Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Andreas Lykke

Construction and characterisation of artificial promoters in *Bacillus subtilis*.

Master's thesis in Biotechnology
Supervisor: Rahmi Lale, Lisa Tietze
August 2019

NTNU
Norwegian University of
Science and Technology

Andreas Lykke

Construction and characterisation of artificial promoters in *Bacillus subtilis*.

Master's thesis in Biotechnology
Supervisor: Rahmi Lale, Lisa Tietze
August 2019

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Abstract

Bacillus subtilis is a gram-positive model organism used in both academic research and industrial applications. Due to its efficient protein expression and secretion capacity it is widely considered to be one of the most important bacterium for the industrial production of bio-products.

Despite its importance in academic research and industry, there are few well-characterised promoters available.
An efficient and easy to use way of creating artificial promoters for *B. subtilis* could therefore have implications for the biotechnology industry, as well as leading to the discovery and availability of many new promoters, benefiting research science.

In this master's thesis, a novel method, Gene Expression Engineering (GeneEE), has been used for generating artificial promoters in *B. subtilis*. The GeneEE method relies on the insertion of stretches of DNA with random DNA composition into the promoter regions of genes, after removal of the native promoter.
In this thesis three different random DNA inserts were used: 200N, consisting of 200 nucleotides (nt) random DNA base pairs; 50N, consisting of 50 nt random DNA base pairs; and 200N-SD, consisting of 211 nt random DNA base pairs harbouring a Shine-Dalgarno sequence, GGAG, which is thought to be necessary for translation in *B. subtilis*.

This thesis contains the first published description of the GeneEE-method in *B. subtilis*. In the thesis, the construction, identification and *in vivo* and *in silico* characterisation of 21 promoters in *B. subtilis* is reported.
The sequences were, using several online tools, analyzed and known promoter motifs were identified, including σ^{70} -10 and -35 elements. The samples that displayed the higher fluorescence generally also had a closer consensus match to known elements, with the exception of the 50N samples, which did not match the expected consensus, but also had more variation in the measured relative fluorescence data.

The method was proven effective in creating functional promoters for *B. subtilis*, with the promoters not matching previously published promoter sequences.

Sammendrag

Bacillus subtilis en en gram-positiv modellorganisme. Den er mye brukt i både forskning og industri.
B. subtilis regnes som en av de viktigste bakteriene for industrien, på grunn av effektivt uttrykk og sekresjon av proteiner.
På tross av den viktige rollen *B. subtilis* har i industri og forskning er det få nøye beskrevne promotorer tilgjengelig.
En effektiv og lett gjennomførbar metode for å lage kunstige *B. subtilis* promotorer kan derfor ha implikasjoner for bioteknologisk industri, og det kan lede til at nye og potensielt viktige promoterer kan oppdages og brukes i forskning.

I denne masteroppgaven blir en ny metode kalt Gene Expression Engineering (GeneEE) metoden brukt til å lage kunstige *B. subtilis* promotorer.
GeneEE metoden basererer seg på bruken av en DNA-sekvens med tilfeldig DNA-innhold av en viss lengde. Denne sekvensen settes inn i promotorregionen til et gen som har fått sin naturlige promotor fjernet.

I denne masteroppgaven ble tre forskjellige typer tilfeldig DNA brukt: 200N, som består av 200 tilfeldige nukleotid basepar, 50N, som består av 50 tilfeldige nukleotid basepar, og 200NSD, som består av 211 tilfeldige nukleotid basepar med en Shine-Dalgarno sekvens, GGAG. Det forventes at *B. subtilis* trenger Shine-Dalgarno sekvensen for translasjon.

I denne masteroppgaven beskrives for første gang bruk av GeneEE-metoden i *B. subtilis*.
Oppgaven beskriver konstruksjonen, identifikasjonen og *in vivo* og *in silico* beskrivelser av 21 *B. subtilis* promotorer.
Promotorsekvensene ble, ved hjelp av flere online verktøy, analysert, og kjente promototermotiver ble identifisert, inkludert -10 og -35 elementer fra σ^{70}.
Det ble vist en viss sammenheng mellom prøvene som utrykte sterkt, og match til konsensussekvenser fra kjente elementer. Et unntak var 50N-prøvene, som ikke matchet konsensus, men allikevel viste tegn til genuttrykk, dog med en del variasjon i dataene.

Det ble bevist at GeneEE-metoden er kapabel til å produsere nye, tidligere ukjente, funksjonable promotorer i *B. subtilis*.

Preface

This thesis concludes my degree Master of Science in Biotechnology at the Norwegian University of Science and Technology (NTNU) in Trondheim.
This thesis was written at the Department of Biotechnology, and the lab work described within was performed from summer 2018 until summer 2019 at the laboratories of the department of biotechnology.
During this time I was a part of the research group PhotoSynLab, where I learned a lot.

I would like to thank my supervisors, Rahmi Lale and Lisa Tietze, for not losing their patience when things got stressful, and for always being available to explain and demonstrate when times were confusing. Also thanks to Martin F. Hohmann-Marriott for giving me the opportunity to join the group together with Rahmi.
I would also like to thank all members of PhotoSynLab for friendship and inspiration, and especially Che Fai Alex Wong, who was always very helpful, no matter the subject or time of day.

Thanks to our collaborators at the University of Newcastle for helping a *Bacillus subtilis* beginner find his way.

I would also like to thank Vetle Simensen, Marie Gulla, Even Moen Kirkholt, Madeleine Gundersen and Idd Andrea Christensen for backup, companionship and general advice on everything. Volvox & Alkymisten should also be thanked for their part in making my years at NTNU very enjoyable.

Finally, I thank my girlfriend, Beate Rose Gjendemsjø for having my back, helping me out when I needed it, and accepting the state of my office, no matter how messy.

Contents

List of Tables

List of Figures

Abbreviations

BB	=	Backbone
BE_{avg}	=	Average base expression
bp	=	Base pair
ccc	=	Covalently closed circular
DBTBS	=	Database of transcriptional regulation in *Bacillus subtilis*
E	=	Expression
EDTA	=	Ethylenediaminetetraacetic acid
F	=	Fluorescence
GeneEE	=	Gene Expression Engineering
GRAS	=	Generally recognized as safe (for human consumption)
iGem	=	International Genetically Engineered Machine
IPTG	=	Isopropyl β-D-1-thiogalactopyranoside
l	=	linearized (plasmid)
LA	=	LB-Agar
LB	=	Lysogeny Broth
nt	=	nucleotides
oc	=	Open Circular
OD_{600}	=	Optical density at 600 nm wavelength
PCR	=	Polymerase Chain Reaction
RE	=	Relative Expression
RNAP	=	RNA polymerase
SD	=	Shine-Dalgarno
SMM	=	Spizizen Minimal Medium
SOB	=	Super Optimal Broth
TAE	=	Tris-acetate-EDTA
TnSS	=	Translation start site
TrSS	=	Transcription start site
UP element	=	Upstream Promoter element
UTR	=	Untranslated region

Chapter 1

Background

1.1 *Bacillus subtilis*

Bacillus subtilis is a gram-positive, aerobic, spore-forming bacterium, of the genus *Bacillus*.

B. subtilis is widely studied for its ability to form endospores that can survive conditions that most other organisms can not, including heat, radiation, and chemical stressing (1), as well as its protein secretion ability (2) and natural competence (3).
These factors, combined with being non-pathogenic (2) is why *B. subtilis* is widely regarded as one of the most important bacteria in science and manufacturing and why it is the most widely studied of all the gram-positive bacteria (4). While not as easily transformable as *E. coli*, its low doubling time of about 120 min in minimal media (5) makes it a good bacteria for synthetic biology experiments.
In laboratory conditions, *B. subtilis* can grow in minimal medium, with only essential salts, as well as sources of nitrogen, phosphorus and carbon. Many types of saccharides, as well as amino acids, peptides, and some other short-chain carbon compounds can serve as carbon sources(6).

Initially discovered in 1835 and named *Vibrio subtilis*, *B. subtilis* has rod-shaped cells and are found in many aquatic and terrestrial environments. Its natural environment is not fully understood, due to the ability of the spores to travel long distances and shifting back to its active state upon analysis in the lab. Therefore, it is hard to determine its actual environment, even though they have been known to associate with plant roots, possibly having some probiotic effect on the plant (7).
It has also been proven to be able to survive inside animal gastrointestinal tracts and has recently been shown to eradicate *Staphylococcus aureus* in the human gut (8).
B. subtilis has a generally recognized as safe (GRAS) status in the U.S, which means it is accepted as safe for human consumption, and is currently being considered for sale as a probiotic. It has long been used in traditional fermented soy products in Japan and is also

used in the production of other fermented foodstuffs (3, 7).

The most commonly used *B. subtilis* laboratory strain is strain 168 (6). *B. subtilis* 168 deviates from wild type *B. subtilis* in that it is highly transformable and auxotrophic, needing tryptophane for growth (9).
While most of the published synthetic biology work on *B. subtilis* is done in strain 168, this is not the preferred strain for industrial manufacturing (6).
Industrial applications also utilize other *Bacillus* species in addition to *B. subtilis*, like *B. amyloliquefaciens* and *B. licheniformis*.

B. subtilis, especially strain 168, is well characterized in scientific literature, with both its complete genome (10), proteome (11) and transcriptome, under different environmental conditions (12) published.
It has a single circular chromosome, with the sequenced genome containing 4.2 million base pairs (bp) coding for 4100 proteins (6).
In starvation conditions, a sub-population of *B. subtilis* can become competent, with the other sub-population sporulating. The competence allows for the uptake of external DNA that shares homology with the *B. subtilis* chromosome. This form of uptake of DNA is known as integration(6).

1.1.1 Synthetic biology and industrial applications of *B. subtilis*

B. subtilis has many characteristics that make it an ideal bacterium for synthetic biology work, even though there are some challenges.
While *B. subtilis* is naturally competent, the uptake of plasmid DNA is inefficient, and its comparatively low transformation rate with plasmid DNA often requires *E. coli* to be used as a shuttle organism (2).
Because of its excellent secretion capabilities and its suitability to large-scale fermentations, *B. subtilis* is a suitable and sometimes preferred organism for the production of many products. The products include recombinant proteins like α-amylase and Proinsulin (13) as well as platform chemicals like ethanol, lactic acid and 2,3-butanediol. *B. subtilis* is also an effective producer of bio-products like riboflavins and N-acetylglucosamine. Some of these products, like 2,3-butanediol are naturally produced in *B. subtilis*, while the others have been adapted to *B. subtilis* production by recombinant DNA technologies (4).

While secretion of homologous proteins is highly efficient, the secretion rate for heterologous proteins in *B. subtilis* is low. Considerably less than 1g/L is secreted of heterologous proteins, compared to 20-25 g/L of homologous proteins (14).
The low level of heterologous protein secretion is in part due to a high level of extracellular proteases that degrade heterologous proteins (14). However, there has been significant work in increasing the heterologous protein secretion. One approach is engineering of the secretion pathways and using various signal peptides, which allow for better secretion (4). Another approach is the construction of *B. subtilis* strains with deletions in the genes coding for extracellular proteases and other genomic manipulations that reduce unwanted protease activity (13).

Several vector systems for *B. subtilis* have been developed, including plasmid vectors, integrative vectors, and bacteriophage vectors (2).
There are also inducible and constitutive promoter systems available (4), though in much lower numbers than for *E. coli*. Many modern synthetic biology tools, including the CRISPR-Cas9 system (15) has proven its usefulness as tools in the genome manipulation of *B. subtilis*.

B. subtilis is naturally competent, but the rate of acquisition is low for plasmid DNA. To increase competence, some manipulation of the cells before transformation is required, for example, starving the cells.
When *B. subtilis* 168 experience starvation conditions at the end of the exponential phase, the cell population differentiates into a sporulating and a competent sub-population (16), with the competent sub-population being utilized for transformation.
Due to the lower transformation efficiency of *B. subtilis*, shuttle plasmids in *E. coli* are often used. Here, the recombinant plasmids are being amplified in *E. coli*, before being transformed into *B. subtilis* (2).
Other methods used for *B. subtilis* transformations include electroporation, protoplasts, and mobilization from *E. coli* (2).

1.2 Bacterial Promoters

A bacterial promoter is a stretch of DNA upstream of the transcription start point, usually 100-1000 bp long, that is recognized by, and bind to bacterial RNA polymerase (RNAP) (17).
There is, however, considerable variation in the sequences and placement of promoters, and their identification can prove challenging (18).
To initiate mRNA synthesis, the σ subunit of the bacterial RNAP recognizes the promoter sequence, and the RNAP binds to the promoter sequence, leading the RNAP to initiate transcription. Different σ proteins recognize different sequences in the promoter, called sigma-factors, described in more detailed in section 1.2.1.

The first nucleotide transcribed, at the transcription start site (TrSS), is generally referred to as point +1 of the transcript. DNA before the transcription start point (at the $5'$ end) is generally identified as "upstream", and its location noted as a negative number referring to its relative distance to the transcription start site (17).
A general depiction of bacterial promoter structure, based on the *E. coli* σ^70 recognition elements, can be seen in Figure 1.2.1.
Studies on the promoters of *E. coli* revealed consensus elements, preserved sequences, at -35 bp upstream of the transcription start point with the consensus sequence (TTGACA) and another at -10 (TATAAT) with a 17 bp spacer between the two elements (19). The -10 element is also known as the Pribnow box (17).
A third recognition element, called the upstream promoter (UP) element is an AT-rich section of the promoter, usually found at -60 (17). In *B. subtilis*, a -16 "TG" motif is also conserved (20).

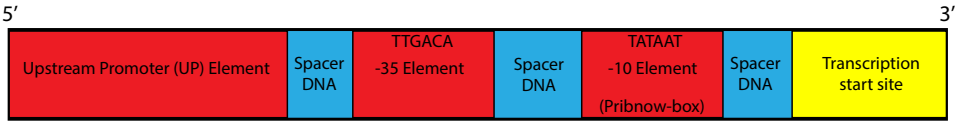


Figure 1.2.1: The basic structure of a bacterial promoter (not to scale). The bacterial promoter typically has a consensus sequence at -10, and -35, upstream of the transcription start point, in addition to an AT-rich Upstream Promoter (UP) element further upstream.

Mutations in the recognition elements have been shown to affect the RNA polymerase binding strength and thereby the rate of transcription. The length of the spacer between the -10 and -35 elements is also implicated in promoter strength (19).

For most genes, under normal condition, the sigma factor σ^{70}, or its *B. subtilis* homologue σ^{A} recognizes the -10 and -35 elements and enables transcription. However, there are other sigma-factors and consensus sequences that enables transcription under specific circumstances, such as under heat-stress or starvation (21), described closer in the next section. A single promoter can be activated by multiple sigma-factors, as the commonly used *B. subtilis* promoter P_{43} , which is activated by both σ^{A} and σ^{B} (22).

Promoters can be either regulatable or constitutive, with regulatable promoters being either inducible or repressible.
An inducible promoter requires an inducer to initiate transcription. The inducer is usually a molecule small enough to enter through the plasma membrane, like Isopropyl β-D-1-thiogalactopyranoside (IPTG), or a stress factor, such as changes in temperature or pH (2). In contrast, repressible promoters respond to signals by turning off or lowering the rate of transcription (17).
Constitutive promoters are in an "always-on" state, continually expressing the gene under most conditions. The genes controlled by these constitutive promoters are sometimes called housekeeping genes (17).

Tandem promoters, promoters where multiple promoters are present in the same upstream region, with the possibility for the binding of several sigma-factors, have also been found in the regulation of wild-type *E. coli* genes (23). An example of an artificial tandem promoter is the promoter P_{43}

1.2.1 The sigma factors

There are a variety of known sigma factors, and the most studied are the *E. coli* sigma factor σ^{70} or σ^{54}, and their homologues.
The σ^{70}-related group is responsible for most expression during growth, and the σ^{54} group is responsible for transcription in response to environmental factors (21).
Most basic descriptions of bacterial promoters are based on σ^{70}.

The σ^{70}-related sigma factors

The σ^{70}-related sigma factors are typically grouped into four groups, with group 1 sigma factors most related to σ^{70}.
The full σ^{70} consensus sequence is seen below in Figure 1.2.2. In *B. subtilis*, the σ^{A} has the same -35 and -10 consensus sequence as *E. coli* σ^{70}, and the two share extensive homology overall (24).

```
      -35          -16  -10
5'- TTGACA -14N -TGNTATAATGGG-3N-TSS  -3'
```

Figure 1.2.2: The consensus sequence of the bacterial sigma factor σ^{70}, adapted from (21)

Group 2 sigma factors are also related to σ^{70}, and have similar consensus sequences, except for a C nucleotide at -13 between the -16 TG motif and the Pribnow box. Group 2 factors are non-essential and are active in stationary phase survival and stress responses. σ^{38} is an example of a group 2 sigma factor (21).

Group 3 sigma factors are structurally related to group 1 and 2, but usually have distinct

-10 and -35 elements, with some having extended -10 elements (21).
An example is σ^{28} and its *B. subtilis* homologue σ^{D}, which is related to the expression of flagella genes (21).

Another important group 3 factor is the *E. coli* heat-shock factor σ^{32}, and the general stress response factor in gram-positives, the σ^{B} from *B. subtilis* (21).

The largest and most diverse group is group 4, the ExtraCytoplasmic Function, sigma factors, which often respond to extracellular signals. The genes encoding for the group 4-factors are often found clustered alongside the genes they regulate (21).

σ^{54} element

The σ^{54} promoters do not have homology to σ^{70}, instead recognizing a -24 and -12 element within the consensus sequence displayed in Figure 1.2.3.
The σ^{54} can not spontaneously initiate transcription, requiring additional activators (25).
The *B. subtilis* homologue of σ^{70} is σ^{L} (26).

```
       -24           -12
5'- TGGCACG 4N TTGC T/A NNA 7N TSS  -3'
```

Figure 1.2.3: The consensus sequence of the bacterial sigma factor σ^{54}, adapted from (27).

The σ^{54} element is activated in response to cellular signals and is involved in the synthesis of pili and flagella, nitrogen synthesis, alginate production and more (27).

1.2.2 5′ Untranslated region

Downstream of the transcription start site and the promoter region, but upstream of the translation start site (TnSS), is a region known as the $5'$ untranslated region (UTR).
The $5'$ UTR is an often short stretch of DNA that is transcribed, but not translated into polypeptides (18). The elements within the $5'$ UTR plays a role in gene expression in many genes. In *B. subtilis*, elements in the $5'$ UTRs of mRNA are known to affect translation efficiency and the rates of mRNA degradation (28). $5'$ UTR elements can lead to improved stability of mRNAs, allowing for higher translation rates, and thereby higher expression of proteins. Typically, a bacterial mRNA has a half-life of around 2 minutes, but some mRNAs have been shown to be active for up to 30 mins, due to stabilizing $5'$ elements, which affect nucleolytic activity (2).

Some prokaryotes, including *B. subtilis*, has been shown to require a short sequence called the Shine-Dalgarno (SD) sequence, found within the $5'$ UTR (29). several SD-sequence variations are known, with most variations having a central GGAG-motif, and often found with a 7 bp spacer between the end of the SD-sequence and the start-codon. The SD-sequence serves as a binding site for the ribosomes and the SD-sequence is complimentary

to the "anti-SD" sequence found in the $3'$ end of the 16s rRNA (30).
In *B. subtilis*, if there is no SD sequence, translation is not expected to occur at high rates (29). The amount of translation is regulated by the SD-sequence, the initiation codon, and the spacer between the initiation codon and the SD-sequence. The SD-sequence affects transcription rates both by influencing the ribosome assembly rate and by stabilizing the transcript (2).

1.2.3 Artificial promoters

As with all industrially useful bacteria, there is a need for highly efficient and regulatable promoters for *B. subtilis* (4). While there are some similarities between promoter sequence of different organisms, there are no guarantees that a highly effective promoter system for *E. coli* would work as effectively in *B. subtilis*, with evidence to the contrary (31).
There are some available *B. subtilis* promoters in widespread use: among the most used are the constitutive promoters P_{43} and P_{veg}, and the inducible Pspa and PxylA promoters (20).
There are also several known *B. subtilis* promoter sequences available online, in various databases:
DBTBS, the database of *B. subtilis* promoters and transcription factors (32) is a compilation database, summarizing experimentally discovered promoters.
The iGem registry of standard biological parts also has a list of both constitutive and inducible *B. subtilis* promoters (33).

There have been several attempts at generating random promoter libraries for *B. subtilis*. Most of these inserted randomized sequences in between and upstream of sigma^{70} conserved elements, or used some method of mutating already know promoters to achieve higher gene expression.
Liu et al. (20) analyzed transcriptomic data from *B. subtilis* in search of promoter consensus sequences. These consensus sequences were used to produce artificial promoters, conserving the -35,-16,-10 and UP-elements.
The experiment resulted in 214 promoters of varying strength, generated using Polymerase chain reaction (PCR) with a primer containing random nucleotides, which were then analyzed using a GFP assay. Even though many of the promoters showed stronger expression than the native P_{veg}-promoter, it did not outperform the P_{43} promoter (20).

Another approach to increase gene expression, this time in *E. coli*, was employed by Li et al. (34), who used tandem promoters. By increasing the number of promoters by copying the existing promoter region, they saw an increase in gene expression.
They did, however, find that while increasing the number of promoters, increasing the number above 5, did not further increase expression, leading to believe that an increase in promoters alone is not the only factor needed to increase gene expression (34).

There have also been developed broad-spectrum artificial promoters, that work in multiple species, including *E. coli*, *B. subtilis* and *Saccharomyces cerevisiae*. The promoter was based on the -35 and -10 elements of the σ^{70}, with an added SD-sequence downstream of the TrSS. Through fine-tuning of spacer-sequences, they found the promoter

sequence that gave the highest expression in all three organisms, which resulted in the P_{bs} promoter, alongside a shuttle-vector containing the promoter sequence and ampicillin resistance genes (31).

1.3 GeneEE method for artificial promoter production

The information presented in this section is based on the at this time unpublished paper by Lale et al. (35), preprint edition available called "A universal method for gene expression engineering". The method is patent pending.
The GeneEE method is a way of producing artificial promoter libraries using random insert DNA. An overview of the method is shown in Figure 1.3.1.
The method is based on replacing the promoter and $5'$ upstream of the gene TnSS with a random DNA insert, containing a stretch of N-nucleotides, and then screen for successful mutants.
The N-nucleotide is a degenerative nucleotide, representing a random nucleotide, with 25% chance of it being A, T, C or G (20). Due to the random nature of the inserts, there are N^4 possible combinations of nucleotides, leading to a vast number of possible promoters within. For 200N, for example, there are $200^4 = 1.6 * 10^9$ possible sequences.
The GeneEE-insert is comprised of a BioBrick prefix + BsaI site, followed by 200 'N' nucleotides, with a BsaI site and BioBrick suffix. The sequence and layout of the GeneEE inserts with the pre- and suffixes can be seen in Appendix B.2.

The BioBrick pre- and suffixes are standard parts used by the iGem standard biological parts library (36). For the experiments in this thesis, a 50N, 200N, and 200NSD stretch of random base pairs were used as the insert. The number before 'N' denotes the number of random nucleotides. The 200NSD insert has 200 N nucleotides, followed by a Shine-Dalgarno sequence, and then 7 N nucleotides.
The 200NSD insert is especially relevant when using *B. subtilis* as the final host organism, due to its requiring the Shine-Dalgarno sequence for translation (29).

Due to the random nature of the inserts, the inserts will sometimes lead to gene expression, with a functional promoter region in the insert, as well as a $5'$ UTR that facilitates translation. The further upstream the promoter region and transcription start site, the longer the $5'$ UTR will be.
By design, in the 200NSD inserts, there will always be a Shine-Dalgarno sequence 7 bp upstream from the TnSS.
There is also the probability that several promoter regions will be present in the insert. This could perhaps give stronger gene expression, working as a tandem promoter.
Lale et al. showed that the method could be used to produce both constitutive and inducible promoter systems (35). However, in this thesis, it will only be screened for constitutive promoters.

The method is performed by simultaneous restriction digest and ligation using the Golden Gate Assembly method (37).
After removal of the wild-type promoter using PCR, a linearized backbone is formed,

which contains the target gene without its native promoter and $5'$ UTR, with a BsaI site directly in front of the start codon, ATG and another on the other end of the PCR product, see Figure 1.3.1(B). These ends will also be complementary to the ends of the GeneEE insert when digested with BsaI, and not to each other, to avoid the backbone self-ligating during the ligation phase of the Golden Gate Assembly.

When designing the primers to create the backbone, any necessary resistance genes and origin of replication should be maintained from the plasmid, while still reducing the size enough to distinguish between the original plasmid and the backbone + insert using agarose gel electrophoresis. This allows for successful identification of hybrid plasmid mutants without needing to send for sequencing.
In the ligation step of the Golden Gate Assembly, the hybrid plasmid with the GeneEE insert directly in front of the TnSS is ligated and results in a closed plasmid which can be cloned into an appropriate host organism. The organisms can then be screened for expression using a suitable method.
If the insert is not in a gene coding for antibiotic resistance, most of the clones will not have functional promoters, due to it not being necessary for their survival. However, by the law of large numbers, some of the inserts can produce functional promoters.

The gene with the random insert upstream should be measurable quantitatively so that it can be compared to the wild-type promoter or another working promoter. When creating the backbone, a closed backbone plasmid should also be created and used as an additional control, to ensure that no gene expression is expected without the inserts.

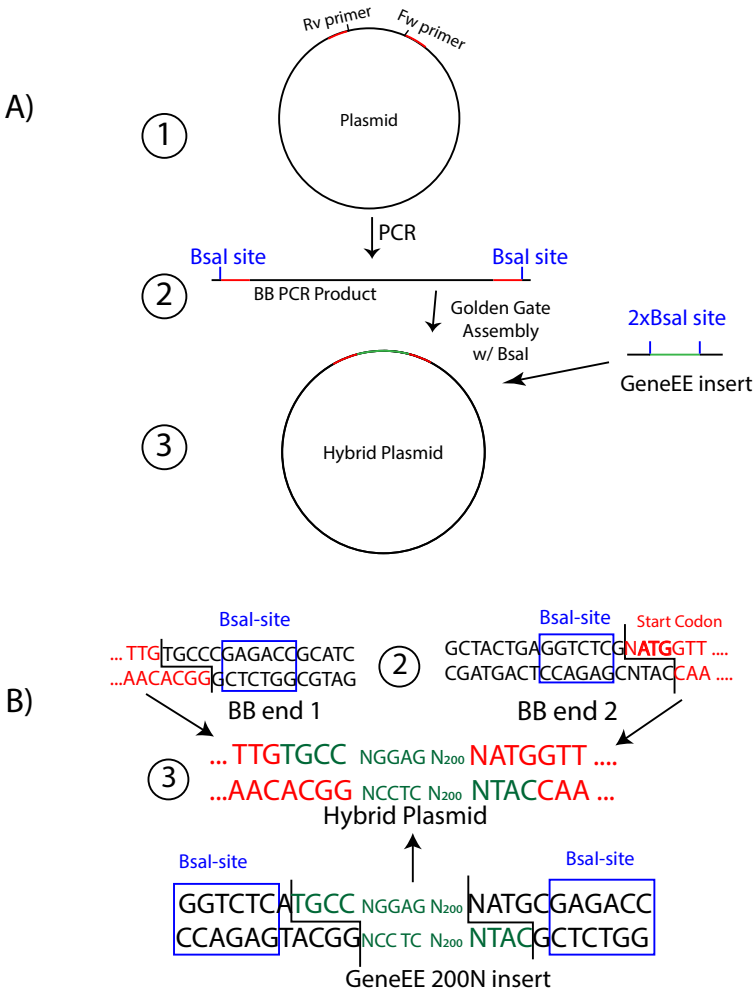


Figure 1.3.1: A: Overview of the GeneEE method using a 200N fragment.
1: To remove the native promoter, $5'$ UTR and other upstream elements from the gene of interest, PCR is used. The PCR uses a forward primer that binds directly in front of the translation start site and a reverse primer further upstream. The primers create a linearized PCR product with BsaI sites at either end. **2:** Using Golden Gate Assembly, The backbone and GeneEE insert is cut using BsaI, creating 'sticky ends' that overlap between the insert and backbone. By design, the backbone sticky ends do not fit to each other, preventing accidental backbone ligation. **3:** The parts are ligated together during the Golden Gate Assembly using T4 ligase, resulting in a closed hybrid plasmid, with the GeneEE insert (here 200N) directly upstream of the translation start site of the gene of interest. Correct ligation removes the BsaI-sites, protecting the hybrid plasmid from further cutting.
B: A more detailed view of the segments that are cut and ligated during the Golden Gate Assembly.

Chapter 2

Materials and Methods

2.1 Experimental design

The purpose of this master thesis and the experiments described within is to demonstrate that the GeneEE-method can produce novel, functional promoters for *Bacillus subtilis*.

To accomplish this the plasmid pHT10-mCherry, containing the mCherry fluorescent protein, had the native inducible promoter and $5'$ UTR removed from the mCherry gene and replaced using Golden Gate Assembly, by a GeneEE insert. These inserts were random DNA insert of either 200 bp, 50 bp, or 200 bp followed by an SD-sequence and another 7 bp random DNA insert. The insert sequences can be seen in Appendix B.2.
This new random promoter library was then amplified in *E. coli*, then isolated and used to transform *B. subtilis* 168 cells.
A randomly selected sample of the transformed *B. subtilis* cells, 84 colonies per plate, was then grown in 96-well plates, and fluorescence and absorbance were analyzed using a Tecan infinite m200 plate reader.
If the process was successful, one or more *B. subtilis* cells should express the mCherry protein, and therefore fluoresce at higher levels than the uninduced native promoter.
After fluorescence measurements, the best candidates were picked and sent for sequencing to analyze the promoter sequences.
Finally, after sequencing, the sequences were analyzed using various software tools, to search for binding sites and known motifs for various sigma-factors.

In this section, the materials and general methods used for these experiments will be explained.
Explained first is the transformation of *E. coli* and *B. subtilis*, then how the random promoter libraries are created, how relative gene expression was measured, and finally how the final *B. subtilis* clones were analysed.

2.2 Materials

2.2.1 Media and antibiotics

Antibiotic stock solutions

Ampicillin 100 mg/mL in dH_2O, final concentration 100 µg/mL
Chloramphenicol 100 mg/mL in 70% ethanol, final concentrations 10 µg/mL in liquid medium or 15 µg/mL in agar plates.

Transformation media

***E. coli* transformation media:**

yB-Medium
2.5g Yeast extract (Oxoid)
10g Tryptone (Oxoid)
0.38g KCl (Merck)
dH_2O to 483 mL
Ingredients mixed, pH adjusted to 7.6 using KOH and autoclaved.
17 mL sterile 1M $MgSO_4$ added and stored at room temperature

SOC solution
20g Tryptone (Oxoid)
5g Yeast extract (Oxoid)
0.584g NaCl (VWR)
0.186g KCl (VWR)
dH_2O to 1 L, autoclaved and stored at 4 °C in the dark.

TfBI Solution
1. 47g potassium acetate
4. 95g $MnCl_2$ (VWR)
6. 05g RbCl (Alfa Aesar)
0. 74g $CaCl_2$ (Sigma-Aldrich)
75 mL glycerol (VWR)
dH_2O to 500 mL
pH adjusted to 5.8 using 0.2 M acetic acid
Filter sterilized, stored at 4 °C

TfBII solution
10 mL 100 mM MOPS (Fisher) (pH 7.0, adjusted using 1M KOH)
1.10 g $CaCl_2$ (Sigma-Aldrich)
0.12 g RbCl (Alfa Aesar)
15 mL glycerol (VWR)
dH_2O to 100 mL
Autoclaved and stored at 4 °C, wrapped in foil.

***B. Subtilis* transformation media:**

SMM medium
2.0 g Ammonium sulphate (Merck)
14.0 g dipotassium hydrogen phosphate (Sigma-Aldrich)
6.0 g potassium dihydrogen phosphate (Merck)
1.0 g sodium citrate dihydrate (Sigma-Aldrich)
0.2 g magnesium sulfate (VWR)
dH_2O to 1L
Autoclaved

MM competence medium, prepared on the day of transformation
10 mL SMM medium
125 µL 40 % glucose (VWR)
100 µL 2mg/mL tryptophan (Sigma-Aldrich)
60 µL 1M $MgSO_4$ (VWR)
10 µL 20% casamino acids (DIFCO)
5 µL 0.22 & $Fe-NH_4$ citrate (Fluka)

Starvation medium, prepared on the day of transformation and warmed to 37 °C
10 mL SMM medium
125 µL 40% glucose (VWR)
60 µL 1M $MgSO_4$ (VWR)

General growth media:

Lysogeny broth (LB) Medium
5 g NaCl (VWR)
5 g Tryptone (Oxoid)
2.5 g Yeast extract (Oxoid)
dH_2O to 500 mL
Autoclaved

LB agar (LA)
5 g NaCl (VWR)
5 g Tryptone (Oxoid)
2.5 g Yeast extract (Oxoid)
7.5 g Agar (Oxoid)
dH_2O to 500 mL
Autoclaved and antibiotics added (if used).

Gel electrophoresis solutions

The TAE buffer and Agarose solution were stock solutions in the lab, not made by the candidate. **50xTris-acetate-EDTA (TAE) buffer**
pr. 1 L:

242 g Tris
57.1 mL acetic acid
100 mL 0.5M Ethylenediaminetetraacetic acid (EDTA)
dH_2O to 1 L

Agarose solution
3.2 g Agarose (Cambrex)
400 mL 1x TAE solution
microwaved, 5 min
20 µL GelRed

Loading dye
20 µL NEB Gel Loading Dye, Purple (6X)
100 µL dH_2O

1 kb DNA ladder
10 µL NEB Gel Loading Dye, Purple (6X)
3 µL NEB 1 kb DNA Ladder
47 µL dH_2O

2.2.2 Bacteria and plasmids

Presented in tables 2.2.1 and 2.2.2 is a list of all bacteria and plasmids used for this experiment.
All bacterial cells were grown at 37 °C incubation. Liquid cultures were grown in sterile 13 mL plastic tubes with 225 RPM shaking.
When *E. coli* DH5α cells were cultured for plasmid production, a single colony was incubated in $\approx$ 3 mL LB-medium with 100 µg/mL ampicillin in a 13 mL tube overnight.
When the cells were grown on LA-plates, they were grown without shaking on plates containing ampicillin, 100 µg/mL, or if untransformed cells were grown, without antibiotics.

When transformed *B. subtilis* 168 cells were grown on plates, it was on LA-plates containing 15 µg/mL chloramphenicol, or without antibiotics for untransformed cells.
In liquid medium transformed *B. subtilis* 168 cells were grown in 3 mL LB-medium in 13 mL tubes with 10 µg /mL chloramphenicol and untransformed *B. subtilis* 168 cells were grown without LB.
When incubated in 96-well plates, *B. subtilis* was grown in 200 µL LB with 10 µg/mL chloramphenicol for 20 hours at 37°C, 700 RPM shaking.

Table 2.2.1: Bacterial strains used in this thesis

Strain	Description
E. coli $DH_5\alpha$	Bacterial cloning host for shuttle vector
B. subtilis 168	Final cloning host

Table 2.2.2: Plasmids used in this thesis

Plasmid	Description
pHT10-mCherry	Original plasmid, expressing mCherry constitutively in *E. coli* and under IPTG (1:1000) induction for *B. subtilis*.
pHT10-mCherry-ΔBsaI	pHT10-mCherry with BsaI site in Ampicillin resistance gene removed
pHT10-mCherry-ΔBsaI-517-BB	Backbone plasmid used as a negative control
pHT10-mCherry-ΔBsaI-50N	Hybrid plasmid with random 50N GeneEE insert
pHT10-mCherry-ΔBsaI-200N	Hybrid plasmid with random 200N GeneEE insert
pHT10-mCherry-ΔBsaI-200NSD	Hybrid plasmid with random 200NSD GeneEE insert

2.2.3 pHT10-mCherry plasmid

The pHT10-mCherry plasmid, based on the pHT10-plasmid (38) is a *E. coli*/*B.subtilis* shuttle vector, which carries an ampicillin resistance gene for *E. coli*, as well as a *B. subtilis* chloramphenicol resistance gene. Using a shuttle vector means that a fairly low amount of PCR-generated plasmid DNA can be greatly amplified in *E. coli* before isolating the plasmids and using it to transform *B. subtilis*.
For this master thesis, $\approx$ 100 ng plasmid DNA was used for transformation of *E. coli* and 3 µg for transformation of *B. subtilis*.
The mCherry protein is a red fluorescent protein created by Schaner et al. as part of the mFruit family of fluorescent proteins (39). mCherry is based on the protein DsRed, from the sea anemone *Discosoma sp*. The promoter of the mCherry gene used in these experiments is the Pgrac promoter, inducible by IPTG in *B. subtilis* and expressed constitutively in *E. coli*. It is used in this experiment as the reporter gene, with higher fluorescence readings indicating a stronger promoter.
The pHT10-mCherry plasmid map is displayed in Figure 2.2.1.

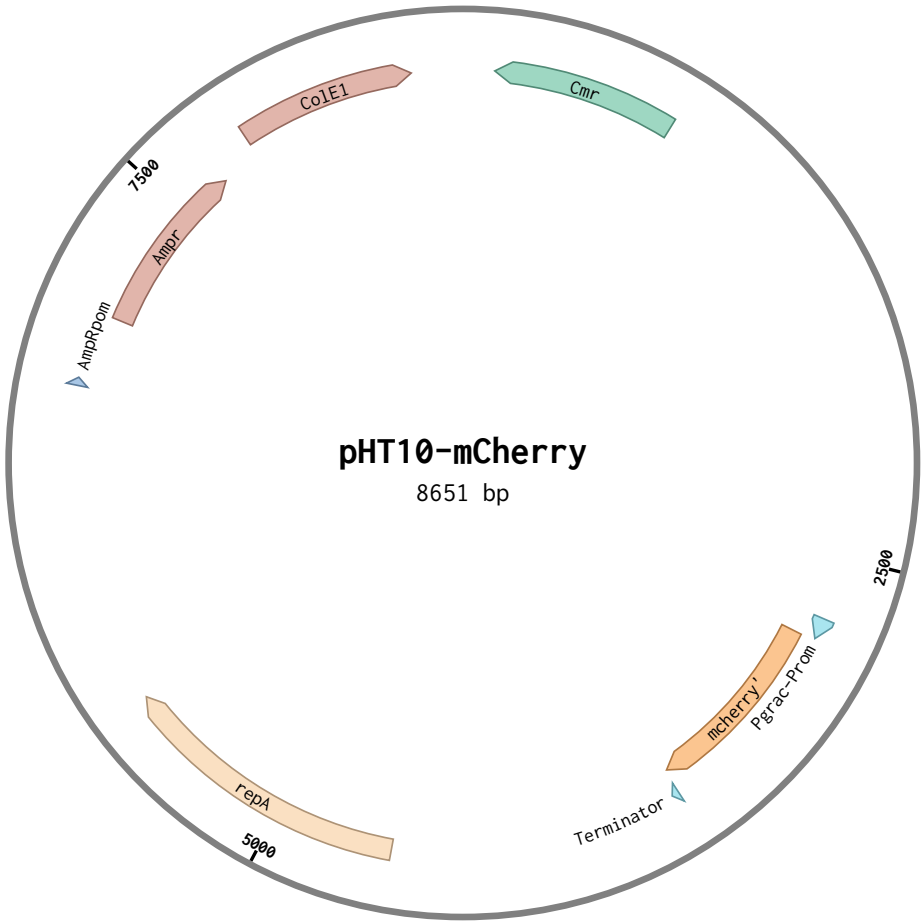


Figure 2.2.1: pHT10-mCherry plasmid map. Ampr gene gives ampicillin resistance, Cmr gives chloramphenicol resistance, mCherry' is the gene encoding for the mCherry red fluorescent protein, under control of the Pgrac IPTG-inducible promoter

2.2.4 Software tools

Benchling

Benchling is, in the words of the website, a ”Cloud-Based Informatics Platform for Life Sciences R&D” (40).
It is made for biotechnological research and allows for several synthetic biology tasks to be either simulated, as PCRs, restriction digests and ligations, or performed, like making annotated plasmid maps, designing primers, and aligning DNA-sequences.
It also works as a storage platform for documents, plasmid maps, primers and other relevant data. Benchling was used for sequence alignment and *in silico* cloning to make the same backbone as was expected from the golden gate cloning, so the sequenced cells could be aligned.

Graphic tools

To capture gel electrophoresis images, the software Image Lab by Bio-Rad was used, and the images exported to Adobe Photoshop and Adobe Illustrator for clarification (setting brightness/contrast levels, inverting colors) and annotation. All images were captured in the .TIFF-format and exported in the .PDF-format.
All figures used in this thesis was created by the author, when noted adapted from figures in the literature, in Adobe Photoshop, Adobe Illustrator or Adobe InDesign.

Data analysis tools

To calculate relative expression values, Microsoft Excel 2019 was used, based on data captured by the Tecan infinite m200 Pro plate reader.
To calculate the possibility of sigma factor binding, the online tool BacPP (41) was used. The online tool BPROM by Softberry (42) was used to score the promoter sequences in relation to known σ^{70} consensus sequences. The improbizer motif matcher tool (43) was used to find -35 and -10 elements in the sequencing data for the 50N insert samples since the BPROM tool did not recognize these elements in the 50N samples.
The improbizer (43) tool was used to search for reoccurring motifs in the sequencing data, analyzing the 50N and 200N/200NSD data separately, because the improbizer algorithm is affected by the size of the sequence. The improbizer produces a consensus sequence, alongside a probability matrix. The probability matrix lists the probability that each nucleotide will appear in each place in the found motif. For some, there might be very close probabilities between two nucleotides, so the consensus sequence should be read in the context of the probability matrix.

2.3 General methods

This section and its subsections describe the general methods used in the experiments in this thesis, and the next section describes how the data generated were analyzed.

2.3.1 Production of *E. coli* competent cells

The method is based on a method by Green et al. (44).
For media and solution recipes, see Section 2.2.1. For preparation of *E. coli* DH5α cells for transformation, a stock of competent cells were prepared and used as the initial cloning hosts for the experiments in this thesis:

In a 50 mL Falcon tube with 20 mL of SOC Media, a single colony of *E. coli* $DH_5\alpha$ cells was incubated overnight.
After incubation, 20 mL of the overnight culture was transferred, at 2x10 mL into 2x150 mL pre-warmed (37°C) yB-media in shake flasks.
The cells were grown for 2 hrs at 37°C, 225 RPM, until OD_{600} reached 0.35.
The cells were then chilled for 5 minutes on ice, and transferred into six chilled 50 mL Falcon tubes and centrifuged for 10 mins at 4000 RPM, 4°C in a table-top centrifuge. The supernatant was discarded, and the pellets resuspended in 15 mL chilled TfBI medium.
The resuspended cells were then pelleted by centrifugation for 10 mins at 4000 rpm, 4°C and each pellet was resuspended in 1 mL chilled TfBII.
After resuspension, the cells were snap-frozen by adding 100 mL cell suspension to 60 pre-frozen (-80°C) microcentrifuge tubes and stored at -80°C.

The next day, the transformation efficiency was calculated by heat-shock transforming one aliquot of cells using 2 µL plasmid DNA, containing ampicillin resistance genes.

2.3.2 *E. coli* heat-shock transformation protocol

The following method was used for all transformations of *E. coli* in this thesis:
Frozen competent *E. coli* DH5α cells were thawed, on ice, for 5 minutes.
DNA was added to cells; for plasmid DNA, 100 ng was used, for Golden Gate Assembly, 10 µL ligation mix was used.
The cells and DNA were pipette-mixed gently and incubated on ice for 30 minutes. The cells were then heat-shocked at 42 °C for 45 seconds and put on ice for 3 minutes.
900 µL LB (room temperature) was added to the tubes, and the cells were placed horizontally in the incubator with shaking for 90 minutes.
After shaking, the cells were plated out on LA-plates. For libraries, the whole tube of transformed *E. coli* was used, for production of plasmids, 200 µL was used.
Finally, the plates were incubated overnight at 37°C.

2.3.3 *B. subtilis* 168 transformation protocol

The following method, based on methods provided by a collaborator at Newcastle University, was used for all transformations of *B. subtilis* in this thesis:

In a sterile 100 mL flask, a colony of *B. subtilis* was added to 5 mL fresh MM competence medium and incubated overnight at 37 °C with 225 RPM shaking.
The next day, the OD_{600} was checked, at 1:10 dilution. The overnight culture was transferred into freshly made 5 mL competence medium in a new sterile 100 mL flask, enough volume for the OD_{600} of the new medium to be ≈ 0.1; usually 3-400 µl.
The fresh media with overnight culture was incubated for a further 3-4 hrs until OD_{600} reached ≈ 0.4-0.5.
When the desired OD_{600} was reached, 5 mL pre-warmed starvation medium was added to the culture and incubated for another 2 hours. The cells were then competent and ready to be transformed.

To transform, 400 µL competent cells were added to 1.5 mL microcentrifuge tubes for each transformation, in addition to the appropriate controls.
2-3 µg DNA was added to the tubes and gently mixed.
The microcentrifuge tubes were then incubated at 37 °C with 225 RPM shaking for 60 mins while being kept horizontally.
After incubation, the cells were plated out on LA-plates containing chloramphenicol, 15 µg/mL, and incubated overnight.

2.3.4 PCR-reactions

Three different types of PCR reactions were used for this thesis. For plasmid creation/amplification, the Takara CloneAmp™ HiFi PCR method was used.
For amplification of promoter libraries, and for making DNA for sequencing from *B.subtilis*, Q5 PCR was used.
For colony PCR, to confirm successful GeneEE insertion, *Taq* PCR was used.
Annealing temperatures varied depending on the promoters and were found using the NEB Tm calculator, available at https://tmcalculator.neb.com. The primer sequences used can be seen in Appendix B.1.
Q5 and *Taq* PCR protocols are based on protocols found on NEBcloner.com, while the Takara CloneAmp™ HiFi PCR protocol is based on the protocols available on the Takara web site.

Takara CloneAmp™ HiFi PCR protocol

The PCR reagents in Table 2.3.1 were combined on ice, pipette-mixed and quickly spun down on a tabletop centrifuge right before the reaction. The PCR reaction was run in a thermocycler according to the setup in Table 2.3.2.

Table 2.3.1: Takara CloneAmp^{TM} HiFi PCR reagents

Reagent	Volume
CloneAmp^{TM} HiFi PCR Premix	12.5 µL
10 µM forward primer	0.5 µL
10 µM reverse primer	0.5 µL
DNA template, <100 pg	1 µL
dH_2O	10.5 µL

Table 2.3.2: Thermocycler settings for Takara CloneAmp^{TM} HiFi PCR

Step	Temperature	Duration
1	98°C	20s
2	98°C	10s
3	55°C	15s
4	72°C	5s/kb
	Repeat 2-4	x30
5	72°C	1 min
	4°C	HOLD

Q5 PCR

In a PCR-tube, on ice, the reagents in Table 2.3.4 were combined and pipette-mixed gently. The PCR was run in a thermocycler with the settings described in Table 2.3.4. The tubes were briefly spun down in a tabletop centrifuge right before they were put in the thermocycler.

Table 2.3.3: Q5 PCR reagents

Component	Volume
5x Q5 Reaction buffer	10 µL
10 mM dNTPs	1 µL
10 µM forward primer	2.5 µL
10 µM reverse primer	2.5 µL
Template DNA, <100 pg	1 µL
Q5 High-Fidelity DNA Polymerase	0.5 µL
dH_2O	32.5 µL

Table 2.3.4: Thermocycler settings for Q5 PCR (Annealing temperature depends on primer melting temperatures)

Step	Temperature	Duration
1	98°C	30s
2	98°C	10s
3	50-72°C	20s
4	72°C	20s/kb
	Repeat 2-4	x30
5	72°C	2 min
	4°C	HOLD

Taq PCR

Taq PCR was used for colony PCR to confirm successful GeneEE inserts.
The *Taq* polymerase does not, however, have proof-reading, and was therefore not suitable for amplifying DNA for sequencing, where Q5 was used.
On ice, the reagents in Table 2.3.5 were combined and pipette-mixed.
When the colony PCR required more than 5 samples to be analyzed, a master mix was set up in a 2 mL centrifuge tube which scaled up the reagents to match the number of samples, and then divided into PCR-tubes before the colony to be analyzed was added into the appropriate tube.
The PCR-reaction was then run according to the setup in 2.3.6, and analyzed using agarose gel electrophoresis.

Table 2.3.5: *Taq* PCR mix

Component	**Volume**
10x *Taq* Reaction Buffer	2.5 µL
10 mM dNTPs	0.5 µl
10 µM forward Primer	0.5 µL
10 µM reverse Primer	0.5 µL
Template DNA	(solid colony)
Taq DNA Polymerase	0.125 µL
dH_2O	20.75 µL

Table 2.3.6: Thermocycler setup for *Taq* Polymerase PCR.

	Temp	**Time**
1	95 °C	30 s
2	95 °C	15 s
3	45-68 °C	30 s
4	68 °C	1 min/kb
	Repeat 2-4	x15
5	72 °C	5 min
6	4 °C	HOLD

2.3.5 Agarose gel electrophoresis

For analysis of PCR-products and to confirm sizes and purity, agarose gel electrophoresis was used.
All electrophoresis reactions presented in this thesis used the same setup: 90V, 40 min, in 0.8% Agarose (recipe in Section 2.2.1).
5 µl 1 kb DNA ladder mix was used, and the for the samples that were analyzed, 1 µl sample was added to 4 µl loading dye. For higher concentration samples, a diluted sample was used.
After the gels had run, they were visualized and stored in digital and physical formats, for analysis and size confirmation:

Plasmid DNA is usually found in one of three forms:
the covalently closed-circular/supercoiled(ccc) form, open circular/relaxed(oc) form, or linearized(l) form.
Agarose gel electrophoresis and image analysis of the gel allows for identification of these three forms by their migration distance on the gel: the oc-form is expected higher up, having migrated the shortest distance, with the l-form migrating a little farther, and the ccc-form migrating the longest (45).
The DNA ladder used in these experiments, seen in Appendix B.3, use linearized DNA, which is why the closed plasmids can appear to be of a smaller size than they really are, and why other plasmids of known size should be used as a comparison, in addition to the ladder.

2.3.6 BsaI digestion

BsaI is a type II restriction enzyme, which means it cuts outside its restriction sequence, leaving ”sticky ends”. The recognition sequence and method of cutting is displayed in Figure 2.3.1

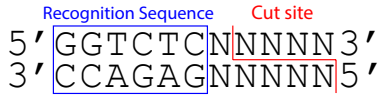


Figure 2.3.1: BsaI enzyme cuts outside its recognition sequence (blue), cut site marked by a red line.

The enzyme was used in Golden Gate Assembly, and also to show successful deletion of BsaI-site in the pHT10-mCherry-ΔBsaI-plasmid.
For BsaI-digestion, the reagents in Table 2.3.7 was combined in a 1.5 mL tube and incubated at 37 °C for 30 min.

Table 2.3.7: Reagents used for BsaI digestion

Reagent	Volume
Plasmid DNA	0.5 µg
5X CutSmart buffer	2.5 µL
BsaI-HF v2	0.5 µL
dH_2O	to 25 µL

2.3.7 Golden Gate Assembly

Golden Gate Assembly is a simultaneous, one-pot restriction digest/ligation reaction, developed by Engler et al. in 2008 used to insert DNA into a backbone. It has been shown to work with multiple DNA fragments (37). Golden gate assembly uses type II restriction enzymes, like BsaI, which cut outside their recognition sequence, and leaves ”sticky-end” overhangs, which can be ligated together using a ligation enzyme, for example, T4 ligase (37). The method is time and cost-efficient and simple to use, while still resulting in a high number of successfully ligated plasmids.
There are several golden gate assembly protocols available, and the one used in this experiment is based on an article by Potapov et al. (46).

The reaction is carried out in a PCR-tube in a thermocycler, cycling between 5-minute reactions at 37 °C, where the restriction enzyme is activated, and 5 minutes at 16 °C where the ligation enzyme is activated. After repeating these two steps 30 times, the restriction enzyme is heat-inactivated at 60 °C.
To ensure the highest possible transformation efficiency, the Golden Gate assembled plasmid should be transformed into an appropriate host organism as soon as possible after the

reaction is completed.

In a PCR-tube on ice, the ingredients were combined as displayed in Table 2.3.8, and the thermocycler setup in Table 2.3.9.

Table 2.3.8: Reagents used in Golden Gate Assembly

Reagent	Volume
T4 ligation buffer	2 µL
T4 DNA ligase	1.25 µL
BsaI-HF v2	0.75 µL
Insert DNA	75 ng
Backbone DNA	75 ng
dH_2O	to 20 µL

Table 2.3.9: Thermocycler settings for Golden Gate Assembly

Step	Temperature	Duration
1	37 °C	5 min
2	16 °C	5 min
	Repeat 1-2 x30	
3	60 °C	5 min
4	4°C	HOLD

2.3.8 DpnI digestion

The DpnI enzyme is used to remove all traces of the original plasmid after a PCR-reaction. The DpnI enzyme cuts methylated adenine at its recognition site GATC, as displayed in Figure 2.3.2.

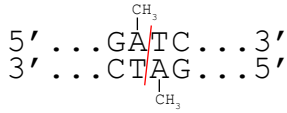


Figure 2.3.2: DpnI enzyme cuts methylated adenine at its recognition sequence, cut site marked by a red line.

Since PCR-reactions do not produce methylated DNA, only the template DNA will be cut by the DpnI enzyme.
The protocol on the NEB website has a 5-15 minutes incubation time, but for this thesis,

many hours were used for incubation, as it was important to avoid any original plasmid in the screening process, and because the DpnI enzyme does not have detrimental effects at prolonged incubation.
Whenever DpnI digestion was used for this thesis, after a PCR reaction, the reagents in Table 2.3.10 was added to the PCR-tube and incubated overnight at 37 °C. After incubation, the PCR product was purified using Qiagen QIAquick PCR purification kit, details in Appendix A.1.

Table 2.3.10: Reagents used for 100 µL DpnI digestion, added to 50 µL PCR product in a PCR-tube.

Reagent	Volume
PCR-product	50 µL
5X CutSmart buffer	10 µL
DpnI	1 µL
dH_2O	to 100 µL

2.3.9 Blunt end ligation and phosphorylation

For creating the closed backbone plasmid, the PCR product was phosphorylated and blunt-end ligated. Phosphorylation was done by combining the reagents in Table 2.3.11 in a PCR-tube. The tube was then incubated at 37 °C for activation, and then 20 minutes at 65 °C for heat-inactivation.
Blunt end ligation was then done by adding the ingredients in Table 2.3.12 to the same PCR-tube and incubating for 2 hours at room temperature.

Table 2.3.11: Reagents used for phosphorylation

Reagent	Volume
pHT10-mCherry-Δ-BsaI-BB PCR product (DpnI-treated)	8.5 µL
T4 ligase buffer	1 µL
T4 PNK	0.5 µL

Table 2.3.12: Reagents used for blunt end ligation

Reagent	Volume
T4 ligase buffer	2 µL
T4 DNA ligase	1 µL
dH_2O	7 µL

2.4 Data analysis

2.4.1 Analysis of plate reader data and selection of colonies

To calculate the expression level E of the picked *B. subtilis* colonies with the GeneEE inserts, as well as the controls, the measured fluorescence F, with the blank subtracted, was divided by the OD_{600} with the blank value subtracted to ensure that higher measured fluorescence was not merely an effect of higher cell densities.
The method is displayed in Equation 2.4.1.

$$E = \frac{F - F(blank)}{OD_{600} - OD_{600}(blank)} \tag{2.4.1}$$

A baseline expression level was calculated as a comparison to identify cells expressing mCherry strongly.
Since the original pHT10-mCherry-Δ-BsaI should have little to no uninduced expression of mCherry, the baseline chosen was the average expression for both uninduced samples.

To get the average base expression BE_{avg}, the average of the two E-values of the uninduced pHT10-mCherry-Δ-BsaI was calculated. This value was unique for each plate, as both the measured fluorescence and the OD_{600} of the controls varied between plates.

To calculate the relative expression, RE, of each well compared to the average base expression, the expression level for each well was divided by the average base expression BE_{avg} as displayed in Equation 2.4.2. Using Equation 2.4.2, each well gets a value that displays their relative expression compared to the BE_{avg}, with the value 1 being the exact same level of expression as measured by fluorescence relative to cell density, and 2 being twice the amount.

$$RE = \frac{E}{BE_{avg}} \tag{2.4.2}$$

2.4.2 Sequence analysis

After samples were sent to sequencing at Eurofins Genomics, they arrived back in a .zip-file containing the sequences in .ab1, .phd.1 and .seq file format. The ab1 format contains the sequence, in addition to some information about the quality of the reads, visible through chromatographic data.
The sequences were aligned to the backbone plasmid with the expected inserts using Benchling. After alignment, the inserts were copied to a new document, all of which can be seen in the result section tables 3.6.2, 3.6.3, 3.6.4, 3.6.5.
The insert-sequences were then copied into online tools BacPP, BPROM and improbizer, and the output data was gathered and analyzed. The BacPP data were analyzed using MS Excel, where the highest value for each sigma factor was identified for each clone.
The BPROM tool was used to find the promoter regions within the random inserts. -35 and -10 elements were identified and scored from 0-100 according to its match to the expected consensus. For the 50N samples, this did not result in any matches, and therefore the improbizer motif matcher was used.

This tool takes a motif from the user, here TTGACA(−35 element) and TATAAT(-10 element), and looks for partial or complete matches within the sequences provided.
To look for other, possibly unknown, motifs in the dataset, the improbizer (43) tool was used. The tool finds motifs that appear in sequences at higher rates than are expected by chance. The user defines the number of motifs to be found and initial motif size. Here, the number of motifs to be found was set to 3 and the initial motif size to 6.

Chapter 3

Results

3.1 Creation of pHT10-mCherry-ΔBsaI plasmid

For Golden Gate cloning using BsaI, the BsaI-site in the ampicillin gene of the plasmid pHT10-mCherry was removed using CloneAmp^{TM} HiFi PCR.
Using the pHT10-mCherry plasmid, diluted 1:1000, a Takara CloneAmp^{TM} HiFi PCR was set up as described in Section 2.3.4 using BsaI-primers. The primers bind to the BsaI site in the ampicillin gene, changing the codon TCT, to TCC, removing the BsaI-site without altering the coding sequence. This PCR product also has overlapping ends in the altered region, which allows the PCR product to be, by homologous recombination in *E. coli*, closed into a circular plasmid.

After the PCR was run, the size of the PCR-product was confirmed using gel electrophoresis, with the original pHT10-mCherry plasmid as a positive control. Gel picture is visible in Figure 3.1.1
The PCR-product has some byproduct, but the majority of the DNA is concentrated around 8600 bp. After confirmation of size, the PCR-product was DpnI-treated and incubated overnight.
After incubation, the DNA was purified using the Qiagen PCR purification kit, as described in Appendix A.1.
Nanodrop measurement of PCR product showed a concentration of 27.3 $\frac{ng}{\mu L}$.
After purification, the pHT10-mCherry-Δ PCR product was used to transform *E. coli* using heat-shock transformation and plated out on LA plates with 100 µg/mL ampicillin.
5 colonies were picked from the plate and grown overnight in liquid LB with 100 µg/mL ampicillin. From the overnight cultures, the closed plasmids were isolated using the Qiagen Qiaprep spin Miniprep kit, full protocol in Appendix A.2.
The plasmids were digested with BsaI, alongside the original plasmid as a control, to test for successful deletion of the BsaI-site, and incubated at 37 °C for 30 min.

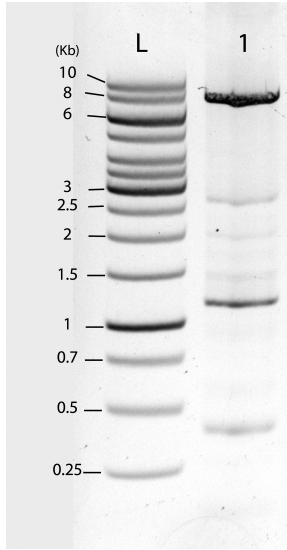


Figure 3.1.1: 1=pHT10-mCherry-ΔBsaI plasmid PCR product. Expected size=8651 bp. L=Thermo Scientific 1 kb GeneRuler DNA Ladder.

For confirmation of succesful deletion, the samples were analyzed using agarose gel electrophoresis.
The gel image is seen in Figure 3.1.2, where the BsaI-digested plasmids are compared to the non-digested plasmids. The plasmids appear to be of the same size, and of the same amount of ccc to l and oc form.

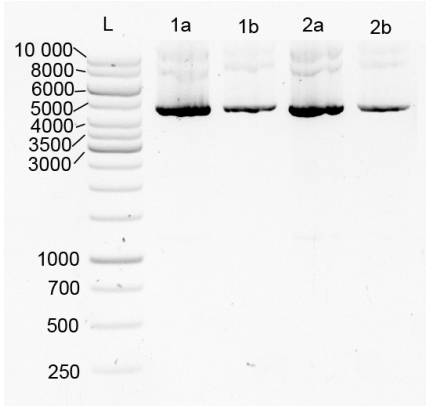


Figure 3.1.2: BsaI-digestion of isolated mCherry plasmids. 1a-2a are the original pHT10-mCherry-ΔBsaI plasmids, while 1b and 2b are digested with BsaI. The strongest bands are at about 5 kb, which is the ccc form of the plasmid that has migrated faster than the l and oc form seen closer to 8-10 kb.

3.2 Creation of pHT10-mCherry-ΔBsaI-BB plasmid

A backbone PCR-product was created, and a closed backbone plasmid was made by phosphorylation and blunt-end ligation before cloning into *E. coli*.

First, the backbone plasmid PCR product was created using Takara CloneAmp™ HiFi PCR, as described in Section 2.3.4, with the forward primer 517, and the reverse primer 518 (primers in Appendix B.1). Using these primers cuts out 1569 bp between the chloramphenicol resistance gene and the start of the mCherry gene, removing the native promoter and 5′ UTR.
For confirmation of correct PCR-product, agarose gel electrophoresis was run with the pHT10-mCherry-ΔBsaI plasmid as a control.

In Figure 3.2.1, the PCR product can be seen alongside the template plasmid for comparison. The BB-plasmid seem to be closer to 7000 bp, with the template (l form) at around 8600 bp.The ccc form of the template plasmid has migrated further on the gel.

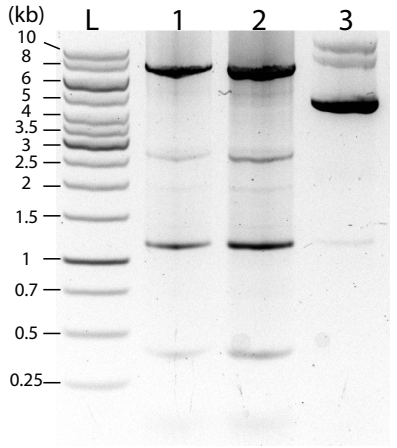


Figure 3.2.1: pHT10-mCherry-ΔBsaI-BB PCR product compared to pHT10-mCherry-ΔBsaI plasmid. L: 1kb GeneRuler™ Ladder. 1-2: pHT10-mCherry-ΔBsaI-BB PCR-product. 3: pHT10-mCherry-ΔBsaI plasmid. Expected sizes: 1-2: 7054, 3:8651 bp.

Nanodrop measurement of PCR DNA showed a concentration of 150.4 $\frac{ng}{\mu L}$.

After size confirmation, the plasmid was DpnI treated overnight as described in Section 2.3.8 and purified using the Qiagen QIAquick PCR Purification Kit.
8.5 µL was taken out and placed in a PCR-tube, for production of the closed backbone plasmid, where it was first phosphorylated, and then blunt-end ligated, as described in Section 2.3.9.
The closed backbone plasmid was then heat-shock transformed into *E. coli* and incubated overnight on LA-plates with ampicillin.

The transformation was successful, resulting in many white colonies. After overnight incubation, a colony was picked and incubated in liquid LB overnight and isolated using the Qiagen QIAprep Spin Miniprep Kit.
The closed plasmid can be observed on the gel in Figure 3.2.2. As a positive control, the original pHT10-mCherry-ΔBsaI plasmid was used.
The positive control resulted in red colonies, while the closed backbone plasmid resulted in white colonies.

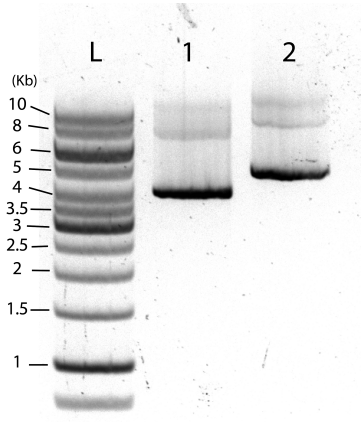


Figure 3.2.2: Comparison of pHT10-mCherry-ΔBsaI-BB plasmid and pHT10-mCherry-ΔBsaI plasmid. L: 1kb GeneRuler^{TM} Ladder. 1:pHT10-mCherry-ΔBsaI-BB. 2:pHT10-mCherry-ΔBsaI. The closed plasmids, in ccc form, have migrated farther than the l and oc form DNA, observable higher up in the gel. Expected sizes: 1:7054 bp, 2:8651 bp.

3.3 Generation of promoter library in *E coli*

3.3.1 GeneEE insert amplification

First, the GeneEE random DNA inserts were generated using PCR, based on templates ordered from Integrated DNA Technologies (IDT).
For the sequences of the 200N/50N/200NSD GeneEE insert templates, see Appendix B.2. The sequences were amplified using Q5 PCR, as described in Section 2.3.4, with the exception that the thermocycler reaction was only repeated 10 times for the 200N and 200NSD inserts, and 5 times for the 50N inserts, to avoid by-products. The primers used were BB-prefix and BB-suffix primers, sequences in Appendix B.1.

After the PCR-reaction, the PCR-products was confirmed for size: ≈100 bp for 50N, ≈250 bp for 200N/200NSD using agarose gel electrophoresis. Gel picture can be seen in Figure 3.3.1.
After sizes were confirmed, the PCR-products were purified using a Qiagen QIAquick PCR Purification Kit as described in Appendix A.1.
The GeneEE PCR product concentrations were measured using Nanodrop One spectrophotometer and the concentrations are presented in Table 3.3.1.

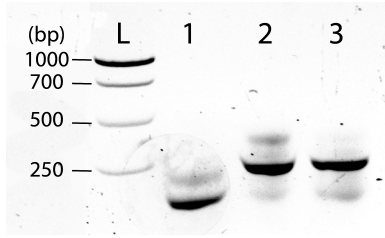


Figure 3.3.1: Q5 PCR amplified GeneEE insert segments. L=1kb ladder. 1=50N, 2=200N, 3=200NSD. Expected sizes: 50N: ≈100 bp, 200N/200NSD: ≈ 250bp

Table 3.3.1: GeneEE PCR product insert concentrations after insert library amplification, measured using a Nanodrop ONE spectrophotometer

GeneEE insert	Concentration $\frac{ng}{\mu L}$
50N	21.8
200N	37.3
200NSD	29.4

3.3.2 Golden Gate Assembly and transformation of *E. coli*

After the PCR backbone and GeneEE inserts were generated and verified, the inserts were ligated into the backbone using Golden Gate Assembly, as described in Section 2.3.7.

Three golden gate reactions were set up, using the PCR amplified 50N, 200N, and 200NSD GeneEE as inserts, and the pHT10-mCherry-Δ-BsaI-BB DpnI-treated and purified PCR-product as the backbone.

Immediately after the golden gate reaction finished, the ligated plasmids were used to transform *E. coli* competent cells using heat-shock transformation:
10 µL x2 from each PCR-tube was used to transform the *E. coli* cells, alongside a positive control; the pHT10-mCherry-ΔBsaI plasmid, and negative control; untransformed *E. coli* cells. The transformed cells were plated out on LA plates with ampicillin and incubated overnight at 37 °C.

All transformations were successful, with visibly red and white colonies on every plate. An example of how the plates looked can be seen in Appendix C.1.
Both red (probable mCherry expression) and white colonies (no expression) were counted. The number of red and white colonies from each plate and the percentage of red colonies are displayed in Table 3.3.2. 5 colonies from each promoter library (50N, 200N, 200NSD

Table 3.3.2: Number of *E. coli* transformants after transformation with hybrid plasmids created by Golden Gate Assembly of pHT10-mCherry-ΔBsaI-BB PCR product with GeneEE inserts.

Insert & Plate nr.	White colonies	Red colonies	% of red colonies
50N 1	3816	236	5.8
50N 2	4536	193	4.1
200N 1	2104	483	18.7
200N 2	1573	213	11.9
200NSD 1	3280	927	22.0
200NSD 2	4543	1151	20.2

inserts), of increasing colour intensity, was picked, and *Taq* colony PCR was run as described in section 2.3.4 to confirm successful random promoter insertion. The expected sizes of the insert PCR-products were 489 bp for 50N and ≈ 650 bp for 200N/200NSD. the pHT10-mCherry-ΔBsaI plasmid had an expected size of 2038 bp, and should be distinguishable using agarose gel electrophoresis.
After the colony PCR had run, the PCR products were analyzed using agarose gel electrophoresis, gel picture in Figure 3.3.2.
While there are some minor variations in insert size, there is no present original plasmid, indicating that all colonies have a GeneEE insert, and since they had a red color, successful expression due to a functional promoter.

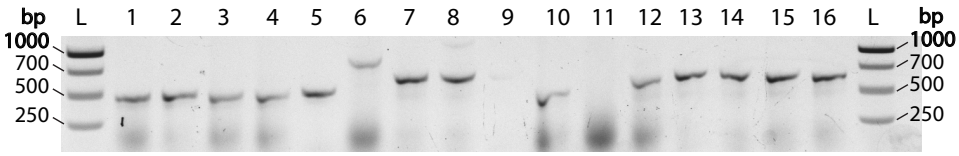


Figure 3.3.2: *Taq* colony PCR of golden gate ligated pHT10-mCherry-ΔBsaI-BB with 50N, 200N and 200NSD GeneEE inserts. L=1kb GeneRuler^{TM} DNA ladder. 1-5: 50N inserts, 6-10: 200N inserts, 11-15, 200NSD inserts. Expected sizes: 50N: 489 bp, 200N/200NSD: ≈ 650 bp, original plasmid: 2038 bp.

For isolation of the *E. coli* random promoter library, the cells were scraped off the plate:
2 mL dH_2O was added to the *E. coli* plates with the random promoter plasmid clones, and the plates were scraped and the resulting water/cell mixture, ≈ 1.5 mL was transferred to 1.5 mL tubes and frozen at -80 °C.
Before freezing, 200 µL was taken out from each tube for isolation of promoter library. The random promoter library was then isolated using the Qiagen QIAprep Spin Miniprep Kit, and the resulting plasmid-random promoter libraries were named according to their GeneEE inserts. Names and concentrations in Table 3.3.3.

Table 3.3.3: Concentrations of isolated random promoter plasmid libraries, isolated from *E. coli* using miniprep.

Plasmid library	**Concentration** $\frac{ng}{\mu L}$
pHT10-mCherry-ΔBsaI-50N	518
pHT10-mCherry-ΔBsaI-200N	627
pHT10-mCherry-ΔBsaI-200NSD	450

3.4 Transformation of *B. subtilis* with random promoter-plasmid library

To establish the random promoter library in *B. subtilis*, the isolated library from *E. coli* was used to transform *B. subtilis* 168 cells.
To get more transformants for screening, 2 transformations were done for each of the GeneEE insert plasmids. 3 µL plasmid DNA was used for transformations, and the cells were transformed as described in Table 3.4.1.

Table 3.4.1: Setup of *B. subtilis* transformation with random promoter library and controls. Each transformation was plated out on an LA-plate with chloramphenicol, 15 µg/mL

Plasmid	Nr. of transformations	Purpose
pHT10-mCherry-Δ-BsaI-BB	1	Screening control
pHT10-mCherry-Δ-BsaI	1	Positive control
pHT10-mCherry-Δ-BsaI-50N	2	Screen for promoters
ppHT10-mCherry-Δ-BsaI-200N	2	Screen for promoters
pHT10-mCherry-Δ-BsaI-200NSD	2	Screen for promoters
Untransformed *B. subtilis* 168	1	Negative control

After overnight incubation, the cells were successfully transformed, and all the expected plates had colonies, with no visible colonies on the negative control plate.
The number of colonies on each plate are displayed in Table 3.4.2. In contrast to the *E. coli* transformants, no visible colour difference could be observed on the plate.
To confirm correct sizes, a *Taq* colony PCR was set up. as described in Section 2.3.4, using the original pHT10-mCherry-ΔBsaI from the positive control plates as reference. The gel picture can be seen in image 3.4.1 To screen for strong promoters, 84 colonies were

Table 3.4.2: Number of transformed *B. subtilis* colonies pr. plate. The cells were transformed with random promoter-plasmid libraries isolated from *E. coli*. The plates are named after the plasmids they were transformed with.

Plasmid library & Plate nr.	Colonies
pHT10-mCherry-ΔBsaI-50N 1	590
pHT10-mCherry-ΔBsaI-50N 2	483
pHT10-mCherry-ΔBsaI-200N 1	299
pHT10-mCherry-ΔBsaI-200N 2	342
pHT10-mCherry-ΔBsaI-200NSD 1	564
pHT10-mCherry-ΔBsaI-200NSD 2	399

picked from each plate, picking colonies of all sizes, from across the plate. The picked colonies were transferred to 96-well plates with 200 µL LB, 10 µg/mL chloramphenicol.

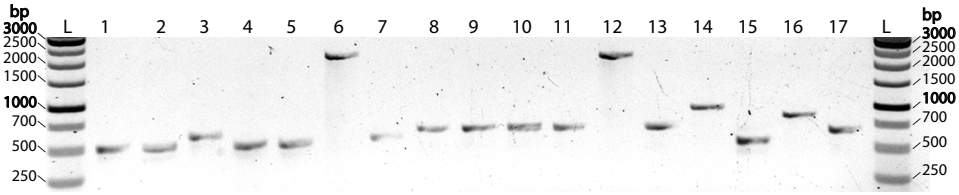


Figure 3.4.1: *Taq* colony PCR of *B. subtilis* 168 colonies transformed with random promoter-plasmid library isolated from *E. coli*, with 50N, 200N and 200NSD GeneEE inserts.
L=1kb GeneRuler^{TM} DNA ladder. 1-5: 50N inserts, 6+12:pHT10-mCherry-ΔBsaI plasmid, 7-11: 200N inserts, 13-17: 200NSD inserts

3.5 Plate readings and picked colonies

In addition to the picked colonies, some controls were set up, as described in Table 3.5.1. All control plasmids were in *B. subtilis* 168.

Table 3.5.1: Controls used for absorbance readings of 96-well plate in Tecan infinite M200 Pro plate reader. All controls in *B. subtilis* 168. Induction of pHT10-mCherry-ΔBsaI was done with 1:1000 (final concentration) IPTG.

Plasmid	Control wells	Purpose
LB-medium	1	Blank
LB-medium with chloramphenicol	1	Blank
Untransformed *B. subtilis* 168	2	Negative control
pHT10-mCherry-Δ-BsaI-BB	2	Negative control
pHT10-mCherry-Δ-BsaI (uninduced)	2	Control
pHT10-mCherry-Δ-BsaI (induced)	2	Positive control

The 96-well plates were incubated overnight for ≈20 hours at 37 °C, shaking at 700 RPM.
After overnight incubation, the samples were diluted 1:10 by transferring 20 µL overnight culture into 180 µL LB with 10 µg/mL chloramphenicol on a Falcon 96-well black clear bottom plate.
The diluted samples were then analyzed using a Tecan infinite M200 Pro plate reader to capture mCherry fluorescence and OD_{600}. mCherry fluorescence was measured by excitation at wavelength 584 nm and emission at 620 nm with a gain value of 100. The cell optical density was measured at 600 nm.

As described in Section 2.4, the relative expression, RE, of each well compared to the uninduced pHT10-mCherry-ΔBsaI was calculated, and the strongest expressions were picked for Q5 PCR and sequencing. Table 3.5.2 show the RE-values of all picked wells, named after the plate and well they were picked from. The data is also displayed graphically in Figure 3.5.1
To select colonies, the ones with the highest RE-values pr. plate was picked, with a low

cut off point of BE=2.5.
One plate, 200N 2, did not have any wells with BE values greater than 2.5, and therefore, the two closest wells were chosen.

To see if the data could be replicated, the picked cells (in Table 3.5.2 were transferred to a new plate, grown overnight and measured at 1:10 dilution again.
This was done 2 times for a total of 3 RE-values for each colony. For these two replications, another 200NSD insert clone, 200NSD2 F10-OLD, which showed promise in an earlier round of cloning was also included, since it had earlier displayed high fluorescence

All RE-values, including those from non-picked wells, can be seen in Appendix C.2.

Table 3.5.2: Relative expression, RE, values from fluorescence and OD_{600}-measured *B.subtilis* 168. The values represent the relative fluorescence pr. cell density for the clones relative to uninduced pHT10-mCherry-ΔBsaI plasmid controls. SD=standard deviation. For values 2 and 3 all clones were on the same plate, and compared to the same control, while the first RE-values are compared to controls on the same plate as the clones. '0'-values indicate that there was not measured fluorescence above the background/blank level.

Plate	Well	RE-value	RE-value 2	RE-value 3	SD
50N 1	D8	2.67	0	0.58	1.40
50N 2	C4	3.49	1.93	0.11	1.69
	E7	5.02	2.06	0	2.52
	F3	3.59	11.46	0	5.86
	G4	4.29	1.47	1.41	1.65
	G8	5.98	2.51	1.70	2.27
200N 1	F7	2.38	3.84	1.45	1.20
	G4	2.27	3.62	0.46	1.59
200N 2	F5	2.42	3.35	0.63	1.38
	F6	2.48	0	2.37	1.40
200NSD 1	D4	5.43	9.61	8.05	2.11
200NSD 2	C3	3.83	0	0	2.21
	D1	3.32	0	1.05	1.70
	E2	3.48	2.48	1.94	0.78
	F2	2.91	0.94	0.95	1.13
	F3	3.31	1.5	0.2	1.56
	F7	2.77	1.98	1.39	0.69
	F8	3.53	6.10	1.13	2.49
	G3	3.00	0	0.59	1.59
200NSD 2 (28.05)	F10		21.43	6.47	10.58

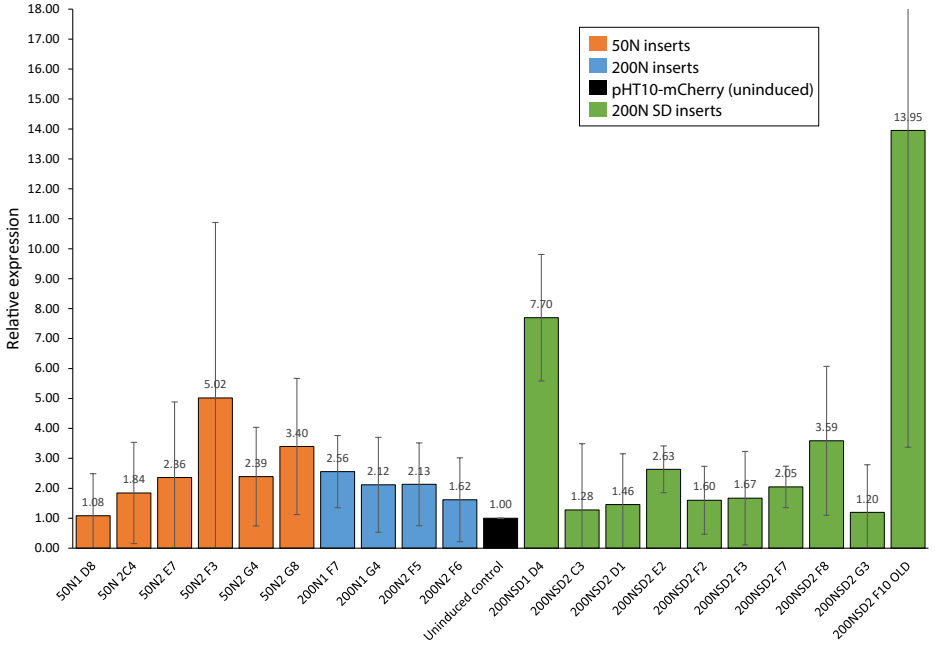


Figure 3.5.1: Visualization of RE-values of the GeneEE insert hybrid plasmids, showing average RE-values based on the values in Table 3.5.2. Uninduced control value set to 1, all other values relative to this.

To prepare the cells for sequencing and Q5 PCR, 1 µL of cell suspension was transferred to a PCR-tube, and 9 µL dH_2O was added. The cells were then lysed by heating to 95 °C for 10 minutes in a thermocycler, to ensure a cleaner template for PCR, and thereby a cleaner PCR-product.
1 µL lysed cells were then used as the template for a standard 50 µL Q5 PCR with the sequencing primers 521 and 522, see Appendix B.1 for primer sequences, and Section 2.3.4 for Q5 protocol.

The PCR products were inspected using agarose gel electrophoresis, cleaned up using the Qiagen PCR purification kit, and measured using nanodrop, before being sent to sequencing at Eurofins Genomics.
The gel pictures from the PCR can be seen in Figure 3.5.2(50N and 200N insert clones) and 3.5.3(200NSD insert clones). The expected sizes for clones with the correct inserts were 489 bp for 50N, 639 bp for 200N and 650 bp for 200NSD.
After purification, sample 50N2 F6 had too low PCR yield to be sent for sequencing along-

side sample 200NSD2 C1.

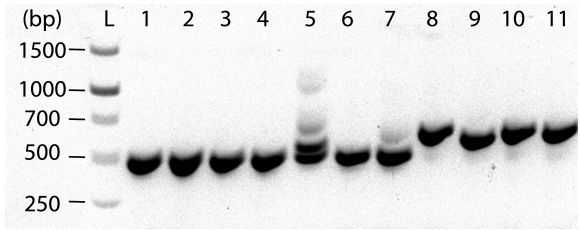


Figure 3.5.2: Q5 PCR of *B. subtilis* 168 transformed with pHT10-mCherry-ΔBsaI-50N(1-7) and pHT10-mCherry-ΔBsaI-200N (8-11) to sequencing. The template was 1 µL 1:10 diluted lysate of liquid culture from 96-well plate, corresponding to the plates in Appendix C.2, with the samples named after the plate and well from which it was picked.
Expected sizes: 1-7= 489 bp 8-11: 639 bp.
L=1kb GeneRuler^{TM} DNA ladder. |**1**=50N1 D8| |**2**=50N2 C4| |**3**=50N2 E7| |**4**=50N2 F3| |**5**=50N2 F6| |**6**=50N2 G4| |**7**=50N2 G8| |**8**=200N1 F7| |**9**=200N1 G4| |**10**=200N2 F5| |**11**=200N2 F6|

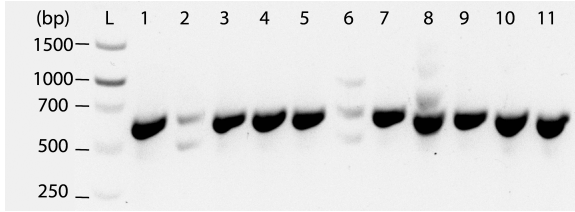


Figure 3.5.3: Q5 PCR of *B. subtilis* 168 transformed with pHT10-mCherry-ΔBsaI-200NSD to sequencing. The template was 1 µL 1:10 diluted lysate of liquid culture from 96-well plate, corresponding to the plates in Appendix C.2, with the samples named after the plate and well from which it was picked.
Expected sizes: 650 bp.
Lane 1 is from the plate 200NSD1, while 2-11 are from 200NSD2
L=1kb GeneRuler^{TM} DNA ladder. |1=(200NSD1) D4| |2=C1| |3=C3| |4=D1| |5=E2| |6=E8| |7=F2| |8=F3| |9=F7| |10=F8| |10=G3|

3.6 Sequence analysis

In silico construction of Golden Gate Assembled plasmid and Q5 PCR-product, using Benchling for sequence alignment, produced the expected sequences, the insert part of which can be seen in Figure 3.6.1. These were used as the basis for sequence alignment in Benchling. Some samples had either too low-quality sequencing data to be included in

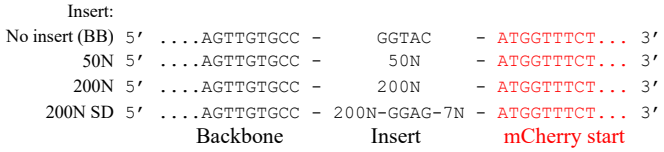


Figure 3.6.1: Expected sequences for GeneEE inserts, found using *in silico* cloning in Benchling (40). If the ligation was unsuccessful, the BB plasmid was expected to be seen instead of the inserts. N denotes any nucleotide (ACTG)

the results:
Sample 200NSD2 E8 had a low-quality PCR product, and the sequenced DNA was not of high enough quality to be included in the rest of the experiments.

The cells that were sent for sequencing mostly came back with the expected backbone sequence and insert size, with some variations:
The expected insert sizes were 50, 200 and 211 (200N + SD sequence(GGAG)+7N).
Sample 200NSD2 F3 only had an insert size of 201, but with an intact SD-sequence(190N+SD+7N).
Sample 200NSD2 E2 had the expected insert size, and good quality of reads, but with a substitution in the SD-sequence, where GGAG had been changed to GGAA.
All 50N samples had the expected size inserts.

The sequencing data were analyzed using three different online tools:
The BacPP tool (41) was used for the prediction of recognition from different sigma-factors. BacPP only supported 80 bp segments, so the results were split into 80 bp segments, which was scored individually. Presented in Table 3.6.1 is the highest values for each sigma factor pr. sample.

The tool BPROM (42) was used to identify possible -35 and -10 elements for the sequences, based on the *E. coli* σ^{70}, and thereby the *B. subtilis* σ^{A}, which shares consensus sequences at -35 and -10. Each of these elements was given a score from 0-100 for how much they matched the consensus sequences.
The sequences, alongside the identified -35 and -10 elements, and its identified transcription start site and if any, Shine-Dalgarno sequences, are presented in tables 3.6.2, 3.6.3, 3.6.4, 3.6.5. Here the whole insert sequence is displayed, and the BPROM score, alongside the average RE-values from Table 3.5.2. The higher expressing cells generally had a closer consensus match and a higher BPROM-score, except for in the 50N samples: The BPROM tool did not find any promoter elements for any of the 50N samples. Instead, the Improbizer motif matcher tool was used to find hits close to the consensus sequences -35: TTGACA and -10: TATAAT. The 50N sequences and consensus matches are listed in

Table 3.6.6.
Also of note is that the BPROM tool is able to find multiple promoter sequences in the same dataset. However, for each of the 200N/200NSD samples, only 1 promoter was identified.

Since some of the samples did not match any of the predicted sigma-factors bindings (in *E. coli*), the improbizer tool was used to look for motifs that appeared more than expected by chance in the sequences. For accuracy, the 50N samples and 200N/200NSD samples were analyzed separately, because the length of the sequences affects how the tool performs the analysis. The results are displayed in Figures 3.6.2 and 3.6.3.

The sequencing analysis data presented in the tables below will first be briefly summarized.
All of the 200N and 200N SD samples were recognized by the BPROM promoter prediction software, with a -10 and a -35 element being recognized and scored, alongside the TrSS.
The -10 element had the lowest score of 31 for the sample 200N2F5 and the highest score of 83 for the sample 200NSD2E2.
The -35 element had scores ranging from 2 in sample 200NSD2E2 to 56 in sample 200N2F6.
The Shine-Dalgarno sequence was intact in all the 200NSD samples, with a mutation in 200NSD2E2 changing GGAG to GGAA. All 200NSD samples also had the expected 7 bp spacer after the SD-sequence. Some of the samples had additional GGAG motifs present, but further upstream from the TnSS.

In the 200N samples, some SD-sequences were found downstream of the predicted promoter sites. Sample 200N2F5 had the shortest distance between The SD sequence and the TnSS, with the SD-sequence 13 bp upstream of the TnSS.
Also noted in the tables below are the length of the 5' UTRs from the identified TrSS. These ranged from 22 for sample 200NSD2E2 to 140 in sample 200NSD2E8.

As can be seen in Table 3.6.6, in the 50N samples, motifs were found that could resemble the -10 and -35 elements in almost all samples, with only 50N2G8 lacking a -35 element match.
Only sample 50N2F3 had an SD-sequence consensus match, with the SD-sequence found within a partial -35 element match. As seen in Table 3.6.1, All of the 200N and 200NSD-sequences were predicted to be able to bind σ^{70}, or its *B. subtilis* homologue, σ^{A}, with sample 200N1F7 having an 81 % probability, and all other samples at 99 or 100% probability. As expected, there was a clear correlation between predicted binding for σ^{70} and σ^{38}, especially for the 200NSD samples. 200N2F5 had only 50% predicted binding for σ^{38}, despite having a 99 % chance of binding σ^{70}.
There was no clear correlation between predicted binding to σ^{54} and σ^{70}, but several samples, including all 200N samples, were predicted to have about 50% chance of binding to σ^{54}. None of the 50N samples were predicted to bind to σ^{54}.

There was the least predicted possibility of binding to σ^{28}, with none of the 50N samples, half of the 200N samples and 7/11 200NSD samples being predicted to bind.
Two 50N samples were not predicted to have any σ-factor binding sites. One of them, 50N1D8 had the lowest average RE-values, that averaged almost at the base level. The other sample, 50N2G8 was one of the higher average expressing samples, with only 1 50N sample, 50N2F3, and 3 200NSD samples, SD1D4, SD2F8 and SD2F10-OLD having a higher average expression.

Table 3.6.1: Results from BacPP (41) promoter prediction related to sigma factors 24,28,32,38,54 and 70 for all sequenced samples from *B. subtilis*. The scores for each σ-factor is the BacPP prediction that this sigma factor could bind to the sequence.
Sequences over 80 bp have several scores for each 80 bp segment. Shown here is the highest scores. For comparison, the average relative expression (RE) values of measured mCherry fluorescence are also displayed.

Promoter	σ^{24}	σ^{28}	σ^{32}	σ^{38}	σ^{54}	σ^{70}	**Avg. RE-value**
50N1 D8	0	0	0	0	0	0	1.08
50N2 C4	66	0	45	35	0	50	1.84
50N2 E7	28	0	0	0	0	22	2.36
50N2 F3	0	0	0	0	0	31	5.02
50N2 G4	0	0	25	0	0	50	2.39
50N2 G8	0	0	0	0	0	0	3.40
200N1 F7	46	0	83	90	51	81	2.56
200N1 G4	97	45	57	99	62	100	2.12
200N2 F5	58	0	66	50	65	99	2.13
200N2 F6	58	46	73	99	60	99	1.62
200NSD1D4	92	58	80	99	61	100	7.70
200NSD2C3	69	51	61	99	56	100	1.28
200NSD2D1	69	55	54	99	56	100	1.46
200NSD2E2	58	55	52	99	51	100	2.62
200NSD2E8	76	46	59	100	56	100	1.70
200NSD2F2	58	0	45	99	72	100	1.60
200NSD2F3	56	49	42	99	0	99	1.67
200NSD2F7	56	0	59	99	60	100	2.05
200NSD2F8	84	0	69	99	70	100	3.59
200NSD2G3	56	0	73	90	0	100	1.20
200NSD2-F10-OLD	64	51	45	95	0	99	13.95

Table 3.6.2: GeneEE Random insert sequences from *B. subtilis* samples with promoter region and $5'$ UTR. The name of the promoter corresponds to the well from which it was grown and measured. The BPROM (42) bacterial promoter predictor was used to identify -35 and -10 elements. Presented in the table are also the scores for consensus match of the -35 and -10 elements. The matches are scored 0-100, with higher scores indicating closer match to consensus. The -35 and -10 elements are marked in red and the transcription start sites, first nucleotide (+1) is marked in green. In some sequences, a -16 'TG'-motif was also observed, marked in **bold**. If present, SD-sequences (GGAG) is marked in blue. Only SD-sequences downstream from the TrSS is marked. Also presented is the length of the $5'$ UTR

Promoter	**Insert sequence** $5' \rightarrow 3'$	**insert size**	**-10**	**-35**	$5'$ **UTR**	**RE**
200N1G4	AGGGTTTGTGGGATCCGTTGTCTACTTGCCTATACGCTGCTAGCAT GGAGAGATTTTTCGAGTATGTGATATCTGGCCAGTTGCGAAACAGT TCAGACTTAATTCAGCCAAGCTTATACATTAAA TTGAATTGATGCAGAGGATAGAGCTAACTTGTTTTGG AGAAGACTTGTGCGCTACCGGTTCCTGGGCTTGCCGGGA	200	38	35	39	2.12
200N1F7	TTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAG GCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCG AATGGCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGTCGGAA TGGACGACGGCCATAGTTACCCTTATTATCAAGATA AGAAAGAAAAGGATTTTTCGCTACGCTCA	200	61	23	30	2.56
200N2F5	AGACTTAGAACAGAATGAGTATATGGGCCGGTGTTGAGCGTTGATCGTATT TAGACATTGGCCCGTAAGGTGAAGATGAGCTGT CGCTCGATGGTAATGATCACGCGTCGACTTGGACGTAACTATGGGGGAGTAGGT TCGTGAGAGCGTTTGTATACCACGGGGTTGTATGGCAGTATGGAGCATGCTAGTCTGCCTTA	200	31	29	117	2.13
200N2F6	GAGGTCTTTTTGCCAATTTACATCGTGCCTATAGGTCAT TTGCCTGCGTCACGGAGTTCACGCATTATTTGTGGT GTTATCTAATAAGTTGGGCCCGGCTTAGGCATTAAGACACGGGACTCTT GCACTATGGAGGAAATGGTTTTCTATCTGAACGTCTAAGTGTA AAAGTTGCGGAGGTCATCATGAGCGGTGTATTG	200	35	56	126	1.62

Table 3.6.3: GeneEE Random insert sequences from *B. subtilis* samples with promoter region and $5'$ UTR.
The name of the promoter corresponds to the well from which it was grown and measured. The BPROM (42) bacterial promoter predictor was used to identify -35 and -10 elements. Presented in the table are also the scores for consensus match of the -35 and -10 elements. The matches are scored 0-100, with higher scores indicating closer match to consensus.
The -35 and -10 elements are marked in red and the transcription start sites, first nucleotide (+1) is marked in green. In some sequences, a -16 'TG'-motif was also observed, marked in **bold**.
If present, SD-sequences (GGAG) is marked in blue. Only SD-sequences downstream from the TrSS is marked. Also presented is the length of the $5'$ UTR

Promoter	**Insert sequence** $5' \rightarrow 3'$	**insert size**	**-10**	**-35**	$5'$ **UTR**	**RE**
200NSD1D4	CCACTGCGGGATTCTTAGAAATAAGTGGCCTCGAATAATC ACGTGCGTTATTTTGATGTAAATACGTAATACTCGTAGTCGGG TAATGAGGTAAGTGTGTAGTATTGCTGCATATATTTTT TTGAAGTGTAGAGATCTCAGGCTTTCTTGAATAATACCAGA GGTTACACACCAATTGTTTATATATTGGTTATTAGAAAGGAGTTTATGT	211	51	54	46	7.70
200NSD2C3	CTAATAAAATCTAAAAGTGCCCCAGTCATTAAGGTTCA TTTACTGGAAAATAGAGGGTTAATTTGGCAGACTATCAAGC GGTAGTACCGTTATATGCTTCCATTCATTCGAGAGTTTCTTTGG TCGAGCGGCGTTGATAGCTATTGGTGGGTTGAATATGTTCCTGG GGCCTCACGAGCTGTCTTCCGGGAGGAGAGCGAGGAGCGTGCGG	211	41	42	133	1.28
200NSD2D1	AAAGCTGGAGGTCCTTAATTCAGAATGTTGTTCTATTTATTCAATAAGCTTCATCGTTTG TTGACCTTTTCGTTTCCGTTTTCGCTGTTATCAAGAGTCTA GGAGTATACCTTCCGGGGTTGCAGGTGGTTTGGTCGTGT AAGACGGTGCGAGGATCTGATACTATGAGTCTCAATAGTTGACTG GGGGTTGGCATCATTGGAGTTGCGCG	211	37	44	111	1.46
200NSD2E2	TTGATAATAATTTCAGACGGTTGGTCGTAGGAAGAAAATCGCTTCATGGTAA GATGGCGTTATTACGTATGGATCTTGTGTCGGGGTTAGTTTTGGACGTTTGGG GGAGTCGTGGTTGGGAATTAAGTTACGTTTCTTTGGAACAAGTGACGG ATAAATACGCTTCTGGGTTGTAGCTATAATGAAGCCT TTTGTTTGTGGGAAACATAGG	211	83	2	22	2.63

Table 3.6.4: GeneEE Random insert sequences from *B. subtilis* samples with promoter region and 5′ UTR.
The name of the promoter corresponds to the well from which it was grown and measured. The BPROM (42) bacterial promoter predictor was used to identify −35 and −10 elements. Presented in the table are also the scores for consensus match of the −35 and −10 elements. The matches are scored 0-100, with higher scores indicating closer match to consensus.
The −35 and −10 elements are marked in red and the transcription start sites, first nucleotide (+1) is marked in green. In some sequences, a −16 'TG'-motif was also observed, marked in **bold**.
If present, SD-sequences (GGAG) is marked in blue. Only SD-sequences downstream from the TrSS is marked. Also presented is the length of the 5′ UTR

Promoter	**Insert sequence** 5′ → 3′	**insert size**	**-10**	**-35**	5′ **UTR**	**RE**
200NSD2E8	CGATTCATACTGACCACTTACCGGCTCCGT TTTACAAGAAATGTGTATTTTATCTTCCGGGGCACTC TAATTAGCTTGATTACTTCCTTATGCGGTTCGTAGCTGAATTTCCGGG ATGTGGTACAGGGTCCATAAAGGTGCGGCTGAAGTTGCAGCTAGTGAC GCTGAATTGAGGTGTTGAGGTGCTATGACAGCGGAGGTAGGGG	206	48	47	140	1.70
200NSD2F2	TACGTTAGTTTTTGGCATTATGATTGACGCTTTCTCAACAGAGGCACTAGAG TCGCGGGTGCGTTTATATCATTTGGCACTTGCTGCGTTAAGGGTACTGG TTACGTAGTAGTATATCGGTTCTACAATAATGTTG CACAGGTGGTAGTTTAGAGCCGGCCGAATGGGTGGGGAG CGTTTACGTGGGAAGGTCTAGTGCGGGAGCGCCATG	211	65	10	76	1.60
200NSD2F3	CTTAATTGGTAATGATGAGAGTGTAGTATATAACTGTTACGAGTACG TGTAGGAGAGAGGTTTCTGGTTTACGGGACTATGGAGTATAGTAACGT TTGTCTTTCGACGTCTTCAGCGTTCTGTTAATAGTTAGAG TCGCTTTTGCTGGGTTTCGCGTAAGTTGCGCAGCCGGTGCTGATAAGC GGAAGAGGGAGGGGGCGA	201	34	48	66	1.67
200NSD2F7	AACTCTGTTGGGGGGCGGTGCGGCTAACGGTAGC CATGGTCCACGGACACTCTCACCTCCATTCTTT TTGGTTCGGGGTCTGAAA**TG**CTATCGTCAATACCA TTGGGTTTTTTTTTGATGGATGCGTATATTGAGTGATGGGTGTGTTTCAGCTTTA TGGATGGAGGTCTAGTGTATCAACGTGCGACGTGGTTTGTGTAGGAGTTGCAAG	211	46	25	111	2.05

Table 3.6.5: GeneEE Random insert sequences from *B. subtilis* samples with promoter region and $5'$ UTR.
The name of the promoter corresponds to the well from which it was grown and measured. The BPROM (42) bacterial promoter predictor was used to identify -35 and -10 elements. Presented in the table are also the scores for consensus match of the -35 and -10 elements. The matches are scored 0-100, with higher scores indicating closer match to consensus.
The -35 and -10 elements are marked in red and the transcription start sites, first nucleotide (+1) is marked in green. In some sequences, a -16 'TG'-motif was also observed, marked in **bold**.
If present, SD-sequences (GGAG) is marked in blue. Only SD-sequences downstream from the TrSS is marked. Also presented is the length of the $5'$ UTR

Promoter	**Insert sequence** $5' \rightarrow 3'$	**insert size**	**-10**	**-35**	$5'$ **UTR**	**RE**
200NSD2F8	ATGTGAATTTTTGGTGTAAGCGTACTTTGCCGTGTG TTCATGGGTGGATGGGTTTAACAACCTATTCTGATGGGA AGAATCAGATAGAACGTGCATACTGGATATCTTGAGAAATGGGAG TCTTTTTGATTAGGCGTATGTGTGTTATTTATGCGAGGTTATCTCACGGGAG TGTATGGTCAGGACTTTCGTTGGGAGCTGGAGAGGCAAA	211	45	32	137	3.59
200NSD2G3	ATTGATCTTTGGTGGGGATTCTTGGATCGCGTGTTATGCGAT TAGAGACGGGTTTGTATTCATTTAACGTGGTTGTGGAATAGT GCTCTTTCTAGTCTCAACTGCGGACGGAGGAATTATGT TTGTCGAGTGTGGTTTGGGATATACAATTTTGTCC TGTATTATTGCGGGGGCAATGGGGGATAGGCGGGTACGATGTAGGAGGCAGGCT	211	54	47	55	1.20
200NSD2F10 old	AAATGGCTGTGTTTTCGGGCGTTTTTGGTTCTCACGGCTTGT GTGTACGTGTGAGGTCGACGTGGGCCACGAATGGTCATTAAA TCCTTTTGGTTAAAGGGATTATCGTACTGTTTGATTGGG TTTAAGGGGTTGTCAGTA**TG**TTATAGTGGTGAAG GGGGTGCTTTAAAGATTTTTCTACCGGTCACCGGCATGGGCATGGAGTGTCTGA	211	69	35	55	13.95

Table 3.6.6: 50N Promoter(random insert) sequences from *B. subtilis* samples. The name of the promoter corresponds to the well from which it was grown and measured. The Improbizer motif matcher was used to identify -35 and -10 elements from consensus sequences TTGACA(-35) and TATAAT(-10). The -35 and -10 elements are marked in red. Only -35 elements upstream of the -10 elements were marked. The locations of the -10 and -35 elements are also marked

Promoter	**Insert sequence** $5' \rightarrow 3'$	**insert size**	**RE**
50N1D8	TGTAGGCGATCGTCCGCCAATCGGTAGTGGGCGTCAGGGTGGGGCTAACA	50	1.08
50N2C4	GTCAACTTAGAATGAAGTGTCACGTCTGAATTTTTTAATGATAAAGTGGG	50	1.84
50N2E7	CCGACAGGTTCGCTGAGTTCTTATTTGGGACTGCTCTTCAATGATAAAGA	50	2.36
50N2F3	TGTATAAGCCTGGCGGAGACTTGGATTATTTAGTATTGGGTTCAATAGCG	50	5.02
50N2G4	TATGTATCTGGCGTCAGCTCGTATATTCTGTTGACAGTGATCTATATTTT	50	2.39
50N2G8	TGCGAGTCATGGTGTATGGCCTTTAGGCTTCAGGGTGCGGGTCTTCCACT	50	3.4

In figures 3.6.2 and 3.6.3 the improbizer-detected motifs are displayed.
For both sets of sequences (50N and 200N+200NSD) the improbizer was set to look for 3 sequences with a 6 bp initial motif size.
For the 50N samples, the identified motifs were:
CGTCTG furthest upstream, followed by TATTT, and TGATCAA closest to the TnSS. However, only the CGTCTG motif had a score associated with consensus match.
While not fully preserved, the TGATCAA motif was distributed in a small area, indicating that the placement might have some benefit.

In the 200N/200NSD samples, the three motifs identified was: AATGGT the furthest upstream, but widely distributed in the first two thirds of the sequence. The AAT was specially preserved. In the similar upstream region, a CGTTTT motif is also present, but more concentrated around 100 bp. Closest to the TnSS, and also distributed over a smaller area than the two other motifs is a CGGGGGT motif. In this motif, the SD-sequences are also included as a partial match for some of the 200NSD-sequences. This motif is also present many times over in the end of the sequence, with several sequences having 2 or 3 matches. As visible in the weight matrix in table 3.6.3, there is a 35 % chance of the fourth 'G' in the motif and a 27 % chance of the third G being an A-nucleotide. Due to this, several GGAG sequences is also recognized by the improbizer.

To check that the promoter/insert sequences were in fact previously unpublished, all insert sequences were BLAST searched, and they were compared to previously published *B. subtilis* promoters, resulting in no matches.

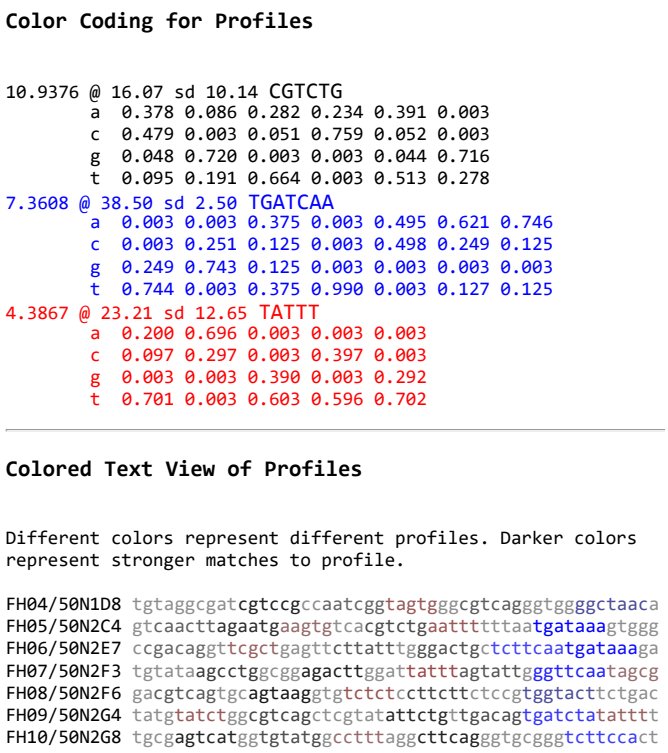


Figure 3.6.2: Reoccurring motifs in the 50N samples found by the improbizer online tool (43). The motifs are scored, with higher scores indicating closer matches to the motif, and higher probability of the motif not being present by chance. Also noted is the placement of the motif with the sd-value for the placement position. For all nucleotides in the motif, a possibility of its occurrence is noted, with a higher value indicating that the nucelotide is more conserved in the motif. A value of 1 means that all occurrences will be the nucleotide, and 0 means none.

Color Coding for Profiles

13.9436 @ 78.47 sd 31.99 CGTTTT

a	0.003	0.290	0.022	0.312	0.357	0.309
c	0.990	0.157	0.003	0.069	0.178	0.112
g	0.003	0.529	0.003	0.003	0.003	0.069
t	0.003	0.024	0.972	0.616	0.462	0.510

13.9934 @ 163.61 sd 26.72 CGGGGGT

a	0.465	0.046	0.090	0.269	0.354	0.003	0.024
c	0.528	0.003	0.003	0.069	0.003	0.024	0.268
g	0.003	0.549	0.903	0.639	0.640	0.577	0.202
t	0.003	0.401	0.003	0.024	0.003	0.396	0.506

12.6208 @ 75.14 sd 45.64 AATGGT

a	0.838	0.710	0.003	0.333	0.397	0.025	
c	0.155	0.025	0.003	0.135	0.046	0.178	
g	0.003	0.003	0.003	0.508	0.553	0.223	
t	0.003	0.262	0.990	0.024	0.003	0.574	

Colored Text View of Profiles

Different colors represent different profiles. Darker colors represent stronger matches to profile.

Figure 3.6.3: Reoccurring motifs in the 200N/200NSD samples found by the improbizer online tool (43). The motifs are scored, with higher scores indicating closer matches to the motif, and higher probability of the motif not being present by chance. Also noted is the placement of the motif with the sd-value for the placement position. For all nucleotides in the motif, a possibility of its occurrence is noted, with a higher value indicating that the nucelotide is more conserved in the motif. A value of 1 means that all occurrences will be the nucleotide, and 0 means none.

Chapter 4

Discussion

4.1 Sequence analysis

The insert sequences were analyzed using several online tools:
BacPP (47) searches for promoters that bind to the *E. coli* sigma factors 24, 28, 32, 38, 54 and 70.
Even though a promoter would be able to bind to a sigma factor, this is not enough to express the gene in *B. subtilis* which is expected to require the SD-sequence, properly spaced, to have successful gene expression (29, 30). So a high likelihood of sigma recognition does not automatically mean a high level of gene expression, as is evident in the measured fluorescence of the samples since all 200N/200NSD samples had high percentages of recognition, but a highly variable relative expression value.

Under normal conditions, σ^{70}/σ^{A} was expected to be contributing to most gene expression under normal conditions, which was why the BPROM tool was used to find possible promoters.
Based on existing knowledge, it is to be expected that clones with promoters closer to consensus, and with a Shine-Dalgarno sequence, would have a higher expression in *B. subtilis* (20, 29). The samples that had the highest consistent gene expression did also, in fact, have at least partial matches to -10 and -35 elements, as well as SD-sequences. The BPROM tool did also identify promoter regions in all the 200N/200NSD samples, but none of the 50N-samples. This could be due to the lower number of nucleotides in the 50N samples not working with the BPROM algorithm. Currently available online are only tools that search for *E. coli* motifs, which might contribute to some *B. subtilis* motifs being overlooked. There is, however, especially in the σ^{70} related sigma factors, extensive homology between *E. coli* and *B. subtilis*

In this dataset, two clones have consistently high expression: the clone SD1D4 and SD2-F10-OLD. It would be expected that they should both have an intact SD-sequence, and be relatively close to consensus for one or more bacterial sigma factors, probably σ^{70}. They

do both, by design, have an SD-sequence, with a 7 bp spacer to the TnSS, that should ensure efficient translation. They are also both high scoring for the -10 element for the BPROM promoter recognition, among the highest. Sample SD1D4 has a BacPP prediction of 50% and above for all sigma factors, as seen in 3.6.1, which could indicate that several sigma factors contribute to the expression.
This could allow the promoter to enable transcription under many conditions.

Interestingly, the strongest promoter 200NSD2-F10-OLD, does not score as high on the BacPP promoter recognition for other sigma-factors than σ^{70}. However, since the experiments were done at 37 °C, with little stressing, and no induction of any other promoters than the two positive control wells, the *B. subtilis* σ^{A} is probably contributing the most to the expression.

It is accepted that *B. subtilis* should not have expression without an SD-sequence (29, 48). The data partially support this, with the 200NSD samples generally having the most consistent expression as viewed in table 3.5.2. However, several of the 200N and 50N samples also had high expression, though with higher variability between parallels, without having clearly identified SD-sequences. This suggests that some binding occurs, even in the absence of the core SD-sequence, with only one sample, 200N2F5 having an SD-sequence with the expected spacing between the SD-sequence and the TnSS (30).

Only one of the 50N samples, 50N2F3 had the GGAG motif, but 32 bp upstream of the TnSS, at which distance the ribosome should not be able to translate (30). Two of the samples, 50N1D8 and 50N2G8 did have a GGTG motif within the distance of the TnSS where it could give some translational benefit, and this might explain the observed gene expression. Sample 50N1D8 did also have a GGGG motif 6 bp upstream of the TnSS. Even though the absence of an SD-sequence should disable translation, some translation might still occur, at low levels. For the 50N samples, no TrSS could be identified, making it hard to identify whether or not these motifs are present in the $5'$ UTR, or if they are upstream from the TrSS, and are not transcribed. If they were not transcribed, they should affect translation.

The BacPP promoter prediction did not recognize any known sigma factor binding sites in sample 50N2G8, despite the observed fluorescence. However, since BacPP recognizes *E. coli* sequences, a *B. subtilis* sigma factor could perhaps recognize the sequence. The absence of an SD-sequence, except for a GGAG motif 35 bp upstream of the TnSS, which would be too far to facilitate translation (30).
This and the other 50N samples would nonetheless be interesting to investigate, to search for known or unknown motifs and binding sites that the online tools did not identify.

Using the Improbizer tool to identify motifs in the 50N samples revealed the motifs in Figure 3.6.2. The elements did, however, score low, which reduces their significance. The CGTCTG motif probability matrix reveals there is a relatively high probability of the first and sixth nucleotide being exchanged for A, giving the motif AGTCAG. None of these motifs match any of the known sigma factor consensus sequences.
The sample size should be bigger to correctly identify reoccurring motifs.

The sample size for the 200N/200NSD samples was larger, perhaps indicating more substantial reoccurring motifs. The probability matrices reveal a large variation between nucleotides here as well. The CCGGGGGT motif however, is clearly identifiea the SD-sequences, as visible in figure 3.6.3. The fact that this motif is also recognized in the 200N samples, at similar locations, could indicate partial SD-matches.
The other two improbizer-recognized motifs have some homology with the σ^{70} -10 element, and recognizes some of the same sequence segments as the BPROM tool.

While examining the sequence for known or unknown motifs is important for describing the promoter sequences, the GeneEE method does not require this step to be efficient. If the only goal is to create a strong promoter for a given product in a given organism, the method could simply be applied, and the output measured. If promoters of satisfactory strength are found, they can then be sequenced and checked to make sure it is indeed a previously undescribed promoter. The promoter sequences could then also be further tweaked using normal synthetic biology methods for tuning of gene expression.
While not used in all organisms, the GeneEE method has been demonstrated in multiple microorganisms, including gram-positives, gram-negatives and yeast (35). Therefore, if a new organism was to be investigated, this method could be a helpful supplement to systems biology measures in identifying σ factors and other genomic motifs in organisms that have not had its entire genome sequenced.

Lale et al. showed that the GeneEE inserts were not only capable of producing constitutive promoters but also inducible promoters, so treating the GeneEE insert clones with inducers could perhaps result in a different pattern of expression, which could be interesting to try in later experiments.

Some samples indicate that just a high consensus match does not automatically lead to high expression. An example is the sample 200NSD2G3, which does not have a high expression, except for in the initial reading. It does, however, have some of the higher BPROM-scores. What sets it apart from the strongest expressing sample is on the surface only a slightly lower consensus match and a lack of the -16 'TG' motif, with similar $5'$ UTR lengths and intact SD-sequences.
Sample 200NSD2E2 had a mutated SD-sequence, with the sequence GGAA instead of GGAG. Since *B. subtilis* is known for its reliance on a conserved SD-sequence (29), this might very well explain the lower expression level compared to other SD-samples.

It is possible that this, and other samples without consensus SD-sequences are actually transcribing at much higher rates than the cells that have the highest expression, but that the transcripts are being quickly degraded, or just not translated, due to them lacking a proper ribosome binding site (29).
It would also be interesting to have data on transcription rates for the samples, and compare the samples with intact SD-sequences to those without.
Some samples had multiple SD-sequences, like sample 200NSD2C3. This sample had a low expression, so the multiple SD-sequences might interfere with the proper initiation of translation.

Of the 200N-insert clones, the expression was relatively equal, with the highest expression in 200N1F7. This sample scores the highest of all the 200N samples for the −10 element but does not have the −16 'TG'-motif, and no Shine-Dalgarno sequence. It does, however, have an SD-like motif downstream from the TrSS, with the partial match GGAT, which might partially facilitate translation. At 11 bp upstream of the TnSS it should still be able to facilitate translation if ribosome binding occurs.
Overall, the 200N-samples had the lowest average RE-values, and also low reproducibility.

For the 50N-samples, the improbizer motif matcher had to be used, since the BPROM promoter predictor did not recognize them as promoters. This also means that they are not scored in the same way as the 200N samples.
The 50N Samples had high variance between samples and low reproducibility, and interestingly the average RE-values do not seem to have a clear connection between avg. RE-values and predicted sigma factor binding. This can be due to how the tools handle shorter sequences, or it could be an indication of its usefulness in identifying short *B. subtilis* promoters.
However, the high variability of the data makes it hard to come to any conclusions about the actual gene expression. Nonetheless, the 50N insert plasmids could be interesting to investigate further, to find out what contributes to the gene expression.

Of all the samples, it is the two samples that scored the highest RE-values that also had the most of the expected elements, with intact SD-sequences, partial −35 and −10 element matches, and a −16 'TG'-motif. However, several samples deviated from what was expected and still showed some of the higher relative expression values, including samples without SD-sequences, and samples that did not match consensus.

Because of the wide number of sigma factors predicted to be recognized, with more time and resources available, it would be fascinating to see how these same promoters perform under different conditions.
It should be noted that under other conditions, other clones than the ones picked here could respond differently, and the picked colonies might not be the ideal candidates since these are only the ones expressing strongest under the given conditions in this experiment: 37 °C overnight incubation without induction.
It would also be interesting to analyze the transcription in the samples with only partial SD-sequence matches, compared to the two strongest promoters, to see if the observed difference in gene expression is both at the level of transcription and translation.

4.2 Future work

The methods outlined in this thesis have produced constitutive promoters for *B. subtilis* that produces fluorescence, in all probability from mCherry, at a higher than non-induced control level, from random gene inserts, though with some samples having an SD-sequence inserted by design.

This is, as far as the author is aware, the first published example of this being done in *B. subtilis*, and opens the door to much fascinating research.
There is much more work that can and should be done, using the experimental methods described here. A variety of *B. subtilis* promoters that perform at a high rate using several different induction methods and under several environmental conditions should be created. Also, the promoters described here could be further investigated, to see if some unknown transcription or translation elements are contributing to the gene expression.

If the only goal is to produce stronger generic promoters for *B. subtilis*, using only the 200N SD GeneEE inserts would be recommended, since these samples generally expressed most strongly, and since *B. subtilis* is believed to require an SD-sequence for translation. Experiments can be done with longer SD-sequences since these will have an even closer anti-SD match.
However, for characterization of new, unknown transcriptional elements, using a 200N or 50N insert could give exciting results if strong gene expression could be found in the absence of SD-sequences.

If new strong promoters are indeed discovered, they should be compared to the popular constitutive promoters P_{veg} and P_{43}.
Analysis of transcription levels of both the stronger promoters presented here, and any future promoters, could be interesting, and should perhaps be included in further studies to have a more robust set of data and not to overlook potentially highly transcribed mRNAs that are not being translated.

If development of new *B. subtilis* promoters is a priority in the scientific community, more software tools should be created and made available to predict promoter sequences, not only in *E. coli*, but also in *B. subtilis*.

The true power of the method, however, could lie in the production of promoters for genes encoding for industrially important bio-products.
B. subtilis is among the most important bacteria for the production of bio-products, contributing to much of the bacterial production of bio-products and recombinant proteins (4). Using this method, custom promoters could be created that are specifically designed for expression of the product in question, under environmental conditions that simulate the factory-conditions where the proteins are made. A targeted approach could be used, with the strongest commercially available promoters as a control. Since these promoters are mostly generic promoters, that are not tailored to any gene in specific, the targeted approach that the GeneEE-method allows could potentially result in promoters that consistently outperform the ones that are commercially available today.
While this might require more lab work than just using a strong generic promoter, if it could increase yields, even by a little, the cost of the lab work could be quickly recuperated and could potentially create cleaner products that require less biomass, also giving some potential environmental benefits, as less waste is formed.

4.3 Conclusion

The goal of this master thesis was to present and demonstrate a method of generating novel constitutive promoters for use in *B. subtilis* using the GeneEE random insert method.
21 promoters were identified, with several having consistently higher than background levels of expression, indicating that they are indeed functional.
The promoters found here have not been previously described or characterized, and are therefore new *B. subtilis* promoters.
Also, several of the sequences described here could be studied further, possibly revealing new information about gene regulation in *B. subtilis*.

This thesis provides evidence that the GeneEE method can be used to create functional promoters in *B. subtilis*, and sets the stage for further experiments and possible scientific and industrial applications.

Bibliography

[1] Setlow P. Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals. *Journal of Applied Microbiology*. 2006;101(3):514–525.

[2] Schumann W. Production of Recombinant Proteins in *Bacillus subtilis*. *Advances in Applied Microbiology*. 2007 jan;62:137–189.

[3] Ila S. *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends in Biotechnology*. 1992;10(Vol 10):247–256.

[4] Gu Y, et al. Advances and prospects of *Bacillus subtilis* cellular factories: From rational design to industrial applications. *Metabolic engineering*. 2018;50:109–121.

[5] Burdett IDJ, et al. Growth kinetics of individual *Bacillus subtilis* cells and correlation with nucleoid extension. *Journal of Bacteriology*. 1986;167(1):219–230.

[6] Piggot PJ. *Bacillus subtilis*. *Encyclopedia of Microbiology*. 2009;p. 45–56.

[7] Earl AM, et al. Ecology and genomics of *Bacillus subtilis*. *Trends in microbiology*. 2008;16(6):269–275.

[8] Piewngam P, et al. Pathogen elimination by probiotic Bacillus via signalling interference. *Nature*. 2018;562(7728):532–537.

[9] Zeigler DR, et al. The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *Journal of Bacteriology*. 2008;190(21):6983–6995.

[10] Borchert S, et al. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*. 1997;390(6657):249–256.

[11] Becher D, et al. From the genome sequence to the protein inventory of *Bacillus subtilis*. *Proteomics*. 2011;11(15):2971–2980.

[12] Nicolas P, et al. Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*. *Science (New York, NY)*. 2012;335:1103–1106.

[13] Westers L, et al. Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 2004;1694(1):299 – 310. Protein Export/Secretion in Bacteria.

[14] Harwood CR, et al. Bacillus protein secretion: an unfolding story. *Trends in Microbiology*. 2008;16(2):73–79.

[15] Altenbuchner J. Editing of the *Bacillus subtilis* Genome by the CRISPR-Cas9 System. *Applied and Environmental Microbiology*. 2016;82(17):5421–5427.

[16] Hamoen LW, et al. Controlling competence in *Bacillus subtilis*: Shared use of regulators. *Microbiology*. 2003;149(1):9–17.

[17] Nelson DL. *Lehninger principles of biochemistry*. 6th ed. New York: W.H. Freeman; 2012. :1156–1157.

[18] Clark D, et al. *Molecular biology*. 2nd ed. Oxford: Elsevier Academic Press; 2013. :312–313.

[19] Nair TM, et al. On the consensus structure within the E. coli promoters. *Biophysical Chemistry*. 1994;48(3):383–393.

[20] Liu D, et al. Construction, Model-Based Analysis, and Characterization of a Promoter Library for Fine-Tuned Gene Expression in Bacillus subtilis. *ACS Synthetic Biology*. 2018;7(7):1785–1797.

[21] Paget MS. Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. *Biomolecules*. 2015;5(3):1245–1265.

[22] Yang S, et al. Characterization and application of endogenous phase-dependent promoters in *Bacillus subtilis*. *Applied Microbiology and Biotechnology*. 2017;101(10):4151–4161.

[23] Jia J, et al. Three tandem promoters, together with IHF, regulate growth phase dependent expression of the Escherichia coli kps capsule gene cluster. *Scientific Reports*. 2017;7(1):1–11.

[24] Haldenwang WG. The sigma factors of *Bacillus subtilis*. *Microbiological reviews*. 1995;59(1):1–30.

[25] Danson AE, et al. Mechanisms of σ 54 -Dependent Transcription Initiation and Regulation. *Journal of Molecular Biology*. 2019;.

[26] Coelho RV, et al. *Bacillus subtilis* promoter sequences data set for promoter prediction in Gram-positive bacteria. *Data in Brief*. 2018;19:264–270.

[27] Barrios H, et al. Compilation and analysis of σ54-dependent promoter sequences. *Nucleic Acids Research*. 1999;27(22):4305–4313.

[28] Mars RAT, et al. Regulatory RNAs in *Bacillus subtilis*: a Gram-Positive Perspective on Bacterial RNA-Mediated Regulation of Gene Expression. *Microbiology and Molecular Biology Reviews*. 2016;80(4):1029–1057.

[29] Band L, et al. *Bacillus subtilis* requires a "stringent" Shine-Dalgarno region for gene expression. *DNA (Mary Ann Liebert, Inc)*. 1984;3(1):17–21.

[30] Chen H, et al. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. . 1994;22(23):4953–4957.

[31] Yang S, et al. Construction and Characterization of Broad-Spectrum Promoters for Synthetic Biology. *ACS Synthetic Biology*. 2018;7(1):287–291.

[32] Ishii T. DBTBS: a database of *Bacillus subtilis* promoters and transcription factors. *Nucleic Acids Research*. 2001;29(1):278–280.

[33] IGEM. Registry of standard biological parts - *Bacillus subtilis*; 2019. Accessed: 2019-07-19. Available from: `http://parts.igem.org/Bacillus{_}subtilis`.

[34] Li M, et al. A strategy of gene overexpression based on tandem repetitive promoters in Escherichia coli. *Microbial Cell Factories*. 2012;11:1–10.

[35] Lale R, et al. A universal method for gene expression engineering. *bioRxiv*. 2019;.

[36] OpenWetWare. Synthetic Biology:BioBricks/Part fabrication — OpenWetWare{,};. Accessed: 8-July-2019. Available from: `https://openwetware.org/mediawiki/index.php?title=Synthetic{_}Biology:BioBricks/Part{_}fabrication{&}oldid=466093`.

[37] Engler C, et al. A One Pot, One Step, Precision Cloning Method with High Throughput Capability (High Throughput Cloning Method). *PLoS ONE*. 2008;3(11).

[38] MoBiTec GmbH. *Bacillus subtilis* expression system; 2019. Accessed: 2019-07-26. Available from: `https://www.mobitec.com/cms/products/bio/04{_}vector{_}sys/bacillus{_}subtilis{_}expression.html`.

[39] Shaner NC, et al. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology*. 2004;22(12):1567–1572.

[40] Benchling. Benchling[Biology Software]; 2019. Available from: `http://benchling.com`.

[41] de Avila e Silva S, et al. BacPP: Bacterial promoter prediction-A tool for accurate sigma-factor specific assignment in enterobacteria. *Journal of Theoretical Biology*. 2011;287(1):92–99.

[42] Salamov VSA, et al. Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture*;.

[43] Ao W, et al. Environmentally Induced Foregut Remodeling by PHA-4/FoxA and DAF-12/NHR. *Science*. 2004;305(5691):1743–1746.

[44] Green R, et al. Chapter Twenty Eight - Transformation of Chemically Competent E. coli. In: Lorsch J, editor. *Laboratory Methods in Enzymology: DNA*. vol. 529 of Methods in Enzymology. Academic Press; 2013. p. 329 – 336.

[45] Hintermann G, et al. Simple procedure for distinguishing CCC, OC, and L forms of plasmid DNA by agarose gel electrophoresis. *Plasmid*. 1981;5(3):371–373.

[46] Potapov V, et al. Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synthetic Biology*. 2018;7(11):2665–2674.

[47] de Avila e Silva S, et al. BacPP: Bacterial promoter prediction-A tool for accurate sigma-factor specific assignment in enterobacteria. *Journal of Theoretical Biology*. 2011;287(1):92–99.

[48] Abolbaghaei A, et al. How Changes in Anti-SD Sequences Would Affect SD Sequences in Escherichia coli and *Bacillus subtilis* . *G3: Genes—Genomes—Genetics*. 2017;7(5):1607–1615.

[49] McCabe S. QIAquick PCR Purification Kit Protocol. *Benchmarking in Construction*. 2001;(July):277–283.

[50] QIAgen. QIAprep Spin Miniprep Kit QuickStart Protocol; 2012. Available from: `https://www.qiagen.com/us/resources/download.aspx?id=56b0162c-23b0-473c-9229-12e8b5c8d590&lang=en`.

Appendix A

Additional protocols

A.1 PCR purification protocol

For purification of PCR products, the Qiagen QIAquick PCR Purification Kit was used. for more info see the Qiagen website: http://www.qiagen.com.
The method shown is the one used for this thesis, based on the quick-start guide associated with the kit (49).
All centrifugation steps carried out at 13 000 RPM in a table top centrifuge. Several Qiagen buffers were used, listed in table A.1.1.
To bind DNA, QIAquick columns were used.

Table A.1.1: Buffers used for PCR purification

Buffer	Description
PB	Binding buffer
PE	Wash buffer
EB	Wash buffer

Method:
Transfer PCR reaction to a microcentrifuge tube
Add 5 volumes PB buffer to 1 volume of the PCR reaction, mix.
Place QIAquick column in provided collection tube.
Add sample to column, centrifuge for 30 sec and discard flow-through and put the column back in the tube.
Add 750 µL PE buffer to column. Centrifuge for 30 sec and discard flow-through, and put column back in the tube.
centrifuge tube for 1 minute to remove the rest of the wash buffer.
Place the column in a sterile 1.5 mL microcentrifuge tube.
Elute by adding 30 µL elution buffer to the column. Let stand 1 minute and centrifuge for

1 minute. Discard column.
Measure concentration using a Thermo scientific NanoDrop One Microvolume UV-Vis Spectrophotometer.

A.2 MiniPrep protocol

For isolation of plasmids from *E. coli*, the Qiagen QIAprep Spin Miniprep Kit was used. for more info see the Qiagen website: http://www.qiagen.com.
The method shown is the one used for this thesis, based on the quick-start guide associated with the kit (50).
All centrifugation steps carried out at 13 000 RPM, unless otherwise stated, in a table top centrifuge. Several Qiagen buffers were used, listed in table A.2.1.
To bind DNA, Qiagen QIAprep 2.0 spin columns were used.

Table A.2.1: Buffers used for MiniPrep plasmid isolation

Buffer	Description
P1	Resuspension buffer
P2	Lysis buffer
N3	Neutralization buffer
PE	Wash buffer
EB	Elution buffer

Method:
Pellet 1-5 mL overnight culture by centrifugation at 7800 RPM.
Discard supernatant and resuspend pelleted cells in 250 µL buffer P1. Transfer to a 1.5 mL microcentrifuge tube
To tube, add 250 µL buffer P2 and mix by inverting tube 6 times.
To tube, add 350 µL buffer N3 and mix by inverting 6 times.
Centrifuge for 10 min.
Apply 800 µL supernatant to the QiaPrep spin column, without disturbing cell materials.
Centrifuge spin column for 30 sec. discard flow-through
Wash spin column using 750 µL buffer PE. Centrifuge 30 sec, discard flow-through.
Centrifuge spin column for 1 minute to remove the rest of the wash buffer.
Place spin column in a sterile 1.5 mL microcentrifuge tube.
Elute by adding 50 µL buffer EB. Let stand for 1 minute and centrifuge for 1 minute. Discard column.
Measure concentration using a Thermo scientific NanoDrop One Microvolume UV-Vis Spectrophotometer.

Appendix B

Additional DNA sequences

B.1 Primers

All primer sequences used for PCR in this thesis is listed in table B.1.1

Table B.1.1: PCR Primers used in this thesis

Primer name	**Primer sequence** $5' \rightarrow 3'$	**Description**
BsaI-Fwd	AATGATACCGCGGGACCCACGCTCACCGGC	BsaI-site deletion in ampicillin resistance gene
BsaI-Rev	GGGTCCCGCGGTATCATTGCAGCACTGGGG	BsaI-site deletion in ampicillin resistance gene
BB-Prefix-Fwd	GAATTCGCGGCCGCTTCTAGAG	Amplification of GeneEE inserts
BB-Suffix-Rev	CTGCAGCGGCCGCTACTAGTA	Amplification of GeneEE inserts
517-HT10-mCh-fw1	GCTACTGAGGTCTCGNATGGTTT CTAAAGGT-GAAGAAGAC	Creation of pHT10-mCherry-ΔBsaI-BB backbone plasmid.
518-HT10-mCh-Rv1	GATGCGGTCTCGGGCACAACTTAATCGCCTTGCAG	Creation of pHT10-mCherry-ΔBsaI-BB backbone plasmid.
521-seq-200N-mCh-fw	TTCGTTTGTTGAACTAATGGGTGC	For colony PCR using *Taq* PCR and sequencing using Q5
522-seq-mCh-rv	TGCCTTCATACGGACGACCTTC	For colony PCR using *Taq* PCR andsequencing using Q5

B.2 GeneEE inserts

The template for the PCR-amplification of the GeneEE inserts are shown in figure B.2.1. The pre-and suffixes are similarly design, so the same primers could be used for library amplification (primer BB-Prefix-Fwd/BB-Suffix-Rev)

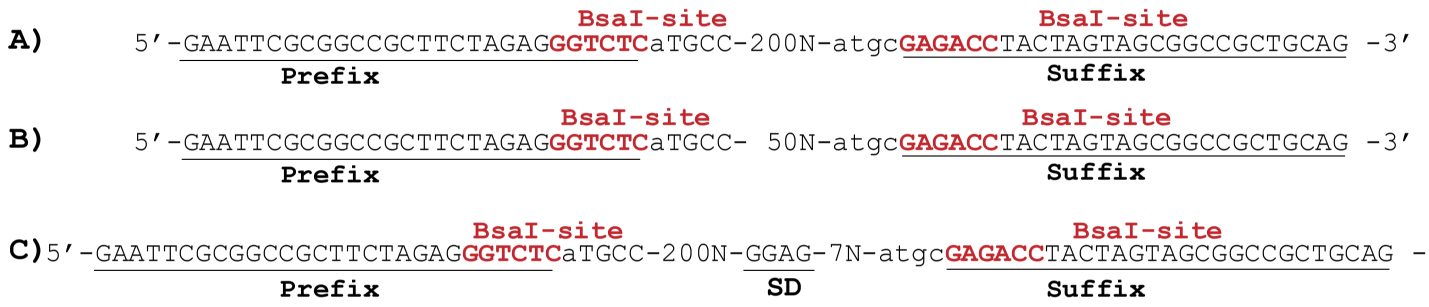


Figure B.2.1: GeneEE insert PCR templates. A) 200N insert, B) 50N insert, C) 200N SD insert:Shine-Dalgarno sequence + 7N inserted after the 200N random DNA stretch.

B.3 DNA Ladder

For all agarose gel electrophoresis in this the Thermo Scientific^{TM} GeneRuler 1 kb DNA Ladder was used.
The size of the ladder parts are shown in figure B.3.1

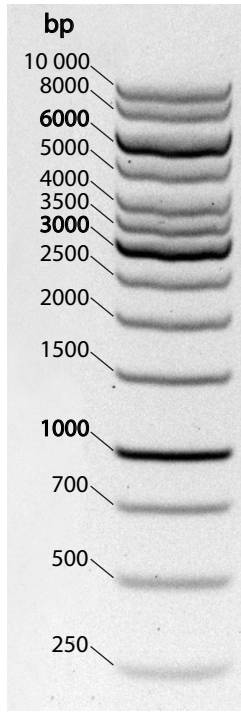


Figure B.3.1: Thermo Scientific ^{TM} GeneRuler 1 kb DNA Ladder. Image from the author, sizes based on information from Thermo Scientific.

Appendix C

Additional Data

C.1 *E. coli* Golden Gate cloning plate image

Presented in figure C.1.1 is a LA-plate with *E. coli* clones, transformed with pHT10-mCherry-ΔBsaI-200NSD random promoter library plasmid. The image has been color-corrected to emphasize the difference between white and red colonies. No color has been added. Circles mark some red colonies.

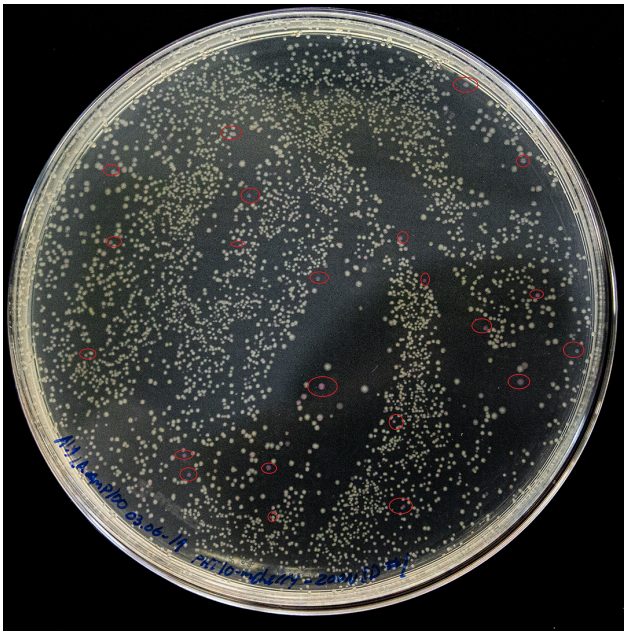


Figure C.1.1: Image of LA-plate with *E. coli* clones, transformed with pHT10-mCherry-ΔBsaI-200NSD random promoter library plasmid

C.2 Tecan-plate reader data

The RE-values from initial Tecan readings are displayed in figures C.2.1 and C.2.2. The values of picked wells are marked in bold.

50N 1:

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	0.39	0.53	BLANK	1.37	0.67	0.52	1.48	197.66	168.50		
B	0.25	0.92	0.93	0.85	0.37	0.46	2.04	1.18	1.64	1.38	0.35	-0.11
C	0.31	0.24	0.67	0.57	0.28	1.71	1.70	0.43	1.56	2.15	0.00	0.00
D	0.41	0.38	1.12	0.68	-0.13	0.43	0.13	**2.67**	0.00	1.37	0.34	1.08
E	0.34	0.46	0.00	0.35	0.78	0.72	0.00	0.46	1.33	0.24	0.22	0.22
F	0.74	0.37	1.97	0.54	0.76	0.59	1.62	0.58	0.00	0.63	0.13	1.96
G	0.88	0.67	1.39	1.59	0.38	0.60	0.00	0.94	0.25	0.69	0.25	-0.12
H	0.10	0.00	0.00	0.80	0.32	0.30	0.00	0.10	0.41	0.00	-0.44	0.00

50N 2:

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	1.55	1.43	BLANK	1.12	1.35	1.00	1.00	152.04	150.49		
B	1.40	0.89	1.71	1.75	0.36	1.09	1.30	0.51	0.82	1.23	0.39	0.63
C	1.35	0.50	1.60	**3.49**	0.52	0.64	1.22	2.25	0.61	2.23	0.28	0.31
D	1.01	1.27	0.83	1.43	1.70	1.37	1.63	1.15	1.55	0.97	0.87	0.64
E	0.87	0.66	0.74	0.51	1.98	0.74	**5.02**	0.66	2.12	1.12	0.10	0.74
F	1.80	2.24	**3.59**	2.17	0.67	**3.52**	0.24	0.38	0.56	0.23	1.39	0.87
G	0.39	2.42	0.73	**4.29**	0.49	1.85	0.32	**5.98**	0.61	0.60	1.29	0.54
H	0.45	1.73	1.54	0.92	1.51	1.49	0.53	2.11	1.67	0.28	1.17	0.70

Figure C.2.1: Tecan plate readings of pHT10-mCherry-ΔBsaI-50N plasmid in *B. subtilis 168*
Lane 'A' are control clones. A1 and A4: blanks of LB-medium, A2-3: *B. subtilis* 168 untransformed cells, A5-6: pHT10-mCherry-ΔBsaI-BB closed backbone plasmid without inserts, A7-8:pHT10-mCherry-ΔBsaI, uninduced, A9-10:pHT10-mCherry-ΔBsaI induced with 1:1000 IPTG.
Numbers are relative fluorescence pr. cell relative to the average fluorescence pr. cell of uninduced pHT10-mCherry-ΔBsaI.
Numbers in **bold** represent clones that were picked for sequencing

200N 1:

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	1.31	1.19	BLANK	1.04	0.75	1.12	0.88	93.30	95.89		
B	0.95	1.62	1.10	0.63	0.78	2.16	2.08	1.26	1.19	0.82	0.77	0.52
C	0.66	0.68	1.10	0.89	0.87	0.89	0.77	0.55	1.34	0.65	1.16	1.03
D	1.42	0.87	1.50	0.82	1.91	1.74	0.79	1.12	0.63	1.63	0.71	0.64
E	0.80	1.38	1.73	1.77	1.18	1.47	1.82	0.51	1.08	0.74	0.69	0.68
F	0.64	1.06	1.63	0.66	1.61	2.11	**2.38**	0.86	1.81	2.12	0.91	1.12
G	0.83	0.87	1.91	**2.27**	1.18	1.43	0.63	1.95	1.49	1.11	0.71	0.73
H	0.53	0.97	0.58	1.11	1.54	1.78	0.81	1.19	1.99	1.06	0.76	0.53

200N 2:

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	0.64	0.18	BLANK	0.79	0.33	1.00	1.00	137.05	83.31		
B	0.18	1.30	0.27	0.46	-0.17	0.39	1.77	0.43	0.16	0.35	0.45	0.47
C	0.49	0.41	1.37	0.87	1.25	1.53	0.97	0.35	0.76	0.43	0.38	0.57
D	0.41	1.82	1.21	0.34	0.51	0.16	0.18	2.29	0.23	0.64	0.09	0.18
E	0.29	0.73	1.05	0.82	0.79	0.50	1.13	2.11	0.00	0.15	0.35	0.27
F	0.29	0.00	0.51	1.93	**2.42**	**2.48**	1.78	1.50	1.37	2.20	0.74	0.51
G	0.16	0.79	0.00	1.55	0.69	1.69	0.08	0.83	0.07	0.62	-0.27	0.17
H	0.31	0.43	0.55	0.49	0.46	1.50	-48.50	1.87	0.69	0.44	0.17	0.27

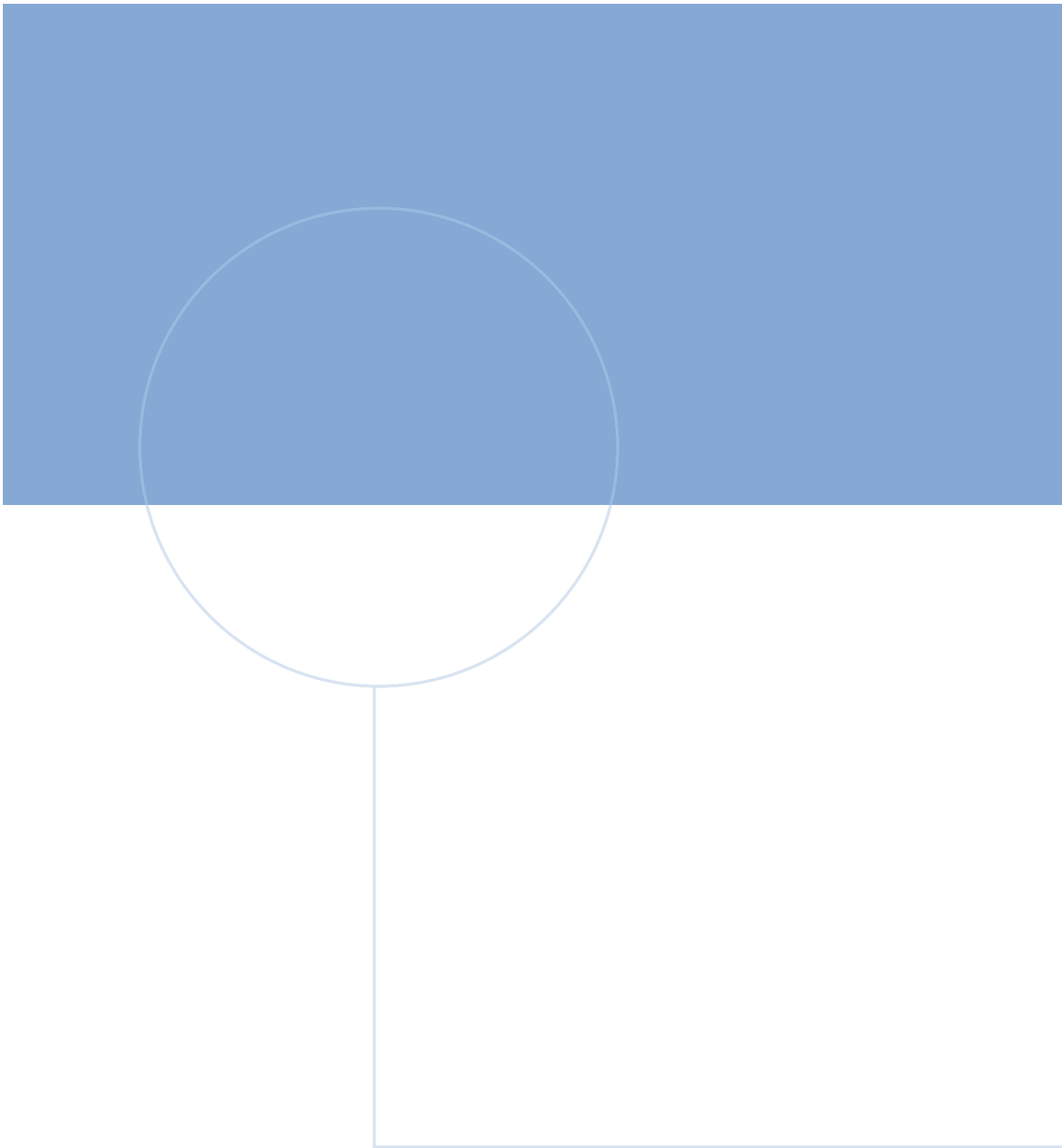
200N SD 1:

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	0.92	0.79	BLANK	0.96	0.92	1.01	0.99	79.87	74.53		
B	0.68	0.55	1.15	0.65	1.25	0.75	1.04	0.70	0.23	0.62	0.59	0.95
C	0.48	1.47	1.56	0.28	0.19	0.82	0.56	0.00	0.58	0.38	0.60	0.45
D	0.57	0.79	2.03	**5.43**	0.68	0.47	0.64	0.49	0.32	0.90	0.70	0.56
E	0.83	0.18	0.47	0.45	2.08	0.78	0.30	0.58	0.46	0.52	0.63	0.58
F	0.36	0.97	0.97	1.07	0.78	0.87	0.80	0.62	1.05	0.56	0.96	0.86
G	0.76	0.83	0.69	0.52	0.86	0.47	0.86	0.47	0.64	0.86	0.61	0.72
H	0.80	0.67	0.56	0.86	1.15	0.89	0.99	0.86	0.86	0.67	0.82	1.06

200N SD 2:

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	1.73	0.73	BLANK	0.74	0.59	0.74	1.26	145.73	177.25		
B	0.77	0.50	0.69	0.81	1.02	1.32	0.74	0.28	0.65	0.83	0.75	0.17
C	2.74	1.34	**3.83**	0.95	1.24	1.01	0.91	0.55	1.44	0.46	1.29	0.83
D	**3.32**	2.39	1.37	0.71	0.73	1.07	1.51	1.62	0.58	0.74	0.34	0.79
E	1.75	**3.48**	0.72	2.22	2.08	1.56	0.50	**2.63**	0.45	0.38	1.19	0.56
F	0.33	**2.91**	**3.31**	1.28	1.22	0.76	**2.77**	**3.53**	1.61	1.17	0.21	1.53
G	1.87	0.56	**3.00**	1.46	1.61	1.92	0.66	1.25	1.02	0.22	1.61	0.37
H	2.46	0.10	1.91	1.61	0.74	0.22	1.71	0.42	2.03	1.88	0.79	0.75

Figure C.2.2: Tecan plate readings of pHT10-mCherry-ΔBsaI-200N and pHT10-mCherry-ΔBsaI-200NSD plasmid in *B. subtilis 168*
Lane 'A' are control clones. A1 and A4: blanks of LB-medium, A2-3: *B. subtilis* 168 untransformed cells, A5-6: pHT10-mCherry-ΔBsaI-BB closed backbone plasmid without inserts, A7-8:pHT10-mCherry-ΔBsaI, uninduced, A9-10:pHT10-mCherry-ΔBsaI induced with 1:1000 IPTG.
Numbers are relative fluorescence pr. cell relative to the average fluorescence pr. cell of uninduced pHT10-mCherry-ΔBsaI.
Numbers in **bold** represent clones that were picked for sequencing



NTNU
Norwegian University of
Science and Technology