000	5	0.12%
R7	793	792	1	0.13%
R41	3246	3240	6	0.18%
R16	1203	1200	3	0.25%
R30	4012	4000	12	0.30%
R33	350	348	2	0.57%
RE	1308	1300	8	0.61%
R13	5033	5000	33	0.66%
R4	4286	4256	30	0.70%
R34	1862	1840	22	1.18%
R29	80	80	1	1.25%
R25	2321	1	2320	99.96%
R18	5001	1	5000	99.98%
R2	5000	0	5000	100.00%
R8	428	0	428	100.00%
R39	2500	0	2500	100.00%
R24	2500	0	2500	100.00%

Only mutants with an activity below half of the *lacZ* reporter strain's activity were assayed and are listed here.

4.3 NGS and genome sequence processing

From all the mutants that fulfilled both requisites, i.e. no linkage and high or low enzymatic activity, 14 R-mutants and 22 M-mutants were cultured in LB medium until the stationary phase was reached. The cultures were pelleted and high-quality gDNA was prepared from each culture, following the CTAB extraction protocol detailed in methods section 3.10.1. After sequencing with Illumina technology and receiving all the reads for each mutant, a manual check of the coverage for each position in the reference genome was done using the Artemis software. There were always at least 200 reads covering each part of the genome. Sequence read assembly, alignment, and SNP calling were performed with the Snippy script taken from GitHub. Snippy finds SNPs between a haploid reference genome (*M. luteus* trpE16 in this case) and the NGS sequences reads. It produces a consistent set of output files in a single folder. All these produced data were used to generate a table containing each SNP position, the original nucleotide, and the new variant present in each mutant. As expected after using EMS as mutagen (Sega, 1984), almost all documented mutations were G:C→A:T transitions. There were on average 9 SNPs per genome in the M-mutants and 8 SNPs per genome in the R-mutants. The effect of the mutation, the affected loci, and their annotations in databases (NCBI) were also listed in separate columns. The table was used as input data for an R script (see Annex I), which takes only the mismatching positions (SNPs) and plots them at their respective place aligned to the reference genome (see Figure 4.8). It was possible to observe the general distribution of the mutations along the genome and to analyze how many SNPs there are per position or in which areas there is a higher density of them. To systematically detect those higher SNP densities, a programmable "sliding window" function was introduced into the R script (see Annex I). This "window", which length can be set (e.g. 5000 bps), screens a defined number of positions (e.g. every 500 bps) along the genome giving a score at each step. This score is a measure of the number of SNPs within the covered area and the distance between them. The more SNPs within the window and the closer their positions, the higher the final score. In this way, two windows with the same number of SNPs will have different scores if the distance between the SNPs is different. The variation of the cluster score throughout the genome can be observed in Figure 4.8B.

A set of primers was designed for each SNP of interest to generate PCR amplicons for sequencing. When a certain SNP was amplified along with its flanking regions, the mutant with the highest enzymatic activity (for the M-mutants) and the one with the lowest (for the R-mutants) was chosen as the template for the reaction. In Tables 4.3 and 4.4., the SNPs are named by showing the original and the new nucleotides with their position in the genome in between, but to make them

easier to identify another nomenclature was used. According to this nomenclature, the SNP names first state the cluster number to which the SNP belongs and then, separated by a hyphen, also include the designation of the parental EMS mutant used as PCR template. As an example, the SNP C1061676T from cluster 7 which is present in mutant M89 and was amplified from this mutant would be named 7-M89. This nomenclature is possible because there are no SNPs from the same cluster in the same mutant.

The goal after having identified all SNPs was to choose the best candidates for linking single mutations as possibly causative for the observed dysregulated reporter strain mutant phenotype. The selected SNPs were checked by amplicon Sanger sequencing and then introduced into *M. luteus* trpE16, which contains the intact *comEA/EC* region and is able to develop competence, by using the *codBA*-based counterselection system (see Methods 3.26). Thus, the transformable wild-type strain with each specific SNP of interest was obtained, and the transformation frequencies of the obtained single-SNP mutants could be determined. Also, it was possible to assess how the implicated single SNPs affect the *comEA/EC* promoter activity in a clean wild-type genetic background without further SNPs (several of which are present in the genomes of each original EMS mutant strain). For this, after introduction of the single SNP mutations, the 7 kbps Δ*comEA*/*EC*:*lacZ-Kan* fragment had to be reintroduced into the genomes of the single-SNP strains. The introduction of the Δ*comEA*/*EC*:*lacZ-Kan* construct was achieved by transformation with gDNA from the *lacZ* reporter strain and plating on LB^{Kan} plates.

4.3.1 M-mutants' SNPs

In the case of the M-mutants, 16 different SNPs clusters were identified, containing either several positions affected and/or in which the same position was affected in a relatively high number of mutants. This last case represents mutations that appeared earlier during the cultivation with EMS. As can be observed in Table 4.3 and Table 4.4, some clusters were selected depending on which genes they affect or if they had a higher number of mutants with one specific mutation. Besides the SNPs belonging to specific clusters, also SNPs affecting transcription factors, proteases, or other genes with annotations of interest were also considered for further steps.

The most outstanding of all M-mutants clusters was cluster 13, which showed the highest density of SNPs. Almost every M-mutant sequenced was found to have a SNP in it. All mutations in cluster 13 are within a region of 550 bps that encompasses 2 genes, Mlut_14660, a DNA binding protein with a Helix-turn-Helix 17 motif domain, and Mlut_14650, an Flp pilus assembly protein RcpC/CpaB with an annotated SAF domain (see zoomed panel in Figure 4.8). This last one belongs to

an operon together with Mlut_14640 and Mlut_14630 which encode the pilus assembly ATPase CpaE and a two-component sensor histidine-kinase containing a PAS domain, respectively. The intergenic region between Mlut_14660 and Mlut_14650 is also affected by 2 different SNPs. Among other affected genes, there are Mlut_09880 and Mlut_09900 which code for the Clp protease proteolytic subunit ClpP and the ATP-binding subunit ClpX, respectively. Interestingly, the only M-mutant which did not carry a SNP within cluster 13 had a SNP in Mlut_02150, which has the same annotation as Mlut_14660. All details are listed in Table 4.3 and Table 4.4. As many SNPs as possible were taken for reinsertion, and the SNPs 13-M27, 13-M89, 13-M1090, 13-M93, 7-M28, 7-M89, 7-M93, 3-M43, and 17-M33 were successfully inserted at the desired position in the genome, resulting in a series of single-SNP mutant strains. Besides the SNP mutants, three clean gene deletion mutants were generated to analyze their effect on *comEA/EC* promoter and transformability (Mlut_03135, Mlut_14650, and Mlut_02150*)*. Mlut_14660 and Mlut_09880 were also subjected to gene deletion attempts with the same deletion techniques, using both the *codBA*-based system and the traditional Kan cassette exchange method (Section 3.25), but despite numerous attempts it was not possible to obtain knock-out strains.

Table 4.3 M-mutants: main clusters, SNP features and affected genes.

Cluster	SNP	Mutant/s	Mut_Type	Locus	ORF orientation	Annotated Product
1	G103107A	M89, M106, MT	Transition	Mlut_00960	>	Cof subfamily of IIB subfamily of haloacid dehalogenase superfamily
1	G103500A	M27, M2	Transition	Mlut_00970	<	Protein of unknown function DUF88
1	G106197A	M33	Transition	Mlut_00990	<	Amino acid transporter
3	G341241A	M43, M47, M101, M1080, M42	Transition	Mlut_03135	<	N6-DNA Methylase
3	G341448A	M96	Transition	Mlut_03135	<	Hypothetical protein
4	C575082T	M93	Transition	-	-	Intragenic sequence
4	G575822A	M28, M28B	Transition	Mlut_05280	<	Hypothetical protein
7	C1061676T	M89, M106, MT	Transition	Mlut_09880	>	ATP-dependent Clp protease proteolytic subunit ClpP
7	C1063999T	M93	Transition	Mlut_09900	>	ATP-dependent Clp protease ATP-binding subunit ClpX
7	G1064341A	M28, M28B	Transition	Mlut_09900	>	ATP-dependent Clp protease ATP-binding subunit ClpX
13	G1606810A	M105, M27, M2	Transition	Mlut_14650	<	SAF domain-containing protein
13	G1607081A	M89, M106, MT	Transition	-	-	Intergenic sequence (Mlut_14660 promoter)
13	G1607117A	M1090	Transition	-	-	Intergenic sequence (Mlut_14660 promoter)
13	C1607204T	M96	Transition	Mlut_14660	>	DNA-binding protein, excisionase family
13	C1607261T	M9, M7, M43, M47, M101, M44, M1080, M42, M6, M3	Transition	Mlut_14660	>	DNA-binding protein, excisionase family
13	C1607314T	M93	Transition	Mlut_14660	>	DNA-binding protein, excisionase family
15	C2060795T	M93	Transition	Mlut_19080	<	Flavodoxin
15	C2061073T	M33	Transition	Mlut_19090	<	Thiol-disulfide isomerase-like thioredoxin
15	T2061826C	M43	Transition	Mlut_19090	<	Thiol-disulfide isomerase-like thioredoxin
17	C226684T	M33	Transition	Mlut_02150	>	DNA-binding protein, excisionase family
18	G1878213A	M42, M1080	Transition	Mlut_17340	<	RNA polymerase sigma factor, sigma-70 family

Sequence positions and locus tags refer to the genome of *Micrococcus luteus* trpE16, NCBI: ASM87779v1. A table with the updated locus tags can be found in Annex 2.

Table 4.4 M-mutants: main clusters and SNP effects.

Cluster	SNP	Mutant/s	Original_codon	Mutated_codon	OriginalAA	NewAA	Mut_effect
1	G103107A	M89, M106, MT	GCC	ACC	Ala	Thr	Missense
1	G103500A	M27, M2	GAC	AAC	Asp	Asn	Missense
1	G106197A	M33	GGC	GAC	Gly	Asn	Missense
3	G341241A	M43, M47, M101, M1080, M42	GGT	AGT	Gly	Ser	Missense
3	G341448A	M96	GCG	ACG	Ala	Thr	Missense
4	C575082T	M93	-	-	-	-	-
4	G575822A	M28, M28B	CGG	CGA	Arg	Arg	Synonymous
7	C1061676T	M89, M106, MT	CCG	TCG	Pro	Ser	Missense
7	C1063999T	M93	GGC	GGT	Gly	Gly	Synonymous
7	G1064341A	M28, M28B	ACG	ACA	Thr	Thr	Synonymous
13	G1606810A	M105, M27, M2	GTC	ATC	Val	Ile	Missense
13	G1607081A	M89, M106, MT	-	-	-	-	-
13	G1607117A	M1090	-	-	-	-	-
13	C1607204T	M96	TCG	TTG	Ser	Leu	Missense
13	C1607261T	M9, M7, M43, M47, M101, M44, M1080, M42, M6, M3	TCC	TTC	Ser	Phe	Missense
13	C1607314T	M93	CGG	TGG	Arg	Trp	Missense
15	C2060795T	M93	ACC	ACT	Thr	Thr	Synonymous
15	C2061073T	M33	TCC	TTC	Ser	Phe	Missense
15	T2061826C	M43	GTG	GCG	Val	Ala	Missense
17	C226684T	M33	CGG	TGG	Arg	Trp	Missense
18	G1878213A	M42, M1080	GGA	GAA	Gly	Glu	Missense

Sequence positions and locus tags refer to the genome of *Micrococcus luteus* trpE16, NCBI: ASM87779v1. A table with the updated locus tags can be found in Annex 2.

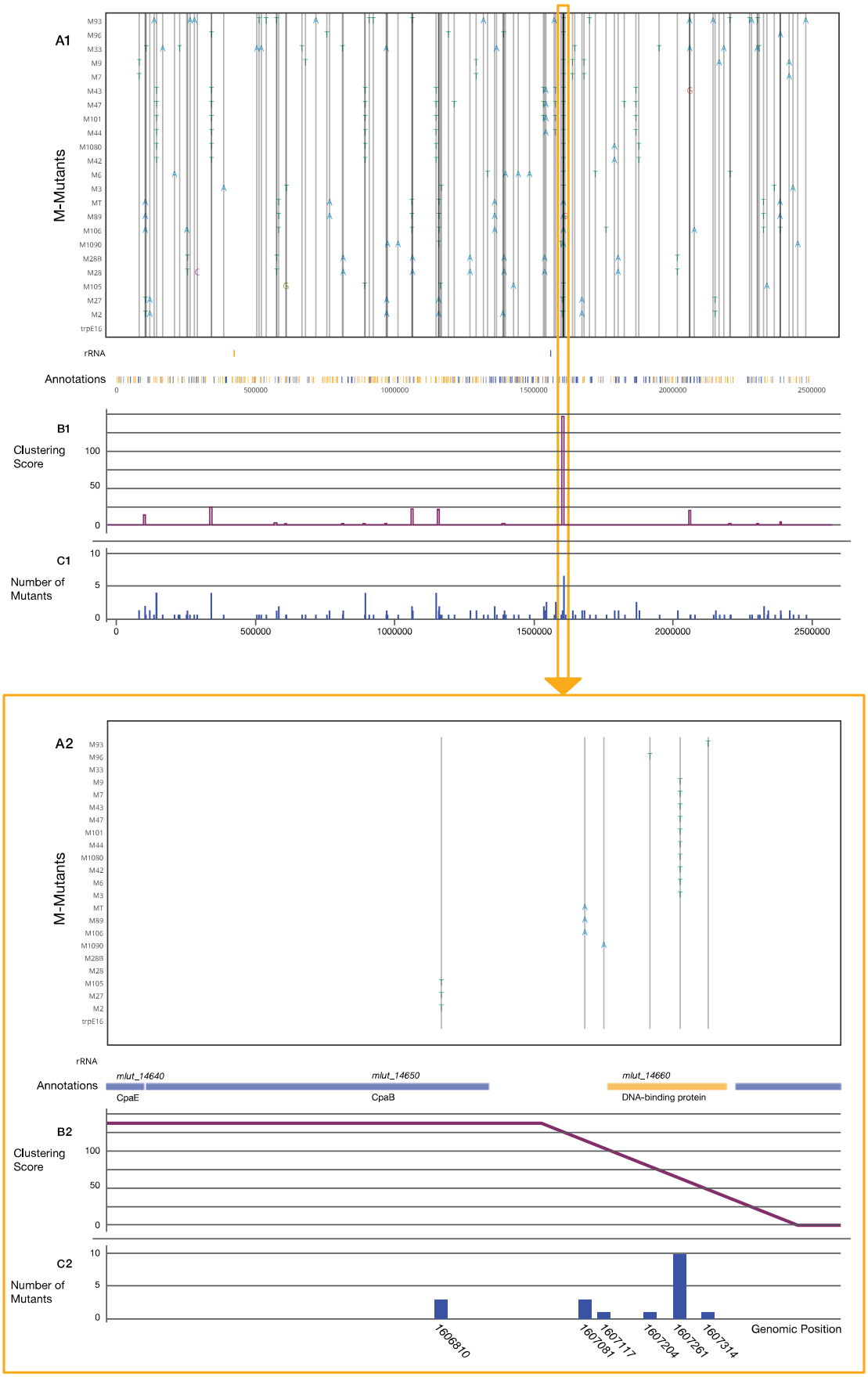

A1
M-Mutants
M93
M96
M33
M9
M7
M43
M47
M101
M44
M1080
M42
M6
M3
MT
M89
M106
M1090
M28B
M28
M105
M27
M2
trpE16
rRNA
Annotations
0
500000
1000000
1500000
2000000
2500000
B1
Clustering Score
100
50
0
C1
Number of Mutants
10
5
0
A2
rRNA
Annotations
mlut_14640
CpaE
mlut_14650
CpaB
mlut_14660
DNA-binding protein
B2
Clustering Score
C2
Number of Mutants
1606810
1607081
1607117
1607204
1607261
1607314
Genomic Position

Figure 4.8 M-mutants: SNP distribution and clustering analysis.
All plots in the figure share the horizontal axis which holds each position within the reference genome (*M. luteus* trpE16). Every SNP of each mutant can be observed at the upper panel and a zoomed view at the lower one (framed in orange). In both panels: (A1, A2) All aligned mutants and their SNPs, along with the annotated genes affected by each one. (B1, B2) Clustering score for each position, pointing out the most affected areas in the genome. It is the result of a programmable sliding window function that runs throughout the genome. The window coverage here was set to 5 kbp. (C1, C2) Number of mutants holding a mutation at each position in the genome. When zooming in the area with the highest clustering score we get the view of the lower panel. Six mutated positions from the M-mutants belonging to cluster 13 are shown here. Directly under the SNPs plot, both affected genes Mlut_14650 and Mlut_14660 are represented in blue and orange according to their orientation in the DNA molecule. Two positions fall at the intergenic area, most specifically within the 5´UTR of Mlut_14660. This figure was created by this thesis author for the publication of this work in MDPI-Genes journal (Torasso Kasem *et al*., 2021)

4.3.2 R-mutants' SNPs

In contrast to the M-mutants, there was not a clear clustering pattern for the R-mutants (see Table 4.5 and Figure 4.9). All SNPs were present in only one of the R-mutants, except for three SNPs present in mutants R3 and R30. To select for interesting putative causative SNPs, the sliding "window" function of the already mentioned R script was also used, but a more dedicated visual analysis of the data was carried out. Interestingly, the larger SNPs cluster (number two) of the R-mutants affects three genes belonging to or neighboring the *tad-1* genes cluster. More specifically, Mlut_07490, which encodes an ATPase from the FtsK family involved in DNA segregation; Mlut_07500, coding for an Flp pilus assembly protein, the bifunctional ATPase CpaF; and Mlut_07560, encoding the Flp pilus-assembly TadG-like protein.

Considering the lack of clear clustering of several SNPs, any pair of SNPs affecting neighboring genes was also considered. Examples of this are cluster 6 or cluster 8. In total, 12 clusters were taken for further work (see Table 4.6). Many SNPs didn´t change the amino acid residues of the corresponding protein, further reducing the number of interesting candidates. Curiously, one isolated mutation belonging to the mutant R4, which has a strongly repressed phenotype, also affects Mlut_14650, encoding the Flp pilus assembly protein RcpC/CpaB with an annotated SAF domain that was one of the interesting candidate genes identified in the M-mutant screening (cluster 13, see above, Table 4.3).

After observing the annotated proteins of all affected genes, the following SNPs were selected for reinsertion into the mutant strains' parental strain *M. luteus* trpE16: 2-R33, which affects the ATPase from the FtsK family; 8-R34, affecting an annotated response regulator; and 13-R4, which affects Mlut_14650.

Table 4.5 R-mutants: main clusters, SNP features and affected genes.

Cluster	SNP	Mutant/s	Locus	Orientation	Annotated product
1	C9672T	R42	Mlut_00060	>	DNA gyrase subunit A
1	G10325A	RE	Mlut_00060	>	DNA gyrase subunit A
2	G810889A	R33	Mlut_07490	<	DNA segregation ATPase, FtsK/SpoIIIE family
2	G812767A	R13	Mlut_07500	>	Flp pilus assembly protein, ATPase CpaF
2	C812940T	R16	Mlut_07500	>	Flp pilus assembly protein, ATPase CpaF
2	G816694A	R34	Mlut_07560	>	hypothetical protein
3	C934586T	R21	Mlut_08640	>	glutamyl-tRNA synthetase
3	G935940A	R16	Mlut_08660	>	tRNA-Gln
4	G1011921A	R4	Mlut_09340	<	2-oxo-acid dehydrogenase E1 component, homodimeric type
4	C1014740T	R13	Mlut_09380	>	4-azaleucine resistance probable transporter AzlC
5	G1029395A	R7	Mlut_09550	<	peptide deformylase
5	G1031740A	R13	Mlut_09580	>	translocating P-type ATPase, Cd/Co/Hg/Pb/Zn-transporting
6	G1249315A	R42	Mlut_11600	<	phosphoserine phosphatase SerB
6	C1250564T	R42	Mlut_11610	>	ABC-type molybdenum transport system
7	G1460599A	RE	Mlut_13330	>	2-oxoglutarate dehydrogenase E2 component
7	C1462336	R4	Mlut_13350	<	protein kinase family protein
8	G1541663A	R34	Mlut_14110	>	response regulator of citrate/malate metabolism
8	C1542830T	R41	Mlut_14120	<	acyl-CoA synthetase/AMP-acid ligase
9	G1814683A	R3, R30	-		Intergenic sequence, Mlut_16600 promoter
10	C1962653T	R4	Mlut_18140	>	helicase family protein with metal-binding cysteine cluster
10	C1964134T	R13	Mlut_18140	>	helicase family protein with metal-binding cysteine cluster
11	C2027721T	R3, R30	Mlut_18760	>	tRNA (guanine-N(7)-)-methyltransferase
11	C2028876T	R34	Mlut_18770	<	Alpha/beta hydrolase of unknown function (DUF1023)
12	G2427638A	R3, R30	Mlut_22740	<	hypothetical protein
13	G1606821A	R4	Mlut_14650	<	SAF domain-containing protein

Sequence positions and locus tags refer to the genome of *Micrococcus luteus* trpE16, NCBI: ASM87779v1. A table with the updated locus tags can be found in Annex 2.

Table 4.6 R-mutants: main clusters and SNP effects.

Cluster	SNP	Mutant/s	Original_codon	New_codon	OriginalAA	NewAA	Mut_effect
1	C9672T	R42	CGA	TGA	Arg	Stop	Nonsense
1	G10325A	RE	CGG	CGA	Arg	Arg	Synonymous
2	G810889A	R33	GGG	GAG	Gly	Glu	Missense
2	G812767A	R13	CGG	CAG	Arg	Gly	Missense
2	C812940T	R16	CGC	AGC	Arg	Cys	Missense
2	G816694A	R34	GGC	GAC	Gly	Asp	Missense
3	C934586T	R21	CCC	CCT	Pro	Pro	Synonymous
3	G935940A	R16	GTA	ATA	Val	Ile	Missense
4	G1011921A	R4	GTG	GTA	Val	Val	Synonymous
4	C1014740T	R13	CGC	CGT	Arg	Arg	Synonymous
5	G1029395A	R7	GAG	AAG	Glu	Lys	Missense
5	G1031740A	R13	GGG	GAG	Gly	Glu	Missense
6	G1249315A	R42	CGG	CGA	Arg	Arg	Synonymous
6	C1250564T	R42	CTC	CTT	Leu	Leu	Synonymous
7	G1460599A	RE	AGC	AAC	Ser	Asn	Missense
7	C1462336	R4	TCC	TCT	Ser	Ser	Synonymous
8	G1541663A	R34	GAC	AAC	Asp	Asn	Missense
8	C1542830T	R41	CGC	TGC	Arg	Cys	Missense
9	G1814683A	R3, R30	-	-	-	-	-
10	C1962653T	R4	CGC	CGT	Arg	Arg	Synonymous
10	C1964134T	R13	GCC	GTC	Ala	Val	Missense
11	C2027721T	R3, R30	CGA	TGA	Arg	Stop	Nonsense
11	C2028876T	R34	CAC	TAC	Hys	Tyr	Missense
12	G2427638A	R3, R30	TAC	TAT	Tyr	Tyr	Synonymous
13	G1606821A	R4	GGT	GAT	Gly	Asp	Missense

Sequence positions and locus tags refer to the genome of *Micrococcus luteus* trpE16, NCBI: ASM87779v1. A table with the updated locus tags can be found in Annex 2.

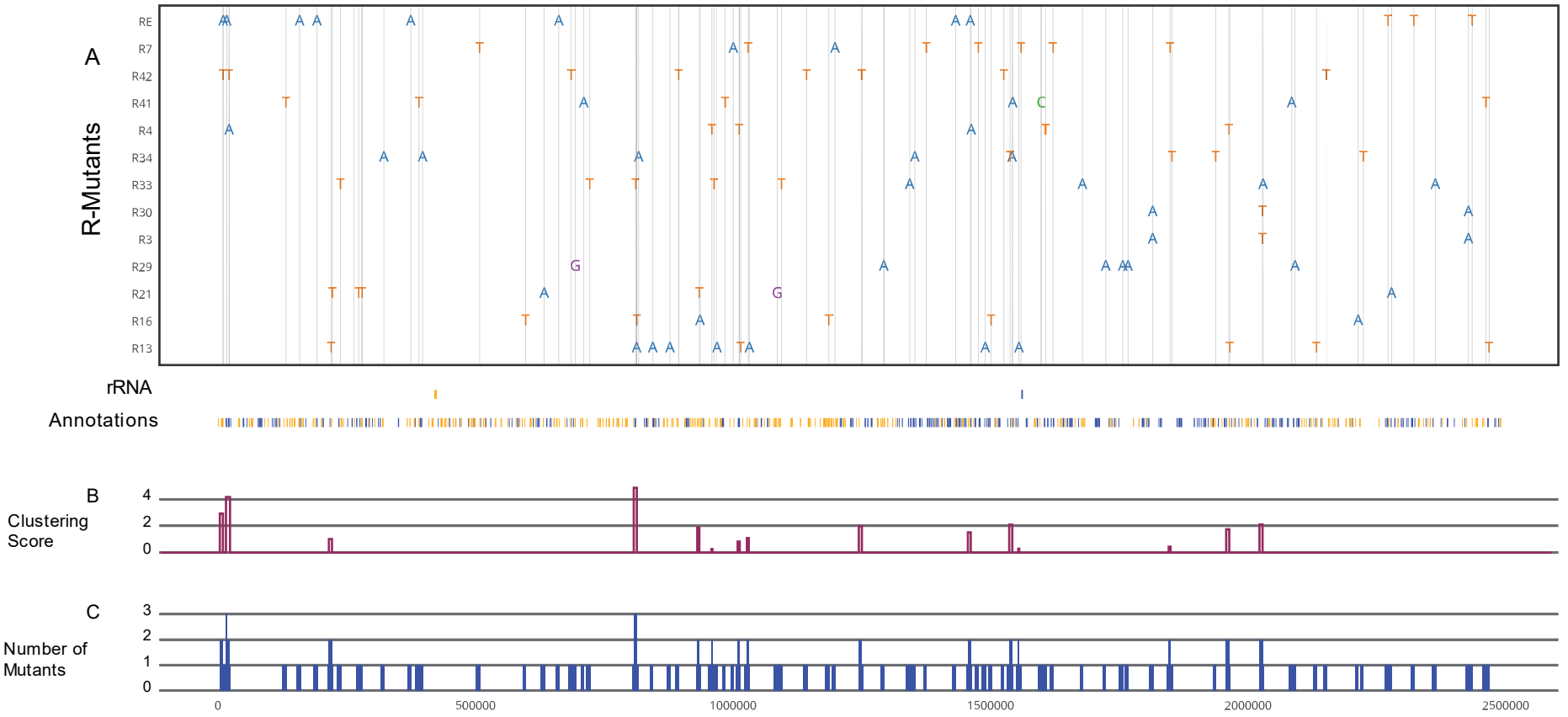


Figure 4.9 R-mutants' SNP distribution and clustering analysis.
All plots in the figure share the horizontal axis which holds each position within the reference genome (*M. luteus* trpE16). All SNPs of each mutant can be observed at the upper panel. A) All aligned mutants and their SNPs, along with the annotated genes affected by each one. B) Clustering score for each position, pointing out the most affected areas in the genome. It is the result of a programmable sliding window function that runs throughout the genome. The window coverage here was set to 5 kbp. C) Number of mutants holding a mutation at each position in the genome. The mutations don´t seem to be particularly clustered at any specific position.

4.4 The effect of SNPs of cluster 13 on the amino acid sequences of Mlut_14650- and Mlut_14660-encoded proteins

Both Mlut_14650 and Mlut_14660 were affected in more than one position giving rise to missense mutations and significant changes in the amino acid sequence of their proteins. In the case of the transmembrane CpaB-like protein, there was one SNP identified from the R-mutants and one from the M-mutants affecting codons 27 and 31, respectively. Codon 27 (GGT) normally codes for Gly, which turns into Asp (GAT) after the transition; and codon 31, a GTC which codes for Val, turns into an Ile (ATC). When observing the pHMM of ChapFlgA, those positions do not seem to be highly conserved, but for the other member of the SAF-like β-clip superfamily: RcpC, codon 31 shows higher conservation, with Gly as the most common amino acid in that position (see Figure 4.10).

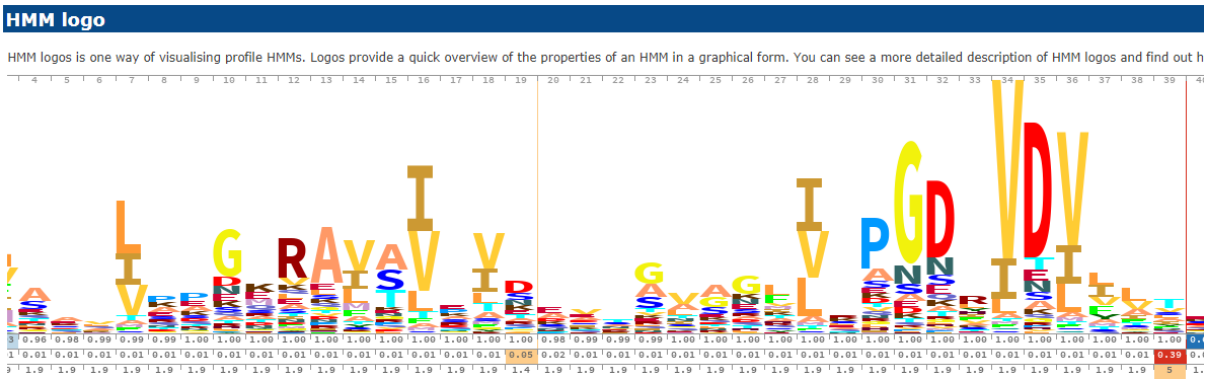


Figure 4.10 Profile Hidden Markov Model for the RcpC/CpaB domain annotated in Mlut_14650.
HMM logo obtained after pfam search for the RcpC/CpaB domain present in Mlut_14650. It shows the conservation of each residue in an amino acid sequence (stack height) and which amino acid is the most frequent for that position. This plot encompasses 40 amino acids that belong to the RcpC/CpaB domain.

For Mlut_14660 two nucleotide positions upstream of the start codon and three within the coding region were affected. Codons 27 (TTG), 46 (TTC), and 64 (TGG), coding for Ser, Ser, and Arg, respectively, were turned into Leu, Phe, and Trp. The first two of the changed amino acid residues lie within the Helix-turn-Helix 17 domain at the N-terminus, and the third one is downstream of it. When observing the pHMM of the HTH17 domain, codon 46 seems to code for the most conserved residue, and codon 27 is among the most conserved, although Ser is not the most common amino acid found at those positions (see Figure 4.11).

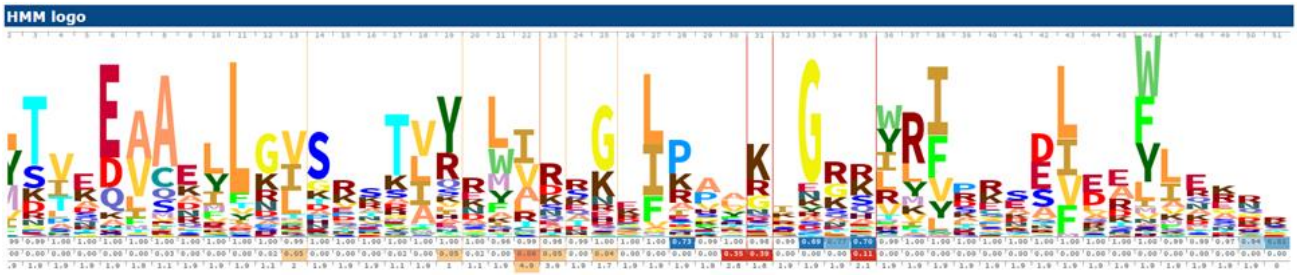


Figure 4.11 Profile Hidden Markov Model for Helix-Turn-Helix 17 domain annotated in Mlut_14660.
HMM logo obtained after pfam search for the HTH 17 domain found in Mlut_14660. It shows how well conserved each residue in a protein chain is (stack height) and which amino acid is the most frequent at that position. This plot shows 47 of the 51 amino acids that belong to the HTH 17 motif.

4.5 Detection of causative SNPs

To find some of the causative mutations that apparently affect the expression levels of *comEA/EC,* a *codB/codA* based counter-selection system was used to cleanly reintroduce the SNPs of interest into the *M. luteus* genome. Because of the lack of transformability of the reporter strain, the genomic editions had to be carried out in the trpE16 wild-type strain. After confirmation of successful SNP reinsertion by amplicon sequencing, the single-SNP mutant strains were transformed with the *ΔcomEA/EC:lacZ-Kan* construct in order to get this *lacZ* reporter construct recombined in each single-SNP strain at the same position as in the original reporter strain. Therefore, for each single-SNP genome edition, one strain was obtained with the reporter construct and another one without it. In total, 9 SNPs belonging to 4 different clusters were re-inserted successfully, and additionally 3 of the affected genes were knocked out by clean deletion (for details see Table 4.7). Once the SNPs of interest were individually and properly reintroduced into the genome along with the *lacZ* reporter construct, the *comEA/EC* promoter expression levels of those strains were measured through the β-galactosidase/MUG activity assay (section 3.20.2.2). Each single-SNP mutant was evaluated along with its parental mutant (original EMS mutant selected during the screening procedure) and a knockout strain of the affected gene (when available).

The β-galactosidase enzymatic activity can be modeled by the classical Michaelis-Menten equations which assume that initially the reverse reaction can be neglected, and the maximal velocity of product synthesis can be measured. Moreover, since substrate was added in excess, the initial linear increase in relative fluorescence units (RFU) per unit time only depends on the concentration of the enzyme, which is the property of interest:

(4) $v_{max} = k_{cat}\, x\, [E]_0$

(v_{max}=maximal velocity, k_{cat}=rate constant, $[E]_0$=total enzyme concentration)

The cells were lysed before the addition of the substrate, avoiding the lag phase, and the measurements started immediately after substrate addition. Therefore, the comparison of the initial velocities may serve as a measure of the individual amounts of enzyme in the lysates of the different mutants, which are dependent on the respective activation of $P_{comEA/EC}$.

Table 4.7 List of all strains of *M. luteus* constructed in this study.

Name	Genotype and relevant phenotype	Source
Δ14650	trpE16 *Δ*Mlut_14650	This study
Δ02150	trpE16 *Δ*Mlut_02150	This study
Δ03135	trpE16 *Δ*Mlut_03135	This study
3-M43	trpE16 G341241A (Mlut_03135*)*	This study
7-M89	trpE16 *C*1061676T (Mlut_09880*)*	This study
7-M93	trpE16 C1063999T (Mlut_09900*)*	This study
7-M28	trpE16 G1064341A (Mlut_09900*)*	This study
17-M33	trpE16 C226684T (Mlut_02150*)*	This study
13-M27	trpE16 G1606810A (Mlut_14650*)*	This study
13-M89	trpE16 G1607081A (Mlut_14660*)*	This study
13-M1090	trpE16 G1607117A (Mlut_14660*)*	This study
13-M93	trpE16 C1607314T (Mlut_14660*)*	This study
2-R33	trpE16 G810889A (Mlut_07490*)*	This study
13-R4	trpE16 G1606821A (Mlut_14650*)*	This study
8-R34	trpE16 G1541663A (Mlut_14110*)*	This study

A table with the updated locus tags can be found in Annex 2.

4.5.1 Determination of the dynamic range of the MUG/β-galactosidase activity assay

Maintaining an appropriate relationship between catalytic enzyme concentration, substrate concentration and incubation time is paramount in obtaining linearity in the assay. The reaction should not be too fast to determine the initial velocity, nor should the enzyme concentration be so low, that no reliable signal can be obtained.

These parameters restrict the range within which mutants with different *comEA/EC*-promoter activities and therefore enzyme concentrations can be compared. Differences between two mutants with lower expression levels might not be detectable below a certain threshold. Further, two highly active mutants may display kinetics too fast to be distinguished fluorometrically. When considering a one-point method, the respective time point should be in the linear part of the reaction for all mutants. Otherwise, the initial velocity for highly active mutants may be underestimated, as at the time of measurement the respective reaction has already reached its plateau phase due to depletion of the substrate.

Results

To estimate the signal range within which measurements can be compared, as well as the time-period feasible for the β-galactosidase/MUG activity assays, a dilution series of the *lacZ*-reporter mutant was prepared (OD_{600}=5 to 0.08). The different cell densities of *M. luteus ΔComEA/EC:lacZ-Kan* were used to generate a broad spectrum of enzymatic activities which mimic the different M-mutants.

As shown in Figure 4.12, the signal shows a linear dependence on the change in cell density and therefore also overall enzyme activity within a certain range of time points. This does not apply to the beginning of the reaction (approximately for the first twenty minutes). This may be due to a higher error of measurement at a low fluorescence of the probe. Also, after approximately one hour, the signals stemming from probes with high cell densities displayed a smaller signal than would be expected. This might be due to the depletion of the substrate and therefore the decrease and finally stagnation in reaction velocity. According to these findings, a suitable time point for an end-point measurement under the assay conditions employed were found to be between 30 min and 1 h.

It was also of interest whether an initial velocity could still be trusted if it was higher or lower than the one of the *lacZ* reporter by a certain factor, that is if it was still in the linear range as noted above. Assuming an OD_{600} = 1.5 for the reporter strain, results may be trustworthy if they are twice as high or ~ 30 times as low. Low activities could probably be determined with greater accuracy since the reaction curves were very flat and therefore many measurements can be taken during the reaction as opposed to a very fast reaction. Especially, the upper border is problematic since the measurement was only conducted with a maximal OD_{600} of 5. Additional measurements with higher cell concentrations might reveal a much larger linear range. Therefore, the same assay was conducted with the mutant *M. luteus* M89 (Figure 4.13). Here, a roughly linear dependence could still be observed for measurements taken for the different cell concentrations between 30 min and 1 h, thus confirming the previous observations.

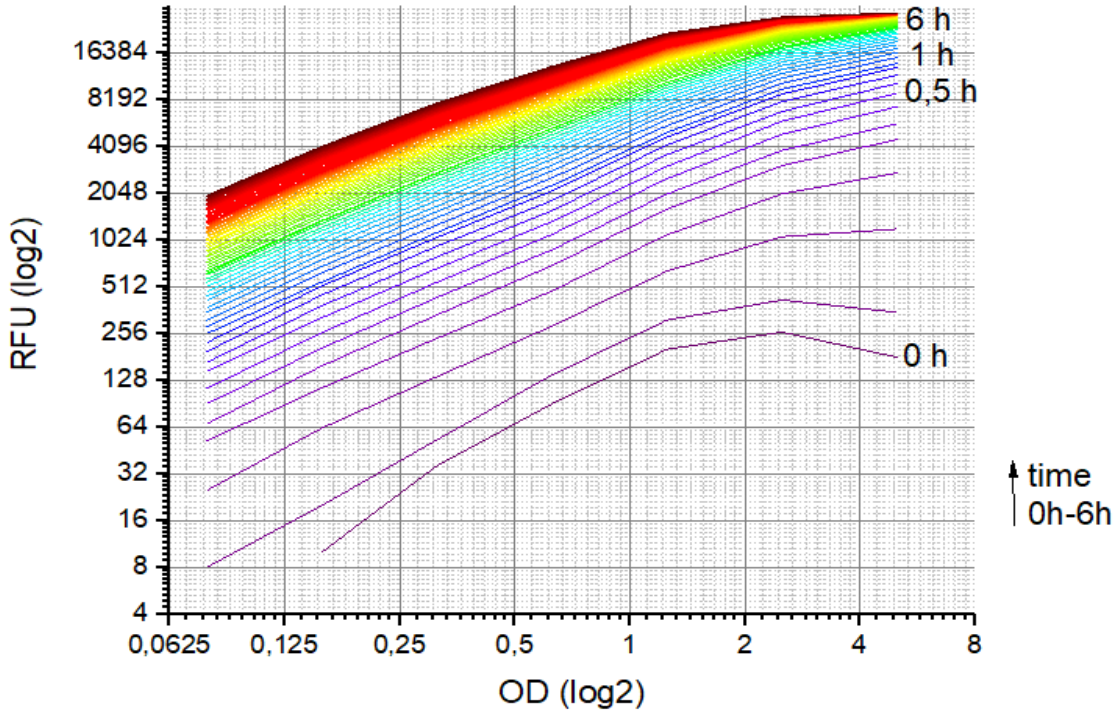


Figure 4.12 β-Gal/MUG activity assays of a series of dilutions of *M. luteus lacZ* reporter strain grown under repressive conditions.

The cells were harvested and resuspended in 1 mL of buffer with a final concentration of 1 µg/mL lysozyme and 250 µg/mL MUG. The OD_{600} values (log2-scale) of the bacterial dilutions are plotted against the RFU values (log2-scale). The curves represent each RFU value for each bacterial density (OD_{600}=5, 2.5, 1.25, 0.63, 0.31, 0.16, 0.08) at different time points of the reaction (represented in different colors). Measurements were taken every five minutes for 6 hours. Two biological replicates were evaluated.

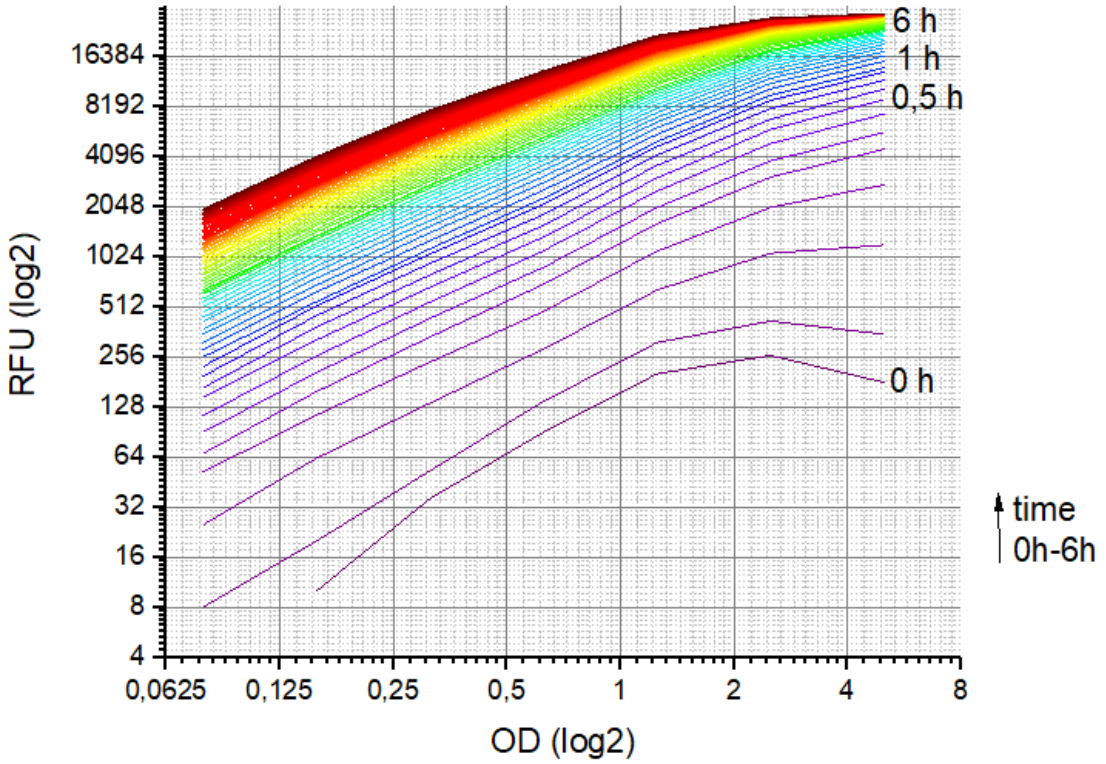


Figure 4.13 β-Gal/MUG activity assays of a series of dilutions of the M-mutant *M. luteus* M89 grown under repressive conditions.
The cells were harvested and resuspended in 1 mL of buffer with a final concentration of 1 µg/mL lysozyme and 250 µg/mL MUG. The OD_{600} values (log2-scale) of the bacterial dilutions are plotted against the RFU values (log2-scale). The curves represent each RFU value for each bacterial density (OD_{600}=5, 2.5, 1.25, 0.63, 0.31, 0.16, 0.08) at different time points of the reaction (represented in different colors). Measurements were taken every 10 minutes for 6 hours. Two biological replicates were prepared.

4.5.2 Causative SNP determination by comparative β-galactosidase activity assay

All single-SNP mutants and their respective parental EMS mutant strains were assayed according to the MUG activity assay protocol detailed in section 3.20.2.2. First, all measurements taken every 10 min were plotted against time (min). These plots are shown in Figure 4.14 to Figure 4.20. For the single-SNP mutants 7-M89, 13-M89, 7-M93, 3-M43, 7-M28, Δ02150, and Δ03135, the phenotype of their parental mutants was not reestablished. In the case of mutant 17-M33, there were significant differences compared with the mutant M33, but it did show in average a slightly higher activity level than the *lacZ* reporter strain *M. luteus* Δ*comEA*/*EC*:*lacZ*-Kan. For the mutants 13-M27, 13-M93 and, 13-1090, there were no significant differences compared with their respective parental mutants, and the phenotypes were reestablished after SNP reinsertion. In the case of the knockout strain Δ14650, overexpression of the promoter was also observed after the deletion.

After determining all kinetic curves, a four-parameter logistic function was fitted to them. The linear range of the kinetic curves prior to substrate depletion was used for evaluation. As stated previously, by calculating the initial velocity for each strain, it is possible to compare their enzyme

amounts and thus, their promoter activity levels. The initial velocity for each technical replicate was calculated individually and then used to create a bar chart where the mean values and standard deviations for each mutant are shown (see Figure 4.21).

The *lacZ* reporter strains with the SNPs 2-R33 and 8-R34 in their genomes were plated on GMM + X-Gal 80 µg/mL and Kan 60 µg/mL to qualitatively check their activity. Both reporter mutants yielded intense blue colonies. Single-SNP mutant 13-R4 could not be transformed with the *lacZ* reporter construct.

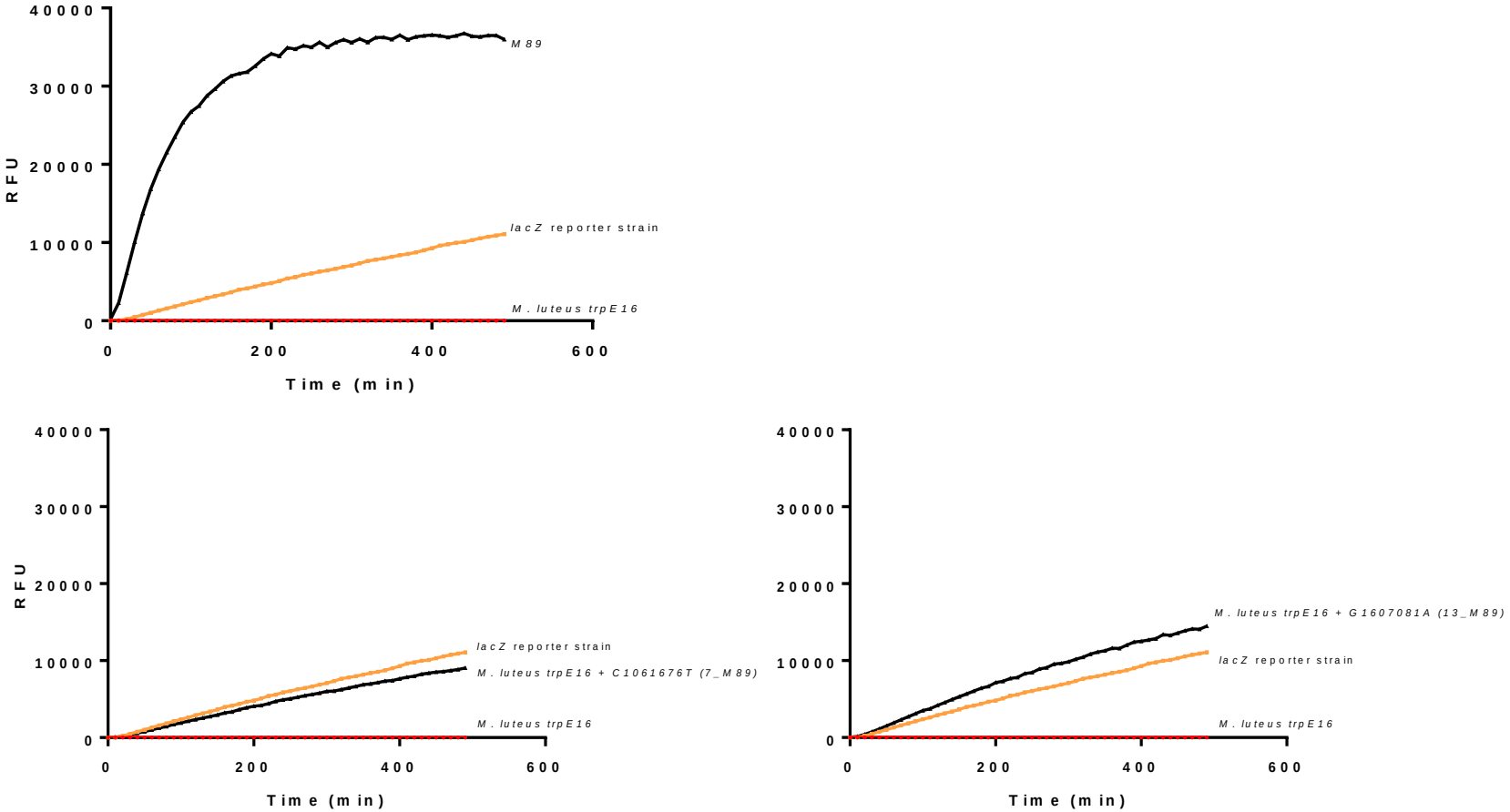


Figure 4.14 β-Gal/MUG activity assays of *M. luteus* M89 and its corresponding single-SNP mutants 7-M89 and 13-M89.
The measured samples contained lysed cells of the parental mutant *M. luteus* M89 or of its derivative SNP-mutants *M. luteus* 7-M89 or 13-M89 with an OD_{600} = 1.5 and a MUG concentration of 250 µg/mL. The RFU value was detected every 10 min for 8 h. *M. luteus* Δ*comEA*/*EC*:*lacZ*-Kan (*lacZ* reporter strain) was utilized as a reference and *M. luteus* trpE16 (wild-type) as a negative control. At least three individual biological replicates were prepared and three technical replicates from each one were measured in each assay.

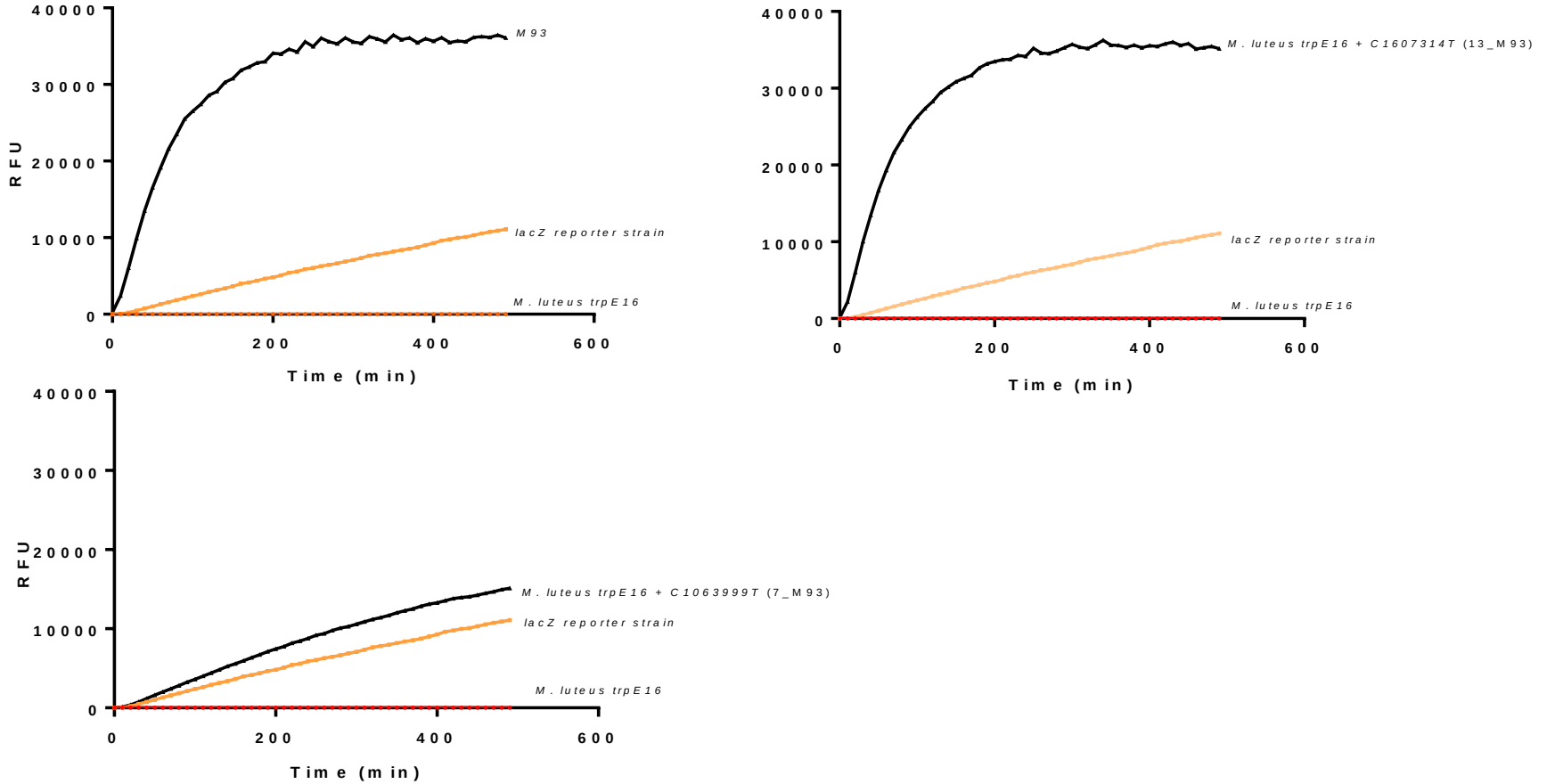


Figure 4.15 β-Gal/MUG activity assays of *M. luteus* M93 and its corresponding single-SNP mutants 7_M93 and 13-M93.
The measured samples contained lysed cells of the parental mutant *M. luteus* M93 or of its derivative SNP-mutants *M. luteus* 7_M93 or 13-M93 with an OD_{600}=1.5 and a MUG concentration of 250 µg/mL. The RFU value was measured every 10 min for 8 h. *M. luteus* Δ*comEA*/*EC*:*lacZ*-Kan (*lacZ* reporter strain) was utilized as a reference and *M. luteus* trpE16 (wild-type) as a negative control. At least three individual biological replicates were prepared and three technical replicates from each one were measured in each assay.

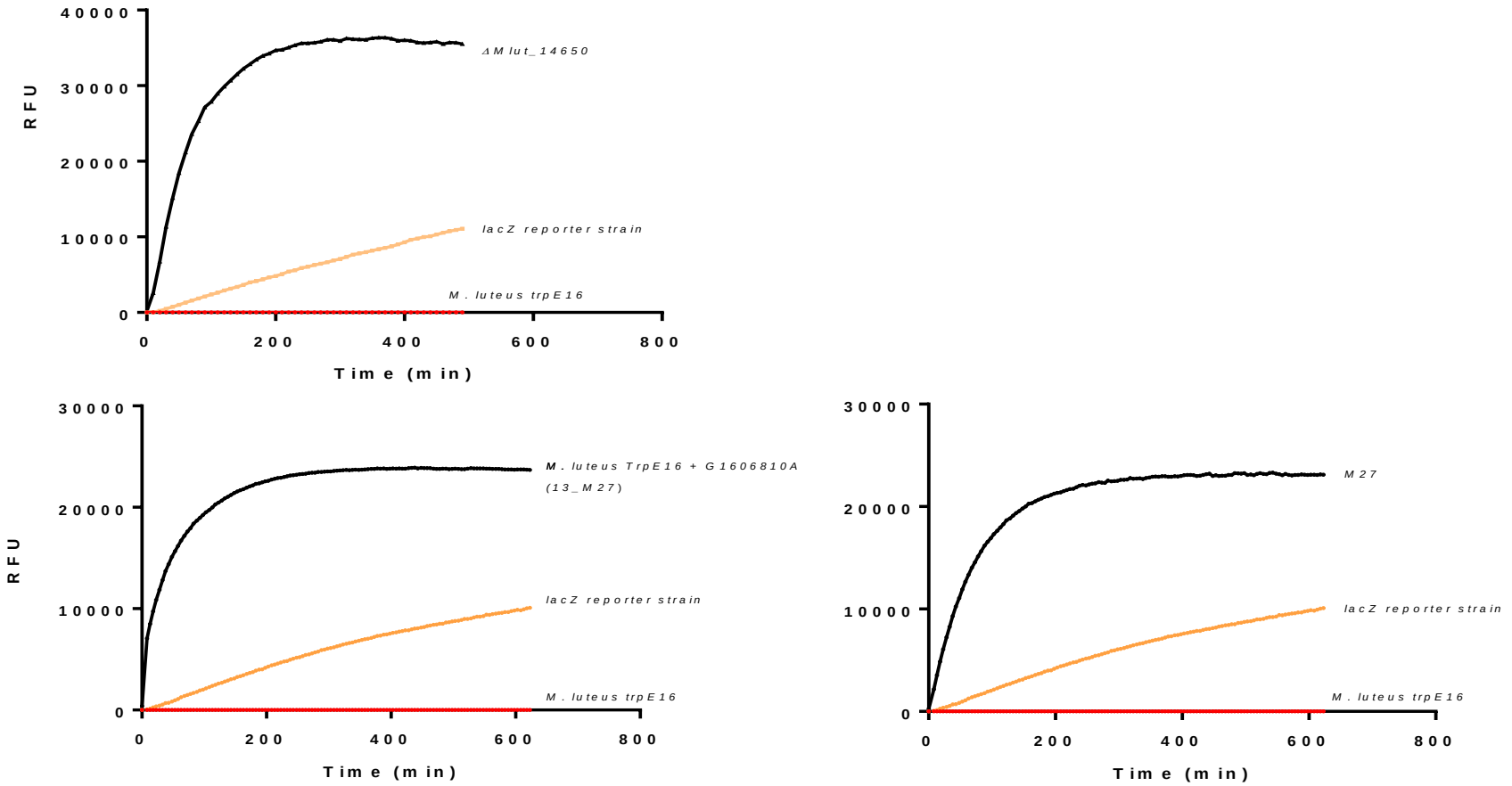


Figure 4.16 β-Gal/MUG activity assays of *M. luteus* M27, its corresponding single-SNP mutant 13-M27, and the knockout strain ΔMlut_14650.
The measured samples contained lysed cells of the parental mutant *M. luteus* M27, of its derivative SNP-mutant *M. luteus* 13-M27 or of the clean knockout strain ΔMlut_14650 with an OD_{600}=1.5 and a MUG concentration of 250 µg/mL. The RFU value was detected every 10 min for 8 h. *M. luteus* Δ*comEA*/*EC*:*lacZ*-Kan (*lacZ* reporter strain) was utilized as a reference and *M. luteus* trpE16 (wild-type) as a negative control. At least three individual biological replicates were prepared and three technical replicates from each one were measured in each assay.

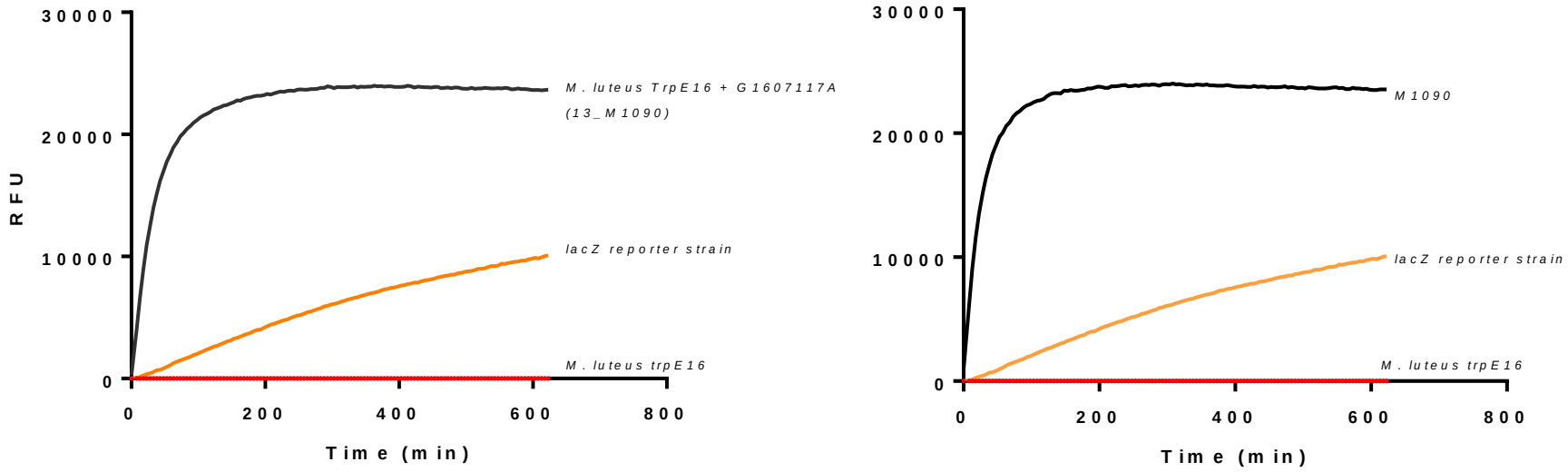


Figure 4.17 β-Gal/MUG activity assays of *M. luteus* M1090 and its corresponding single-SNP mutant 13-M1090.
The measured samples contained lysed cells of the parental mutant *M. luteus* M1090or of its derivative SNP-mutants *M. luteus* 13-M1090 with an OD_{600}=1.5 and a MUG concentration of 250 µg/mL. The RFU value was detected every 10 min for 8 h. *M. luteus* Δ*comEA*/*EC*:*lacZ*-Kan (*lacZ* reporter strain) was utilized as a reference and *M. luteus* trpE16 (wild-type) as a negative control. At least three individual biological replicates were prepared and three technical replicates from each one were measured in each assay.

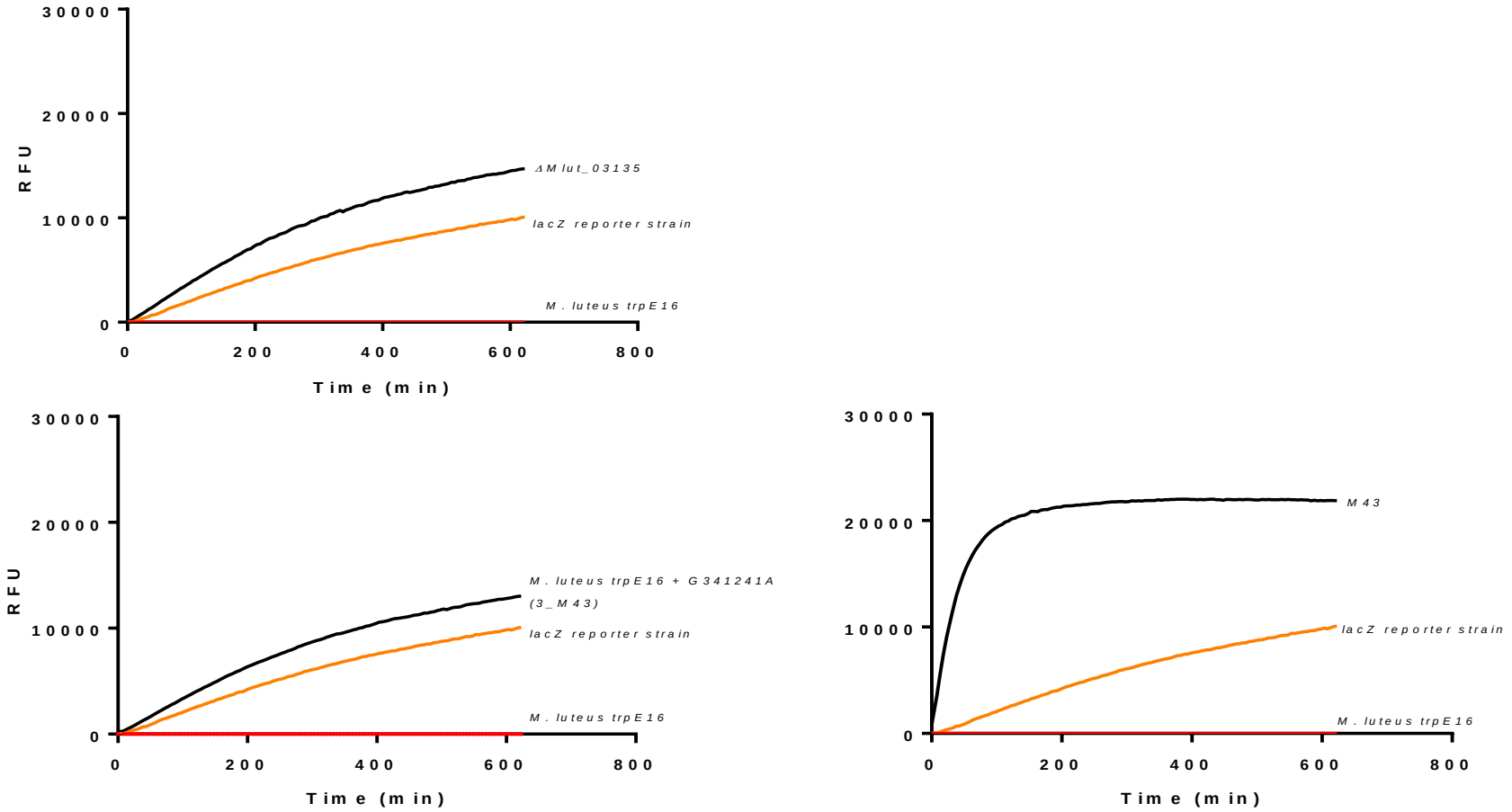


Figure 4.18 β-Gal/MUG activity assays of *M. luteus* M43, its respective corresponding single-SNP mutant 3-M43, and the knockout strain ΔMlut_03135.
The measured samples contained lysed cells of the parental mutant *M. luteus* M43, of its derivative SNP-mutant *M. luteus* 3-M43, or the clean knockout strain ΔMlut_03135 with an OD_{600}=1.5 and a MUG concentration of 250 µg/mL. The RFU value was detected every 10 min for 8 h. *M. luteus* Δ*comEA*/*EC*:*lacZ*-Kan (*lacZ* reporter strain) was utilized as a reference and *M. luteus* trpE16 (wild-type) as a negative control. At least three individual biological replicates were prepared and three technical replicates from each one were measured in each assay.

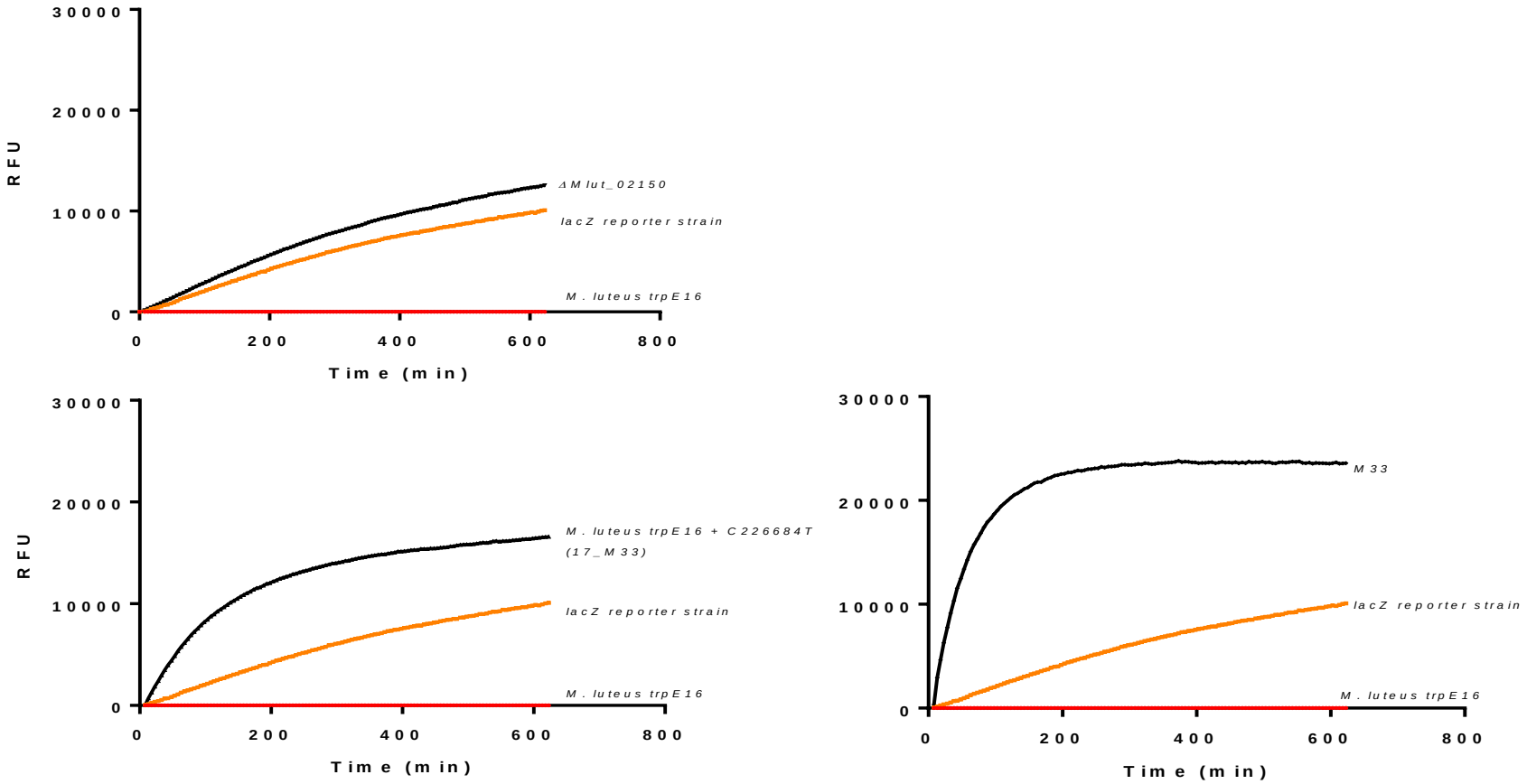


Figure 4.19 β-Gal/MUG activity assays of *M. luteus* M33, its respective derivative single-SNP mutant 17_M33, and the knockout strain ΔMlut_02150.
The measured samples contained lysed cells of the parental mutant *M. luteus* M33, of its derivative SNP-mutant *M. luteus* 17_M33, or of the clean knockout strain ΔMlut_02150 with an OD_{600}=1.5 and a MUG concentration of 250 µg/mL. The RFU value was detected every 10 min for 8 h. *M. luteus* Δ*comEA*/*EC*:*lacZ*-Kan (*lacZ* reporter strain) was utilized as a reference and *M. luteus* trpE16 (wild-type) as a negative control. At least three individual biological replicates were prepared and three technical replicates from each one were measured in each assay.

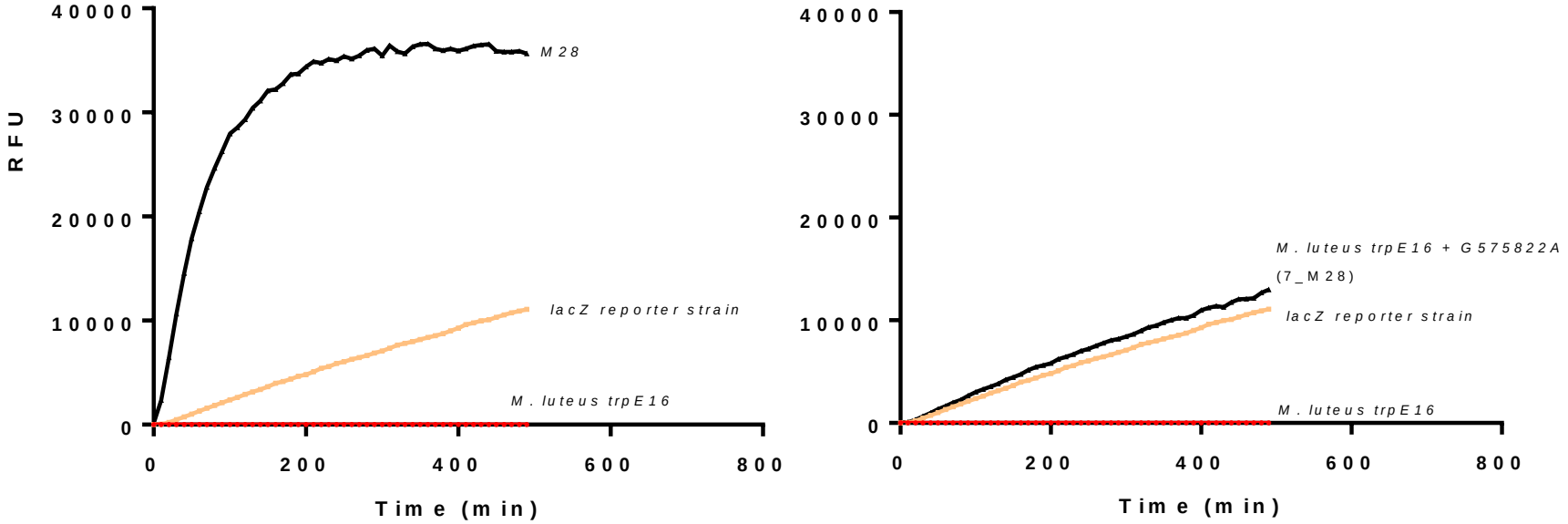


Figure 4.20 β-Gal/MUG activity assays of *M. luteus* M28, and its derivative SNP mutant 7_M28.
The measured samples contained lysed cells of the parental mutant *M. luteus* M28, or of its derivative SNP-mutants *M. luteus* 7_M28 with an OD_{600}=1.5 and a MUG concentration of 250 µg/mL. The RFU value was measured every 10 min for 8 h. *M. luteus* Δ*comEA*/*EC*:*lacZ*-Kan (*lacZ* reporter strain) was utilized as a reference and *M. luteus* trpE16 (wild-type) as a negative control. At least three individual biological replicates were prepared and three technical replicates from each one were measured in each assay.

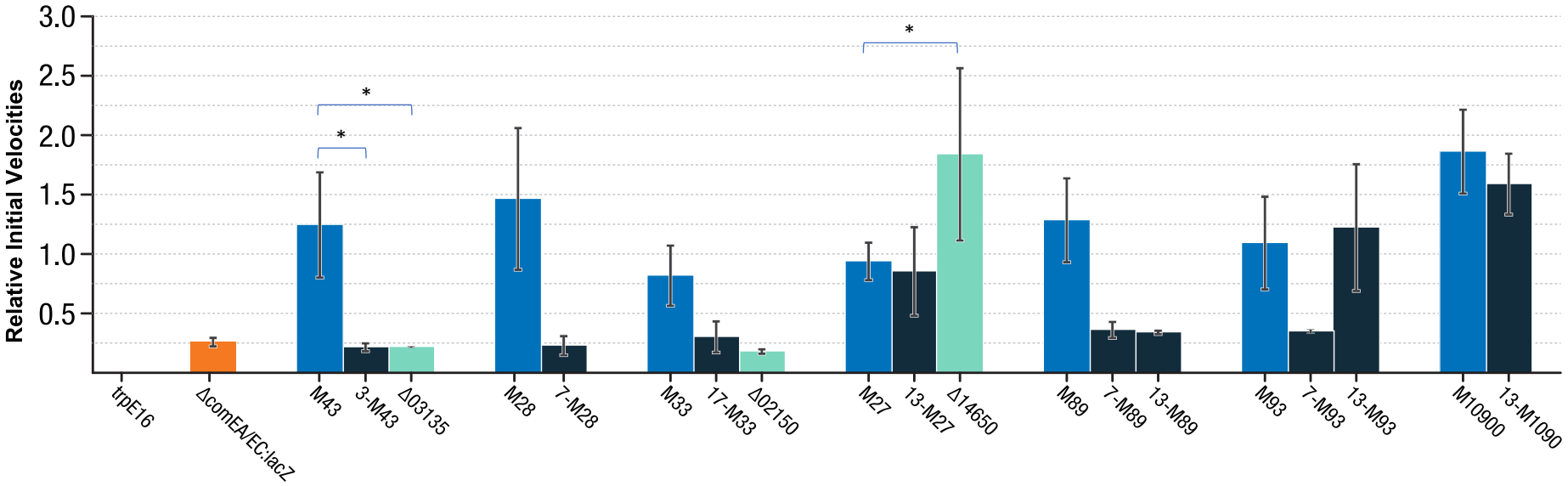


Figure 4.21. Relative initial velocities of *M. luteus* trpE16 and its derivative EMS- und single-SNP-mutants.
ComEA/EC promoter transcriptional levels for *M. luteus* trpE16, *ΔcomEA/EC:lacZ* reporter strain (orange) and all its derivative mutants under normally repressive conditions (LB). All cultures were normalized to cell density prior lysozyme treatment. LacZ substrate MUG was added and kinetic data for the promoter activity was recorded over 8 hours by measuring the fluorescence generated from the substrate cleavage. A four parameter logistic model was fitted to the kinetic curves (n ≥ 3). The slope of the logistic curve served as a measure for the relative promoter activity. Expression levels of the M-mutants are shown in blue, next to them the single SNP mutants in dark blue, and the gene knockouts in turquoise. The mean and standard deviations of all experiments are represented. An unpaired, two-tailed Student's t-test showed no difference between M27, M93, M1090 and 13-M27, 13-M93 and 13-M1090 respectively. This figure was created by this thesis author for the publication of this work (Torasso Kasem *et al.*, 2021)

4.6 *M. luteus* transformation frequency variations

An important question about the selected and re-inserted mutations was if they had a direct impact on *M. luteus* transformability. To answer this question, transformation frequency assays were carried out. The wild-type strains with the SNPs or the clean deletions were cultured on GMM and transformed with pJET plasmid. *M. luteus* 13-M89 systematically showed a 100-fold decrease in transformability compared to the trpE16 strain. Mutants *M. luteus* 2-R33 and *M. luteus* 8-R34 had even lower frequencies of transformation, but not as low as the knockout *M. luteus* Δ14650 or the mutant *M. luteus* 13-R4 (See Figure 4.22). This last mutant is the only one that seems to have a transformation frequency well below the detection limit of the assay. Because of this, it was not feasible to construct a Δ*comEA/EC*:*lacZ-Kan* reporter strain after introducing the mutation. Mutants *M. luteus* 13-M93, 13-M1090, 13-M27, Δ02150, and Δ03135 didn´t shown a significant difference to the reference strain trpE16.

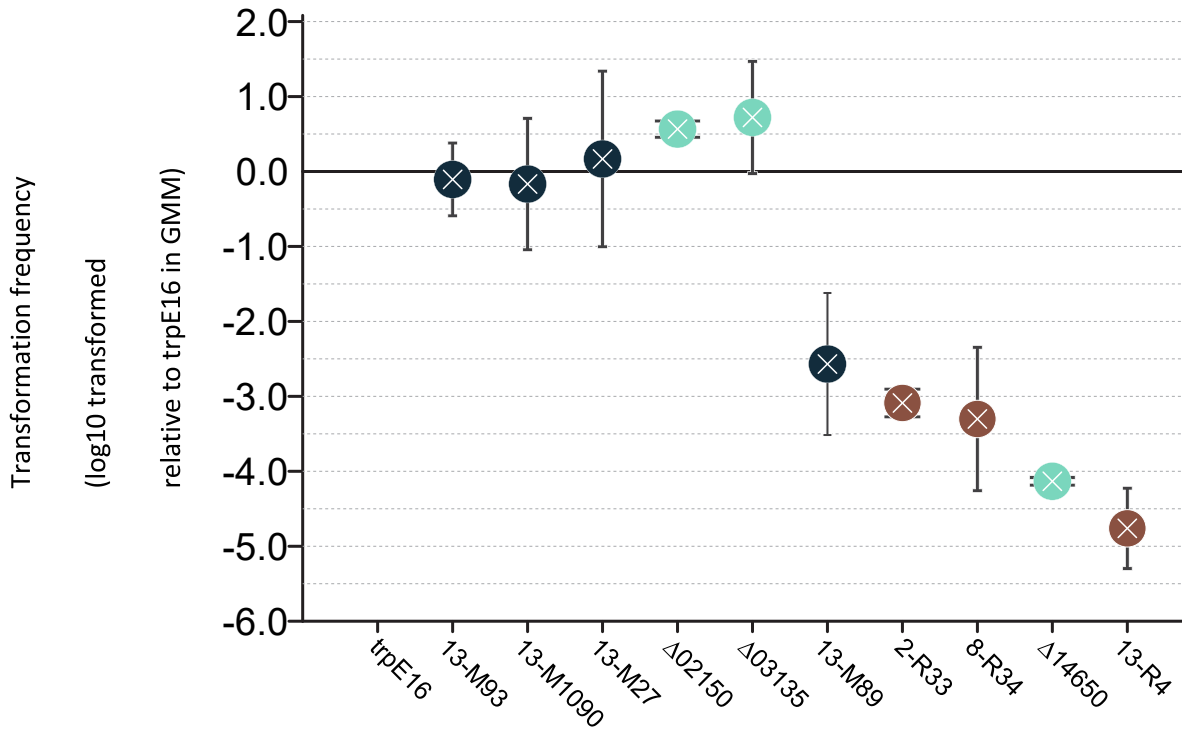


Figure 4.22 Effect of specific genome editions on the transformability of *M. luteus*.
Transformation frequencies of clean gene deletion mutants are shown in turquoise, while single-SNP mutants with SNPs from the M-mutants and the R-mutants are shown in dark blue and brown respectively. Each dot represents the $\log_{10}$-transformed mean transformation frequency relative to the mean of the trpE16 strain. All strains were grown in GMM for 20 hours until OD_{600} reached 1. The error bars indicate the SD of the biological replicates. An ordinary one-way ANOVA test was carried out to compare all means with each other ($p<0.005$, $n \geq 3$). This figure was created by this thesis author for the publication of this work (Torasso Kasem *et al.*, 2021)

4.7 5´UTR determination through 5´RACE

Given the fact that Mlut_14650 and Mlut_14660 are divergently overlapping (← →), it is difficult to tell if the affected positions disturb the usual functioning of Mlut_14660 only or of both genes. To shed more light on this, a 5'RACE experiment was performed to determine where the transcriptional start site. Figure 4.23 shows the peaks of the Sanger sequencing result. As a T-tail is added during the 5'RACE preparation, the 5'UTR starts immediately thereafter. The 5'UTR of Mlut_14650 has a length of 76 bp and the one of Mlut_14660 has a length of 51 bp. Considering that the intergenic region between the 5'UTRs of both genes is 97 bp long and that the DNA bound RNApol footprint covers around 70 bp (from -50 to +20) (J. Chen, Darst, & Thirumalai, 2010; Sneppen *et al*., 2005), it seems feasible that relevant parts of the promoter and operator sequences of each transcriptional unit indeed could interfere with the smooth transcription of the other.

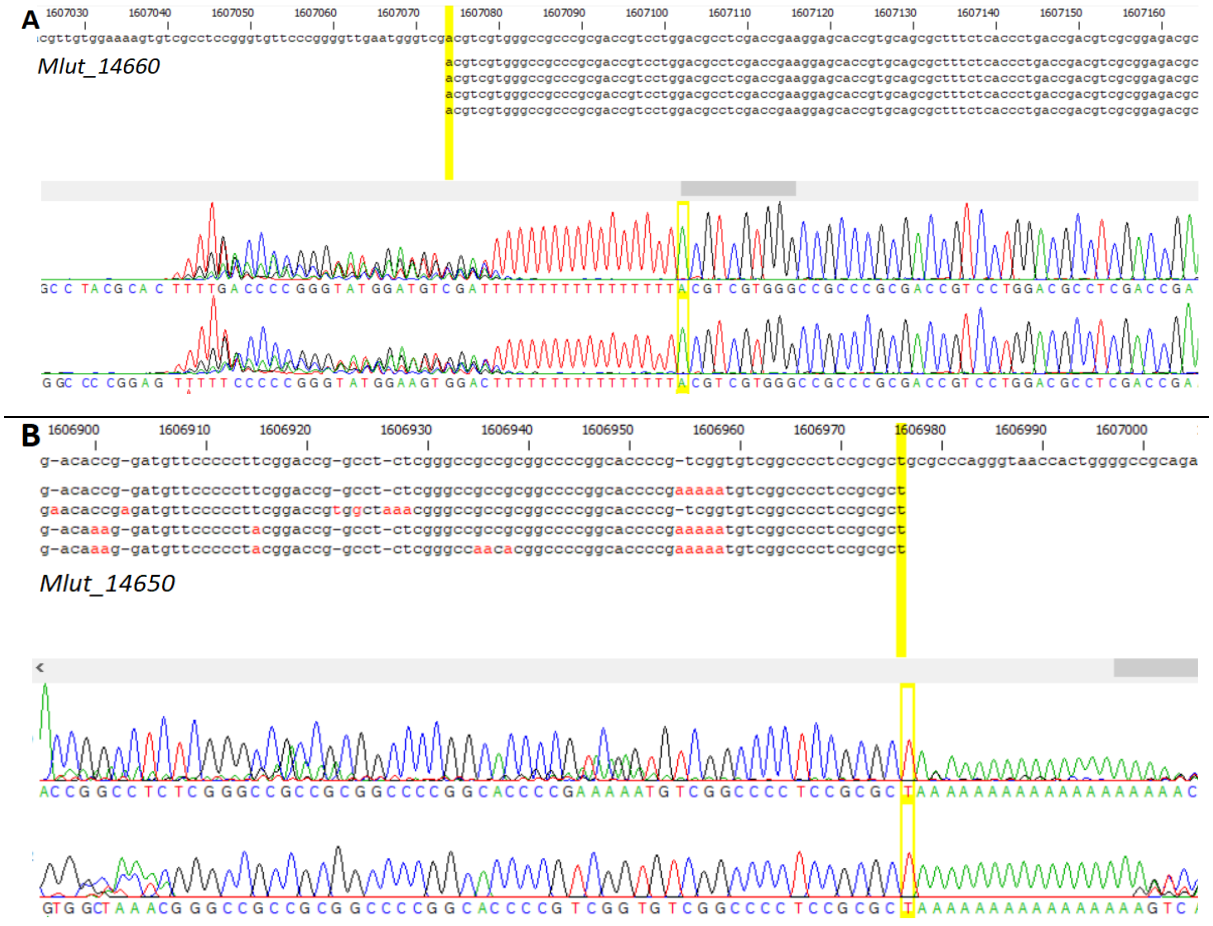


Figure 4.23 Mlut_14660 and Mlut_14650 5´UTR determined through 5´RACE.
After total RNA preparation from *M. luteus* trpE16 grown in GMM, cDNA was synthesized and the Sigma-Aldrich 5`RACE Kit 2nd generation protocol was carried out. After obtaining the expected size amplicons, these were sent for sequencing. (A) Four individual biological replicates for Mlut_14660 transcript. The 5´UTR starts at position 1607074, 51 bp upstream of the starting codon. (B) Four individual biological replicates for Mlut_14650 transcript. The 5´UTR starts at position 1606976, 76 bp upstream of the start codon.

4.8 Secondary structure and translation initiation rate prediction of the Mlut_14660 transcript

To predict the effect of the SNPs 13-M89 and 13-M1090 in Mlut_14660, its original and mutated 5´UTR sequences were analyzed with the secondary structure prediction program (see Table 3.20) and with the RBS calculator. With the first of these tools, it was possible to observe 3 loops forming, with a Gibbs free energy of -19.1. After insertion of SNP 13-M1090, there seems to be no significant change generated. But with the SNP 13-M89, the whole conformation changes to one long loop structure, that has a higher free energy level (Figure 4.24).

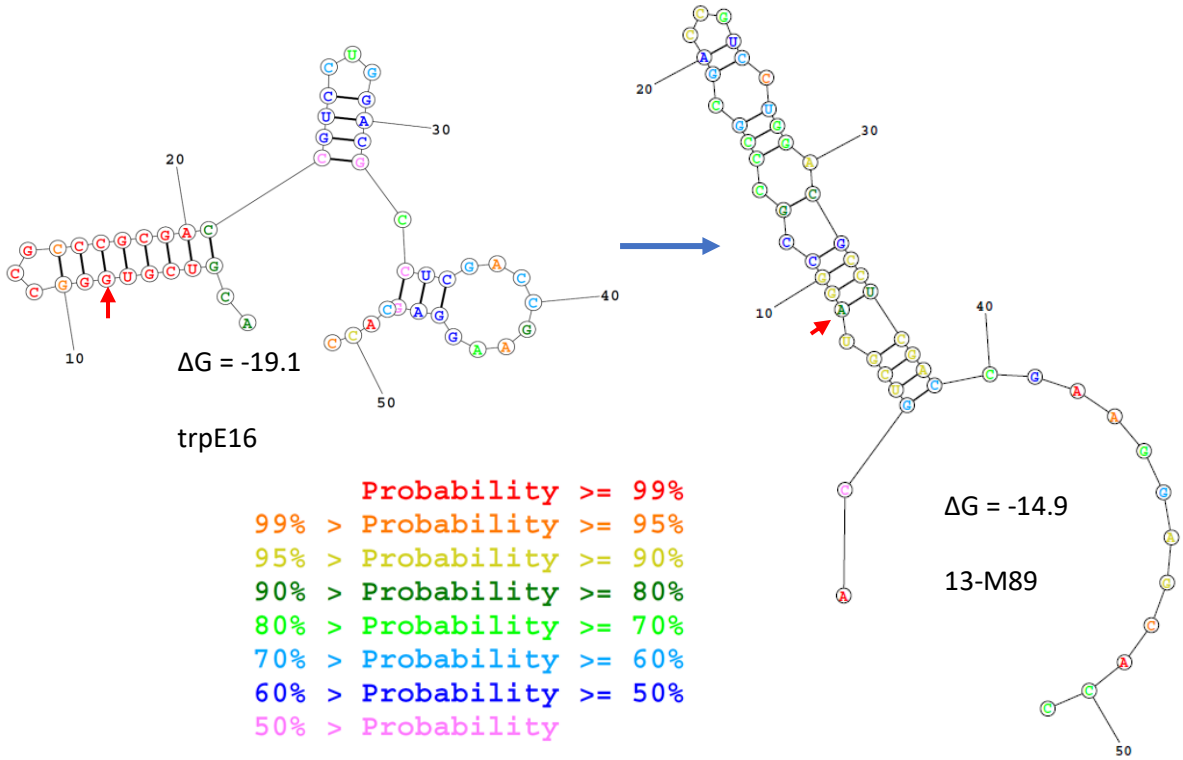


Figure 4.24 Output of Secondary structures prediction program for Mlut_14660 transcript.
Most probable secondary structures predicted through RNA secondary structure prediction tool (https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/AllSub/AllSub.html). Three loops can be observed for the original nucleotide sequence, having the one at the beginning of the transcript the highest probability of formation. When the SNP 13-M89 is inserted, the previous structure is lost and another one forms. This figure was created by this thesis author for the publication of this work (Torasso Kasem *et al*., 2021)

By analysis of the sequences in the RBS calculator program for translation initiation rates a decrease was predicted: from 74.96 arbitrary units (au) to 24.74 au for the transcript with the SNP 13-M89 and 17.13 au for the one of the 13-M1090. The ΔG_{total} for the translation initiation increased from 6.22 Kcal/mol to 8.69 Kcal/mol for 13-M89 and 9.50 Kcal/mol for 13-M1090. When the secondary

structures were predicted by this program, the same loop in the 5´end was observed, being disrupted again when the SNP 13-M89 was inserted.

According to these predictions, it seems that SNP 13-M89 affects the standby site (ribosome preinitiation landing pad) of the mRNA, while the SNP 13-M1090 changes the Shine-Dalgarno sequence that turns from 5´-AAGG-3´ to 5´-AAAG-3´ (see Figure 4.25).

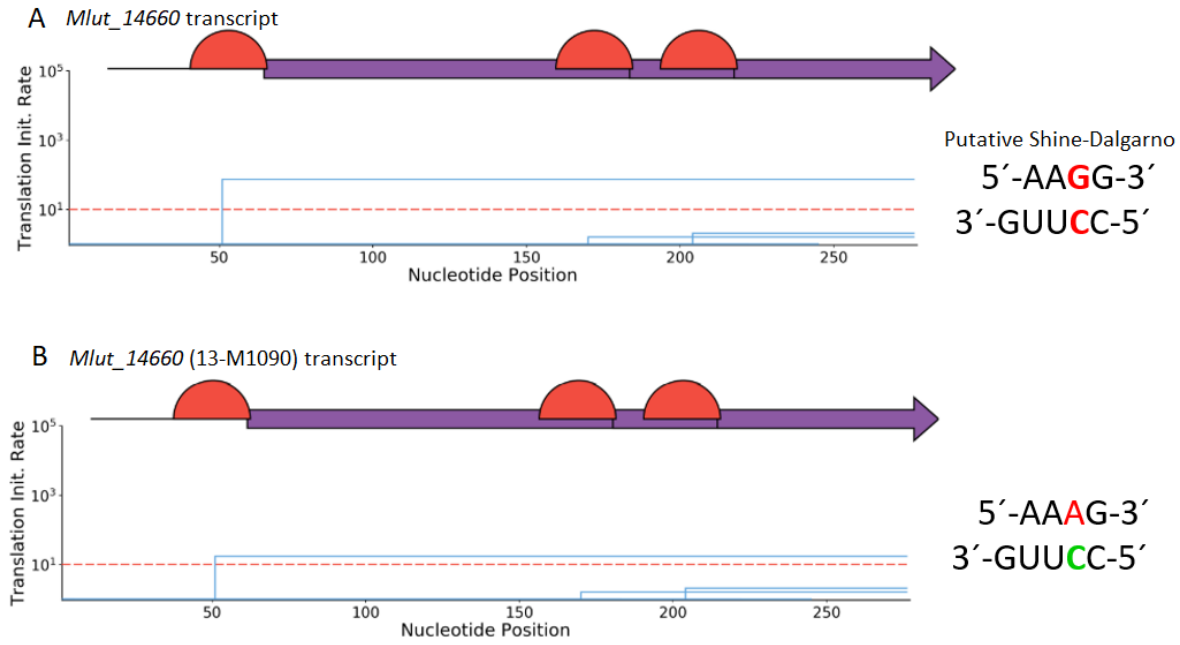


Figure 4.25 Translation initiation rates of *M. luteus* trpE16 and the mutant 13-M1090.
Translation initiation rates for each starting codon found at the mRNA sequence of Mlut_14660. The RBS calculator program was used (https://salislab.net/software/predict_rbs_calculator) to compare the values of (A) *M. luteus* trpE16 and (B) the mutant 13-M1090. A significant decrease in translation initiation is predicted after mutation G1607117A occurs. Red semicircles represent putative ribosome binding sites.

4.9 Mlut_14660 overexpression and purification

To investigate the putative function of the protein encoded by Mlut_14660, the Mlut_14660 ORF was cloned into the expression plasmid pMAL-c2x. After 4 h of protein expression, the cells were harvested, sonicated and the MBP-Mlut_14660 fusion protein (Mlut_14660-encoded polypeptide sequence fused to the C-terminus of MBP) was purified from the crude extract with an amylose resin. SDS PAGE of the crude extract and the amylose eluate (see Figure 4.26) shows the ~50 KDa band corresponding to the MBP-Mlut_14660ORF fusion protein. To separate the Mlut_14660 encoded protein from MBP, a Factor Xa protease digestion was performed. Tris-tricine PAGE demonstrated that the fusion protein digestion was successful. The bands corresponding to the MBP and the Mlut_14660 encoded protein (42KDa and ~8 KDa) were visible. Due to the small size of the target protein, it cannot be clearly said from PAGE alone if all the fusion protein was digested. In order to separate the MBP from the putative DNA binding protein, the mixture was run through an affinity chromatography column (see Methods 3.32). However, it was not possible to separate the proteins. A VivaSpin Turbo with a size cut-off of 30 KDa was further used to separate the MBP protein from the protein of interest, but upon separation of the flow-through on a Tris-Tricine-PAGE gel, no band was detected showing the Mlut_14660 encoded protein. This suggests that the protein may be degraded after being separated from the MBP. This is the reason why all subsequent analyses were performed with the fusion protein and appropriate controls.

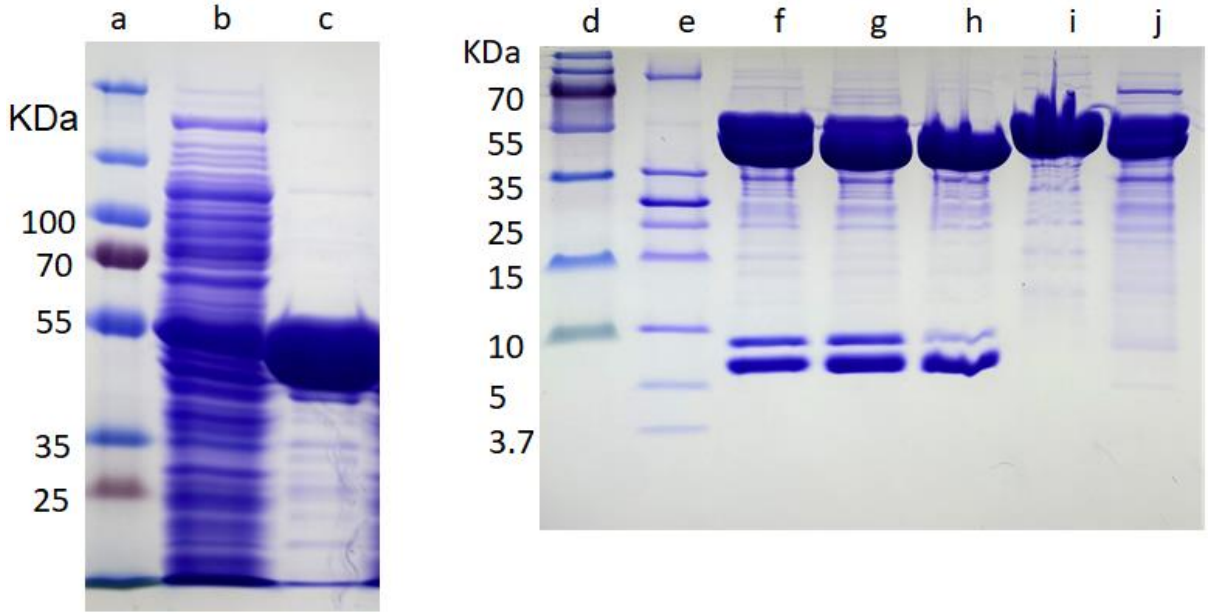


Figure 4.26 SDS-PAGE of each step in the purification of Mlut_14660 protein.
The left-hand panel shows a Coomassie-stained 10% SDS-PAGE gel. (a) (d) PageRuler™ Plus Prestained Protein Ladder was used as standard (5 µL). (b) Crude extract 4 h after IPTG induction and sonication. (c) Eluate after amylose resin purification with MBP-spin protein Miniprep kit. The panel on the right shows a low range Tris-tricine 18% SDS-PAGE gel. (e) PageRuler™ Unstained Low Range Protein Ladder (5 µL). (f) MBP/Mlut_14660 fusion protein 3 h after Factor Xa digestion. (g) MBP/Mlut_14660 fusion protein 6 h after Factor Xa digestion. (h) Fusion protein after O/N Factor Xa digestion. (i) Control: MBP/Mlut_14660 fusion protein without FactorXa digestion. (j) Control: MBP alone after O/N FactorXa digestion.

4.10 Affinity of the MBP-Mlut_14660 fusion protein to the *comEA/EC* promoter sequence

As a putative regulator/repressor of the *comEA/EC* promoter, the Mlut_14660 encoded protein was hypothesized to specifically interact with this promoter. This was tested by EMSA in agarose gels. Given that it was not easy to separate the MBP and the putative DNA binding protein Mlut_14660 the EMSA was carried out with the fusion protein and the MBP alone control. The assay was done with a 1.2 Kbp fragment containing the *comEA/EC* promoter. Controls without protein, without DNA, and with the MBP alone behaved as expected (see Figure 4.27 and Figure 4.28). When mixing the fusion protein with the DNA of the promoter, there was significant retention of the amplicon until around 4 µM of protein. The concentration at which 50% of the protein is bound to the DNA seems to be below 3.7 µM (lane 9 in Figure 4.27). Therefore, another assay was carried out with a narrower range of protein concentrations around the Kd, starting with 2 µM until 0.5 µM. In this second assay, two DNA sequences were used as controls. These were expected not to interact with the putative repressor protein; one was from the *M. luteus* gene Mlut_02900 encoding a

dehydrogenase, and the other was from the kanamycin resistance gene. The fragments had sizes of 929 bp, 706 bp, and 911 bp respectively. Results were similar for all sequences, showing a shift at around 1.5 µM. For the *Kan* fragment, it seemed to be a bit lower than for the *M. luteus* sequences (see Figure 4.28). This shows that, at least under these conditions, the protein was not specifically binding to a particular sequence.

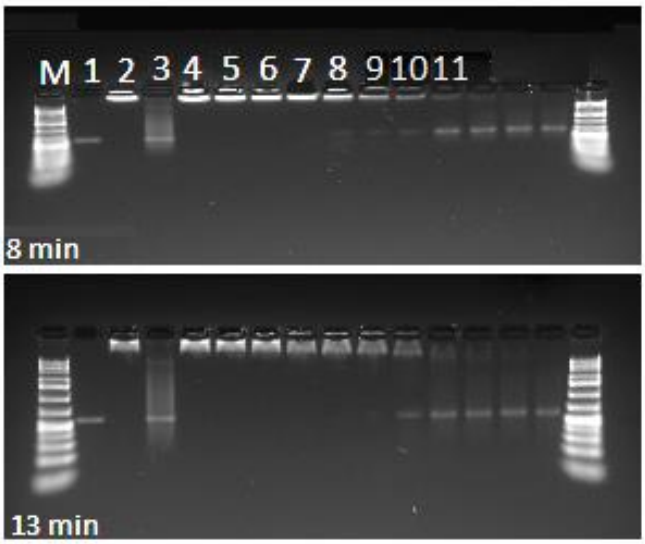


Figure 4.27 EMSA of amplicons encompassing the *comEA/EC* promoter and the MBP-Mlut_14660 fusion protein.

Both panels show a 2% agarose gel run at 180V, the upper one was run for 8 min, and the lower one for 13 min. From left to right: (M) DNA ladder; (1) No protein control, only the 1.2 kbp fragment; (2) No DNA control, only the fusion protein [9.82 µM]; (3) MBP alone [9.82 µM], no fusion protein present; (4) Amplicon containing the *com* promoter mixed with the MBP-Mlut_14660 fusion [9.82 µM]; (4-10) Amplicon and the fusion protein pipetted in a concentration decreasing gradient of 1.23 µM steps. For more details on the set-up of the assay see 3.33 of the Methods section.

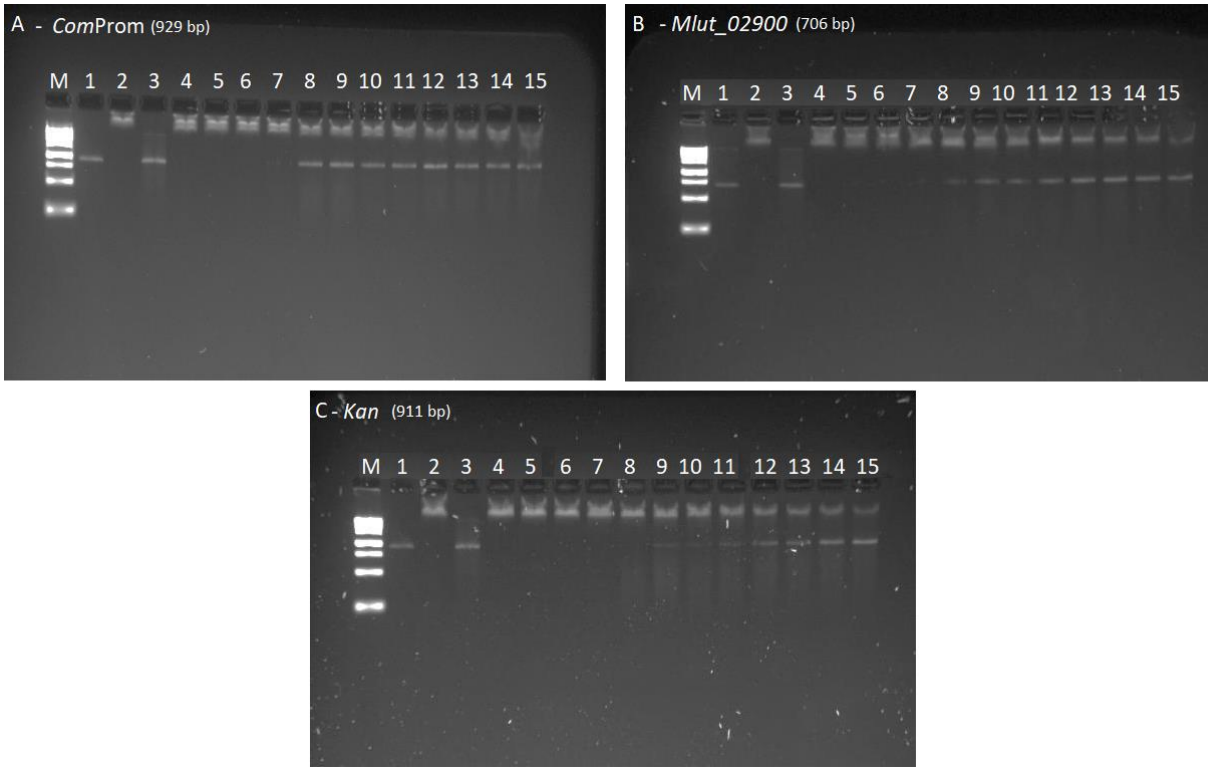


Figure 4.28 EMSA of different PCR amplicons and the MBP-Mlut_14660 fusion protein.
All agarose gels (2%) were run for 12 min at 180V. Panel A corresponds to the 929 bp amplicon holding the *comEA/EC* promoter, panel B to the 706 bp amplicon holding an upstream region of Mlut_02900, and panel C to the 911 bp amplicon containing the *Kan* resistance gene. Lanes from left to right: (M) DNA ladder; (1) No protein control, only the PCR amplicon; (2) No DNA control, only the fusion protein [9.82 µM]; (3) MBP alone [9.82 µM], no fusion protein present; (4) An amplicon containing the *com* promoter was mixed with the MBP-Mlut_14660 fusion [9.82 µM]; (5-15) All the pockets left also had the amplicon and the fusion protein in a concentration decreasing gradient in steps of 0.12 µM. For more details on the set-up of the assay see Methods 3.33 Table 27.

4.11 Gene architecture conservation throughout actinobacterial species

To evaluate the presence and disposition of the studied genes and those flanking them in other members of *Actinobacteria* we used the Microbial genomic context viewer. This software allows the user to see if there are homologs of genes of interest, how many copies, and how they are organized in the genome. The software's database consists of type strains, and it does not hold all actinobacterial organisms. Therefore, these results might ignore other organisms which also possess the genes under consideration. BLAST was used to obtain the E-values of the homologs by using *M. luteus* genes sequences as queries. In species of *Geodermatophilus, Rothia*, *Jonesia,* and *Cellulomonas* there seems to be the same genomic organization of these transcriptional units as in *M. luteus*. In *Nocardioides* the genes are also present, although they have a different orientation relative to each other. For *Gordonia* and *Kytococcus* species, the program showed that some homologs of the genes are present, but organized differently (see Figure 4.29).

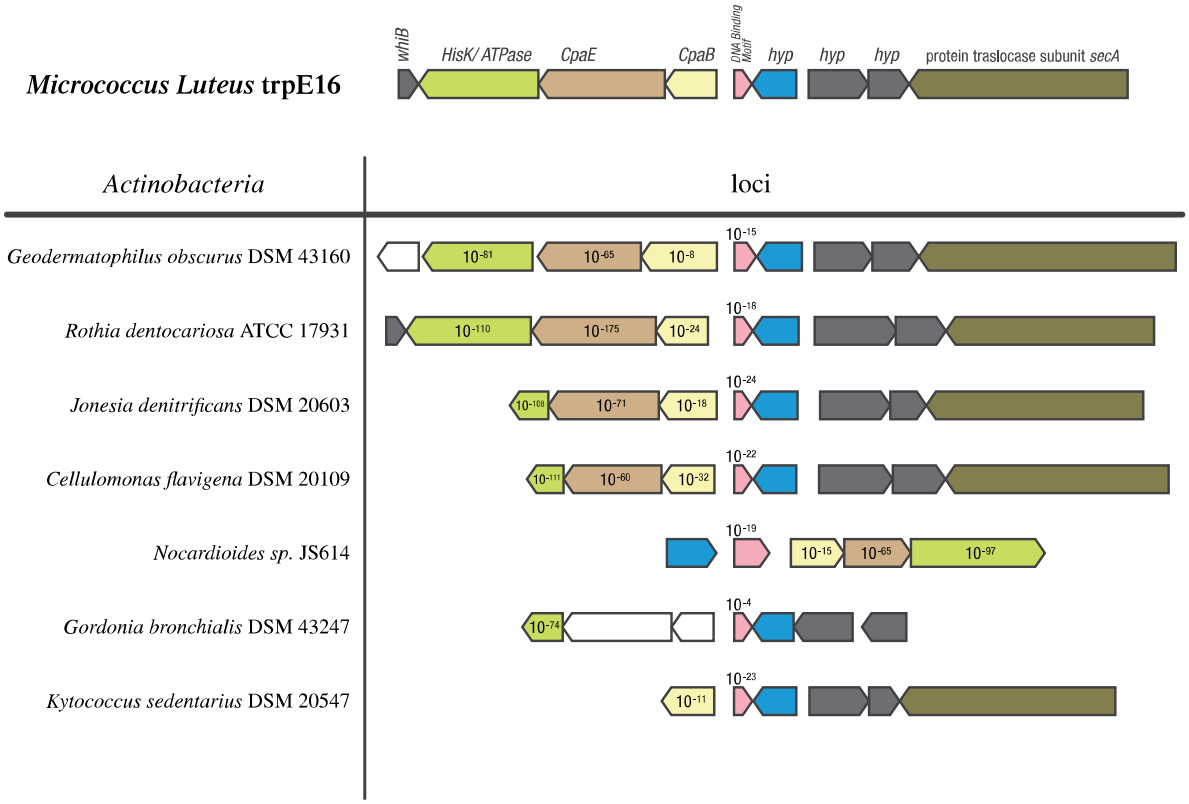


Figure 4.29 Physical maps of Mlut_14660, Mlut_14650/40/30 loci and their genomic surroundings from the genomes of a variety of actinobacterial representatives (high-GC Gram-positive) and our model organism *M. luteus.*

ORFs of the same color encode proteins with significant similarity (BLAST E < 10^{-4}) to the respective gene products of *M. luteus*. The predicted gene products are shown and the numbers in the arrows representing homologs are E-values of a BLAST search using the full-length *M. luteus* genes as queries. This figure was created by this thesis author for the publication of this work (Torasso Kasem *et al*., 2021)

5 Discussion

Natural genetic transformation enables bacteria to acquire exogenous DNA molecules, thus, this means of horizontal gene transfer plays an essential role in the evolution of bacterial populations. Among the most fundamental and conserved proteins involved in the process are the multiple membrane-spanning ComEC and the DNA receptor ComEA. Their presence and importance in any of the described naturally transformable organisms, including the actinobacterial model organism *Micrococcus luteus*, is highly conserved.. In the present study, we aimed to find out more about how the co-transcribed genes *comEA-comEC* are regulated. To achieve this goal, we decided to take a rather classical approach, based on a sequence of studies done on *Bacillus subtilis* by Dubnau and other authors between 1986 and 1990 who investigated competence regulation in this Gram-positive model organism from the phylum *Firmicutes.* Dubnau and his colleges isolated 28 mutants of *B. subtilis* that were deficient in competence acquisition by using *Tn917-lacZ*, a transposon derivative that can generate transcriptional gene fusions. After mapping the mutations, the mutants were characterized according to their abilities to bind and take up DNA. While all mutants were defective in DNA uptake, only some showed a lower performance for DNA binding (Perkins and Youngman, 1986; Hahn *et al.*, 1987). The expression of β-galactosidase under the control of the *com* promoter was studied as a function of growth stage, medium, and genetic background (Albano *et al.*, 1987). Later in 1990, Dubnau and Roggiani used EMS-induced chemical mutagenesis to generate mutants that would overexpress *com* genes in normally repressive conditions, like in full medium. The identified mutations were designated *mec* (for medium-independent expression of competence), and two loci were mainly affected by them, named *mecA* and *mecB.* In subsequent studies, it was determined how they affected the regulation of medium-dependent expression of competence (Turgay *et al.*, 1998).

Similar to what was done for *B. subtilis*, a *lacZ* reporter strain for the late *com* genes *comEA/EC* was constructed for *M. luteus* in our group (Lichev *et al*., 2019). In this study, EMS random chemical mutagenesis was used as a valid and efficient method for introducing mutations into the *M. luteus* trpE16 genome. On average, one stable transition was obtained every 3×10^5 bp. This is a convenient number given that, if it were much lower, the frequency of appearance of the desired phenotype would be too low. On the other hand, if it were much higher, it would be very unlikely to find the causative mutation for a specific trait. Regarding the mutagenesis efficiency, the values obtained in this work match with the ones obtained by Dubnau and Roggiani in their study. For the screening of mutants, the lactose analog X-Gal included into the agar medium plates was used instead of the MUG plate-spraying technique (David Dubnau & Roggiani, 1990). With the MUG plate-spraying technique, when looking for M-mutants, it was not easy to qualitatively distinguish up-regulated colonies from

regular ones. The fluorescence was too homogenous when many colonies were close to each other, and this would prevent them to be distinguishable. It was also not convenient for the screening of R-mutants, given the lower sensitivity of the method. With relatively higher X-Gal concentrations (80 µg/mL), it was possible to differentiate strongly repressed phenotypes from others in *M. luteus*. In both types of mutants ("M" and "R"), the expression of the essential late competence genes of interest was affected. Even though the expression of *comEA/EC* is normally inhibited in full medium, M-mutants exhibited a clear and strong expression. Both groups of mutants failed to show the usual response to nutrient availability, the R-mutants displaying a repressed phenotype when the competence-inducing minimal medium was used.

β-Galactosidase activity can be analyzed by several different chromogenic or fluorogenic substrates, and is a versatile reporter for bacterial gene expression. I used the first standard assay for quantification of β-galactosidase activity in bacterial cells which was described by Miller (Miller, 1972). It involves spectrophotometric measurements of the generation of the yellow chromophore p-Nitrophenol (pNp), which is the hydrolytic product of the β-galactosidase on the colorless substrate *p*-nitrophenol-β-D-galactopyranoside (see Methods 3.20.2.1). Arbitrary units were calculated as function of the amount of *p*-nitrophenol (*p*Np) produced, the reaction time, the volume of cell culture, and its optical density. The permeabilization of cells is a critical step in the assay process. Instead of using toluene or chloroform and SDS, or sonic disruption, it was decided to use lysozyme for obtaining crude extracts. This *p*Np adapted method was used for the M-mutants. Although giving reliable results, this method has some disadvantages. First of all, it cannot be easily performed in high throughput. Working with 1.5 and 2 mL tubes made sample handling more laborious and less efficient for obtaining optimal results in less time. Timing at each step of the procedure becomes less uniform as the number of samples increases, which limits the number of samples that can be evaluated. Volumes used were relatively high and lysis degree seemed to vary a lot. Additionally, the measurements were individually performed in 2 mL quartz cuvettes. Therefore, the assay had to be carried out more often to increase the reliability of the observed outcome, but still there were some fluctuations in the results. Because of these drawbacks, we decided to change the β-galactosidase substrate to fluorogenic 4-methylumbelliferyl-β-D-galactopyranoside (MUG), which can easily permeate through the cellular membranes, leading to results that better reflect the total enzyme present in a sample. Due to the high number of samples and time points, the assay was adapted to a 96-wells plate format, employing a microtiter plate reader and multichannel digital pipettes. Manipulations are more simple by using this methodology (see Methods 3.20.2.2), the plate reader permits immediate and accurate absorbance determination, and the data are easily exported and

analyzed with Microsoft Excel (K. L. Griffith & Wolf, 2002). In conclusion, this way of performing the assay yields more reproducible results with low standard errors and was used for the quantitative screening of the R-mutants (see Results 4.7) and for comparing enzyme kinetics between parental M-mutants and the single-SNP mutants (see Results 4.5.2) to determine if certain mutations were causative for the phenotype of the mutants.

It was crucial to filter out mutants that, despite displaying the desired phenotype, may have *cis*-acting *comEA/EC* promoter or *lacZ* mutations, generating an overexpression or repression phenotype that was not due to the change of an unlinked regulatory sequence. By transforming the wild-type strain with gDNA of the M or R mutants, it was possible to analyze the degree of genetic linkage between the causative SNP and the 7 kbps Δ*comEA*/*EC*:*lacZ-Kan* construct. Results were presented as linkage percentages, which were obtained by calculating the percentage of colonies that after transformation with gDNA from the M or R mutants showed the same phenotype as the mutant from which the gDNA was extracted. The LOD score (Morton, 1955) normally used to quantitate genetic linkage for genetically more complex organisms like eukaryotes was not necessary in this particular case. A similar experiment was carried out by Dubnau and Roggiani to study *cis*-acting mutations in *comG* of *B. subtilis* (David Dubnau & Roggiani, 1990).

The genomes of 22 M-mutants and 14 R-mutants were sequenced and variant calling was performed by using the "Snippy" software tool, which finds high confidence differences (indels or SNPs) between a known reference genome and the reads of interest. This software aligns the reads to the reference genome, decides if any of the resulting discrepancies are real variants or technical artifacts that can be ignored, and lastly checks which effect these differences have on the predicted genes (stop mutations, frameshifts, etc.) (Bagnacani *et al*., 2018). Sixteen SNP clusters were identified for the M-mutants, whereas for the R-mutants there was not a clear clustering. This could mean that there may be more than one region in the genome that can generate the partial or total repression of the *comEA/EC* promoter. It would be interesting to see if a more clear clustering shows up for that phenotype by increasing the EMS exposure time or the number of R-mutants sequenced and aligned against the reference genome. Of the 16 M-mutant clusters, the ones with larger numbers of positions affected and of mutants carrying one of them were selected. Another selection criterion was the annotation of the genes being affected by mutations, prioritizing DNA-binding proteins, transcription factors, proteases, response regulators, or any other known to participate in a known competence regulatory network.

Discussion

In principle, the most outstanding clusters were number 3, 7 and 13. Cluster number 3 encompasses 6 mutants with 2 positions affected within the same gene, Mlut_03135, which codes for an N6-adenine DNA methyltransferase of unknown recognition sequence. This type of enzyme is widely spread in prokaryotes, and besides their role in restriction-modification systems (Bickle & Kruger, 1993), they represent an important way of regulation of gene expression. Changes in promoter methylation status have been shown to control several genes in different model organisms like *E. coli* and *Caulobacter sp.* Although both adenine and cytosine methylations are found in both bacteria and eukaryotes, adenine methylation seems to have a greater regulatory role in bacteria (Adhikari & Curtis, 2016). A transcriptomic study carried out in *E. coli* reported many deregulated genes in DNA adenosine methyltransferase mutants, including global regulators like catabolite activator protein or fumarate nitrate reductase (Oshima *et al*., 2002). Cluster number 7 has 3 positions affected in 6 different mutants. The affected loci, Mlut_09880 and Mlut_09900, encode the AAA+ ATP-dependent Clp protease proteolytic subunit ClpP and the ATP-binding subunit ClpX respectively. ClpX recognizes unstructured peptide tag sequences in protein substrates, proceeds to unfold stable tertiary structure in the protein, and then spools or translocates the unfolded polypeptide chain into a sequestered proteolytic compartment in ClpP for degradation into small peptide fragments (Baker and Sauer, 2012). The levels of the main transcriptional regulator ComK of *B. subtilis* are controlled by this kind of proteolytic system, composed by the previously mentioned adapter protein MecA (see above, and Introduction) and the ATP-binding subunit ClpC (formerly known as MecB) (David Dubnau & Roggiani, 1990; Turgay *et al*., 1998). Mutants in *mecA, clpC*, and *clpP* experience an overexpression of *com* genes and others activated by ComK. This scenario seems to match what has been observed in the M-mutants carrying changes in the region of cluster 7. And finally, cluster number 13, which holds mutations at 6 different positions, with one of them found in almost every M-mutant (except for M33). As can be observed in Figure 3.8, the highest density of SNPs in the scatter plot corresponds to this region of the genome. ORFs of two divergently overlapping transcriptional units are affected, i.e., SNP G1606810A/V31I (13-M27) lies within the Mlut_14650 ORF, which codes for a SAF-domain containing Flp pilus assembly protein RcpC/CpaB, SNPs G1607081A (13-M89) and G1607117A (13-M1090) putatively affect the 5´UTR of Mlut_14660, and SNPs C1607204T/S27L (13-M96), C1607261T/S46F (13-M9) and C1607314T/R64W (13-M93) affect at least the Mlut_14660 ORF. This small gene encodes a DNA-binding protein with an HTH-17 domain annotated as member of the excisionase protein family. The operon upstream of Mlut_14660 encompasses 3 structural genes, Mlut_14640 and Mlut_14630 downstream of the already mentioned Mlut_14650. The first gene codes for a *tadZ*/*CpaE* pilus assembly protein, and the second one for a PAS domain-containing sensor histidine kinase. The RcpC/CpaB homolog in *M. luteus* possess a signal peptide (predicted by SignalP 5.0 software) and no

transmembrane domains, suggesting that the protein is located outside the membrane like its homologs in other organisms (Tomich *et al.*, 2007). *RcpC* present in the prototypical *tad* locus of *Aggregatibacter actinomycetemcomitans, cpaB* in *Caulobacter crescentus* and all its known homologs, including those in *H. ducreyi*, *P. multicoda*, and *Pseudomonas* spp. among others, contain two β-clip motifs involved in a variety of functions including binding carbohydrate moieties to assemble structures like Flp pili (Clock, Planet, Perez, & Figurski, 2008; Iyer & Aravind, 2004). There is also another hypothesis that proposes that one or both β-clip motifs glycosylate the pilin subunits to facilitate its assembly. As pointed out by Tomich *et al.,* 2007, the *rcpC/cpaB* gene is absent in the *tad* loci of Gram-positives, which is also the case for *M. luteus* (Angelov *et al.*, 2015). This does not mean that the genes could not still be present under the control of a different transcriptional unit. The same applies for *tadZ*/*cpaE*, although orthologous genes are indeed present in the *tad* clusters of other actinobacteria (Tomich *et al.*, 2007). This cytosolic soluble protein is an essential component of the Tad secretion system in many organisms, and its particular molecular architecture combines components from the bacterial cytoskeleton (MinD/ParA/Soj ATPases) and two-component signal transduction response regulator. Studies in *A. actinomycetemcomitans* and *C. crescentus* showed that they probably function as localization factors (Perez-cheeks *et al*., 2012; Xu *et al*., 2012). In *C. crescentus,* CpaE mediates positioning of the pilus secretin protein CpaC, as well as the histidine kinase PleC. Lastly, Mlut_14630 codes for a putative PAS-domain histidine kinase . Proteins containing PAS-domains are usually modular, functioning as sensors that detect a variety of stimuli and regulate, in response, the activity of effector domains (output) with catalytic or DNA binding functions (Möglich *et al.,* 2009). In some organisms, like *P. aeruginosa* (Bernard *et al*., 2009)*, A. actinomycetemcomitans*, *A. aphrophilus* (Tomich *et al*., 2007), and *C. crescentus* (Viollier *et al.*, 2002a, 2002b), the environment-dependent modulation of pilin production through a two-competent regulatory system of this kind seems to be a common feature. Curiously, the only M-mutant that didn't hold a mutation of cluster 13 was also the only one with the transition C226684T affecting Mlut_02150 ORF (R121W). This gene also encodes a DNA-binding protein from the excisionase family, as Mlut_14660 does, and therefore it was also selected for evaluation.

When selecting for the most interesting putative causative mutations of the R-mutants the clustering was not as clear as with the M-mutations (see Figure 4.9 R-mutants' SNP distribution and clustering analysis.). There were almost no positions affected in more than one R-mutant. Besides the regions with higher clustering scores, a visual analysis based on gene annotations was carried out. Mutation G810889A/G327E affecting Mlut_07490 was among the selected ones. This gene codes for an AAA+ ATPase DNA translocase of the FtsK/SpoIIIE family, which is interestingly right next to the

tad-1 cluster (Angelov *et al.,* 2015). Members of this family are involved in a wide variety of processes in many species, such as conjugation, DNA packaging, DNA segregation, and maintenance (Iyer & Aravind, 2004). Mlut_07490 seems to share some structural features with FtsK and SpoIIIE, like a well-conserved DNA translocation module and transmembrane helices, being four in the particular case of FtsK and SpoIIIE, but only two for Mlut_07490 (sequence alignment in www.uniprot.org/align Identifiers: P46889, P21458, and C5C9X9). These transmembrane domains are localized at the proteins' N-termini, usually followed by a sequence of variable length that interacts with other proteins. At the highly conserved C-termini there is a DNA-binding module whose specific role in *M. luteus* has not yet been studied (Barre, 2007; Crozat, Rousseau, Fournes, & Cornet, 2014). Mlut_07490 has the typical highly conserved P-Loop consensus amino acid sequence "GTTGSGKS", called Walker A motif. The Walker B motif is usually located downstream of Walker A and has a much higher variability. The only invariant features of this motif are a negatively charged residue following a stretch of bulky, hydrophobic amino acids, like the "AAAAPAVAPWE" present in *M. luteus* Mlut_07490 (Koonin, 1993). In a similar study carried out on *B. subtilis* (Foulger & Errington, 1989), a *spoIIIE* mutation was the only one that significantly decreased the *lacZ* reporter gene expression (in that case under the *spoIIIG* promoter). This suggests that *spoIIIE* product is directly involved in a mechanism of differential gene expression. The membrane-spanning and mononucleotide-binding domains suggest that the protein might function as part of a transduction pathway based on phosphoryl group transfer, similar to the NtrB/NtrC system in other bacteria (Pelton *et al.*, 1999). Within cluster number 2 of the R-mutants, there are also 3 other positions affected, 2 missense transitions within Mlut_07500 encoding the Flp pilus assembly protein/NTPase TadA1/CpaF and another one in Mlut_07560, a hypothetical Flp pilus assembly TadG-like protein. Both proteins belong to the *tad-1* cluster and are essential for transformation/pili assembly (Angelov *et al.,* 2015). The second R-mutation selected for re-insertion was G1541663A/D149N (8-R34) which affected Mlut_14110, encoding an annotated response regulator, co-transcribed along with Mlut_14100, encoding a putative histidine kinase which may belong to a two-competent system signal transduction cascade similar to *sivS/R* of *Streptococcus iniae*, according to a BLAST search (Bolotin, Fuller, Bast, Beveridge, & De Azavedo, 2007). The third R-mutation selected for re-insertion was G1606821/G27D (13-R4) affecting the already mentioned Mlut_14650. This mutation is 10 bp upstream of the SNP G1606810A/V31I (13-M27) from cluster 13 of the M-mutants.

All the selected mutations were re-introduced individually as single-SNP mutations into the chromosome of the naturally transformable wild-type *trpE16* strain (more details in section 3.5). Some of the affected genes were deleted to observe the resulting phenotype. In the case of Mlut_14660, it

was not possible to obtain a knockout strain, even after having tried different techniques like *codBA* clean deletion or *Kan* cassette exchange. Also, different lengths of fragments were used as homologous flanking regions to promote recombinative chromosomal integration, but in any of those cases was possible to get colonies with a deleted Mlut_14660. It is fair to infer that the protein encoded in this locus might be essential for *M. luteus* and may be involved in the regulation of more than one cellular process like natural competence or transformation. On the other hand, the ORF Mlut_02150 was knocked out without any trouble, meaning that even though they have similar annotations, they are not interchangeable or have the same cellular task.

After generating the desired editions in *M. luteus* genome, it was necessary to reinsert the 7-kbp *ΔcomEA/EC:lacZ-Kan* reporter construct. Because this succeeded well, it was clear that at least the re-introduced M-SNPs did not completely turn off transformability. This was later validated with the transformation frequency assays, in which only 13-R4 showed more than 10^5-fold decrease in transformation frequency in comparison to the wild-type.

From the 16 SNPs found in the M-mutant M33, only two fell within a cluster and a third one into a particularly interesting DNA-binding protein-encoding gene (Mlut_02150*)*. For both mutants 17-M33 and Δ02150, there were no significant changes in *comEA/EC* promoter activity and transformation frequencies. This means that probably some of the other 15 SNPs in the M33 chromosome are causative for the observed *comEA/EC* promoter overexpression. In the cases of the mutants derived from M43, 3-M43, and the knockout Δ03135, none showed overexpression of the *comEA/EC* promoter, making a regulation controlled by promoter adenine methylation as previously hypothesized unlikely. Similarly, neither of the single-SNP mutants of cluster 7 regained their respective parental mutants' promoter expression levels, meaning that at least those single mutations in Clp proteases are not enough to affect *comEA/EC* promoter expression rates. All of the previously mentioned single-SNP mutants and knockouts did not significantly change their transformation frequencies in comparison to the wild-type in the late exponential growth phase in GMM. There exists the possibility that the resulting phenotypes of at least some M- or R-mutants are the outcome of more than one mutation adding up. As can be observed in Figure 3.21, neither the single-SNP mutant 7-M89 nor 13-M89 displayed the phenotype of the parental EMS mutant, meaning that maybe the *comEA/EC* expression level of M89 needs both SNPs to be present. The mutation G1606081A of 13-M89 falls into the 5´UTR of Mlut_14660*,* and this results in a 100- to 1000-fold decrease in transformability. According to what has been predicted by several mRNA secondary structure prediction programs, this transition affects a highly conserved loop that may be necessary for the correct translation of the encoded DNA binding protein (see Figure 4.24).

Discussion

As expected for the single-SNP R-mutants, which showed a strongly repressed *com* promoter under competence-inducing conditions, transformation frequencies were much lower. For SNPs 2-R33 and 8-R34, which give rise to significant changes in the amino acid sequence of the affected Mlut_07490 and Mlut_14110 gene products (ATPase, FtsK/SpoIIIE family, and response regulator, respectively) , transformation frequencies were 1000-fold lower than the wild-type in late-exponential phase grown in GMM (see Figure 4.22). Both changes, G327E and D149N, respectively, are significant considering the biochemical features of the amino acids involved. Interestingly, when transforming these mutants with the *lacZ* reporter construct, blue colonies were obtained. This means that even though these SNPs affect transformability, they do not directly repress *comEA/EC* promoter expression, suggesting that they may play a role in the process at a different level or that one or more other mutations are required for *comEA/EC* promoter repression.

Mutant 13-R4 transformability was the lowest and below the detection limit of the assay. This R-mutant has the missense mutation G27D in the protein Mlut_14650 which is close to the V31I mutation present in the same protein of mutant 13-M27. Even though they are close together, the generated effects of these mutations are opposite. In the first, transformability was completely repressed along with the *comEA/EC* genes, while in the second, the overexpression of the *comEA/EC* promotor didn't seem to affect transformability. This phenotype repeats itself in mutants 13-M1090 and 13-M93, suggesting that there may be some sort of posttranscriptional regulation. Some sort of buffering of gene expression diminishing either mRNA concentration, protein abundance, ribosomal concentration, or translation initiation rate may be occurring (Blevins *et al*., 2019). A decrease in any of these biomolecules may result in the stabilization of ComEA and ComEC proteins abundance and lead to the observed phenotype. Contrary to the other single-SNP mutants of its cluster, mutant 13-M89 experienced a decrease of between 100- and 1000-fold in transformability. As already pointed out, *comEA/EC* expression levels in this mutant were not affected, although there seems to be an alteration at the translational level for Mlut_14660. The highly probable predicted loop forming within the standby site of the transcript (see figure 3.24) could be important for optimal translation initiation of the protein, which may play a key role in transformability regulation. On the other hand, while mutant 13-M1090 is also located in the 5´UTR of Mlut_14660, it doesn't seem to change any secondary mRNA structure according to the prediction programs used. Instead, it seems to modify the Shine-Dalgarno sequence, affecting ribosome binding and translation initiation rates (see Figure 4.24Figure 4.25). It remains to be understood why two mutations that affect translation generate such different phenotypes. Mutation C1607314T in mutant 13-M93 generates a R64W exchange directly affecting the structural part of the protein. It is worth mentioning that M93 is the only mutant to hold

this particular mutation and that after its insertion in the reporter strain, the parental mutant phenotype was fully regained (see Figure 3.21). A clean knockout of Mlut_14650 was done through the *codBA* deletion system, and a transformability loss of 10^4-fold was determined for this mutant (see Figure 4.22). At the same time, it was one of the mutants with the highest *comEA/EC* expression levels measured (see Figure 4.21). It is possible that the mentioned gene expression buffering occurs until a certain point, and that above that threshold this effect cannot control protein abundance anymore, somehow being detrimental for transformation frequencies.

Considering genes disposition and how close Mlut_14660 and Mlut_14650 are, it is difficult to assess whether the mutations falling so close by or even within the regulatory parts of the genes are in fact affecting one gene or the other, or both, and in which way. To get more information on this matter, the starting point of each transcript unit was determined (see Figure 4.23). Considering that there is a distance of 97 bp between the 5'UTR of each transcriptional unit, it seems that a divergent overlap of the transcriptional units exists and that the promoter and/or operator regions of each unit may share certain stretches of sequence, even with the structural part of the genes (Felder, 2007). This kind of arrangement opens up the door to the hypothesis of transcriptional interference (TI) as a way of gene regulation of Mlut_14660 and the operon Mlut_14650*/40/30*. TI was defined as the suppressive influence of one transcriptional process, directly and *in cis* on a second transcriptional process by Shearwin *et al*., 2005. A model like this one was also proposed for the divergently overlapping *comEA/EC* and *comER* genes in *B. subtilis* (Inamine & Dubnau, 1995). In more recent work (Yan *et al*., 2016), it has been shown that *comER* indeed plays an important role in biofilm formation and sporulation, and that competence has a negative interplay with biofilm formation (She *et al*., 2020), supporting the stated hypothesis. When observing if the gene Mlut_14660 and the operon Mlut_14650*/40/30* and their specific disposition is conserved throughout the phylum (see Figure 4.29 Physical maps of Mlut_14660, Mlut_14650/40/30 loci and their genomic surroundings from the genomes of a variety of actinobacterial representatives (high-GC Gram-positive) and our model organism *M. luteus*.) it came out that they are indeed present and in the same or similar disposition in other *Actinobacteria* genera. On the other hand, this was not the case for many other genera in which the genes are present although not in the same manner. It is worth mentioning that only a small group of representative organisms, usually the type species, are available in the context visualizer program. For example, all other species of *Micrococcus* are not available, so there is probably a broader presence of these genes organized in that way than what we could observe.

In the proposed hypothetical model (see Figure 5.1), the expression of one of the transcripts may generate a negative superhelical turn in the opposite sense, impeding the expression of the other

transcript. The steric hindrance derived from the RNA polymerase bound to the promoter can also play a role or a mix of both phenomena. Data published by our group in 2018 showed that the Mlut_14650*/40/30* operon is among the 25 most highly expressed genes in *M. luteus* genome in the mid-late exponential phase when grown in GMM. This was not the case for the same strain (trpE16) when cultured in LB (Surger *et al.*, 2018). The direct implication of Mlut_14650*/40/30* in pili formation still needs to be fully verified. The same is necessary for the binding of the Mlut_14660 product to the *comEA/EC* promoter. EMSA assays have shown not to be suitable for such an analysis in this case, given that is not possible to determine a reliable Kd value. A better and more specific strategy to address this issue would be to use surface plasmon resonance (Majka & Speck, 2006), e.g. with a Biacore apparatus.

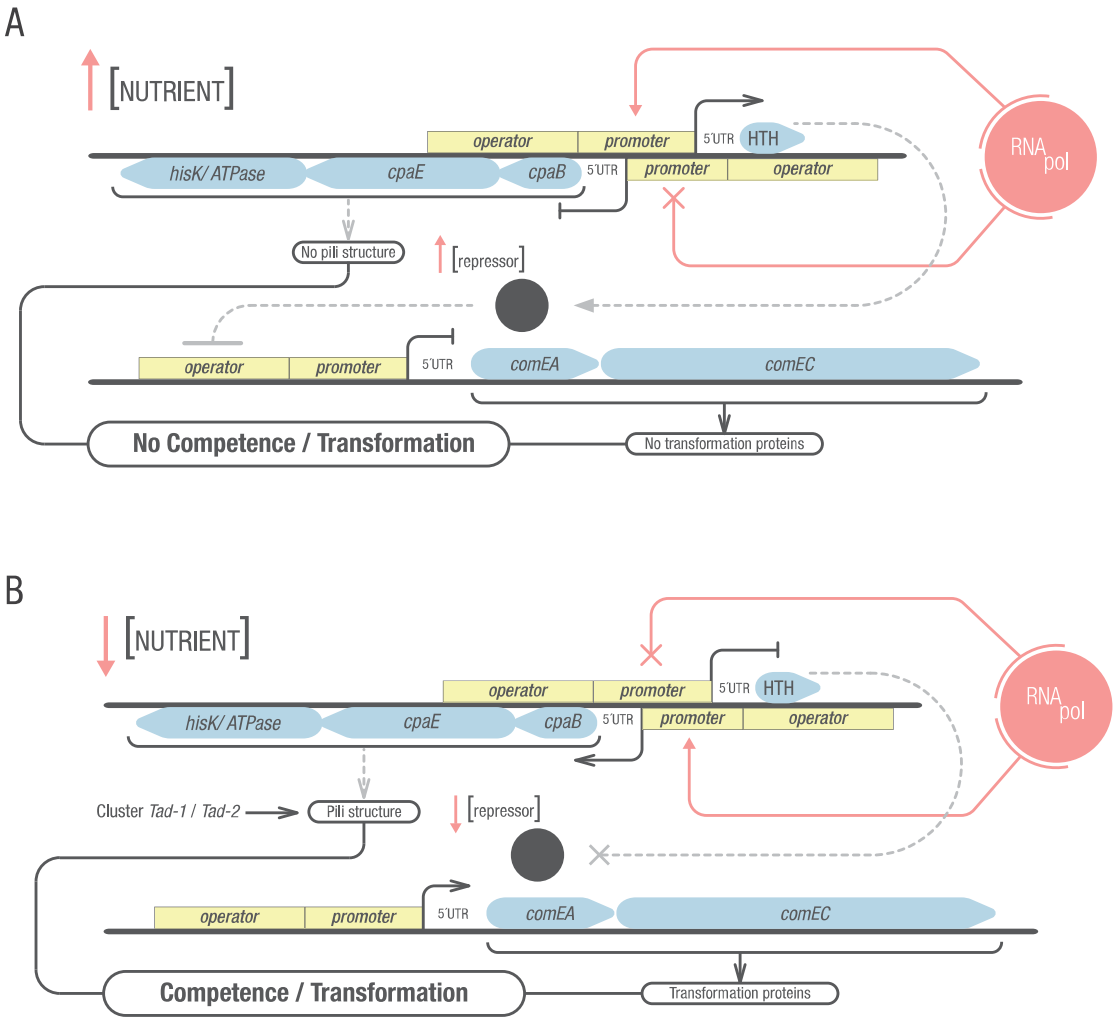


Figure 5.1 Transcriptional regulation model proposed for Mlut_14660 and Mlut_14650*/40/30*.
Transcriptional interference of the divergently overlapping transcriptional units shown above. Grey intermittent lines represent yet unproven but probable effects. (A) Under non-competence inducing conditions like high nutrient concentrations, Mlut_14660/HTH is expressed, impeding the expression of

Mlut_14650(*cpaB*)/Mlut_14640(*cpaE*)/Mlut_14630(*hisK*) operon which could play a role in pili structure building. Mlut_14660 product may exert the repression of essential *com* genes. (B) Opposite to the first scenario, under competence inducing conditions like low nutrient availability, Mlut_14650/40/30 operon is expressed, not letting Mlut_14660 be transcribed. This enhances *com* genes expression, resulting in a higher abundance of transformation proteins and pili structure both necessary for DNA uptake. This model is based on the results of this and previous work, but it still needs to be completely verified. This figure was created by this thesis author for the publication of this work (Torasso Kasem *et al*., 2021)

6 References

Adhikari, S., & Curtis, P. D. (2016). DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiology Reviews*, (July), 575–591. https://doi.org/10.1093/femsre/fuw023

Albano, M., Hahn, J., & Dubnau, D. (1987). Expression of competence genes in *Bacillus subtilis*. *Journal of Bacteriology*, *169*(7), 3110–3117. https://doi.org/10.1128/jb.169.7.3110-3117.1987

Alcázar, R., Altabella, T., Marco, F., Bortolotti, C., Reymond, M., Koncz, C., … Tiburcio, A. F. (2010). Polyamines: Molecules with regulatory functions in plant abiotic stress tolerance. *Planta*, *231*(6), 1237–1249. https://doi.org/10.1007/s00425-010-1130-0

Angelov, A., Bergen, P., Nadler, F., Hornburg, P., Lichev, A., Ãœbelacker, M., … Liebl, W. (2015). Novel Flp pilus biogenesis-dependent natural transformation. *Frontiers in Microbiology*, *6*(February), 1–11. https://doi.org/10.3389/fmicb.2015.00084

Angelov, A., Li, H., Geissler, A., Leis, B., & Liebl, W. (2013). Toxicity of indoxyl derivative accumulation in bacteria and its use as a new counterselection principle. *Systematic and Applied Microbiology*, *36*(8), 585–592. https://doi.org/10.1016/j.syapm.2013.06.001

Bagnacani, A., Backofen, R., Nekrutenko, A., Cech, M., Wubuli, A., Yusuf, D., … Backofen, R. (2018). Community-Driven Data Analysis Training for Report Community-Driven Data Analysis Training for Biology. *Cell Systems*, *6*, 752–758. https://doi.org/10.1016/j.cels.2018.05.012

Baltz, R. H. (2014). Spontaneous and induced mutations to rifampicin , streptomycin and spectinomycin resistances in actinomycetes : mutagenic mechanisms and applications for strain improvement. *The Journal of Antibiotics*, *67*(9), 619–624. https://doi.org/10.1038/ja.2014.105

Barre, F. (2007). FtsK and SpoIIIE : the tale of the conserved tails. *Molecular Microbiology*, *66*(October), 1051–1055. https://doi.org/10.1111/j.1365-2958.2007.05981.x

Berka, R. M., Hahn, J., Albano, M., Draskovic, I., Persuh, M., Cui, X., … Dubnau, D. (2002). Microarray analysis of the *Bacillus subtilis* K-state: Genome-wide expression changes dependent on ComK. *Molecular Microbiology*, *43*(5), 1331–1345. https://doi.org/10.1046/j.1365-2958.2002.02833.x

Bernard, C. S., Bordi, C., Termine, E., Filloux, A., Bentzmann, S. De, Al, B. E. T., & Acteriol, J. B. (2009). *Organization and PprB-Dependent Control of the Pseudomonas aeruginosa tad Locus , Involved in Flp Pilus Biology* ⍰ ǂ. *191*(6), 1961–1973. https://doi.org/10.1128/JB.01330-08

Bickle, T. A., & Kruger, D. H. (1993). Biology of DNA Restriction. *Microbiological Reviews*, *57*(2), 434–

450.

Bielaszewska, M., Mellmann, A., Zhang, W., Köck, R., Fruth, A., Bauwens, A., … Karch, H. (2011). Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany , 2011 : a microbiological study. *The Lancet Infectious Diseases*, *11*(9), 671–676. https://doi.org/10.1016/S1473-3099(11)70165-7

Blevins, W. R., Tavella, T., Moro, S. G., Blasco-Moreno, B., Closa-Mosquera, A., Díez, J., … Albà, M. M. (2019). Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-47424-w

Blokesch, M. (2016). Natural competence for transformation. *Current Biology*, *26*(21), R1126–R1130. https://doi.org/10.1016/j.cub.2016.08.058

Bolotin, S., Fuller, J. D., Bast, D. J., Beveridge, T. J., & De Azavedo, J. C. S. (2007). Capsule expression regulated by a two-component signal transduction system in Streptococcus iniae. *FEMS Immunology and Medical Microbiology*, *50*(3), 366–374. https://doi.org/10.1111/j.1574-695X.2007.00261.x

Brimacombe, C. A., Ding, H., & Beatty, J. T. (2014). Rhodobacter capsulatusDprA is essential for RecA-mediated gene transfer agent (RcGTA) recipient capability regulated by quorum-sensing and the CtrA response regulator. *Molecular Microbiology*, *92*(6), 1260–1278. https://doi.org/10.1111/mmi.12628

Brimacombe, C. A., Ding, H., Johnson, J. A., & Thomas Beatty, J. (2015). Homologues of genetic transformation DNA import genes are required for Rhodobacter capsulatus gene transfer agent recipient capability regulated by the response regulator CtrA. *Journal of Bacteriology*, *197*(16), 2653–2663. https://doi.org/10.1128/JB.00332-15

Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F., & Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. *FEMS Microbiology Reviews*, *39*(1), 81–95. https://doi.org/10.1111/1574-6976.12085

Cameron, A. D. S., Volar, M., Bannister, L. A., & Redfield, R. J. (2008). *RNA secondary structure regulates the translation of sxy and competence development in Haemophilus influenzae*. *36*(1), 10–20. https://doi.org/10.1093/nar/gkm915

Chen, I., Provvedi, R., & Dubnau, D. (2006). A macromolecular complex formed by a pilin-like protein

in competent *Bacillus subtilis*. *Journal of Biological Chemistry*, *281*(31), 21720–21727. https://doi.org/10.1074/jbc.M604071200

Chen, J., Darst, S. A., & Thirumalai, D. (2010). Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(28), 12523–12528. https://doi.org/10.1073/pnas.1003533107

Chilton, S. S., Falbel, T. G., Hromada, S., & Burton, B. M. (2017). A conserved metal binding motif in the *Bacillus subtilis* competence protein ComFA enhances transformation. *Journal of Bacteriology*, *199*(15), 1–10. https://doi.org/10.1128/JB.00272-17

Chung, Y. S., & Dubnau, D. (1998). All seven comG open reading frames are required for DNA binding during transformation of competent *Bacillus subtilis*. *Journal of Bacteriology*, *180*(1), 41–45. https://doi.org/10.1128/jb.180.1.41-45.1998

Claverys, J.-P., Prudhomme, M., & Martin, B. (2006). Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annual Review of Microbiology*, *60*, 451–475. https://doi.org/10.1146/annurev.micro.60.080805.142139

Claverys, J., & Martin, B. (2003). *Bacterial ' competence ' genes : signatures of active transformation , or only remnants ? 11*(4), 161–165. https://doi.org/10.1016/S0966-842X(03)00064-7

Claverys, J., Martin, B., & Polard, P. (2009). *The genetic transformation machinery : composition , localization , and mechanism*. https://doi.org/10.1111/j.1574-6976.2009.00164.x

Clock, S. A., Planet, P. J., Perez, B. A., & Figurski, D. H. (2008). Outer Membrane Components of the Tad (Tight Adherence) Secreton of *Aggregatibacter actinomycetemcomitans* □ †. *American Society for Microbiology*, *190*(3), 980–990. https://doi.org/10.1128/JB.01347-07

Croucher, N. J., Mostowy, R., Wymant, C., Turner, P., Bentley, D., & Fraser, C. (2016). *Horizontal DNA Transfer Mechanisms of Bacteria as Weapons of Intragenomic Conflict*. 1–42. https://doi.org/10.1371/journal.pbio.1002394

Crozat, E., Rousseau, P., Fournes, F., & Cornet, F. (2014). The FtsK Family of DNA Translocases Finds the Ends of Circles. *Journal of Molecular Microbiology and Biotechnology*, (24), 396–408. https://doi.org/10.1159/000369213

Damke, P. P., Marie, A., Guilmi, D., Varela, P. F., Velours, C., Marsin, S., … Radicella, J. P. (2019). Identification of the periplasmic DNA receptor for natural transformation of Helicobacter pylori. *Nature Communications*, *10*. https://doi.org/10.1038/s41467-019-13352-6

References

Davies, J. (1995). Vicious circles: looking back on resistance plasmids. *Genetics*, *139*(4), 1465–1468.

Diallo, A., Dujeancourt, A., Krasteva, P., Fronzes, R., Falbel, T., Perry, T., … Burton, B. (2017). Bacterial transformation: ComFA is a DNA-dependent ATPase that forms complexes with ComFC and DprA. *J*, *105*(5). https://doi.org/10.1111/mmi.13732

Dib, J. R., Wagenknecht, M., Hill, R. T., Farías, M. E., & Meinhardt, F. (2010). First report of linear megaplasmids in the genus Micrococcus. *Plasmid*, *63*(1), 40–45. https://doi.org/10.1016/j.plasmid.2009.10.001

Dubnau, D. (1991). The regulation of genetic competence in *Bacillus subtilis*. *Molecular Microbiology*, *5*(1), 11–18. https://doi.org/10.1111/j.1365-2958.1991.tb01820.x

Dubnau, David, & Blokesch, M. (2019). Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annual Review of Genetics*, *53*(1), 217–237. https://doi.org/10.1146/annurev-genet-112618-043641

Dubnau, David, & Roggiani, M. (1990). *Growth Medium-Independent Genetic Competence Mutants of Bacillus subtilis*. *172*(7), 4048–4055.

Felder, Y. (2007). Analysis of Biological Networks : Transcriptional Networks - Promoter Sequence Analysis. *Gene*, (Figure 1), 1–15.

Fleming, A., & B, P. R. S. L. (1922). On a remarkable bacteriolytic element found in tissues and secretions. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *93*(653), 306–317. https://doi.org/10.1098/rspb.1922.0023

Fortier, L., & Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, *4*(5), 354–365. https://doi.org/10.4161/viru.24498

Foulger, D., & Errington, J. (1989). *The role of the sporulation gene spolllE ' m the regulation of prespore-specific gene expression in Bacillus subtilis*. *3*(9), 1247–1255.

Gelbart, S. M., & Juhasz, S. E. (1970). Genetic Transfer in Mycobacterium phlei. *J. Gen. Mocrobiol.*, *64*, 253–254.

Greenblatt, C. L., Baum, J., Klein, B. Y., Nachshon, S., Koltunov, V., & Cano, R. J. (2004). *Micrococcus luteus* - Survival in amber. *Microbial Ecology*, *48*(1), 120–127. https://doi.org/10.1007/s00248-003-2016-5

Griffith, F. (1928). The Significance of Pneumococcal Types. *Journal of Hygiene*, *27*(2), 113–159.

https://doi.org/10.1017/S0022172400031879

Griffith, K. L., & Wolf, R. E. (2002). Measuring β-galactosidase activity in bacteria: Cell growth, permeabilization, and enzyme assays in 96-well arrays. *Biochemical and Biophysical Research Communications*, *290*(1), 397–402. https://doi.org/10.1006/bbrc.2001.6152

Hahn, J., Albano, M., & Dubnau, D. (1987). Isolation and characterization of Tn917lac-Generated competence mutants of *Bacillus subtilis*. *Journal of Bacteriology*, *169*(10), 4469–4478. https://doi.org/10.1128/jb.169.10.4469-4478.1987

Hahn, Jeanette, Tanner, A. W., Carabetta, V. J., Cristea, I. M., & Dubnau, D. (2015). ComGA-RelA interaction and persistence in the *Bacillus subtilis*K-state. *Molecular Microbiology*, *97*(3), 454–471. https://doi.org/10.1111/mmi.13040

Haijema, B. J., Hahn, J., Haynes, J., & Dubnau, D. (2001). A ComGA-dependent checkpoint limits growth during the escape from competence. *Molecular Microbiology*, *40*(1), 52–64. https://doi.org/10.1046/j.1365-2958.2001.02363.x

Hamoen, L. W., Venema, G., & Kuipers, O. P. (2003). Review Controlling competence in *Bacillus subtilis: shared use of regulators*. (October 2014). https://doi.org/10.1099/mic.0.26003-0

Håvarstein, L. S., Coomaraswamy, G., & Morrison, D. a. (1995). An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(24), 11140–11144. https://doi.org/10.1073/pnas.92.24.11140

Inamine, G. S., & Dubnau, D. (1995). ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *Journal of Bacteriology*, *177*(11), 3045–3051. https://doi.org/10.1128/jb.177.11.3045-3051.1995

Inaoka, T., & Ochi, K. (2002). RelA protein is involved in induction of genetic competence in certain *Bacillus subtilis* strains by moderating the level of intracellular GTP. *Journal of Bacteriology*, *184*(14), 3923–3930. https://doi.org/10.1128/JB.184.14.3923-3930.2002

Iyer, L. M., & Aravind, L. (2004). The emergence of catalytic and structural diversity within the beta-clip fold. *Proteins: Structure, Function and Genetics*, *55*(4), 977–991. https://doi.org/10.1002/prot.20076

J. Peter Gogarten & J. P. Townsend. (2005). Horizontal gene transfer accelerates genome innovation and evolution. *Molecular Biology and Evolution*, *20*(10), 1598–1602.

https://doi.org/10.1093/molbev/msg154

Jask, M., Stutzmann, S., Stoudmann, C., & Blokesch, M. (2018). QstR-dependent regulation of natural competence and type VI secretion in *Vibrio cholerae*. *46*(20), 10619–10634. https://doi.org/10.1093/nar/gky717

Johnsborg, O., Eldholm, V., & Håvarstein, L. S. (2007). Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology*, *158*(10), 767–778. https://doi.org/10.1016/j.resmic.2007.09.004

Johnsborg, O., & Håvarstein, L. S. (2009). Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. *FEMS Microbiology Reviews*, *33*(3), 627–642. https://doi.org/10.1111/j.1574-6976.2009.00167.x

Johnston, C., Martin, B., Fichant, G., Polard, P., & Claverys, J.-P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology*, *12*(3), 181–196. https://doi.org/10.1038/nrmicro3199

Kachlany, S. C., Planet, P. J., Bhattacharjee, M. K., Kollia, E., Salle, R. O. B. D. E., Fine, D. H., & Figurski, D. H. (2000). *Nonspecific Adherence by Actinobacillus actinomycetemcomitans Requires Genes Widespread in Bacteria and Archaea*. *182*(21), 6169–6176.

Keeling, P. J., & Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, *9*(8), 605–618. https://doi.org/10.1038/nrg2386

Kloos, W. E., & Musselwhite, M. S. (1975). Distribution and persistence of *Staphylococcus* and *Micrococcus* species and other aerobic bacteria on human skin. *Applied Microbiology*, *30*(3), 381–385.

Kloos, W. E., & Schultes, L. M. (1969). Transformation in *Micrococcus lysodeikticus*. *Journal of General Microbiology*, *55*(2), 307–317. https://doi.org/10.1099/00221287-55-2-307

Koonin, E. V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Research*, *21*(11), 2541–2547. https://doi.org/10.1093/nar/21.11.2541

Kostner, D., Peters, B., Mientus, M., Liebl, W., & Ehrenreich, A. (2013). Importance of *codB* for new *codA*-based markerless gene deletion in *Gluconobacter* strains. *Applied Microbiology and Biotechnology*, *97*(18), 8341–8349. https://doi.org/10.1007/s00253-013-5164-7

References

Lang, A. S., Zhaxybayeva, O., & Beatty, J. T. (2013). Gene transfer agents: Phage-like elements of genetic exchange. *Nature Reviews Microbiology*, *10*(7), 472–482. https://doi.org/10.1038/nrmicro2802

Lazazzera, B. A., Kurtser, I. G., Mcquade, R. S., & Grossman, A. D. (1999). An autoregulatory circuit affecting peptide signaling in *Bacillus subtilis*. *Journal of Bacteriology*, *181*(17), 5193–5200. https://doi.org/10.1128/jb.181.17.5193-5200.1999

Lichev, A., Angelov, A., Cucurull, I., & Liebl, W. (2019). Amino acids as nutritional factors and (p)ppGpp as an alarmone of the stringent response regulate natural transformation in *Micrococcus luteus*. *Scientific Reports*, *9*(1), 1–15. https://doi.org/10.1038/s41598-019-47423-x

Magnuson, R., Solomon, J., & Grossman, A. (1994). Biochemical and Genetic Characterization of a Competence Pheromone from *B. subtilis*. *Cell*, *77*, 207–216.

Majka, J., & Speck, C. (2006). Analysis of protein-DNA interactions using surface plasmon resonance. *Advances in Biochemical Engineering/Biotechnology*, *104*(August 2006), 13–36. https://doi.org/10.1007/10_026

Marrs, B. (1974). Genetic recombination in *Rhodopseudomonas capsulata*. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(3), 971–973. https://doi.org/10.1073/pnas.71.3.971

Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C., & Schoolnik, G. K. (2005). Chitin induces Natural competence in *Vibrio cholerae*. *Science*, *310*, 1824–1827.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics*, *7*(3), 277–318.

Norgard, M. V., & Imaeda, T. (1978). Physiological factors involved in the transformation of *Mycobacterium smegmatis*. *Journal of Bacteriology*, *133*(3), 1254–1262. https://doi.org/10.1128/jb.133.3.1254-1262.1978

Ochman, Howard, Lawrence, Jeffrey G., & Groisman, Eduardo A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, *405*(6784), 299–304. Retrieved from http://adsabs.harvard.edu/abs/2000Natur.405..299O

Ohama, T., Muto, A., & Osawa, S. (1989). Spectinomycin operon of *Micrococcus luteus*: Evolutionary implications of organization and novel codon usage. *Journal of Molecular Evolution*, *29*(5), 381–395. https://doi.org/10.1007/BF02602908

References

Omer Bendori, S., Pollak, S., Hizi, D., & Eldar, A. (2015). The RapP-PhrP quorum-sensing system of *Bacillus subtilis* strain NCIB3610 affects biofilm formation through multiple targets, due to an atypical signal-insensitive allele of RapP. *Journal of Bacteriology*, *197*(3), 592–602. https://doi.org/10.1128/JB.02382-14

Oshima, T., Wada, C., Kawagoe, Y., Ara, T., Maeda, M., Masuda, Y., … Mori, H. (2002). Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in *Escherichia coli*. *Molecular Microbiology*, *45*(3), 673–695.

Pelton, J. G., Kustu, S., & Wemmer, D. E. (1999). Solution Structure of the DNA-binding Domain of NtrC with Three Alanine Substitutions. *Journal of Molecular Biology*, *292*, 1095–1110.

Perez-cheeks, B. A., Planet, P. J., Sarkar, I. N., Clock, S. A., Xu, Q., & Figurski, D. H. (2012). The product of *tadZ*, a new member of the *parA*/*minD* superfamily , localizes to a pole in *Aggregatibacter actinomycetemcomitans*. *Molecular Microbiology*, *83*(January), 694–711. https://doi.org/10.1111/j.1365-2958.2011.07955.x

Perkins, J. B., & Youngman, P. J. (1986). Construction and properties of Tn917-*lac*, a transposon derivative that mediates transcriptional gene fusions in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(1), 140–144. https://doi.org/10.1073/pnas.83.1.140

Peterson, S. N., Sung, C. K., Cline, R., Desai, B. V, Snesrud, E. C., Luo, P., … Morrison, D. A. (2004). Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *51*, 1051–1070. https://doi.org/10.1046/j.1365-2958.2003.03907.x

Provvedi, R., & Dubnau, D. (1999). ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Molecular Microbiology*, *31*(1), 271–280. https://doi.org/10.1046/j.1365-2958.1999.01170.x

Roelants, P., Konvalinkova, V., Mergeay, M., Lurquin, P. F., & Ledoux, L. (1976). DNA uptake by *Streptomyces* species. *BBA Section Nucleic Acids And Protein Synthesis*, *442*(1), 117–122. https://doi.org/10.1016/0005-2787(76)90182-9

Salvadori, G., Junges, R., Morrison, D. A., & Petersen, F. C. (2019). *Competence in Streptococcus pneumoniae* and Close Commensal Relatives*: Mechanisms and Implications*. *9*(April), 1–8. https://doi.org/10.3389/fcimb.2019.00094

Santos, J. D., Vitorino, I., De La Cruz, M., Díaz, C., Cautain, B., Annang, F., … Lage, O. M. (2019).

Bioactivities and extract dereplication of actinomycetales isolated from marine sponges. *Frontiers in Microbiology*, *10*(APR), 1–11. https://doi.org/10.3389/fmicb.2019.00727

Scrudato, M. Lo, Borgeaud, S., & Blokesch, M. (2014). Regulatory elements involved in the expression of competence genes in naturally transformable *Vibrio cholerae*. 1–13. https://doi.org/10.1186/s12866-014-0327-y

Sega, G. A. (1984). A review of the genetic effects of ethyl methanesulfonate. *Mutation Research/Reviews in Genetic Toxicology*, *134*(2–3), 113–142. https://doi.org/10.1016/0165-1110(84)90007-1

Seitz, P., & Blokesch, M. (2013). Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiology Reviews*, *37*(3), 336–363. https://doi.org/10.1111/j.1574-6976.2012.00353.x

Shanker, E., & Federle, M. J. (2017). Quorum Sensing Regulation of Competence and Bacteriocins in *Streptococcus pneumoniae* and *mutans*. https://doi.org/10.3390/genes8010015

She, Q., Hunter, E., Qin, Y., Nicolau, S., Zalis, E. A., Wang, H., … Chai, Y. (2020). Negative interplay between biofilm formation and competence in the environmental strains of *Bacillus subtilis*. *BioRxiv*, *5*(5), 17–20. https://doi.org/10.1101/2020.06.15.153833

Sia, B. Y. R. H. P., & Dawson, M. H. (1931). In vitro Transformation of pneumococcal types. 701–710.

Sinha, S., Mell, J., & Redfield, R. (2013). The availability of purine nucleotides regulates natural competence by controlling translation of the competence activator Sxy. *88*(May), 1106–1119. https://doi.org/10.1111/mmi.12245

Skerman, V. B. D., McGowan, V., & Sneath, P. H. A. (1980). Approved lists of bacterial names. *International Journal of Systematic Bacteriology*, *30*(1), 225–420. https://doi.org/10.1099/00207713-30-1-225

Sneppen, K., Dodd, I. B., Shearwin, K. E., Palmer, A. C., Schubert, R. A., Callen, B. P., & Egan, J. B. (2005). A mathematical model for transcriptional interference by RNA polymerase traffic in *Escherichia coli*. *Journal of Molecular Biology*, *346*(2), 399–409. https://doi.org/10.1016/j.jmb.2004.11.075

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics*, *16*(8), 472–482. https://doi.org/10.1038/nrg3962

Surger, M. J., Angelov, A., Stier, P., übelacker, M., & Liebl, W. (2018). Impact of branched-chain amino

acid catabolism on fatty acid and alkene biosynthesis in *Micrococcus luteus*. *Frontiers in Microbiology*, *9*(MAR), 1–14. https://doi.org/10.3389/fmicb.2018.00374

Tomich, M., Planet, P. J., & Figurski, D. H. (2007). The tad locus: postcards from the widespread colonization island. *Nature Reviews. Microbiology*, *5*(5), 363–375. https://doi.org/10.1038/nrmicro1636

Torasso Kasem, E. J., Angelov, A., Werner, E., Lichev, A., Vanderhaeghen, S., & Liebl, W. (2021). Identification of New Chromosomal Loci Involved in com Genes Expression and Natural Transformation in the Actinobacterial Model Organism *Micrococcus luteus*. *Genes*, *12*(9). https://doi.org/10.3390/genes12091307

Treangen, T. J., & Rocha, E. P. C. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genetics*, *7*(1). https://doi.org/10.1371/journal.pgen.1001284

Turgay, K., Hahn, J., Burghoorn, J., & Dubnau, D. (1998). Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *17*(22), 6730–6738.

Underhill, S. A. M., Burne, R. A., & Hagen, S. J. (2018). Intracellular Signaling by the *comRS* System in *Streptococcus mutans* Genetic Competence. *3*(5), 1–21.

Underhill, S. A. M., Shields, R. C., & Hagen, S. J. (2019). Carbohydrate and PepO control bimodality in competence development by *Streptococcus mutans.* *112*(August), 1388–1402. https://doi.org/10.1111/mmi.14367

Viollier, P. H., Sternheim, N., & Shapiro, L. (2002a). A dynamically localized histidine kinase controls the asymmetric distribution of polar pili proteins. *21*(17).

Viollier, P. H., Sternheim, N., & Shapiro, L. (2002b). Identification of a localization factor for the polar positioning of bacterial structural and regulatory proteins. (4).

Von Wintersdorff, C. J. H., Penders, J., Van Niekerk, J. M., Mills, N. D., Majumder, S., Van Alphen, L. B., … Wolffs, P. F. G. (2016). Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Frontiers in Microbiology*, *7*(FEB), 1–10. https://doi.org/10.3389/fmicb.2016.00173

Wozniak, R. A. F., & Waldor, M. K. (2010). Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Reviews Microbiology*, *8*(8), 552–563. https://doi.org/10.1038/nrmicro2382

References

Xu, Q., Christen, B., Chiu, H., Jaroszewski, L., Heath, E., Knuth, M. W., … Wilson, I. A. (2012). Structure of the pilus assembly protein TadZ from *Eubacterium rectale*: Implications for polar localization. *Molecular Microbiology*, *83*(4), 712–727. https://doi.org/10.1111/j.1365-2958.2011.07954.x.Structure

Yamamoto, S., Izumiya, H., Mitobe, J., Morita, M., Arakawa, E., Ohnishi, M., & Watanabe, H. (2011). Identification of a chitin-induced small RNA that regulates translation of the *tfoX* gene, encoding a positive regulator of natural competence in *Vibrio cholerae*. *Journal of Bacteriology*, *193*(8), 1953–1965. https://doi.org/10.1128/JB.01340-10

Yamamoto, S., Mitobe, J., Ishikawa, T., Wai, S. N., Ohnishi, M., Watanabe, H., & Izumiya, H. (2014). Regulation of natural competence by the orphan two-component system sensor kinase ChiS involves a non-canonical transmembrane regulator in *Vibrio cholerae*. *91*(December 2013), 326–347. https://doi.org/10.1111/mmi.12462

Yan, F., Yu, Y., Wang, L., Luo, Y., Guo, J. H., & Chai, Y. (2016). The *comER* gene plays an important role in biofilm formation and sporulation in both *Bacillus subtilis* and *Bacillus cereus*. *Frontiers in Microbiology*, *7*(JUN), 1–16. https://doi.org/10.3389/fmicb.2016.01025

Young, M., Artsatbanov, V., Beller, H. R., Chandra, G., Chater, K. F., Dover, L. G., … Greenblatt, C. L. (2010). Genome sequence of the fleming strain of *Micrococcus luteus*, a simple free-living actinobacterium. *Journal of Bacteriology*, *192*(3), 841–860. https://doi.org/10.1128/JB.01254-09

Yue, J., Hu, X., Sun, H., Yang, Y., & Huang, J. (2012). Widespread impact of horizontal gene transfer on plant colonization of land. *Nature Communications*, *3*, 1152–1159. https://doi.org/10.1038/ncomms2148

ANNEX I – R Scripts

R Script to visualize SNP distribution from the EMS mutagenesis experiment.

```
# script to vizualise SNP data from the EMS experiment

library(tidyverse)
library(plotly)
library(data.table)
source("snp-clustering-score-function.R")

#setwd("C:/Users/Student/Dropbox/TUM Team Folder/EMS-R")
#After procesing all genomes and filtering good quality/high coverage mutations, a table with mutants, mutated positions in the genome,
# and type of mutations is crated

#First, a Data Frame with all nucleotides -ordered by mutant- that differ from the reference is created.
df <- read_delim("results.tab", delim = "\t")
df.long <- df %>% gather("strain", "mutation", M1090:`374-trpE16-out-trp`) %>%
                  mutate(mutation1 = ifelse(mutation == Reference, " ", mutation)) # crazy bug, if mutation1 is NA or "", plotly zoom doesn't

df.cons <- df.long %>% filter(mutation1 != " ") %>% group_by(POS) %>% summarise(n = n())

#A plot is created.The names of the mutants on the Y axe and all the positions of the genome on the X axe, so each mutated position shows up
# by presenting the SNP. In the end we obtain a distribution of all SNPs along the genome, with the relative distance between, therefore
# it can be seen how they are clustered.
p <- df.long %>%
    ggplot() +
    geom_text(aes(POS, strain, label = mutation1, color = SNP), fontface = "bold") +
    geom_vline(aes(xintercept = POS), alpha = 0.2, size = 0.2, data = df) +
    theme_bw() +
    theme(panel.grid = element_blank()) +
    scale_color_brewer(type = "qual", palette = "Set1")


## read and process gff file
gff <- fread("ref.gff") %>% tidyr::separate(V9, into = paste0("gff", 1:9), sep = ";", remove = FALSE)

# Below the graph, additional lines with information are added, first the position of rRNA genes as reference, and second the position of
# each gene, with its name and annotation according to NCBI.
p2 <-gff %>%
      dplyr::filter(V3 == "CDS" | V3 == "rRNA") %>%
      plot_ly() %>%
          add_segments(x = ~ V4, xend = ~ V5,
               y = ~ V3, yend = ~ V3,
               opacity = 0.7,
               color = ~ V7, size = I(14),
               colors = c("red", "forestgreen"),
               hoverinfo = "text",
               text =  ~paste("<b>", gff1, "</b>", "<br />", gff6))

## A function called "slideFunct" was created, it consists of a "window" which length can be set and it runs through every position
# along the genome, giving a score as a result. This score gets plotted and tells us if there are many or few SNPs clustered within the
# coverage of the window for a certain position.
# The slidefunct takes uses the formula: (1/(distance between the two closest SNPs within the window))x100
# For example here the window encompasses 5000 bp, and makes steps of 500 bp.
scores <- slideFunct(1:2501304, 5000, 500, df$POS)

p3 <- scores %>% plot_ly() %>% add_lines(x = ~ spots, y = ~ corrresult, color = I("red"))

#p4 <- scores %>% plot_ly() %>% add_lines(x = ~ spots, y = ~ res, color = I("blue"))

# In a fourth horizontal axe, the amount of mutants that contain a SNP per position is plotted in bars. Therefore, with both the
# cluster score and the number of SNPS in each position, it can be seen which are the most mutated areas which
p4 <- df.cons %>% plot_ly() %>% add_trace(x = ~ POS, y = ~ n, type = "bar", color = I("black"))

subplot(p, p2, p3, p4, shareX = TRUE, nrows = 4, heights = c(0.6, 0.1, 0.15, 0.15), titleX = FALSE) %>%
  layout(#width = "1200",
         #height = "600",
         showlegend = FALSE,
         yaxis = list(showgrid = FALSE,
                      showline = FALSE))
### Cheers!
```

R Script for the “slide-window” function

```
slideFunct <- function(data, window, step, snps){
   total <- length(data)
  spots <- seq(from = 1, to = (total - window), by = step)
  result <- vector(length = length(spots))
  corrresult <- vector(length = length(spots))

  for(i in 1:length(spots)){
    spotarray <- data[spots[i]:(spots[i]+window)] %in% snps

    # which(arr) gives the indices of TRUE values, from where we can get the distances
    # naive approach - sum of element-wise difference times number of TRUE?
    # 1/diff(which(arr))*length(arr) ????? will this work?

    result[i] <- sum(spotarray)
    if (length(which(spotarray)) <= 1) {
      corrresult[i] <- result[i]
    }else {
      corrresult[i] <- length(spotarray)/min(diff(which(spotarray)))
    }
  }
  return(data.table(spots = spots, res = result, corrresult = corrresult))
}
```

ANNEX II - Table with the updated locus tags

Table 6.1. Old and new locus tags for the main genes in this work.

Old Locus Tag	New Locus Tag
Mlut_14660	MLUT_RS18825
Mlut_14650	MLUT_RS18820
Mlut_14640	MLUT_RS18815
Mlut_14630	MLUT_RS18810
Mlut_02150	MLUT_RS12640
Mlut_03135	MLUT_RS13120
Mlut_09880	MLUT_RS16460
Mlut_09900	MLUT_RS16470
Mlut_07490	MLUT_RS15280
Mlut_14110	MLUT_RS18560

Affidavit – Eidesstattliche Erklärung

Ich,_____, (Vor- und Nachname) erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung _____

der TUM zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:

_____

unter der Anleitung und Betreuung durch: _____

ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 7 Abs. 6 und 7 angegebenen Hilfsmittel benutzt habe.

☐ Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuer*innen für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

☐ Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

☐ Teile der Dissertation wurden in _____ veröffentlicht.

☐ Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

☐ Ich habe bereits am _____ bei der promotionsführenden Einrichtung _____ der Hochschule _____ unter Vorlage einer Dissertation mit dem Thema _____ die Zulassung zur Promotion beantragt mit dem Ergebnis:

_____

☐ Ich habe keine Kenntnis über ein strafrechtliches Ermittlungsverfahren in Bezug auf wissenschaftsbezogene Straftaten gegen mich oder eine rechtskräftige strafrechtliche Verurteilung mit Wissenschaftsbezug.

Die öffentlich zugängliche Promotionsordnung sowie die Richtlinien zur Sicherung guter wissenschaftlicher Praxis und für den Umgang mit wissenschaftlichem Fehlverhalten der TUM sind mir bekannt, insbesondere habe ich die Bedeutung von § 27 PromO (Nichtigkeit der Promotion) und § 28 PromO (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Mit der Aufnahme meiner personenbezogenen Daten in die Alumni-Datei bei der TUM bin ich

☐ einverstanden, ☐ nicht einverstanden.

Ort, Datum, Unterschrift