מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

עבודת גמר (תזה) לתואר
מוסמך למדעים

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

מאת
איתי ורקוביצקי

חקר האבולוציה של מנגנוני בקרה גנטיים באופרון הטריפטופאן

Thesis for the degree
Master of Science

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

By
Itay Varkovitzky

Studying the Evolution of Genetic-Regulation Mechanisms in the Tryptophan Operon

מנחה:
פרופ' יצחק פלפל

Advisor:
Prof. Yitzhak Pilpel

חודש ושנה עבריים
שבט, תשפ"ג

Month and Year
May, 2023

Abstract

Gene regulation mechanisms enable organisms to adapt to changing external and internal conditions and shape evolution in many ways. This study investigates the tryptophan attenuator, a canonical regulation mechanism in bacteria that controls the expression of the tryptophan operon. Through experimental and computational approaches, we aimed to gain a better understanding of its evolutionary constraints, their functional implications, and the interplay between the two main components of this regulatory device, the transcriptional repressor and the attenuator, a riboswitch that controls transcription termination during translation. Multiple sequence alignment analysis reveals the conservation of key attenuator features across Gram-negative bacteria, while certain regions show variability among bacterial families and species. I have then designed and synthesized 10 different alternative versions of the attenuator and cloned each of them into *E. coli* that were grown under different regimes of exposure to tryptophan. Among them, a version that was designed to harm the anti-termination activity via decoupling of the anti-terminator stem, a version that was designed to have exactly the same behavior in all regimes with a different sequence with conservation of the RNA's secondary structure, and a version that was designed to act in a reverse logic (terminate expression when tryptophan is absent and express when tryptophan is present), through an addition of a new stem to the secondary structure.

My experimental results demonstrate the repressor's step-like function in response to high tryptophan concentrations and attenuator variants' influence on baseline expression levels. With the necessary additional experimentation, expanded variant library, and fitness assessments through competition essays, we will gain deeper insights into Trp attenuator dynamics under different conditions, and it will be able to continue serving as a key model for adaptive genetic regulation. Overall, this study provides valuable insights into the complexity of biological regulation mechanisms and their fitness landscape, through tryptophan attenuator analysis.

Table of Contents

List of abbreviations

Trp: Tryptophan

AA: Amino Acid

DNA: Deoxyribonucleic Acid

RNA: Ribonucleic Acid

PCR: Polymerase Chain Reaction

FACS: Fluorescence-activated Cell Sorting

GFP: Green Fluorescent Protein

tAI: tRNA Adaptation Index

BP: Base pairs

kb: Kilo Bases

WT: Wild Type

E. coli: *Escherichia coli* bacteria

MSA: Multiple Sequence Alignment

Introduction

Gene regulation plays a central role in shaping the evolution of species by allowing organisms to respond and adapt to changing environments. By controlling the expression of genes, gene regulation provides a mechanism for rapid and efficient responses to environmental cues and enables the evolution of new traits. Evolutionary changes in gene regulation have undoubtedly contributed to the emergence of new species, the adaptation of species to diverse environments, and the development of complex cellular processes.

In this work, we will focus on the regulatory elements of a canonic regulation mechanism in bacteria, the Tryptophan (Trp) operon. We will explore the relationship between the different elements and set the ground towards a better understanding of the processes that govern their evolution.

Using both experimental and comparative genomic approaches, this work will examine mechanisms of gene regulation in bacteria and their contribution to the bacteria's adaptation and survival.

Trp Operon and its regulation

Tryptophan is an essential amino acid that is required for the proper functioning of proteins in all living organisms.
However, the production of tryptophan requires a significant amount of energy and is known to have the highest cost among all amino acids [1] [2] [3], and thus, its synthesis is tightly regulated in order to prevent excess production and waste but ensure to have sufficient amounts. There are two key regulators of the Trp synthesis – a transcriptional repressor and an attenuator.

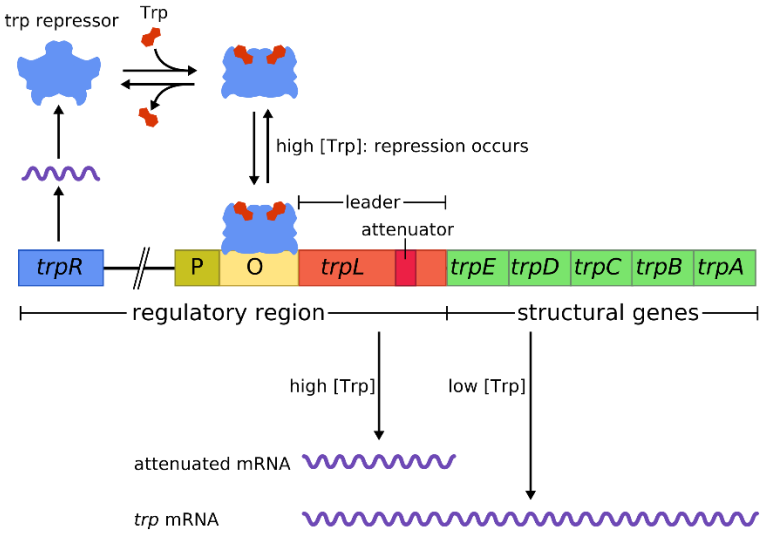


***Figure 1**: Trp Operon regulation scheme [4].*

When the repressor protein binds to the operator site, it blocks the binding of the RNA polymerase, which prevents the transcription of the regulated genes.

The attenuator is a cis-regulatory RNA structure that can fold into two alternative forms, depending on the presence of a regulatory factor. In the case of the Trp operon in *E. coli*, the repressor binding site is located upstream to the Trp operon genes that code for the biosynthesis of Trp, and the attenuator is located downstream to the operator but still upstream to the operon. In the Trp attenuator, the RNA folds into the terminator form when Trp is present, stopping the transcription of the downstream sequence. Therefore, operon genes are not expressed, and the waste of unnecessary resources is avoided. In its alternative form (when Trp is absent), the attenuator takes the Anti-terminator formation, and the downstream operon is transcribed and expressed, so the production of this essential amino acid only happens when it is most needed.

By acting at different stages of gene transcription, the tryptophan attenuator and the repressor provide a two-step regulatory mechanism for controlling tryptophan synthesis. According to the literature [5] [6], the Trp Repressor is altering the expression of the Trp operon over a ca. of 100-fold, while the attenuator has a 6-8 fold effect.

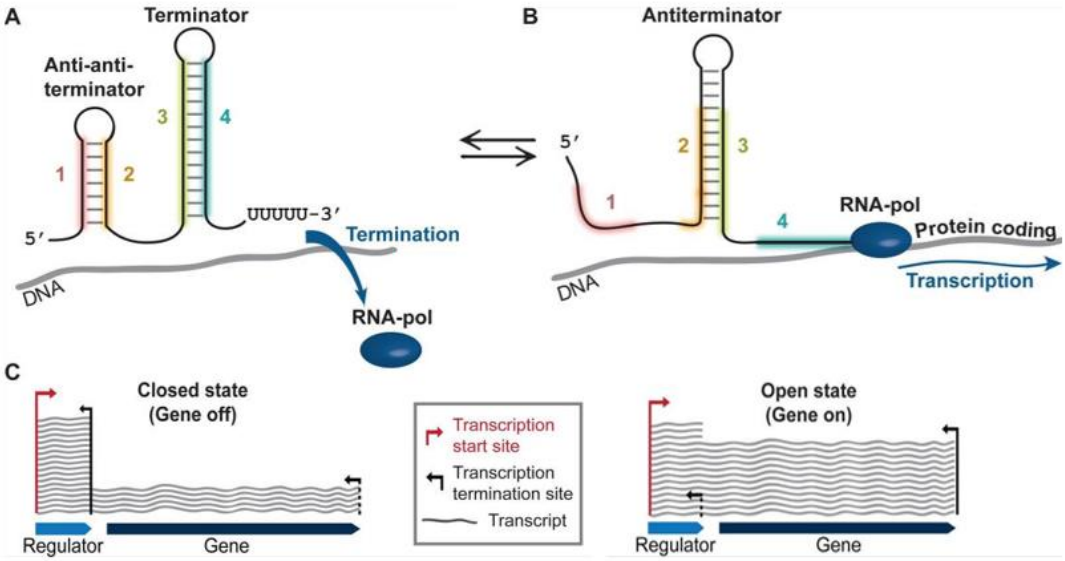


***Figure 2**: Regulation by attenuation – general illustration [7].*

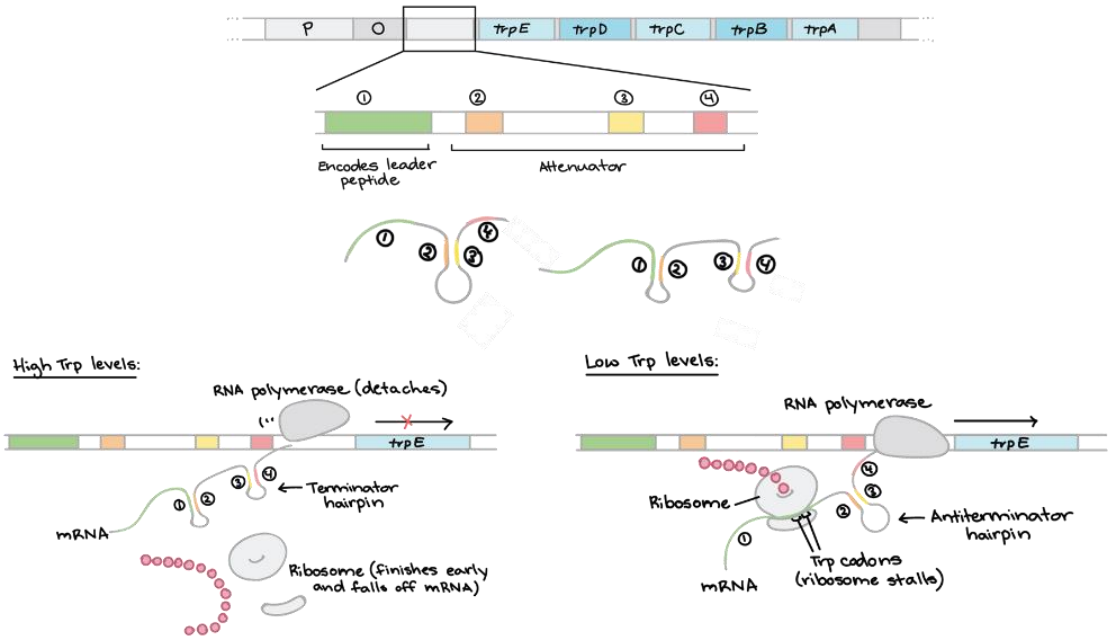


***Figure 3**: Illustration of the Trp Operon attenuation mechanism [8].*

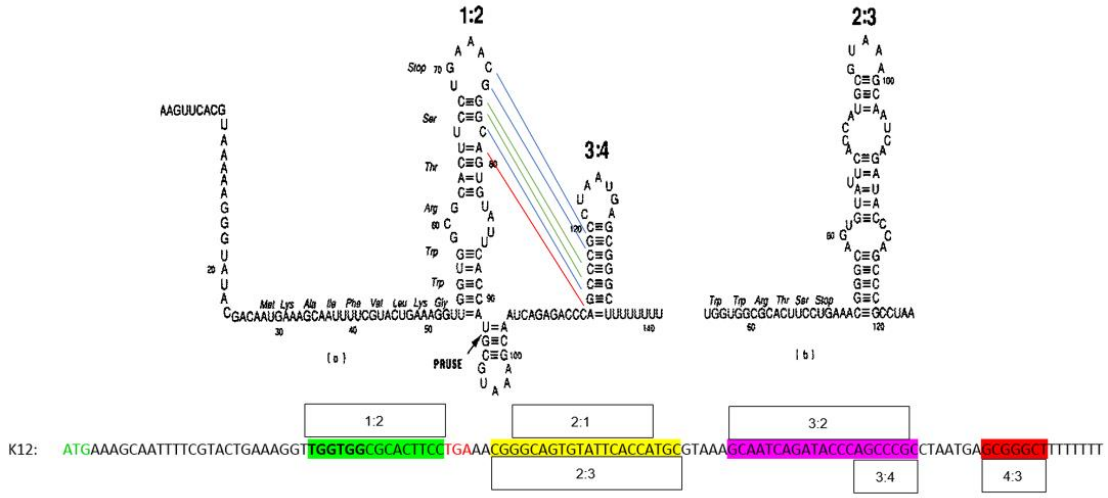


Figure 4*: an Illustration of the alternative formations of the Trp attenuator according to its sequence, adapted from Landick R et al [9]. I added the canonic sequence of the K12 E. coli strain and its corresponding regions to highlight the areas I will focus on later in this work. The lines that connect the 1:2 hairpin with the 3:4 hairpin are demonstrating the base-pairing in the 2:3 formation when a green line represents a base-pair that is part of a "quartet" – our definition of a quartet is a group of nucleotides, in which a region 1 nucleotide can bind a region 2 nucleotide, the region 2 nucleotide can alternatively bind a region 3 nucleotide, and the region 3 nucleotide can alternatively bind a region 4 nucleotide. A blue line represents a triplet, where only 3 bases have such an alternative binding connection. The red line is an example of two couples that do not connect in the 2:3 formation.*

Some of the other AA's biosynthetic operons are also regulated by attenuation (see Table 1). Interestingly, the Trp attenuator of *E. coli* contains only a couple of Trp codons, while others have from 4 to 15 relevant codons. Since Trp levels in cells are known to be very low [7] it is possible that two codons are sensitive enough to serve as sensor in the case of Trp but not in the case of other amino acids that are more prevalent in the cell and thus in these cases you need more codons. It also has the shortest sequence.

Table 1: List of AA's biosynthetic operons controlled by attenuation

Operon name	Att. Length (bp)	Codons & tRNAs	Leader peptide
trp	115	UGG	MKAIFVLKGWWRTS stop
thr	127	ACU, ACC, ACA, ACG	MKRISTTITTTITITTGNGAG stop
Ilv (GEDA)	155	13 different codons	MTALLRVISLVVISVVVIIIPPCGAALGRGKA stop
leu	~120	UUA, UUG, CU**X**	MSHIVRFTGLLLLNAFIVRGRPVGGIQH stop
phe	112	UUU, UUC	MKHIPFFFAFFFTFP stop
his	124	CAU, CAC	MTRVQFKHHHHHHHPD stop

In the table, we listed amino acids (or groups of AAs) with a known attenuator-controlled biosynthetic pathway. The first column represents the length of the attenuator in E. coli, the second column lists the corresponding codons, and the third column shows the AA sequence of the relevant leader peptide.

The current understanding of the tryptophan attenuator and its role in the regulation of tryptophan biosynthesis reflects the contribution of many scientists over the years, mostly during the 70s, 80s, and the 90s, led by Prof. Charles Yanofsky from Stanford and it has been studied primarily in *Escherichia coli* [8]. While much is known about the structure and function of the tryptophan attenuator in this organism, the tryptophan attenuator is not limited to *E. coli* and has also been identified in other bacteria, including Pseudomonas aeruginosa, Vibrio cholera, and many more [9] [10].

The specific mRNA structure and sequence of the tryptophan attenuator vary between different bacterial species, which implies that there might also be some differences in the regulatory mechanisms and function. In B. subtilis it was demonstrated that another trans-regulatory element called TRAP (trp RNA-binding attenuation protein) is also involved in the regulation of the operon expression [11] [12], and other mechanisms were later demonstrated as well in other species [13].

Overall, further research is needed to fully understand the diversity and function of this regulatory element in different organisms.

Fitness Landscape

The concept of a fitness landscape refers to a way of describing the relationship between the genotype of an organism and its phenotype and is attributed to Sewall Wright who introduced it in 1932 [14]. In graphical illustration of fitness landscapes, the genotype of an organism is symbolized by a point in a space (typically a plane), and its fitness is represented by the height of the point on the three-dimensional terrain. The idea behind the fitness landscape is that an organism's genotype determines its phenotype, which in turn determines its fitness. The fitness landscape, therefore, represents the relative fitness of different genotypes.

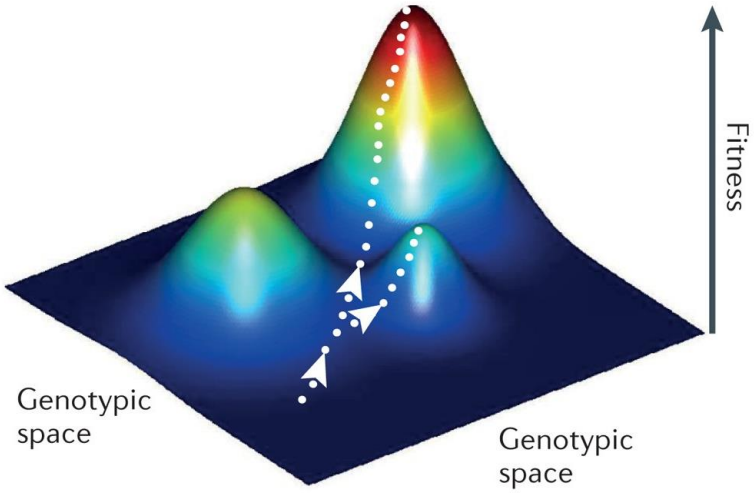


***Figure 5**: an illustration of the fitness landscape concept [18].*

Since 1932 the concept has been further developed and today it is used to study the evolution of many genes. Studies on such landscapes are typically done either by random mutagenesis of the wild-type version of a gene of interest, or by design and synthesis of specific versions of the sequence. Then each variant is typically transformed into another cell in a population, typically of microbes [15] [16] [17], and the fitness of each variant is determined by employing sequencing of the variable gene and following the relative frequency in the population of each variant. Such studies are gaining popularity in the molecular evolution literature, and despite the limitations of experimental exploration of multi-dimensional fitness landscapes and the complexity of epistatic interactions, are providing new and important insights into evolutionary processes [18]. For example, recently a fitness landscape of a yeast tRNA gene was constructed by random mutagenesis [19]. This work revealed positions that are more (or less) sensitive to mutation in this gene, and also the epistatic interactions between them. Likewise, fitness landscape has been mapped in protein-coding genes [17] [20], and in non-coding RNAs [21] [22].

The tryptophan attenuator could be embedded in a complex fitness landscape which would make its evolutionary optimization more challenging. As the sequence needs to adapt to two alternative structures, one in each of two conditions (terminator or anti-terminator conformations), the fitness of the organism would depend on the success of each of the two conformations, and their ability to allow expression when Trp synthesis is needed and to block it when it is not needed. Thus, we recognized that mapping the fitness landscape of the Trp attenuator would represent a uniquely interesting case study for fitness landscape mapping.

For example, if the tryptophan attenuator is not functioning properly, the bacteria may have a lower fitness due to a lack of sufficient tryptophan for the proper functioning of their proteins. On the other hand, if the tryptophan attenuator is functioning optimally, the bacteria may have a higher fitness due to the efficient regulation of tryptophan synthesis. Moreover, in a recent paper from our lab [23], we have also used a GFP model to demonstrate how various elements that control gene expression can impact the fitness, not only by the level of the expression but also by optimizing the cost of expression.

An attenuator may also serve as a model for additional riboswitches, which have an important role in bacterial gene expression regulation. Riboswitches are involved in essential genetic programs including antibiotic resistance [24]. Synthetic riboswitches can be used as biosensors in synthetic circuits [25], or provide an important addition to the toolkit of tunable expression systems, which are an important part of the ability for tunable devices in synthetic biology [26].

Goals

The goal of this project is to improve our understanding and further characterize the fitness landscape of the Trp attenuator, and the Repressor-Attenuator relationship, by developing tools and experimental design using a library of natural and synthetic Trp attenuator variants.

We believed it is worth revisiting the complex regulatory mechanism of Trp synthesis using the advanced tools we have today, and that it may reveal additional characteristics of the relationship between transcription attenuation, repressor regulation, gene expression, and eventually fitness under different conditions.

Gaining a better understanding of the fundamentals of natural diversity together with the wider synthetic potential may help us uncover novel regulatory modalities and, in the future, give rise to novel tools for biotechnological applications, drug targets, and, directed evolution for non-constant environments.

To achieve the above, my thesis focused on the following aims:

Aims:

1. Computational screening of different natural variants of the Trp attenuator in various bacterial species.
2. Computational design of a synthetic library of unnatural variants to explore different hypotheses about the regulation mechanism of the attenuator.
3. Develop an experimental system and measure the expression patterns and fitness of the different variants under different conditions.

Aim 1 – Computational screening of different natural variants of the Trp attenuator in various bacterial species**:**

To achieve this aim, we used computational tools to screen for annotated and unannotated Trp attenuators across sequenced bacteria.

Later, we selected some of them for in-depth inquiry using alignment tools and manual work to map candidate motifs that we found.

Aim 2 – Computational design of a synthetic library of unnatural variants to explore different hypotheses about the regulation mechanism of the attenuator**:**

The work on Aim 1 raised many motifs that may have an impact on the attenuation mechanism. Those motifs were divided into different groups and were used as the backbone for the computational design of many variants of these groups in order to

build a rich and diverse library. The final library was planned to include both synthetic designs and natural ones from various species.

Aim 3 – Develop an experimental system and measure the expression patterns and fitness of the different variants under different conditions**:**

We planned an experimental design where GFP production is regulated by the repressor & attenuation mechanism, instead of the Trp operon. We have designed a plasmid as shown below in the results section, where we planned to clone a library of different attenuator variants and measure GFP expression under different conditions (Trp concentrations, incubation time, etc.).

Furthermore, we planned to transform the plasmid into two different *E. coli* strains. The first one is the background strain of the KEIO collection and the other is a TrpR deletion strain. The comparison between both strains was meant to provide a better understanding of the way both mechanisms (TrpR and the attenuator) control the expression.

Materials and Methods

Please note, all listed primers and PCR protocols can be found in the appendix in Supplementary tables 2 and 3.

Strains, plasmids, and media

Table 2: *List of provided strains*

Species	Strain	Genotype	Reference
E. coli	BW25113	Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) λ- rph-1 Δ(rhaD-rhaB)568 hsdR514	[27]
E. coli	JW4356-2 (ΔtrpR)	Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) λ- rph-1 Δ(rhaD-rhaB)568 hsdR514 ΔtrpR789::kan	[28]
E. coli	JW4356-2_00 (ΔtrpR, Δkan)	Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) λ- rph-1 Δ(rhaD-rhaB)568 hsdR514 ΔtrpR789	Made by our lab
E. coli	DH5α	Δ(argF-lac)169 φ80dlacZ58(M15) ΔphoA8 glnX44(AS) λ- deoR481 rfbC1 gyrA96(NalR) recA1 endA1 thiE1 hsdR17	[29]
E. coli	MG1655	λ- rph-1	[30]

JW4356-2 strain (TrpR-) that was used in this work is based on the BW25113 genetic background, which is the parent strain for the Keio Collection of single-gene knockouts. JW4356-2_00 is based on JW4356-2, but we removed the kanamycin resistance gene for plasmid selection purposes.

DH5alpha is a competent strain, used for plasmid replication and conservation in glycerol stock.

MG1655 is the consensus genome of *E. coli* and was used to amplify the WT attenuator.

pJ251-GERC – a low Copy number plasmid that expresses mCherry and sfGFP from divergent promoters [31].
pVS101 – plasmid we created to measure the expression under Trp attenuator variants regulation in various conditions. The plasmid is containing GFP under the regulation of BsaI sites ready to clone the designed library on a pJ251-GERC backbone, and constitutive expression of mCherry. The different attenuator variants were cloned into

this plasmid to create pVS101_X (while X refers to the variant number) according to the protocol that will be detailed later.

pVS102 – similar to pVS101, containing the WT attenuator sequence without the restriction sites.

pVS103 – similar to pVS101, but designed separately for one of the variants due to the need to use a different restriction enzyme in the cloning protocol (BamHI instead of BsaI).

pCP20 - plasmid containing the yeast Flp recombinase gene, FLP, chloramphenicol, and ampicillin-resistant genes, and temperature-sensitive replication [32]. Used for the removal of Kanamycin resistance from the JW4356-2 (TrpR-) strain.

pZS11_GFP & pZS11_mCherry - two plasmids expressing GFP or mCherry from PLtetO-1 promoter. Used for calibration of the FACS.

LB – Media composed of tryptone 10 g/L, yeast extract 5 g/L, NaCl 10 g/L.

SOB - Rich media composed of 20 g/L Tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$ and 20mM glucose.

M9 – Minimal media composed of 0.1% glucose (w/v), 93.0 mM Sodium (Na^+), 22.1 mM Potassium (K^+), 18.7 mM Ammonium (NH_4), 1.0 mM Calcium (Ca^{2+}), 0.1 mM Magnesium (Mg^{2+}), 29.2 mM Chloride (Cl^-), 0.1 mM Sulfate (SO_4^{2-}), 42.2 mM Phosphate (PO_4^{3-}).

M9 (different Trp concentrations) – Same as M9 above, with the addition of Tryptophan (0.1, 1, 2, 3, 4, 5, 7, 8, 10, 50 µg/ml).

Natural variants' computational screening

Natural sequences of Gram-negative bacteria were collected from natural species as found at the NCBI Gene portal [33]. We chose this group since Gram-positive bacteria use a different mechanism to control Trp operon expression. First, the search included only species with an annotated TrpL region. Later, we also screened for additional, unannotated, natural TrpL variants according to the following rules:

- The presence of an in-frame ATG start codon within a certain distance (above 100bp but no more than 500bp) upstream to the first Trp Operon gene.

- The presence of an in-frame stop codon downstream to the ATG codon in the specified region.
- The presence of at least one in-frame Trp codon (TGG) between the start and stop codon.
- The presence of at least four tandem T's downstream to the stop codon and upstream from the Trp operon gene.

The Jalview software [34] was used for the visualization and analysis of the sequences and comparison between the different sequences. The list of species and their sequences can be found in the appendix in Supplementary table 1.

Predictions of the alternative structures of the attenuator variants were performed using PASIFIC [35] [36].

Setup of the experimental system

Genetic engineering of the JW4356-2_00 strain:

Since the selection of the experimental variants' plasmid was Kanamycin resistance, which was also the selection on the KEIO collection (from which the JW4356-2 was acquired), we had to remove the resistance gene from the strain. To do so, we used the Barrick lab protocol [37] [27] [28] [32] with a pCP20 plasmid (FLP recombination). Results of the protocol were verified on Kanamycin selection plates and Sanger sequencing of the relevant region.

Design of the variant sub-libraries

The designed synthetic sequences were generated by custom codes written in Python or manually.
The designed natural sequences were collected from natural species as found at the computational screening and curated manually to include a variety of families and attenuator types.

Cloning of plasmid sub-libraries
WT attenuator sequence was amplified from *E. coli* MG1655 strain and was added with restriction sites for later use during the cloning process using dedicated PCR primers (see supp. table 2). The purification process was done using the Promega kit for

Genomic DNA purification, DNA concentration was measured using Qubit instrument and kit, and verification was performed using Agarose 1% TAE gel.

The assemblies of pVS101, pVS102, and pVS103 were performed using either Gibson assembly [38] or Golden Gate cloning [39].

The plasmids were then purified using SPRI bids or purification columns.

DNA fragments of up to 200bp that contain the forward and reverse-designed sequences were ordered from Sigma. These fragments were phosphorylated and annealed for the cloning process.

After the transformation and purification of the plasmids (described in the next section), the variant fragments were cloned into their relevant plasmids in either Gibson assembly [38] or Golden Gate cloning [39].

All samples were later verified on an Agarose gel and Sanger sequencing for correct assembly of the plasmid and when the positive results were received, samples were taken for the lab's glycerol stock by suspending the culture to a final concentration of 30% glycerol, then being frozen and stored at -80˚c.

Bacterial transformation of the sub-libraries

The clean product of each of the plasmids was transformed into competent DH5α bacteria using chemical transformation (42˚c heat shock). 2µl of plasmid were added to 100µl of cells and incubated on ice for 30 minutes, before a 1-minute heat shock followed by a 5-minute incubation on ice. Afterward, 900µl of liquid LB were added and incubated at 37˚c for 1 hour before plating on solid LB (+Kan when relevant).

All samples were later verified in an Agarose gel and Sanger sequencing for correct assembly of the plasmid and when the positive results were received, samples were taken for the lab's glycerol stock by suspending the culture to a final concentration of 30% glycerol, then being frozen and stored at -80˚c.

The plasmids were extracted using Promega midiprep kit.

After cloning of the different variant plasmids, they were transformed into the relevant strains (BW25113, JW4356-2, JW4356-2_00) using electrochemical transformation.

Overnight culture of the relevant strain was diluted 1:100 and grown in a shaking incubator at 37˚c until mid-log phase is reached (OD_{600} = 0.4 - 0.6, ~2H).

Then, the cells were washed three times by pipetting 1mL into a centrifuge tube, centrifuge at 13,000g for 30 seconds, remove the supernatant and resuspend in 1mL chilled DDW. This was repeated 2 more times, while at the last time 2ng of the relevant plasmid were added, and the aliquot was transferred to a 1mL chilled electroporation cuvette. After 60 seconds, cells were electroporated using Mage with the following parameters: 1.8 kV, 200 ohms, 25μF.

1mL of warm SOC media was added and cells were moved to a fresh tube to recuperate in a shaking incubator set at 37˚c for one hour. At the end of the hour, cells were plated on LB + antibiotic selection (Kanamycin), and surviving colonies were verified by gel, and samples were taken to the lab glycerol stock.

Growth and expression experiments

Growth experiments

Cells were grown for 48h on the indicated medium until reaching deep stationary phase and then were diluted into fresh medium (dilution factor 1:100). Growth experiments were done in flat-bottom 96-well plates with 150μl per well. In each plate, the strains at the different Tryptophan concentrations were arranged in a checkerboard pattern to cancel out geographical effects by a robotic liquid handler (Beckman-Coulter Biomek i5). Plates were put in a shaking incubator set to 30˚c and OD_{600}, mCherry, and GFP were measured every 1.5 hours for up to 4 days by a plate reader (Tecan Spark). All measurements were executed automatically using a robotic system (Thermo-Fisher Cytomat shaker incubator & Thermo-Fisher Orbitor RS2). The results of the growth experiment were analyzed using in-house MATLAB-GUI and Microsoft Excel.

Expression experiments

This work involved extensive FACS measurements and experiments, which were performed on an Attune NxT Flow Cytometer (Thermo-Fisher, Invitrogen).

The FACS gating was designed to capture all cells and measure their fluorescence in wavelengths 530/30 and 620/15 (GFP and mCherry) as well as forward scatter (FSC) and side scatter (SSC) for gating. Before the first experiment, a calibration was made using BW25113 (WT repressor) with either No plasmid (no fluorescent protein

expression), pZS11_GFP (GFP), pZS11_mCherry (mCherry) or pJ251-GERC (both GFP and mCherry), and the gating was determined according to these strains.

Before every experiment, further validation was performed using the same calibration strains that were grown in tubes at the same time of the experiment variants.

The raw data from the FACS was extracted and went through initial analysis using Kaluza Analysis Software (Beckman-Coulter), and an average score for all cells in each well was calculated and used as data for further analysis that was performed using Microsoft Excel.

The general experimental design can be graphically described as follows:

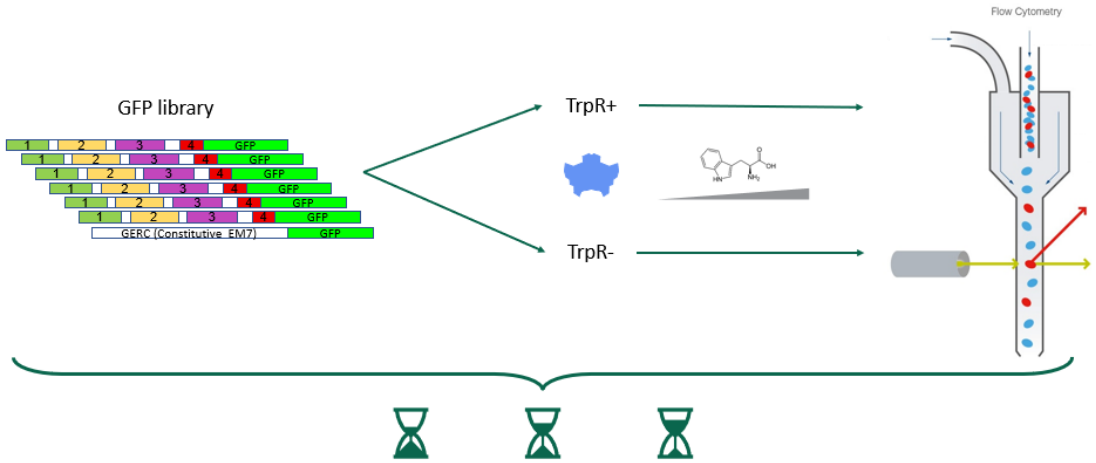


***Figure 6**: in each experiment, some of the designed variants (together with the WT variant of the attenuator and the constitutive expression control) were grown in both strains (with and without the Trp Repressor), in different Trp concentrations. Later, their expression was measured using either FACS or a Plate Reader in different time points.*

The experiments varied in the concentrations of Tryptophan in which the variants were grown, in the time of expression measurement, and additional changes as will be described later.

At the first variants' expression experiment (fig. 18 in the results section), all of the different variants in the TrpR+ strain (two repeats), together with the WT and positive control variants, were first grown in tubes with minimal M9 without Trp for 24 hours. Than, if stationary - dilluted 1:100 and if not stationary just transferred the relevant amount into 96-well plates with Kananmycin, M9, and the different Trp concentrations (0, 1, 5, 10, 25, 50) and grown in a shaking incubator for 2 Hours. After the incubation time, we used FACS to measure the expression.

Plate design (Each of the variant rows (row 1, row 5) is grown at a different Trp concentration):

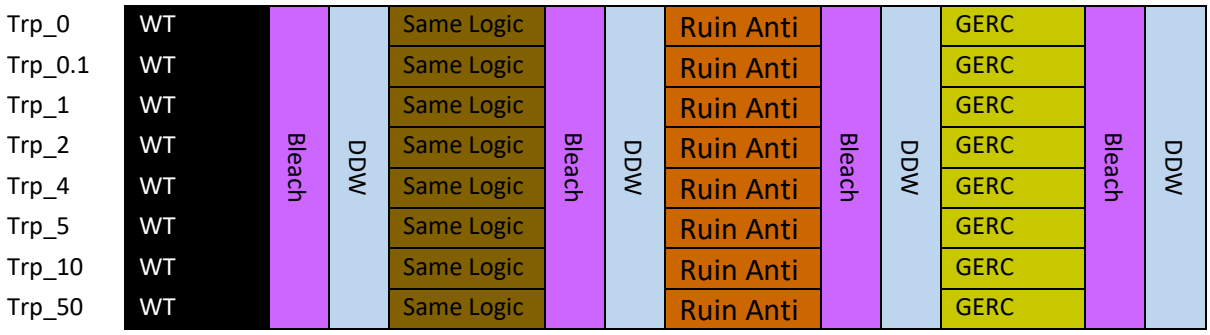
Same Logic	Reverse logic	Ruin terminator	Ruin anti-terminator	No Trp	Trp to stop	5 Trp	No gaps	WT	GERC		
Bleach											
DDW											
DDW											
Same Logic	Reverse logic	Ruin terminator	Ruin anti-terminator	No Trp	Trp to stop	5 Trp	No gaps	WT	GERC		
Bleach											
DDW											
DDW											

In the 72 hours experiment (fig. 21), all of the different variants in the TrpR- strain, together with the WT and positive control variants, were first grown in tubes with minimal M9 without Trp for 48 hours. Than, dilluted 1:100 using a liquid handler into 96-well plates with Kananmycin, M9, and the different Trp concentrations (0, 0.1, 1, 2, 4, 5, 10, 50) and grown in a shaking incubator (total of 10 plates) for 72 Hours.

The same design was used for FACS measurment (fig. 20) after 16 hours of incubation in the different Trp concentrations.

For better control of the time points in which we take measurments, we conducted another calibration experiment, in which we grew a plate with the strains: WT, Same Logic, Ruin Anti-Terminator and GERC (positive control), in the TrpR+ strain. The plate was grown using a shaking incubator and we measured the OD every 30 minutes.

Later, we conducted another expression experiment in different time points (fig. 22-25). We have created 6 plates as described below, 3 identical plate with all variants in TrpR+ strain, and another 3 identical plates with the variants in TrpR- strain. The different variants were first grown in tubes with Kanamycin and minimal M9 without Trp, and then dilluted according to the calubration experiment into the wells with Kan and the different Trp concentrations, and grown on a shaker incubator. For each TrpR strain, the plates were measured using FACS after 2, 16 and 24 hours of incubation in the different concentrations of Trp.



For the final experiment with all variants (fig. 26-27) we used the following plate design and Trp concentrations:

Plate 1:

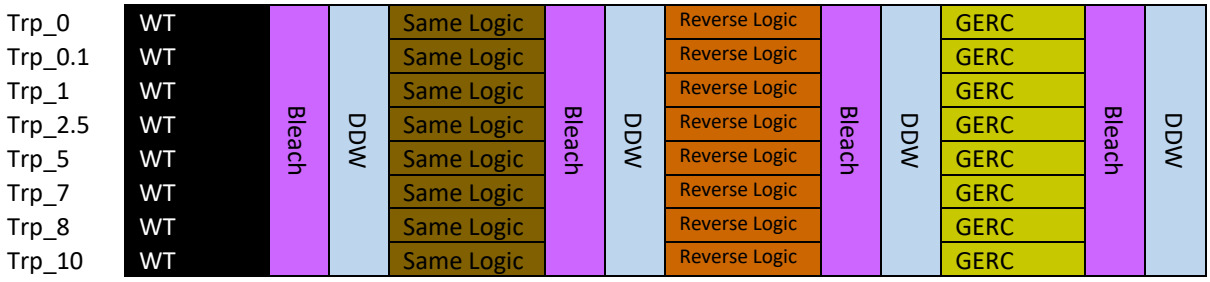


Plate 2:

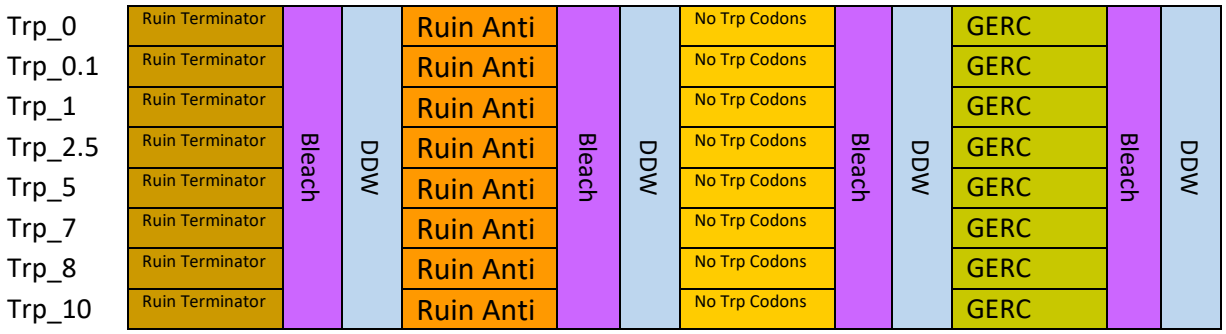
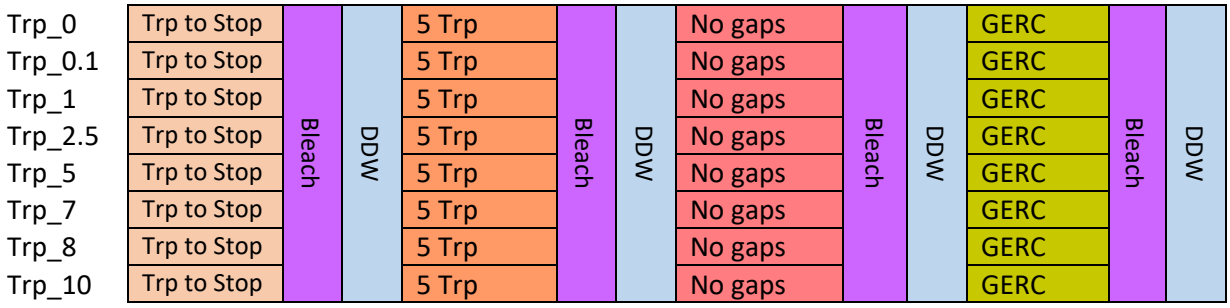


Plate 3:



Results

Natural variants computational screening

First, we compared the Trp attenuator region in various *E. coli* strains:

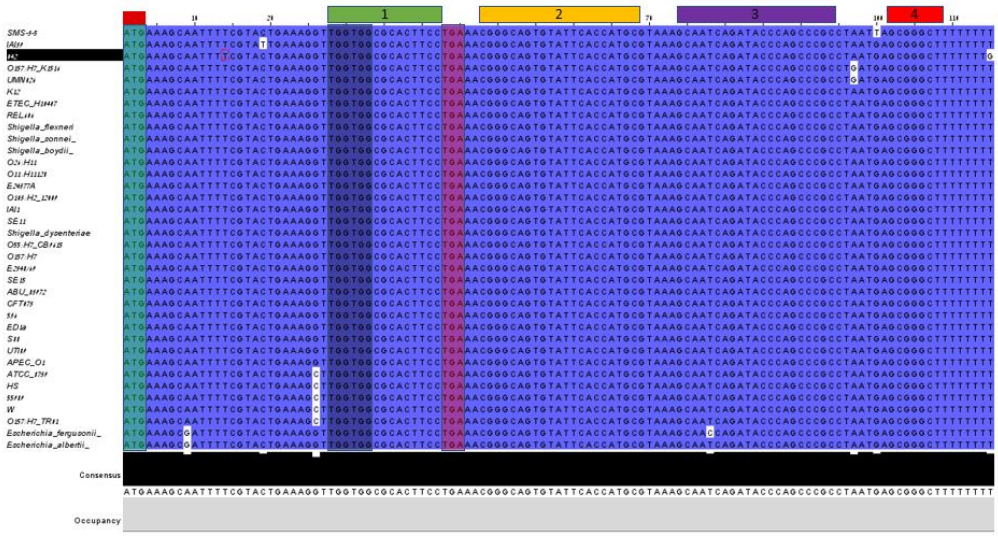


Figure 7: *E. coli strains MSA analysis. In the above figure, there is a representation of the different attenuator regions (1,2,3,4). Marked in green is the ATG (start) codon. Marked in gray are the couple of tandem TGG (Trp) codons. Marked in red is the TGA (stop) codon).*

Among *E. coli* strains, the attenuator sequence has very little variation. Only a few point mutations were found, and not in the known-to-be-important regions.

Later, we compared other species from different families (including *Vibrionaceae*, *Yersiniaceae* and *Enterobacteriaceae*):

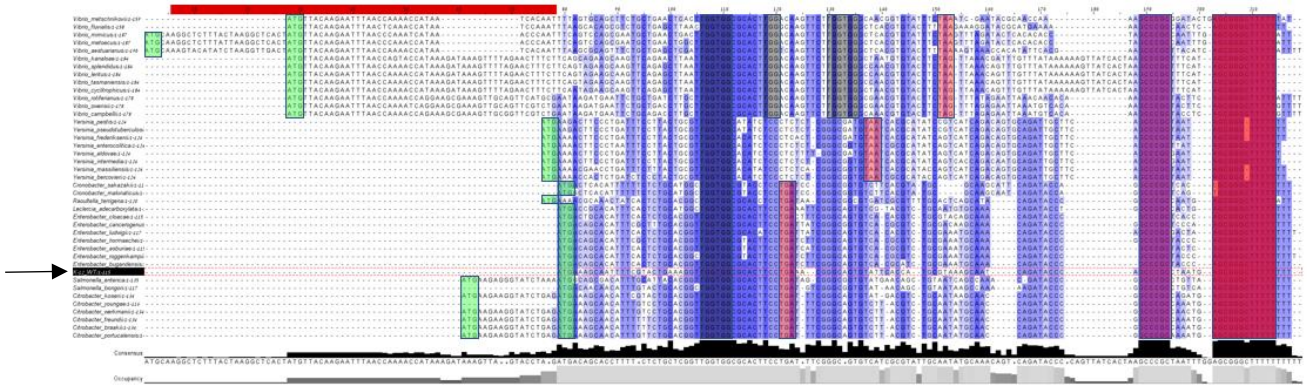


Figure 8: *Cross-species MSA analysis. The color code is the same as the previous figure, and the K12 E. coli strain is marked with a black arrow to the left. Terminator and Anti-terminator regions are also marked in red, on the 3' (right) end of the sequence. The full list of species and sequences is provided at supp. table 1.*

It is clear that the variation between species from different families is much larger, and it is possible to point out a few interesting differences. First is that the distance between the in-frame start codon and the Trp codons vary. In some cases, there are two potential in-frame start codons (9 or 5 AAs apart from each other).

Also interesting to point out is that the leader peptide sequence isn't highly conserved, while the two tandem Trp codons remain the same across all species. In the Vibrionaceae family, we see three additional Trp codons, in a 2-1-2 structure with only a few other amino acids between them.

The distance between the Trp codons and the stop codon also varies between Vibrionaceae, Yersiniaceae and Enterobacteriaceae, and there is also variability of the types of stop codons (TAG, TAA, TGA) that are being used across families.

The structure of the terminator and anti-terminator sequences remains highly conserved among the different species.

These findings guided us towards the design of the library, as we tried to implement some of the variations observed in natural sequences into our synthetic designs.

Later, we carried out a focused analysis of the different families, starting with Yersiniaceae. We also tested if there are differences that correlate with the typical habitats in which each species can be found. We hypothesized that organisms who inhabit different habitats, where Trp availability behaves differently, might be selected to reduce either false-positive activation (i.e., to be ON when Trp is present), or false-negative (i.e., to be OFF when Trp is absent). We thought this approach might provide us with better understanding of the Trp attenuator fitness landscape complexity. For example, soil bacteria usually experience less plentiful nutrition than bacteria who live in the human gut, which may drive their evolution towards the false positive side of this potential tradeoff in comparison with human or animal gut bacteria.

Habitat legend:

Yersiniaceae:

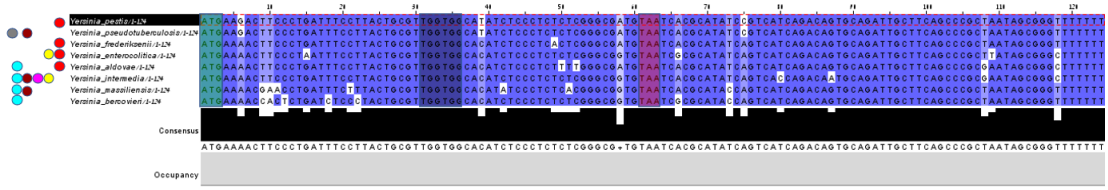


***Figure 9**: Yersiniaceae MSA analysis.*

Within *Yersiniaceae*, the variability is low and similar to the *E. coli* strains comparison. It is worth noting that the distances between the Start codon, Trp codons and Stop codons are conserved, but there is one point mutation that shortens the poly T of their terminator by one nucleotide in some of the species. With such high conservation and ecosystem variability, it is difficult to claim that there is a correlation between the different habitats and the sequence.

Enterobacteriaceae:

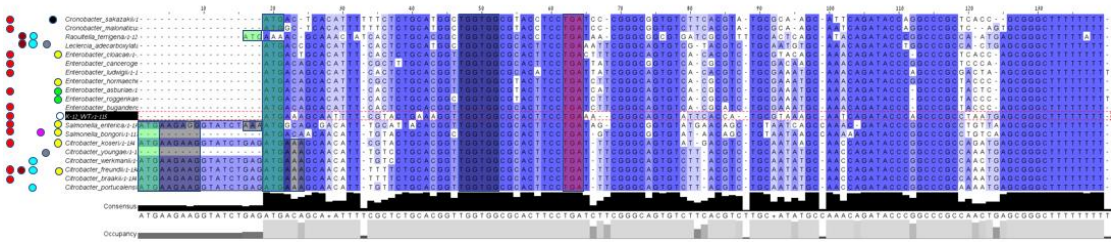


***Figure 10**: Enterobacteriaceae MSA analysis.*

Within *Enterobacteriaceae* we see wider variability. The distance between Trp and stop codons remains constant, but there is an interesting variability among the start codons. In *Raoultella terrigena*, the start codon is located 3 nucleotides upstream to where it is in *E.coli* and the rest of the family, and in the *Salmonella* & *Citrobacter* genera, there is an alternative start codon (in-frame, in addition to the one that other species have) 18 nucleotides (6 AAs) upstream. Also, worth noting that among these two genera, there is an enrichment of Lysine codons (AAA & AAG) that was not demonstrated in other species. Since Trp and Lysine remotely share biosynthetic pathways through Phosphoenolpyruvate [40] (metabolic intermediate of glycolysis), there might be an interesting, non-random, explanation, but we did not pursue this path in this work.

The rest of the attenuator regions and mostly the sequences between its regions (the loops), were not highly conserved amongst the Enterobacteriaceae, who demonstrated high variability.

Vibrionaceae:

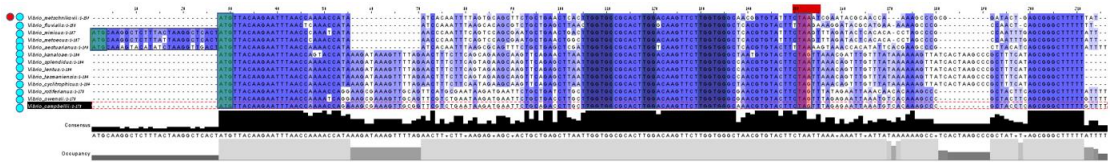


***Figure 11**: Vibrionaceae MSA analysis.*

In the *Vibrionaceae* family we also found considerable variability. As noted earlier, all the species that we found had a 2-1-2 Trp codon motif. The distance from the stop codon is also conserved, as well as the terminator sequence.

We did see another version of alternative start codon in some of the species, but unlike *Salmonella* & *Citrobacter*, there were no Lysine codons in this region.

Also, it is worth noting that in some of the species the length of the leader peptide was 5 AAs shorter than the others, and in some the region that is following the stop codon (regions 3, 4, and the loops) were shorter than others. This family is another example for variation that cannot be explained by habitat, since all species in the list are aquatic.

Set-up of the Attenuator variant library system

Design of a large library of Trp attenuator variants

We have designed a pilot library, containing a combination of synthetic and natural variants. The aim of this design is to provide a conceptual tool to think about the variety of characteristics that may influence the fitness landscape of the Trp attenuator, as well as to set the ground for future experiments. Below is a figure that describes the potential variants of this library, using different types of sub-libraries:

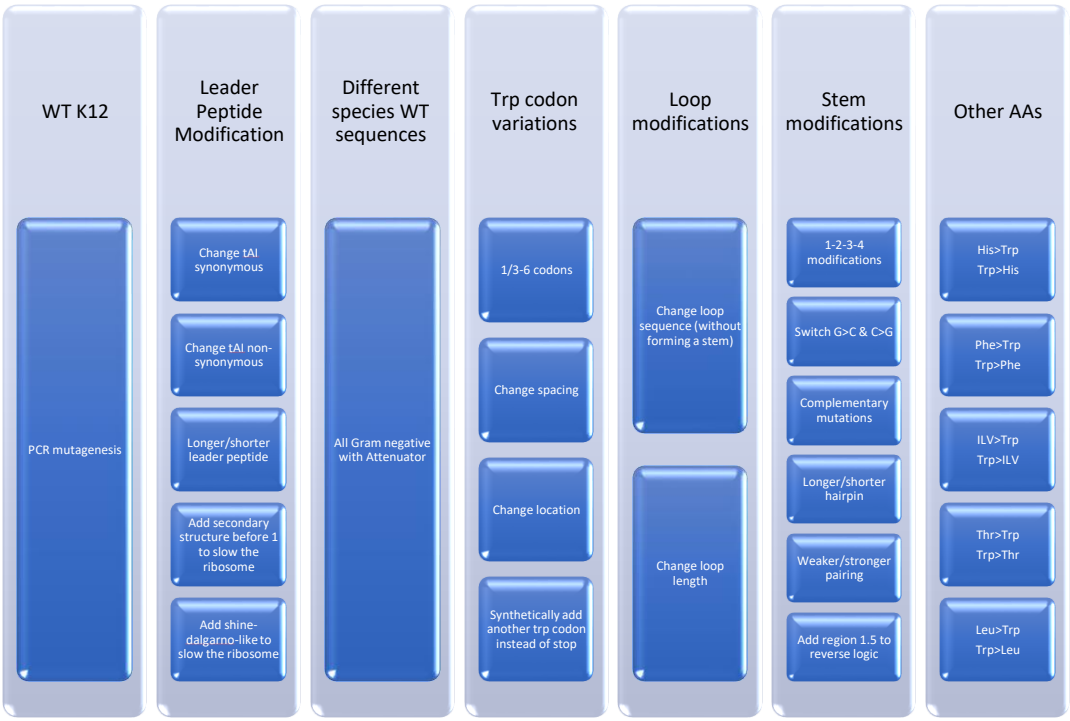


***Figure 12**: a list of Trp attenuator variants sub-libraries.*

WT K12: using PCR mutagenesis, it is possible to create a large variety of variants including most (if not all) one- and two-point mutations, to find the 1-neighbors and two-neighbors as well as subsets of three and four neighbors of the WT on the sequence landscape [41]. This approach was used in other fitness landscape works [42] [43] [44] [45], but so-far not on the Trp attenuator.

Leader Peptide Modification: The leader peptide sequence (TrpL) plays a key role in the attenuation mechanism. We tried to rationalize the main mechanisms that might impact the fitness. It has been demonstrated by our lab that tRNA adaptation index (tAI) of the different codons can influence translation efficiency [46]. Therefore, we wanted to create different changes in the tAI of the leader peptide to explore their impact. Some of the changes were aimed to be synonymous, and some non-synonymous. The rationale for this non-synonymous sub-library was that as far as we know, the leader peptide has no active function other than its impact on the attenuation. A potential result that demonstrates significantly different fitness of synonymous vs. non-synonymous mutations, would suggest otherwise.

Another sub-library was a shorter and longer leader peptide. We hypothesized that different lengths should impact the timing when the RNA-polymerase and the ribosome

reach important codons such as the Trp codons and the stop codon, and therefore impact the expression and fitness.

The leader peptide includes region #1 of the attenuator which, in turn, binds region #2 (the anti-anti-terminator). We planned a sub-library where we will add an RNA secondary structure upstream to region #1 in order to stall the ribosome and interfere with the timing it reaches important regions in relation to the RNA-polymerase. Another approach that was aimed to slow-down the ribosome was to add a Shine-Dalgarno like sequence.

Different species WT sequences: the purpose of this sub-library was to explore the natural diversity of attenuators, in our controlled system, and to study the performance of the different variants outside of their original context. If the attenuator's only mechanism of action is with regard to its secondary structure, our experimental system should give us a measurement of the inherent balance of expression and termination in different conditions and will create a platform for comparison between the variants. Then, it will be interesting to analyze the dynamics each variant will demonstrate, connect it to its original species, typical niche, etc. and contemplate the evolutionary meanings of the results. We chose to use only gram-negative bacteria since they are the only ones who demonstrate this mechanism of attenuation.

Trp codon variations: due to the importance of the tandem Trp codons, we wanted to experimentally test the landscape involving them and their region. Among the changes we planned to add or remove codons (only one, or up to six), to change the spacing between them (add one or two non-Trp codons between the Trp codons), change the location of the Trp codons on the leader peptide, and even add a new synthetic codon for Trp (replacement of one of the stop codons).

Loop modifications: we wanted to study the loops, and to test how different changes of their length and sequence that are hypothesized to have minor or no influence on the expression and fitness, impact the dynamics attenuation and fitness.

Stem modifications: the stems have a complex relationship with each other, and with the attenuation mechanism. Any change in one region of one stem has an influence on at least one more region, if not more than that. Some of the nucleotides are located as part of what we called "quartettes", which are positions in which nucleotide from

attenuator region #1 binds a nucleotide from region #2, which in turn can also bind another nucleotide from region #3, who alternatively binds a nucleotide from region #4. We wanted to test how changing all quartette in a complementary manner (in which they can still bind each other) impacts the overall dynamics. Technically, if the change preserves the GC content (each G is substituted by a C, and vice-versa), we shouldn't see any impact. We can make such changes also outside of the quartettes, while preserving GC content or not preserving it, keeping the complementation as it was with a different sequence or changing the complementation, elongating the stem or making it shorter, decrease the pairing energy (G's and C's into T's and A's) or increase it, etc.

One additional concept we had in mind, was a theoretical reversion of the attenuation logic. We wanted to come up with a change that will cause expression of the downstream elements when Trp is present and avoid expression in its absence. The concept we came up with was the addition of a region dubbed "3.5" which was planned to alternatively bind the terminator and anti-terminator regions.

Other AA's attenuators: here we wanted to test how general is the mechanism of attenuation. Since additional amino acids other than Trp also have similar mechanisms, we planned a sub-library in which we will replace the original AA's codon into a Trp codon, and vice versa, when we change the original Trp codon in the Trp attenuator to another AA codon. It was assumed that the different variants from this library would not have the same expression levels under the same conditions. Both in the case of modified other AA's attenuators to act as a Trp attenuator, and in the case of modified Trp attenuator grown in the relevant AA's absence.

Part of this conceptual exercise was to attempt to reach all the potential extremes of the figures below:

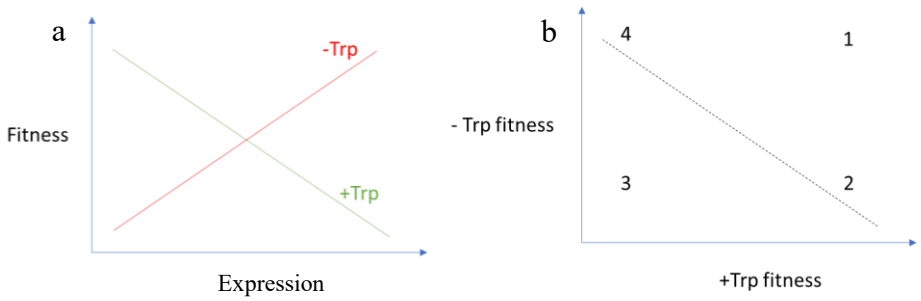


***Figure 13**: Illustrative figures. (a) fitness-expression (GFP as expression-proxy) graph. The red line represents a possible behavior under low Trp concentration, while the green line is in high Trp concentration. (b) theoretical fitness of variants in the presence and absence of Tryptophan. The numbers represent four hypothetical variants that demonstrate different scenarios. The diagonal represents a conceptual linear trade-off between -Trp fitness and +Trp fitness as if it was a zero-sum game.*

In 13 (a) we describe the hypothetical fitness-expression relationship of our library. In the case of low Trp concentration – a variant that expresses the gene strongly would have higher fitness, while in high Trp concentration lower production is expected to gain better fitness. As we suggested in the computational screening section above, relevant scenarios may occur in nutritionally rich environments like the gut (which is likely to also have higher availability of Trp), or nutritionally poor environments like the soil.

In 13 (b), Variant #1 is a variant that has high fitness in both conditions of Trp concentration. The wild type is hypothesized to be above the diagonal line as it will try to balance between both conditions, and ultimately be located where variant #1 is, if there are no additional evolutionary constraints.

Variant #2 has great fitness in the presence of Trp and poor fitness in its absence. For instance, it might have a deletion of the Trp operon, and therefore never produces Trp. When the Trp levels are high, it's fitness is also high since there's no waste of any resource on Trp production, and when Trp levels are low it has poor fitness due to its complete dependence on the extra-cellular resources of Trp.

Variant #3 should have poor fitness in both conditions. Our attempt to design such variant was called "reverse logic", in which we thought of adding a fifth complementary region that will reverse the logic of attenuation (terminate in the absence of Trp and produce in its presence), as explained in the stem modifications sub-library in the previous section.

Variant #4 has great fitness in the absence of Trp and poor fitness in its presence. One option to achieve that might be a constitutive production of Tryptophan, which can be caused by ruining the terminator mechanism.

Creation of the pilot synthetic library of Trp attenuator variants

In order to enable a high-thruput cloning process for the library, we first built a general backbone plasmid.

pVS101 was designed to enable easy cloning of the library variants into a low-copy, two reporter plasmid, in which mCherry will be constitutively expressed to control for cell growth, and GFP will be expressed under the control of the different variants of the Trp attenuator.

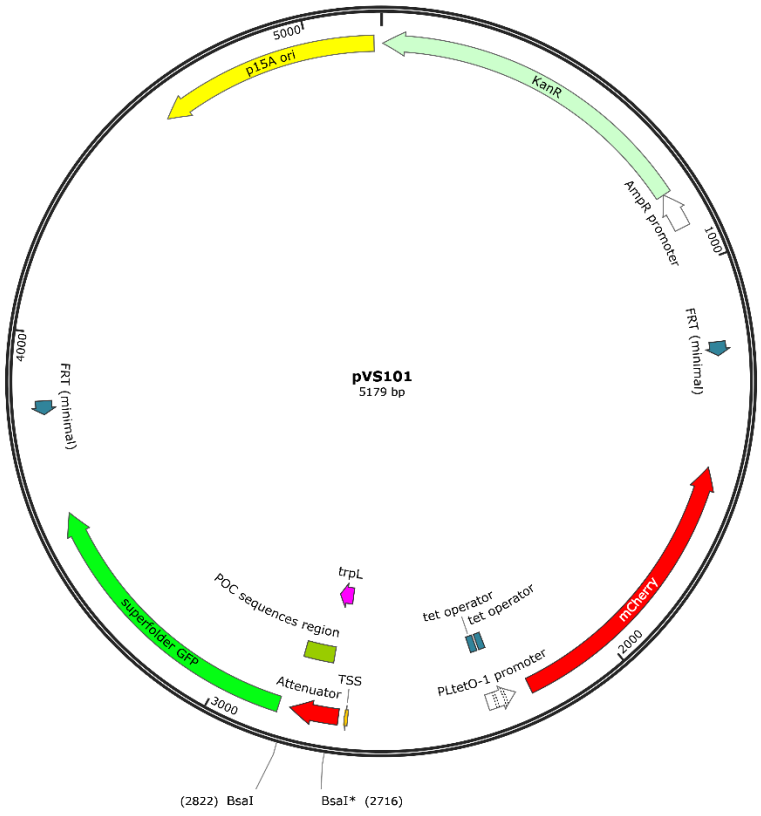


Figure 14: *pVS101* *- The integrative plasmid containing the Trp Attenuator gene upstream to the GFP, with the variable region is cloned into. The plasmid sub-libraries were digested by cloned between the two BsaI sites ("POC sequences region) and were later transformed to the relevant E. coli strains.*

pVS102 was designed to be same as 101, but the expression will be controlled by the WT attenuator sequence.

pVS103 was designed to overcome variant #7 technical cloning challenge but the final variant is identical to pVS101 variants excluding the variable region.

All plasmids were verified by PCR and Gel analysis, as well as Sangar sequencing.

For the expression experiment we chose to begin with 8 variants, which were attempted to be different from each other and interesting in their potential behavior. Below is their illustrated description, together with our hypothesis for their behavior under different Trp concentration conditions, and their computationally predicted behavior.

Design and simulation of a 10 variants' sub-library attenuation potential

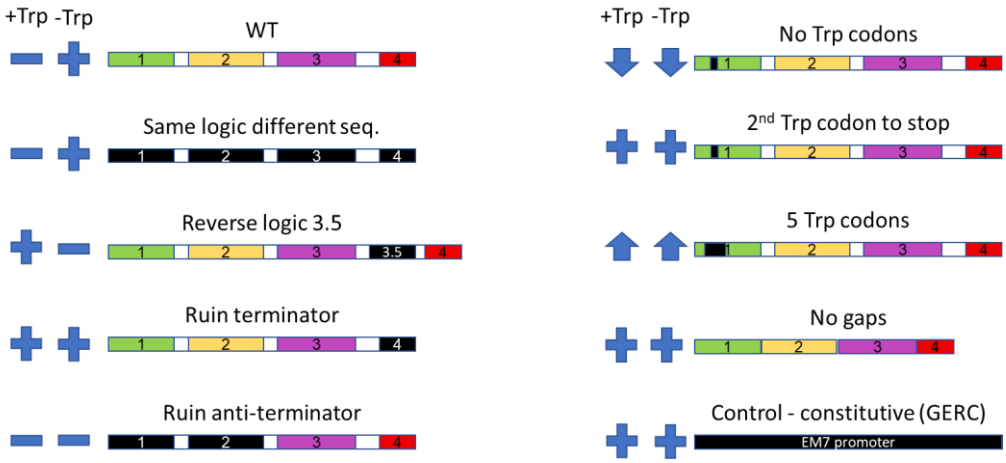


***Figure 15**: Illustration of the 10 chosen experimental variants. The colors illustrate different regions of the sequence as follows: Pink - Stem #1 (First stem of the anti-antiterminator), Orange - Stem #2 (Second stem of the anti-antiterminator, and first stem of the antiterminator), Green - Stem #3 (Second stem of the antiterminator, and first stem of the terminator), Blue - Stem #4 (Second stem of the terminator), Black – the region in which a change was made from the reference WT sequence.*
The "+Trp" and "-Trp" columns on the left of each variants, show a "+" sign when we hypothesized that expression of the downstream elements will take place, a "-" sign when we hypothesized the downstream elements will not be expressed, a "↑" sign when we expect an increase in expression compared to the WT, and a "↓" sign when we expect reduction.

WT – Expression OFF when there is Trp in the media, and ON when it is absent.

Same Logic different sequence – expected to be the same as the WT.

Reverse Logic – expected to be ON when there is Trp in the media, and OFF when it is absent, since we added the "3.5" region as described in the design of the library section above.

Ruin Terminator – expected to be always ON since the termination signal is harmed.

Ruin Anti-Terminator - expected to be always OFF since the anti-termination signal is harmed, and the terminator formation will be favored.

No Trp Codons (Trp to Arg) – expected to reduce the expression in comparison to the WT, both in the presence and the absence of Trp. We expected this outcome because according to what we know on the Trp attenuator mechanism of action, the ribosome is not expected to stall on the region where the two tandem Trp codons used to be, and therefore the anti-terminator formation will not be formed. In our case, we chose to change the Trp codons to Arg codons, which may be also interesting to test for Arginine sensitivity.

Second Trp Codon to Stop – expected to cause constitutive expression, both in the presence and the absence of Trp. When we change the second codon to a stop, we cause the premature termination of the leader peptide translation. When the ribosome drops after it reaches the stop codon, it doesn't reach region #2 of the attenuator, and does not prevent the bonding of the anti-terminator formation. When the anti-terminator formation is favored, expression is expected to be always ON.

Five Trp Codons – expected to increase the expression in comparison to the WT, both in the presence and the absence of Trp. When there are five Trp codons on the leader peptide instead of the original two, the ribosome is expected to stall longer on region #1 of the attenuator, which favors the formation of the anti-terminator. Even in the presence of Trp in the media, we expected to see some increase of expression.

No Gaps – this variant was difficult for us to predict, due to the complexity it brings to the formation of stem & loop structures across the whole mRNA sequence. We hypothesized that it might hamper the formation of both terminator and anti-terminator, and therefore expected that in the absence of termination signal, the RNA polymerase will continue transcribing the downstream elements.

GERC – control variant. Constitutive expression of both mCherry and GFP, expected to have high expression regardless of Trp concentration.

Based on the sequence of each variant, a PASIFIC probability score was also given. According to the creators of PASIFIC [35], the threshold of 0.5 yields a sensitivity of 82.5% and specificity of 80.6% for the prediction, when a high score is predicted to be an attenuator and a lower score is not.

***Table 3**: list of the experiment's variants and their PASIFIC score.*

Variant name	PASIFIC score
WT	0.51
Same logic different sequence	0.47
Reverse logic	0.68
Ruin terminator	0.37
Ruin anti-terminator	0.82
No Trp codons	0.52
Second Trp codon to Stop	0.55
Five Trp codons	0.77
No gaps	0.22

Interestingly, the WT variant got a score only slightly above threshold. The highest scores were achieved by Ruin Anti-Terminator sequence, five Trp codons, and reverse logic. Same Logic variant received a lower score than the WT sequence which brought it below the threshold, together with Ruin-Terminator and no gaps variants.

Growth and expression experiments

In this experiment, we aimed to calibrate our system using the TrpR+ strain and the WT attenuator plasmid. We grew them in different Trp concentrations and measured hourly the GFP expression, mCherry expression and OD for 21 hours.

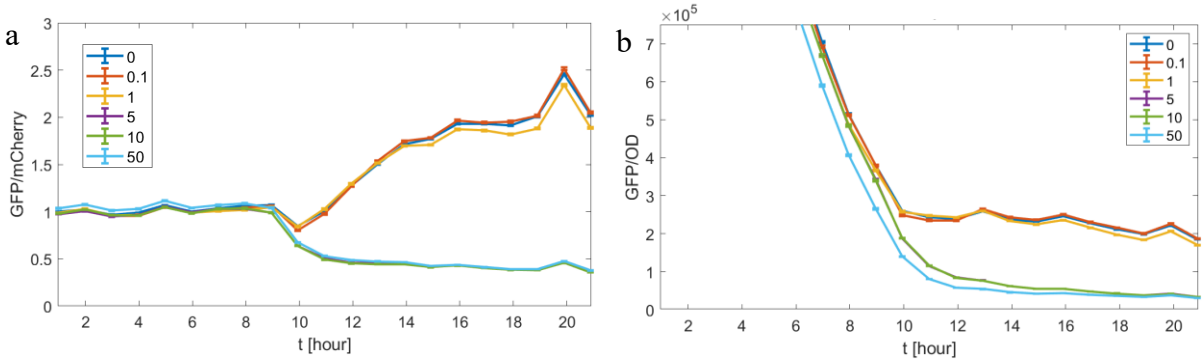


***Figure 16**: the normalized expression of the WT attenuator plasmid in the TrpR+ strain, over time. Each color represents a different concentration of Trp from 0 to 50 µg/mL. (a) normalized expression as a ratio of GFP/mCherry (b) normalized expression as a ratio of GFP/OD.*

From the above figures we can learn that the GFP normalization to mCherry conveys a similar message to the normalization to OD measurement when it comes to the expression-growth dynamics. We can also learn that it takes over 10 hours for the difference between the Trp concentrations to manifest in expression of GFP. After this timepoint, it was very clear that the behavior changes between a concentration of 5 and 10 µg/mL. Concentration from 5 µg/mL and lower demonstrated high expression levels, while higher concentrations seem to reduce the expression by a factor between

4 and 5. Another interesting observation is that there is a basal level of expression that remained static for the first ~9-10 Hours, and later the high [Trp] populations demonstrated a reduction of expression, while the low [Trp] populations demonstrated an increase in expression. None of the populations in this experiment kept the basal level of expression.

In this experiment, we wanted to compare the expression of the constitutive expression control construct, in the TrpR+ strain, across different Trp concentrations over time.

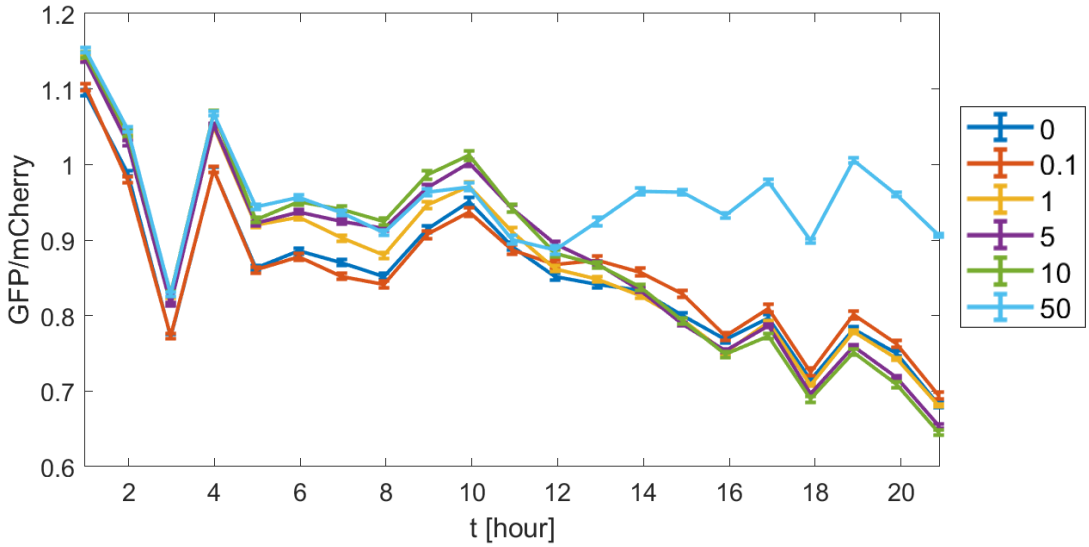


***Figure 17**: the normalized expression of the constitutive expression plasmid in the TrpR+ strain as a ratio of GFP/mCherry, over time. Each color represents a different concentration of Trp from 0 to 50 µg/mL.*

In the results we could see that expression levels didn't remain constant across time. Expression levels fluctuated, but fluctuations remained constant across different concentrations. It was also apparent that expression levels were slightly lower in the 0 and 0.1 concentrations at first ~10 hours of the experiment, and then all concentrations demonstrated similar levels excluding the concentration of 50 which showed high expression levels from 13 hours onwards.

After establishing the time following Trp exposure at which GFP signals begin to reliably emerge in the WT attenuator, we set to measure GFP at such a point in time across all variants in my attenuator collection, each at a range of Trp concentrations. Due to the previous results, we chose the 16-hours' time point as the relevant one to test the impact of the different Trp concentrations on the expression levels. The experiment was performed using a growth robot and a plate reader, in which measurements were automatically taken at the pre-set time point.

In this experiment we made two repeats of all variants, within the Trp+ strain, grown in all Trp concentrations as in the calibration experiment.

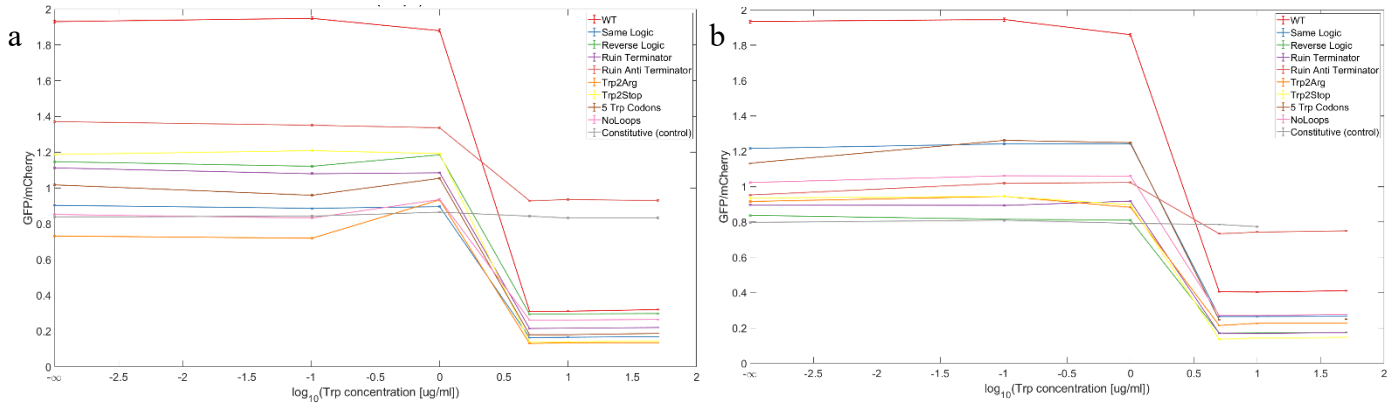


Figure 18: *the normalized expression of the different variants of the attenuator plasmid in the TrpR+ strain as a ratio of GFP/mCherry, on different Trp concentrations. Each color represents a different attenuator variant. The measures were taken after 16 hours of exposure to the different Trp concentrations. (a) first repeat (b) second repeat.*

First, it appears that in both repeats we see a clear drop of expression between a concentration of 1 and 5 µg/mL. this appeared in all variants except the positive control of constitutive expression for both GFP and mCherry, which did not show any sensitivity to Trp concentration in this experiment.

The WT variant had the largest ratio in concentrations lower than 5, while at the higher concentrations the Ruin Anti-Terminator variant, which according to our predictions was supposed to have no anti-terminator activity and therefore to have constant termination, demonstrated the lowest drop in expression ratio and became the variant with the highest expression.

In order to study the dynamics and explore some possible explanations for the measurements from the previous experiment, and to calibrate for the TrpR- strain that wasn't used in previous experiments, we conducted another growth experiment.

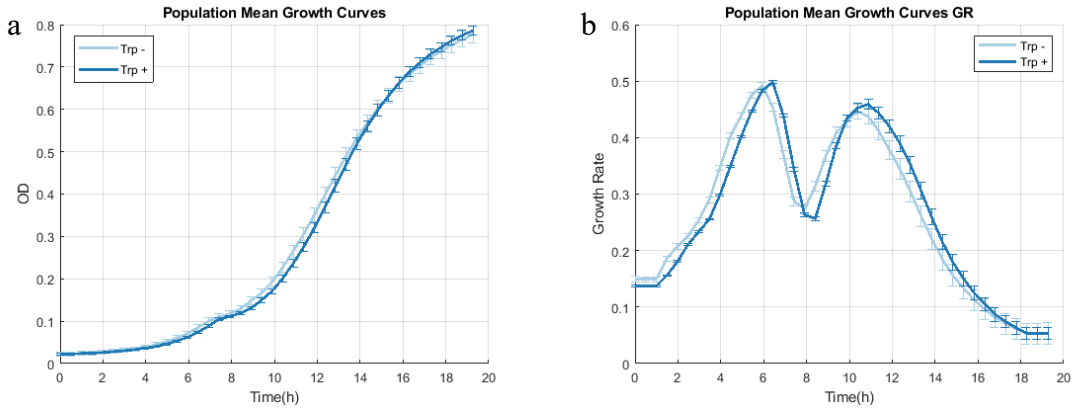


***Figure 19**: the growth curve (a) and growth rate curve (b) of both TrpR+ and TrpR- strains, without any attenuator variant plasmid, over 19 hours in a growth robot.*

Generally, it seems that both curves behave in a very similar manner. We could observe the canonic lag, growth and stationary phase, although after ~6-7 hours it seems like there was a temporary decrease in growth rate. The decrease was recovered quickly, and the growth curve continued as expected with no significant difference between either strains. Maximal growth rate in the experiments' conditions was achieved after 6 hours with a slightly lower pick after 11 hours.

In the following set of experiments, we measured and compared the expression levels in both TrpR+ and TrpR- strains, across all attenuator variants and in different Trp concentration conditions.

The first of this set of experiments examined all variants after 16 hours of incubation in different Trp concentrations:

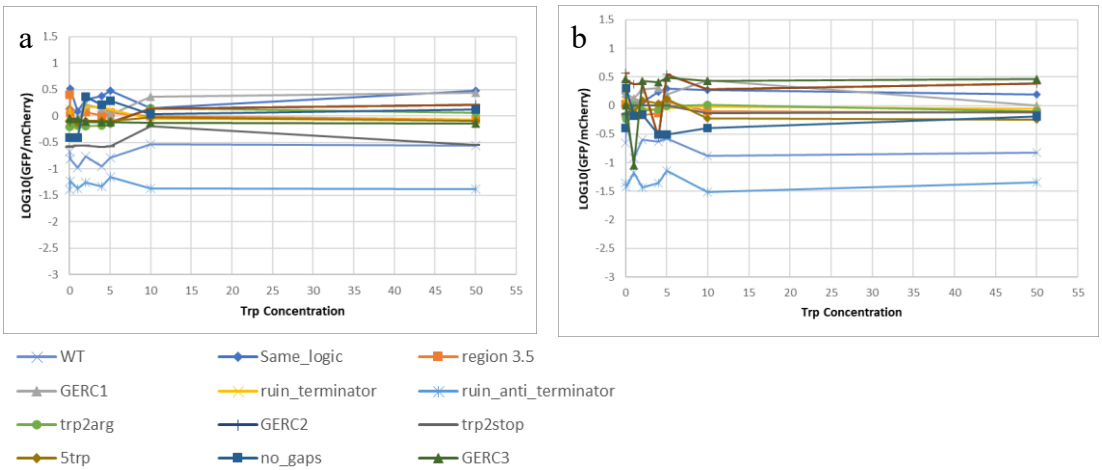


***Figure 20**: the normalized expression of the different variants of the attenuator plasmid in the TrpR+ strain (left) and the TrpR- strain (right) as a ratio of GFP/mCherry on a log scale, on different Trp concentrations. Each color represents a different attenuator variant. The measures were taken after 16 hours of exposure to the different Trp concentrations. (a) TrpR+ strain (b) TrpR- strain.*

In both strains, we see a difference between some of the variants, with Ruin anti-terminator having the lowest expression, and the WT attenuator as second lowest. The constitutive expression controls (GERC at the figure legend) had the highest expression in most of the concentrations in both TrpR+ and TrpR- strains, and the rest of the variants expressed lower than the control but more than the WT throughout the Trp concentration range. We did not observe a clear concentration-dependent step function as we saw in the previous experiments, not in the TrpR+ strain that has already shown this behavior (Fig. 18), and not in the TrpR- strain.

In the following experiment, we wanted to measure the expression over time in the TrpR- strain, using the growth robot and the plate reader. Due to technical reasons, we only have measured well-under 16 hours of incubation with different Trp concentrations (the latest is 5 hours) and another time point after ~70 hours. Worth mentioning that the Trp2Arg plate was stuck in the plate-reader and therefore wasn't incubated in the humid shaking incubator and dried out after the 5H time point.

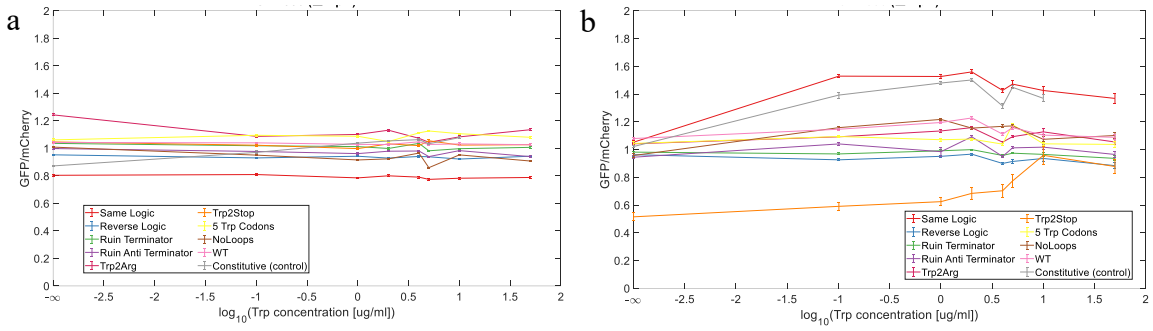


Figure 21*: the normalized expression of the different variants of the attenuator plasmid in the TrpR- strain as a ratio of GFP/mCherry, on different Trp concentrations. Each color represents a different attenuator variant (a) expression after 5 hours of exposure to the different Trp concentrations (b) after 70 hours.*

We can see that after five hours of incubation, most variants have similar expression patterns that do not show great sensitivity to Trp concentration, with Same Logic showing the lowest expression. After 72 hours we see that there is greater variability between the different variants, with Trp to Stop having the lowest expression in concentrations lower than 10 µg/mL, and Same Logic variant just above the constitutive expression control.

The following experiment was performed in order to capture the correct time point to measure both strains using the FACS, and therefore was designed with only four variants (including the WT and control), that were measured in three time points. Ruin Anti-Terminator variant was chosen due to its extremely low expression in the previous FACS experiment, and Same Logic variant was chosen due to its similarity to the WT aiming to test both similar and very different variants within the constraints of the experiment).

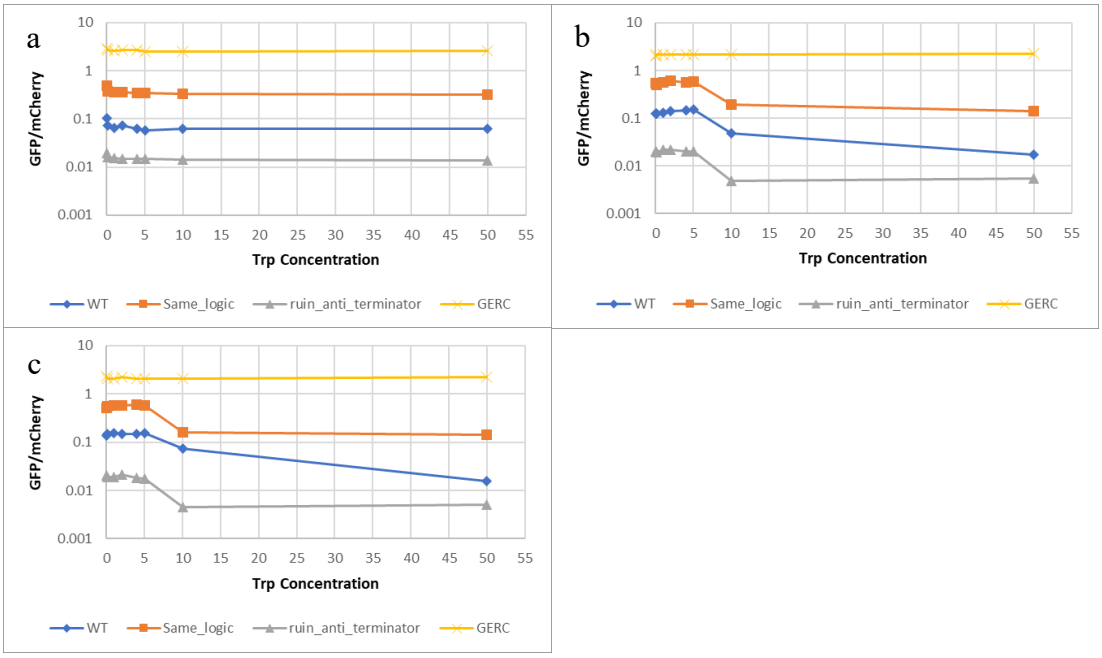


***Figure 22**: the normalized expression of four variants of the attenuator plasmid in the TrpR+ strain as a ratio of GFP/mCherry, on different Trp concentrations. Each color represents a different attenuator variant (a) expression after 2 hours of exposure to the different Trp concentrations (b) after 16 hours (c) after 24 hours.*

From this set we can see that the control variant is not sensitive to Trp concentration, while the other variants demonstrate a clear step function after 16 and 24 hours, but not after 2 hours. Interestingly, each of the variants has a different level of expression, which can go up to an order of magnitude difference. Ruin anti terminator had the lowest expression, the WT had higher expression but still lower than Same Logic variant.

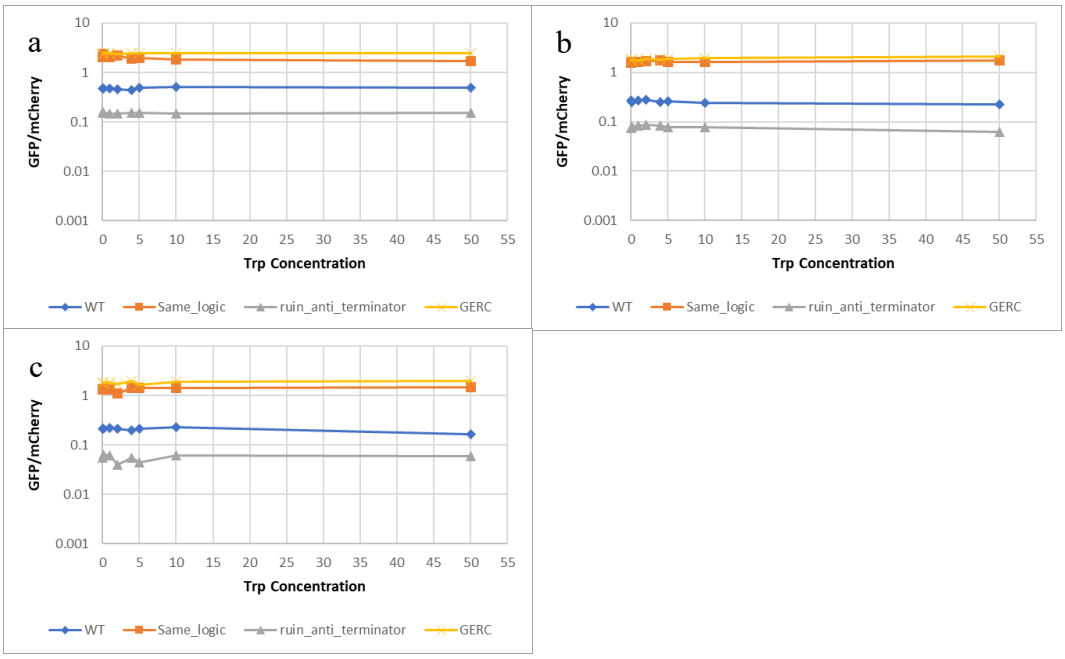


***Figure 23**: the normalized expression of four variants of the attenuator plasmid in the TrpR- strain as a ratio of GFP/mCherry, on different Trp concentrations. Each color represents a different attenuator variant (a) expression after 2 hours of exposure to the different Trp concentrations (b) after 16 hours (c) after 24 hours.*

In the TrpR- set, we can still observe the same ranking between the different variants, with Ruin Anti-Terminator having the lowest expression, WT above it, and Same Logic above both and similar to the control. Here, we do not see the concentration-dependent step function as we saw in the TrpR+ strain, and we see almost no response to the different concentrations, even after 24 hours.

Interestingly, it seems that the absolute values of expression are higher in the second set of measurements, so we chose to create another set of figures with all four variants in both strains together, to observe the results from both strains together:

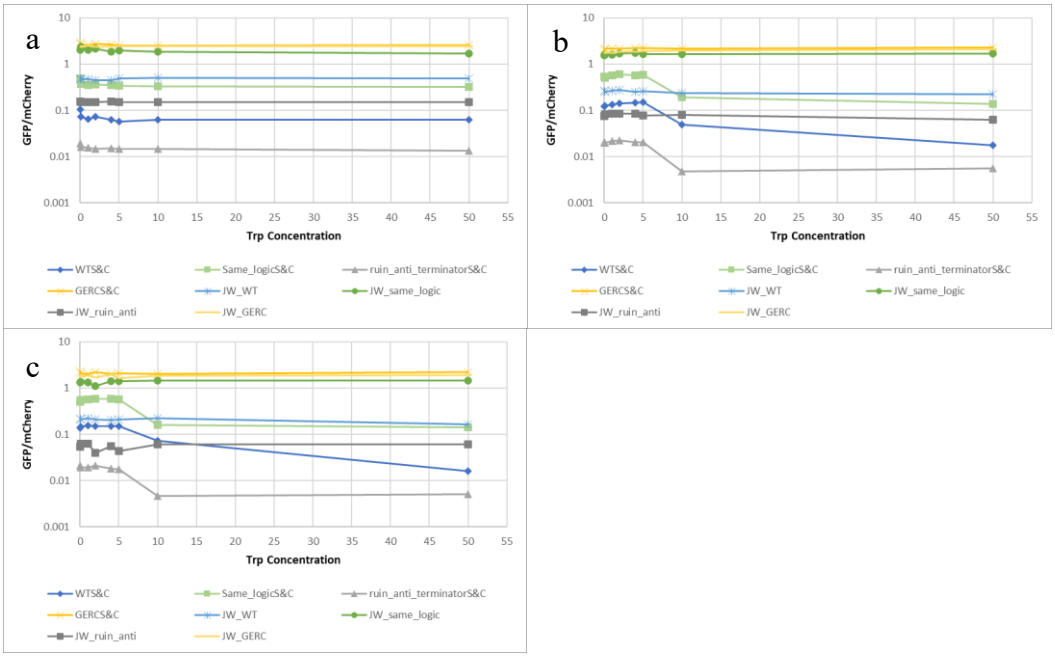


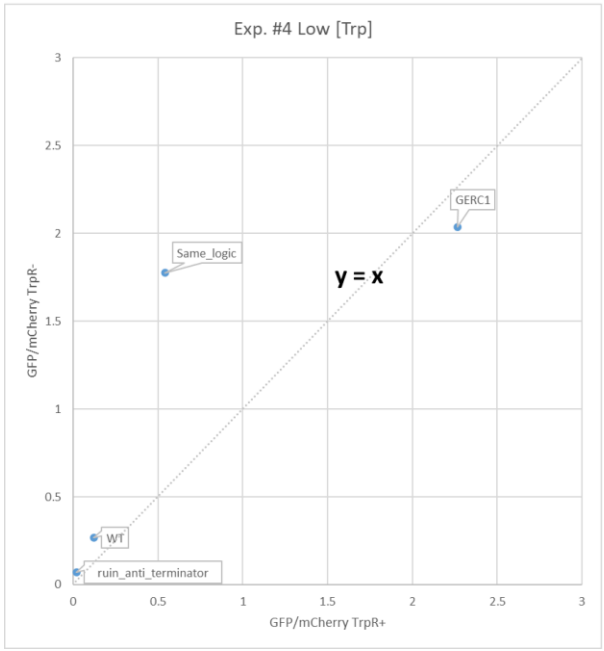
***Figure 24**: the normalized expression of four variants of the attenuator plasmid in the TrpR- strain, and the same variants in the TrpR+ strain, as a ratio of GFP/mCherry, on different Trp concentrations. Each color represents a different attenuator variant, and each attenuator variant has a similar color in both TrpR+ and TrpR- strains: The WT variant is in dark blue in the TrpR+ strain, and light blue in the TrpR- strain. In the legend the S&C stands for the TrpR+ strain, while the JW stands for the TrpR- strain (a) expression after 2 hours of exposure to the different Trp concentrations (b) after 16 hours (c) after 24 hours.*

In the 2 hours figure (a) we can already observe the lower expression levels in the TrpR+ strain when we compare each attenuator variant in both repressor strains. In other words, the TrpR- variants always have a higher expression in comparison to the same variant in the TrpR+ strain, excluding the controls which keep a similar level of expression regardless of the Trp concentration. In the 16 hours and 24 hours we see a similar pattern. It seems that both the repressor and the attenuator impact the level of expression, but while the repressor is causing a step function, the attenuator is impacting the baseline from which the expression is reduced in higher Trp concentrations.

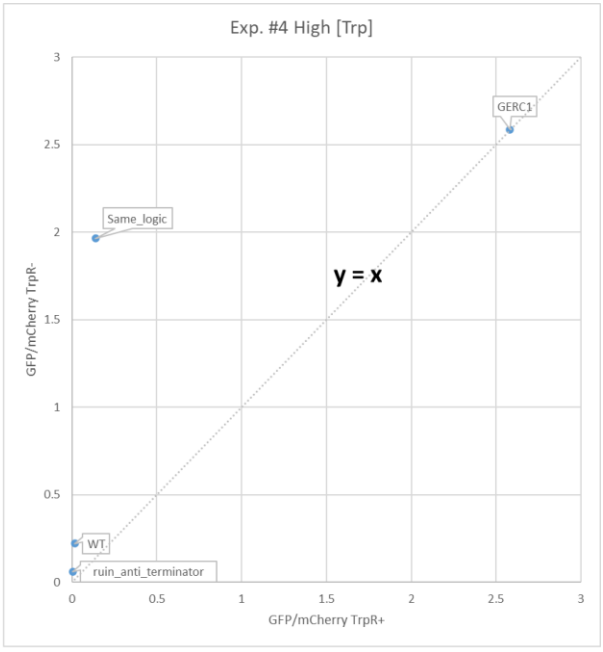
In order to make sure this is not a result of a change in mCherry expression (also expressed constitutively), we have compared the mCherry expression alone between the different variants after 16 hours in the TrpR+ strain (supplementary figure 1). Indeed, we have found that in all variants the expression seems to be similar, with minor changes that cannot explain the previous figures.

Another representation of the experiments' results was created to compare the expression in a certain concentration over the different variants, in both TrpR+ and TrpR- strains. The diagonal represents a hypothetical situation where the expression in both strains was identical for a certain variant. If a variant is above the diagonal, it means that it has a higher expression at the TrpR- strain, while under the diagonal means the expression is higher at the TrpR+ strain.

a



b



c

	Low [Trp]		High [Trp]	
	TrpR+	TrpR-	TrpR+	TrpR-
WT	0.1225	0.2666	0.0173	0.2207
Same_logic	0.5405	1.7765	0.1384	1.9654
GERC1	2.2672	2.0352	2.5805	2.5863
ruin_anti_terminator	0.0202	0.0703	0.0055	0.0611

Figure 25*: the normalized expression of four variants of the attenuator plasmid in the TrpR- strain (y axis) and the TrpR+ strain (x axis), as a ratio of GFP/mCherry. (a) in low Trp concentration (0 µg/mL) (b) in high Trp concentration (50 µg/mL) (c) a table presenting the expression values of each variant in both strains and both concentrations.*

In both high and low Trp concentrations, the WT variant is slightly above the diagonal, while Same Logic is a lot further. Ruin Anti-Terminator variant is also a bit above the diagonal, and the only variant that is on the diagonal or below is the positive control (GERC).

Next, we repeated the last experiment with all 10 attenuator variants (8 designs, 1 WT and 1 control). In this experiment we also gave up the highest (50µg/mL) concentration to gain better understanding of the dynamics in mid-range concentrations:

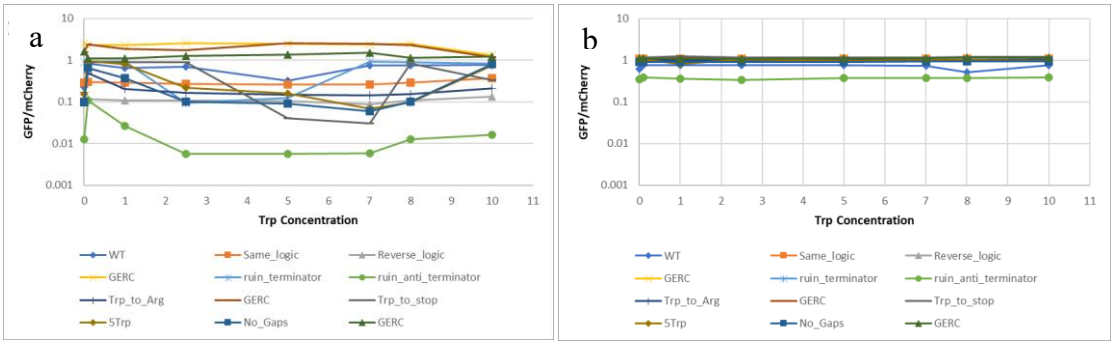
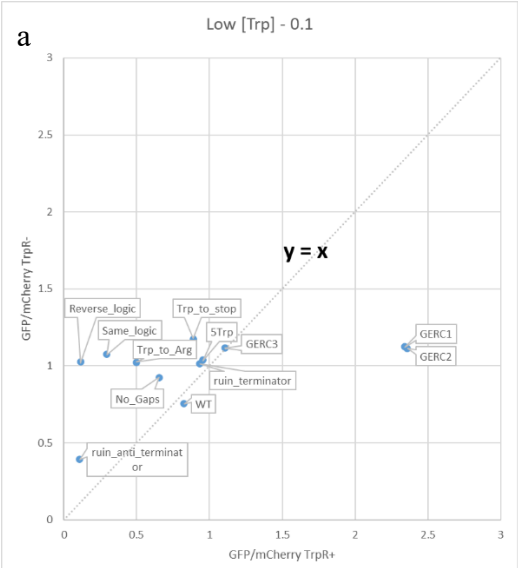


***Figure 26**: the normalized expression of four variants of the attenuator plasmid, as a ratio of GFP/mCherry, after 16 hours on different Trp concentrations. Each color represents a different attenuator variant (a) expression in the TrpR+ strain (b) in the TrpR- strain.*

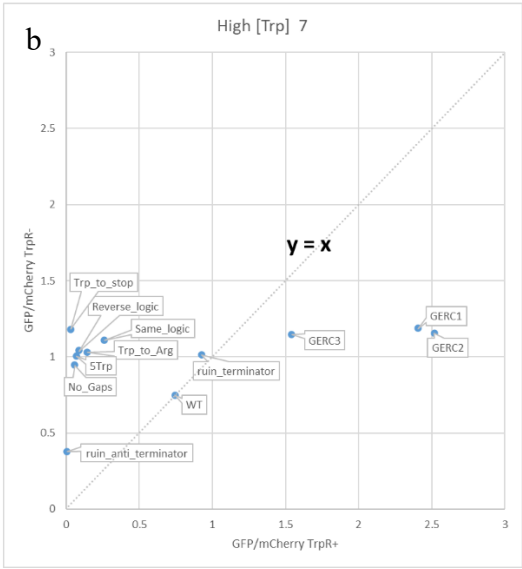
Reproducing previous results, Ruin Anti-Terminator variant has the lowest expression in both strains, as the constitutive expression controls show the highest. In the TrpR+ strain we did not see the clear step function we saw previously. We did observe a general decrease dynamics for some of the variants, but in many cases, we also saw an increase towards the highest Trp concentrations. On the TrpR- strain, it is difficult to see any concentration-dependent dynamics, but the WT variant seems to be lower than the rest who are not Ruin Anti-Terminator or control variants.

Here too, we wanted to compare the expression in a certain concentration over the different variants, in both TrpR+ and TrpR- strains. Due to the fluctuation in expression in the TrpR+ strain and our will to compare an expression level that represents similar behaviors to those of the previous experiments, we chose to use the concentrations of 0.1 and 7 µg/mL for the following analysis:

a



b



c

	Low [Trp]		High [Trp]	
	TrpR+	TrpR-	TrpR+	TrpR-
WT	0.8252	0.7529	0.7426	0.7478
Same_logic	0.2964	1.0753	0.262	1.1076
Reverse_logic	0.1167	1.0263	0.0866	1.0422
GERC1	2.3414	1.1233	2.4059	1.1869
ruin_terminator	0.9326	1.0153	0.9244	1.0133
ruin_anti_terminator	0.1087	0.3915	0.0059	0.3767
Trp_to_Arg	0.4995	1.024	0.1431	1.0294
GERC2	2.358	1.1121	2.5171	1.1539
Trp_to_stop	0.8876	1.1745	0.0304	1.1779
5Trp	0.9533	1.0371	0.0698	1.0056
No_Gaps	0.6546	0.9232	0.0587	0.9457
GERC3	1.1088	1.1167	1.5419	1.1448

Figure 27: *the normalized expression of the attenuator plasmid variants in the TrpR- strain (y axis) and the TrpR+ strain (x axis), as a ratio of GFP/mCherry. (a) in low Trp concentration (0.1 µg/mL) (b) in high Trp concentration (7 µg/mL) (c) a table presenting the expression values of each variant in both strains and both concentrations.*

In these concentrations, we can see again that the three controls are located below or on the diagonal, while almost all the rest of the variants are above it, indicating, high expression of these attenuators in the absence of the repressor (TrpR-).

Ruin Anti-Terminator is again located in the bottom left, separated from other variants as well. The WT variant is located nearly on the diagonal line on both concentrations.

In low Trp concentration, it is interesting to notice that all of the other variants are located on a relatively horizontal cloud of points, which suggests that the governing force in this concentration was the repressor, while at high Trp concentration these variants seem to have similar expression both in the presence and absence of the repressor.

Discussion

More than anything, this work highlights the complexity of biological regulation mechanisms, and the humble approach we should take as we are trying to study these types of mechanisms.

The MSA we performed to compare and study the evolutionary differences between natural variants (fig. 7-11) raised a few insights. First, we have found that there is a large variety among Gram-negative bacteria, even within some of the families, like Vibrionaceae and Enterobacteriaceae. In other families, we have seen a high level of conservation. Yet, the Trp codons, terminator and anti-terminator regions were highly conserved even within the families that had the largest diversity. This might indicate that there are active evolutionary forces that limit the changes in key attenuator features, and that in most cases the attenuator sequence is located on one of the peaks of its fitness landscape. Having said that, we have seen that some regions (loops, non-Trp coding region of TrpL) are less conserved, which may create fitness landscape bridges across some of the fitness valleys.

Close to these conserved and not-conserved regions, there are the semi-conserved parts of this mechanism: the location of the start codon, the location of the stop codon, the number of Trp codons and the spacing between them. These three features were found in many variations among different species, even within the same family, but within certain limits. We did not observe more than 5 Trp codons, leader peptide was not longer than 45 AAs, etc. (see supp. table 1).

These findings may suggest more flexible evolutionary constraints, which might be adaptable to different niches. Although we could not find a clear relationship between these features and the known habitats of the strains that were investigated, we feel strongly that there is more information to uncover in this direction. A future analysis that will investigate a population-wide variety across time and conditions could probably provide additional insights into these dynamics.

The importance of the leader peptide length came up as an extremely interesting topic, since we can speculate on different forces that might act on it. The cost of expression would be higher as the leader peptide is longer, especially since as far as we know, this peptide has no additional function. On the other hand, its length is a good method to

control attenuation timing, since the probability of terminator formation is dependent on the time that the ribosome reaches the Trp codons, the availability of charged Trp-tRNA, and the location of the RNA polymerase at this time. Interestingly, among Vibrionaceae we observed both longer leader peptide, as well as enrichment of Trp codons (five) with some spacing between them. We have speculated that combining these two unique features might offer a quite different phenotype in comparison with the canonic *E. coli* sequence. A more expensive sequence to express, but more flexible in terms of fitness landscape, can indicate a need for rapid evolutionary adaptation for example. Each species needs to find the best balance between the probability of producing unnecessary Trp when it is not needed, the probability of not producing it when needed, and the cost of regulation to achieve the right balance.

Within Enterobacteriaceae, there was no variation in the number or distance of Trp codons from the stop codon, or in the terminator region, but we did see that in many of the species there were two alternative in-frame start codons. Whenever this occurred, the sequence was conserved, which might suggest that it was serving a purpose. We did not examine this further during this work, but this finding resonates with few published findings for alternative translation initiation sites [47], both with the AUG codon [48] and alternative start codons [49] and of the increasing understanding on the importance of alternative splicing [50] and alternative promoters [51]. Moreover, most of the sequences that had an alternative initiation site, also had an enrichment of Lysin codon in the beginning of the leader peptide. Speculatively, this might be a possible sensing mechanism for Lysin, but we did not follow this path in this work.

Among the Vibrionaceae family we also encountered variability in the start codon location relative to the Trp codons location. Interestingly, the 2-1-2 Trp codon structure is highly conserved among all the sequences we aligned, regardless of the location of the start codon. The distance between the last Trp codon to the Stop codon was conserved as well, although two different codons are used among this family to stop the translation of the leader peptide. Taken together, these two notions may suggest that under the natural conditions of this family, the distance from the start codon has weaker selection forces or flatter fitness landscape than the number of Trp codons or the distance between them and the stop codon. This might suggest that there was a selection for timing, since taking the current understanding of the Trp attenuation mechanism,

we could not find another reason why adding AAs “in vain” would be more accepted in the beginning of the leader peptide, rather than its end.

The habitat analysis did not provide high correlation to any of the features we assessed, and it may be that the selective forces are more complex, or that the habitat definition should have better resolution in order to uncover its impact on the sequence.

After we chose representative variants from the synthetic library we have designed (fig. 12 and fig. 15), we evaluated their predicted attenuation activity using the PASIFIC software. The scores (table 3) did not reflect our expectations in some cases, as the Ruin Anti-Terminator variant that got a very high attenuation score. The prediction for the WT attenuator got a score that was only slightly above the threshold, and lower than most other variants. These findings, together with our later achieved results, suggest that the software needs additional data and specifications in order to address this specific use case. For instance, the software is currently forcing a circular RNA structure although in the Trp attenuator case it will be in a linear form. In the future, high throughput experiments with large libraries could potentially improve the software’s ability to predict such structures.

During the first calibration experiment (fig. 16) we measured the GFP/mCherry ratio in the TrpR+ strain with the WT attenuator plasmid across different Trp concentrations which span over 3.5 orders of magnitude. It demonstrated a clear separation between high and low concentrations, with low and high expression of GFP (respectively), but only after 9 hours or more. This result suggests that it takes a long time for the cells to respond to the different concentrations in the conditions of our experiment, and that if we measure earlier, we might not see the impact of the different concentration, rather than the concentration it experienced before the beginning of the experiment.

In our various experiments, we observed many results that were difficult to settle with one another, or with previous publications. When we examined the constitutive expression plasmid in the TrpR+ strain (fig. 17), we have noticed fluctuations in expression levels across different time points, and some separation between the different concentrations after over 10 hours when the highest expression was observed in the highest Trp concentration. This finding contradicts what was expected according to literature. The reason for this conflict is unknown to us, but it seems that we are not

the only ones who encountered it but could not explain it ("the explanation for this difference is unknown") [52], which strengthened our notion that this is not an artifact and further investigation is necessary.

Our continuous expression experiment in TrpR+ strain with all experimental variants (fig. 18) was a great demonstration of the expected repression behavior, which appeared as a step function between 1 to 5 µg/mL. All variants (excluding the constitutive expression control) had higher expression in low Trp concentrations, but not necessarily at the same level. WT attenuator variant demonstrated the highest expression levels in both repeats, while the rest of the variants demonstrated lower expression levels, but their order was not consistent across both repeats and the difference between them was lower than the difference from WT. This interesting result is consistent with later achieved results, in which the difference between variants had an impact on the expression levels even at the presence of the repressor. Interestingly, Ruin Anti-Terminator shows the lowest drop in expression in high Trp concentrations. This might be a result of its designed behavior that causes constant termination as we have seen in additional experiments, but here we see that it maintains an enigmatic, relatively high, level of expression. Also, WT having the highest expression levels at low concentrations was surprising to us, and inconsistent with most of the other results where WT usually had low expression levels in comparison with other variants (excluding Ruin Anti Terminator).

Later we started experimenting with the TrpR- strain and the TrpR+ strain in parallel (fig. 20, 24-27). Surprisingly, we did not see a clear step function of reduced expression in high Trp concentrations in the TrpR- strain, regardless of the attenuator variant or incubation time.

This could not be explained by different growth curves, since in our experiments both strains demonstrated very similar growth (fig. 19). Other possible explanations might be the time of incubation and the Trp concentrations we tested. Trp concentration is a less likely explanation since our experiments span across 2.5 orders of magnitude (0.1 up to 50 µg/mL). As for the incubation time, we have reached as far as 24 hours with the TrpR- variants and did not observe the effect. Since the attenuator is sensitive to tRNATrp and not directly to Trp concentration, it might take a bit longer for the effect to manifest, but 24 hours should have been sufficient to observe the results, especially

since we do see GFP and mCherry expression so we know the cells are still actively expressing genes, and according to our results they express more than their equivalent Trp+ variants. We do not possess any alternative explanation to this phenomenon at the moment, and additional experiments are necessary in order to understand it further. In retrospective, might be recommended that future experiments will use Trp-less GFP [55] and Trp-less mCherry, since they both have one and three Trp codons (respectively) and it might have some impact on the results.

We also saw that the expression level of Ruin Anti Terminator was lower than all other variants, which is an expected result due to its constant termination predicted activity. WT attenuator variant demonstrated a higher expression than Ruin Anti terminator, but Same Logic expressed even more. This result is interesting since Same Logic was expected to act similarly to WT according to its sequence. It might suggest that our hidden hypothesis is not reflecting the reality of attenuation mechanism. We hypothesized that if we do not change the base-pairing between the different regions and just change the nucleotide sequence without harming the start/Trp/stop codons, we will not change the attenuation activity, but the difference between Same Logic and WT suggests otherwise. It may be interesting to conduct a future experiment, where we make an array of changes in a similar manner to Same Logic but change only part of the relevant nucleotides each time to explore the fitness landscape of this feature.

In our final experiment (fig. 26-27) we saw again that Ruin Anti Terminator has the lowest expression, regardless of the repressor strain it was cloned into. These results strengthen our notion that this attenuator variant is indeed causing strong attenuation of transcription in all examined Trp concentrations. As we noted earlier, this result is not in line with one of the previous experiments (fig. 18), but all other experiments yielded the same outcome in which Ruin Anti Terminator consistently showed the lowest expression. Also, we have seen that in the TrpR+ strain the different attenuator variants expressed different GFP levels, while in the TrpR- strain most of them had very similar expression levels. Only Ruin Anti Terminator (lowest) and WT (second lowest) were separated from the rest of the group. It is also shown nicely in fig. 27, where we observed a relatively large separation between the different variants across the X axis (but not Y) in low Trp concentration, while there was large separation on the Y axis (but not X) in high Trp concentration. This result highlights the importance of

measuring expression over various Trp concentrations and in the presence and absence of the repressor, when we are aiming to understand the fitness landscape of the Trp attenuator.

Overall, our experiments suggest that that both the repressor and the attenuator have an impact on expression, but while the repressor is causing a step function in high Trp concentrations, the attenuator variants are impacting the baseline of expression levels. This intriguing and consistent result of this study, has puzzled us since we have first encountered it, and it rises many questions on our understanding of the relationship between the repressor and the attenuator of the Trp operon. Since WT had the lowest expression levels of all the variants except Ruin Anti Terminator, it suggests that the evolutionary forces drove the attenuator to favor low expression in high Trp concentrations over high expression in low Trp concentrations. It is implied that the attenuator's role is to limit the expression levels while the repressor's role is to respond quickly to changes in Trp concentrations. This explanation resonates with some literature findings as well [54].

Future plans

In order to gain confidence in our observations from this study, it is suggested to conduct many more repeats of the major experiments with all variants. It would also be interesting to use the growth robot to measure the expression over time for longer than 24 hours with both TrpR strains. It is also suggested to add a measurement using RT-qPCR, to measure real-time expression levels in a more direct manner, which will be valuable due to the importance of the connection between transcription and translation in the attenuation mechanism.

After achieving consistent results using this system, it is suggested to broaden the concept to a large library containing thousands of variants, conduct competition experiments, and use FACS sorting methodologies to uncover the impact of the different variants directly on fitness, in different conditions. Another possible follow up would be to control the expression of the Trp operon and measure the fitness in different conditions and in alternating conditions to better simulate natural conditions.

As we have conceptualized the library, it is highly recommended that it will have a mixture of synthetic and natural variants. The results of such a series of experiments

could uncover many features in the fitness landscape of the canonic Trp attenuator, and also provide the data to improve computational attenuator-prediction tools.

There is also a possible bioinformatic path, in which it is recommended to conduct a conservation analysis of the Trp attenuator vs. highly and lowly expresses genes to explore the importance of this regulatory element's sequence and structure. Later, it is possible to explore additional attenuators of different AAs and compare between them in different species. It will be interesting to explore all the features of the synthetic library in a bioinformatic manner as well, and maybe later even conduct a lab evolution experiment in different conditions to further see the impact of each condition on the different features.

References

[1] G. T. Akashi H, "Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis," *Proc Natl Acad Sci U S A,* vol. 99, no. 6, pp. 3695-3700, 2002.

[2] J. D. W. R. M. L. R. T. E. D. R. V. M. D. E. K. Esley M. Heizer, "Amino Acid Cost and Codon-Usage Biases in 6 Prokaryotic Genomes: A Whole-Genome Analysis," *Molecular Biology and Evolution,* vol. 23, no. 9, p. 1670–1680, 2006.

[3] C. S. S. R. U. a. S. S. Kaleta, "Metabolic costs of amino acid and protein production in Escherichia coli," *Biotechnology Journal,* vol. 8, no. 9, pp. 1105-1114, 2013.

[4] Histidine, "Wikimedia," 2011. [Online]. Available: https://commons.wikimedia.org/wiki/File:Trpoperon.svg. [Accessed 2018].

[5] H. V. Yanofsky C, "Role of regulatory features of the trp operon of Escherichia coli in mediating a response to a nutritional shift," *J Bacteriol,* vol. 176, no. 20, pp. 6245-6254, 1994.

[6] C. K. R. L. &. H. V. Yanofsky, "Repression is relieved before attenuation in the trp operon of Escherichia coli as tryptophan starvation becomes increasingly severe," *Journal of bacteriology,* vol. 158, no. 3, p. 1018–1024, 1984.

[7] D. D. S. M. S. R. Millman A, "Computational prediction of regulatory, premature transcription termination in bacteria," *Nucleic Acids Res,* vol. 45, no. 2, pp. 886-893, 2017.

[8] K. Academy, "www.khanacademy.org," [Online]. Available: https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/regulation-of-gene-expression-and-cell-specialization/a/the-trp-operon. [Accessed 2018].

[9] Y. C. C. K. P. L. Landick R, "Replacement of the Escherichia coli trp operon attenuation control codons alters operon expression," *J Mol Biol,* vol. 216, no. 1, pp. 25-37, 1990.

[10] S. S. R. U. S. S. Kaleta C, "Metabolic costs of amino acid and protein production in Escherichia coli," *Biotechnol J,* vol. 8, no. 9, pp. 1105-1114, 2013.

[11] Y. C. Jackson EN, "Thr region between the operator and first structural gene of the tryptophan operon of Escherichia coli may have a regulatory function," *J Mol Biol,* vol. 76, no. 1, pp. 89-101, 1973.

[12] J. R. Y. C. Merino E, "Evolution of bacterial trp operons and their regulation," *Curr Opin Microbiol,* vol. 11, no. 2, pp. 78-86, 2008.

[13] K. M. Y. C. H. D. Shimotsu H, "Novel form of transcription attenuation regulates expression the Bacillus subtilis tryptophan operon," *J Bacteriol,* vol. 166, no. 2, pp. 461-471, 1986.

[14] G. P, "Regulation of the Bacillus subtilis trp operon by an RNA-binding protein," *Mol Microbiol,* vol. 11, no. 6, pp. 991-997, 1994.

[15] G. P. McAdams NM, "The Bacillus subtilis TRAP protein can induce transcription termination in the leader region of the tryptophan biosynthetic

(trp) operon independent of the trp attenuator RNA," *PLoS One,* vol. 9, no. 2, p. e88097, 2014.

[16] Y. C, "Transcription attenuation: once viewed as a novel regulatory strategy," *J Bacteriol,* vol. 182, no. 1, pp. 1-8, 2000.

[17] S. Wright, "The roles of mutation, inbreeding, crossbreeding, and selection in evolution," in *Proceedings of the Sixth International Congress on Genetics*, 1932.

[18] K. J. de Visser JA, "Empirical fitness landscapes and the predictability of evolution," *Nat Rev Genet,* vol. 15, no. 7, pp. 480-490, 2014.

[19] L. Y. J. J. H. R. Weinreich DM, "The Influence of Higher-Order Epistasis on Biological Fitness Landscape Topography," *J Stat Phys,* vol. 172, no. 1, pp. 208-225, 2018.

[20] B. D. M. M. e. a. Sarkisyan KS, "Local fitness landscape of the green fluorescent protein," *Nature,* vol. 533, no. 7603, pp. 397-401, 2016.

[21] B. A. D. L. M. L. D. B. C. Fragata I, "Evolution in the light of fitness landscape theory," *Trends Ecol Evol,* vol. 34, no. 1, pp. 69-82, 2019.

[22] Q. W. M. C. Z. J. Li C, "The fitness landscape of a tRNA gene," *Science,* vol. 352, no. 6287, pp. 837-840, 2016.

[23] J. J. B. D. Hietpas RT, "Experimental illumination of a fitness landscape," *Proc Natl Acad Sci U S A,* vol. 108, no. 19, pp. 7896-7901, 2011.

[24] S. J. F. A. W. T. Soo VWC, "Fitness landscape of a dynamic RNA structure," *PLoS Genet,* vol. 17, no. 2, p. e1009353, 2021.

[25] C. B. C. H. T. D. S. G. K. G. Puchta O, "Network of epistatic interactions within a yeast snoRNA," *Science,* vol. 352, no. 6287, pp. 840-844, 2016.

[26] S. D. R. A. e. a. Frumkin I, "Gene Architectures that Minimize Cost of Gene Expression," *Mol Cell,* vol. 65, no. 1, pp. 142-153, 2017.

[27] S. M. M. J. e. a. Dar D, "Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria," *Science,* vol. 352, no. 6282, p. aad9822, 2016.

[28] K. S. A. R. e. a. Köksaldı İÇ, "SARS-CoV-2 Detection with De Novo-Designed Synthetic Riboregulators," *Anal Chem,* vol. 93, no. 28, pp. 9719-9727, 2021.

[29] M. G. d. B. M. G. T. Bartoli V, "Tunable genetic devices through simultaneous control of transcription and translation," *Nat Commun,* vol. 11, no. 1, p. 2095, 2020.

[30] W. B. Datsenko KA, "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products," *Proc Natl Acad Sci U S A,* vol. 97, no. 12, pp. 6640-6645, 2000.

[31] A. T. H. M. e. a. Baba T, "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection," *Mol Syst Biol,* vol. 2, p. 2006.0008, 2006.

[32] D. Hanahan, "iologically pure escherichia coli cell line which is a deoR- mutant and which is more transformation efficient with foreign plasmids than deoR+ escherichia coli cell lines, processes for obtaining these cell lines, methods of use". United States Patent US4851348A, 1986.

[33] R. R. S. J. L. K. Guyer MS, "Identification of a sex-factor-affinity site in E. coli as gamma delta," *Cold Spring Harb Symp Quant Biol,* vol. 45, no. 1, pp. 135-140, 1981.

[34] p.-G. w. a. g. f. G. C. (. p. #. 47441, http://n2t.net/addgene:47441 and RRID:Addgene_47441).

[35] W. W. Cherepanov PP, "Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant," *Gene,* vol. 158, no. 1, pp. 9-14, 1995.

[36] H. V. K. K. e. a. Brown GR, "Gene: a gene-centered information resource at NCBI," *Nucleic Acids Res,* vol. 43, no. Database issue, pp. D36-D42, 2015.

[37] P. J. B. G. Troshin PV, "Java bioinformatics analysis web services for multiple sequence alignment--JABAWS:MSA," *Bioinformatics,* vol. 27, no. 14, pp. 2001-2002, 2011.

[38] adimil, "PASIFIC: PASIFIC v1.1," *Zenodo,* 2016.

[39] B. Lab, "FLP Recombination in E. coli," [Online]. Available: https://barricklab.org/twiki/bin/view/Lab/ProcedureFLPFRTRecombination#:~:text=coli,of%20the%20i%3EDeconvoluter%20library. [Accessed 2021].

[40] N. E. BioLabs, "NEBuilder HiFi DNA Assembly Reaction Protocol," [Online]. Available: https://international.neb.com/protocols/2014/11/26/nebuilder-hifi-dna-assembly-reaction-protocol. [Accessed 2018].

[41] N. E. BioLabs, "Golden Gate Assembly Protocol for Using NEB Golden Gate Assembly Mix (E1600)," [Online]. Available: https://international.neb.com/protocols/2015/03/04/golden-gate-assembly-protocol-for-using-neb-golden-gate-assembly-mix-e1600. [Accessed 2018].

[42] B. A. P. G. H. K. A. a. B. J. Babu, "Identification of candidate gene–based SSR markers for lysine and tryptophan metabolic pathways in maize (Zea mays)," *Plant Breeding,* vol. 131, pp. 20-27, 2012.

[43] A. F. Romero PA, "Exploring protein fitness landscapes by directed evolution," *Nat Rev Mol Cell Biol,* vol. 10, no. 12, pp. 866-876, 2009.

[44] E. A. L. &. W. T. Wrenbeck, "Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded," *Nat Commun,* vol. 8, p. 15695, 2017.

[45] F. A. M. A. e. a. Gonzalez Somermeyer L, "Heterogeneity of the GFP fitness landscape and data-driven protein design," *Elife,* vol. 11, p. e75842, 2022.

[46] J. J. B. D. Hietpas RT, "Experimental illumination of a fitness landscape," *Proc Natl Acad Sci U S A,* vol. 108, no. 19, pp. 7896-7901, 2011.

[47] A. A. D. R. Routh S, "A two-step PCR assembly for construction of gene variants across large mutational distances," *Biol Methods Protoc,* vol. 6, no. 1, p. bpab007, 2021.

[48] D. O. P. Y. Gingold H, "Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome," *Nucleic Acids Res,* vol. 40, no. 20, pp. 10053-10063, 2012.

[49] L. B. L. S. H. S. S. B. Q. S. Lee S, "Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution," *Proc Natl Acad Sci U S A,* vol. 109, no. 37, pp. E2424-E2432, 2012.

[50] K. A. Bazykin GA, "Alternative translation start sites are conserved in eukaryotic genomes," *Nucleic Acids Res,* vol. 39, no. 2, pp. 567-577, 2011.

[51] P. T. e. al, "Alternative Translation Initiation Generates a Functionally Distinct Isoform of the Stress-Activated Protein Kinase MK2," *Cell Reports,* vol. 27, no. 10, pp. 2859-2870, 2019.

[52] J. U. a. B. J. Blencowe, "Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution," *Molecular Cell,* vol. 76, no. 2, p. 329–345, 2019.

[53] K. H. Wrighton, "Shedding light on alternative promoter selection," *Nat Rev Genet,* vol. 19, no. 1, p. 4–5, 2018.

[54] R. L. K. a. V. H. C. Yanofsky, "Repression is relieved before attenuation in the trp operon of Escherichia coli as tryptophan starvation becomes increasingly severe," *J Bacteriol,* vol. 158, no. 3, p. 1018–1024, 1984.

[55] M. A. A. Y. e. a. Kawahara-Kobayashi A, "Simplification of the genetic code: restricted diversity of genetically encoded amino acids," *Nucleic Acids Res,* vol. 40, no. 20, pp. 10576-10584, 2012.

[56] K. A. Bazykin GA, "Alternative translation start sites are conserved in eukaryotic genomes," *Nucleic Acids Res,* vol. 39, no. 2, pp. 567-577, 2011.

Acknowledgments

I want to thank the unique and extraordinary group of people who embarked with me on this journey.

Adi Milman from the Sorek lab, who provided valuable insights on the computational and bioinformatic approach, as well as guidance on the PASIFIC platform and good will to assist and collaborate.

All Pilpel lab students and staff, who contributed their brilliant minds, positive spirits, and collective experience to improve this research and my M.Sc. life in general.

Dr. Dvir Schirman, who took me under his wings and guided me back to pure scientific work after long years of applicative science. You taught me much about many things, took a pivotal role in conceptualizing this study and its milestones, and mentored me along the first and very challenging parts of my thesis, towards my life as a research student in the lab and the scientific community.

Dr. Orna Dahan, who provided wet lab best practices, smart experimental design and extremely valuable insights to the experimental results, as well as her inspiring leadership in science and in life. I admire your genuine passion and pure heart.

Prof. Tzachi Pilpel, for igniting and nurturing my scientific curiosity and teaching me about excellence in research. For connecting matter and spirit, art and mathematics, and so much more. Words cannot describe how your noble personality motivated me and helped me evolve during my time in the lab, vertically and horizontally. I will forever be grateful and do my best to follow your role model.

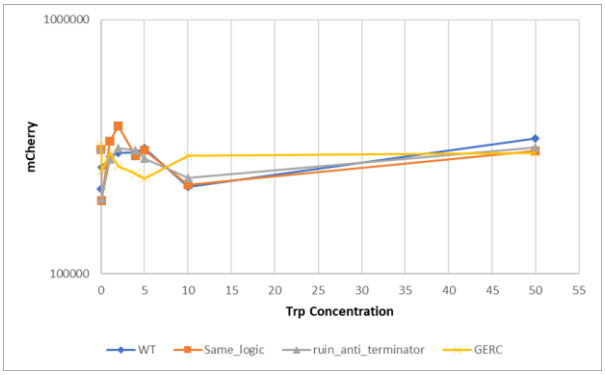
Arch. Rotem Varkovitzky Hausman, for being you. For your sensitive driving forces and caring patience, for being my motivation and endless source of power, and for encouraging me to thrive. I love you.

P.S.

The PCR duckling, who made sure that every reaction would work according to its plan.

Appendix

Supplementary figure 1



The mCherry expression of four variants of the attenuator plasmid in the TrpR+ strain, the measurement was taken after 16 hours of incubation on different Trp concentrations. Each color represents a different attenuator.

Supplementary table 1: List of Species with their TrpL and Trp attenuator sequences:

Species	trpL sequence	Full attenuator sequence
Escherichia coli str. K-12	ATGAAAGCAATTTTCGTACTGAAAGGTTGGTGGCGCACTTCCTGA	ATGAAAGCAATTTTCGTACTGAAAGGTTGGTGGCGCACTTCCTGAAACGGGCAGTGTATTCACCATGCGTAAAGCAATCAGATACCCAGCCCGCCTAATGAGCGGGCTTTTTTTT
Yersinia pestis	ATGAAGACTTCCCTGATTTCCTTACTGCGTTGGTGGCATATCTCCCTCTCTCGTCTCGGGCGATGTAA	ATGAAGACTTCCCTGATTTCCTTACTGCGTTGGTGGCATATCTCCCTCTCTCGGGCGATGTAATCACGCATATCCGTCATCAGACAGTGCAGATTGCTTCAGCCCGCTAATAGCGGGTTTTTTT
Yersinia pseudotuberculosis	ATGAAGACTTCCCTGATTTCCTTACTGCGTTGGTGGCATATCTCCCTCTCTCGTCTCGGGCGATGTAA	ATGAAGACTTCCCTGATTTCCTTACTGCGTTGGTGGCATATCTCCCTCTCTCGGGCGATGTAATCACGCATATCCGTCATCAGACAGTGCAGATTGCTTCAGCCCGCTAATAGCGGGTTTTTTT
Yersinia massiliensis	ATGAAAACGAACCTGATTTCTTTACTGCGTTGGTGGCACATATCCCTCTCACTCTCACGGGCGGTGTAA	ATGAAAACGAACCTGATTTCTTTACTGCGTTGGTGGCACATATCCCTCTCACGGGCGGTGTAATCACGCATACCAGTCATCAGACAGTGCAGATTGCTTCAGCCCGCTAATAGCGGGTTTTTTT
Yersinia aldovae	ATGAAAACTTCCCTGATTTCCTTACTGCGTTGGTGGCACATCTCCCTCTTTTGCTTTTGGGCGATGTAA	ATGAAAACTTCCCTGATTTCCTTACTGCGTTGGTGGCACATCTCCCTCTTTTGGGCGATGTAATCACGCATATCAGTCATCAGACAGTGCAGATTGCTTCAGCCCGCGAATAGCGGGCTTTTTT
Yersinia bercovieri	ATGAAAACCACTCTGATCTCCCTACTGCGTTGGTGGCACATCTCCCTCTCTCTCTCTCGGGCGATGTAA	ATGAAAACCACTCTGATCTCCCTACTGCGTTGGTGGCACATCTCCCTCTCTCGGGCGGTGTAATCGCGCATACCAGTCATCAGACAGTGCAGATTGCTTCAGCCCGCTAATAGCGGGTTTTTTT
Yersinia intermedia	ATGAAAACTTCCCTGATTTCCTTACTGCGTTGGTGGCACATCTCCCTCTCTCGCTCTCGGGCGATGTAA	ATGAAAACTTCCCTGATTTCCTTACTGCGTTGGTGGCACATCTCCCTCTCTCGGGCGGTGTAATCACGCATATCAGTCACCAGACAATGCAGATTGCTTCAGCCCGCGAATAGCGGGCTTTTTT
Yersinia enterocolitica	ATGAAAACTTCCCTAATTTCCTTACTGCGTTGGTGGCACATCTCCCTCTCTCGCTCTCGGGCGATGTAA	ATGAAAACTTCCCTAATTTCCTTACTGCGTTGGTGGCACATCTCCCTCTCTCGGGCGGTGTAATCGCGCATATCAGTCATCAGACAGTGCAGATTGCTTCAGCCCGCTTATAGCGGGCTTTTTT
Yersinia frederiksenii	ATGAAAACTTCCCTGATTTCCTTACTGCGTTGGTGGCACATCTCCCTCACTCCACTCGGGCGATGTAA	ATGAAAACTTCCCTGATTTCCTTACTGCGTTGGTGGCACATCTCCCTCACTCGGGCGATGTAATCACGCATATCAGTCATCAGACAGTGCAGATTGCTTCAGCCCGCTAATAGCGGGTTTTTTT
Citrobacter freundii	ATGAAGAAGGTATCTGAGATGAAAGCAACATTTTTTCTGCACGGTTGGTGGCGCACTTCCTGA	ATGAAGAAGGTATCTGAGATGAAAGCAACATTTTTTCTGCACGGTTGGTGGCGCACTTCCTGATTTCGGGCAGTGTCTTACGTCTGCAATATGCAACCAGATACCCGGCCCGCCAAATGAGCGGGCTTTTTTTT
Citrobacter koseri	ATGAAAGCAACATTCGTACTGCACGGTTGGTGGCGCACTTCCTGA	ATGAAGAAGGTATCTGAGATGAAAGCAACATTCGTACTGCACGGTTGGTGGCGCACTTCCTGATTTCGGGCAGTGTATGACGTCTGCAATAAGCAACCAGATACCCGGCCCGCCAGATGAGCGGGCTTTTTTTT
Citrobacter braakii	ATGAAGAAGGTATCTGAGATGAAAGCAACATTTTTTCTGCACGGTTGGTGGCGCACTTCCTGA	ATGAAGAAGGTATCTGAGATGAAAGCAACATTTTTTCTGCACGGTTGGTGGCGCACTTCCTGATTTCGGGCAGTGTCTTACGTCTGCAATATGCAACCAGATACCCGGCCCGCCAAATGAGCGGGCTTTTTTTT

Citrobacter portucalensis	ATGAAAGCAACATTTGTTCTGCA CGGTTGGTGGCGCACTTCCTGA	ATGAAGAAGGTATCTGAGATGAAAGCAACATTTGTTCTGCACGGTTGGTGG CGCACTTCCTGATTTCGGGCAGTGTCTTACGTCTGCAATATGCAACCAGATA CCCGGCCCGCAAAATGAGCGGGCTTTTTTTT
Citrobacter youngae	ATGAAAGCAACATTTGTCCTGCA CGGTTGGTGGCGCACTTCCTGA	ATGAAAGCAACATTTGTCCTGCACGGTTGGTGGCGCACTTCCTGATTTCGGG CAGTGTCTTACGTCTGCAATATGCAACCAGATACCCGGCCCGCCAAATGAG CGGGCTTTTTTTT
Citrobacter werkmanii	ATGAAAGCAACATTTGTCCTGCA CGGTTGGTGGCGCACTTCCTGA	ATGAAGAAGGTATCTGAGATGAAAGCAACATTTGTCCTGCACGGTTGGTGG CGCACTTCCTGATTTCGGGCAGTGTCTTACGTCTGCAATATGCAACCAGATA CCCGGCCCGCCAACTGAGCGGGCTTTTTTTT
Enterobacter roggenkampii	ATGACAGCACATTTCACTCTGCA CGGCTGGTGGCGTACTTCCTGA	ATGACAGCACATTTCACTCTGCACGGCTGGTGGCGTACTTCCTGATCTTCGG GCAGTGTCACGCGTCTGCGAAATGCAAACAGATACCCGCCCGCTACCCAGC GGGCTTTTTTTTT
Enterobacter ludwigii	ATGACAGCACATTTCACTCTGCA CGGTTGGTGGCGCACATCCTGA	ATGACAGCACATTTCACTCTGCACGGTTGGTGGCGCACATCCTGATTATCGG GCAGTGTCACACGTCTGCGAAATGCAAACAGATACCCAGCCCGCGACTAAG CGGGCTTTTTTTTT
Enterobacter bugandensis	ATGACAGCACATTTCACTCTGCA CGGTTGGTGGCGCACTTCCTGA	ATGACAGCACATTTCACTCTGCACGGTTGGTGGCGCACTTCCTGATCTTCGG GCAGTGTCACGCATCTGCGAAATGCAAACAGATACCCGCCCGCTACCCAGC GGGCTTTTTTTT
Enterobacter cancerogenus	ATGACAGCACATTTCGCTTTGCA CGGTTGGTGGCGCACTTCCTGA	ATGACAGCACATTTCGCTTTGCACGGTTGGTGGCGCACTTCCTGATTATCGG GCGGTGTCACGCGTCTGCGACAAGCAAACAGATACCCGGCCCGCTCCCAAG CGGGCTTTTTTTT
Enterobacter asburiae	ATGACAGCACATTTCACTCTGCA CGGTTGGTGGCGTACTTCCTGA	ATGACAGCACATTTCACTCTGCACGGTTGGTGGCGTACTTCCTGATCATCGG GCAGTGTCACGCGTCTGCGAAATGCAAACAGATACCCGCCCGCTACTCAGC GGGCTTTTTTTT
Enterobacter cloacae	ATGACTGCACATTTCACTCTGCA CGGTTGGTGGCGCACTTCCTGA	ATGACTGCACATTTCACTCTGCACGGTTGGTGGCGCACTTCCTGACTTTCGG GCAGTGTCACACGTCTGCGTACAGCAAACAGATACCCGCCCGCTCACCAGC GGGCTTTTTTTT
Enterobacter hormaechei	ATGACAGCACATTTCGCTCTGCA CGGTTGGTGGCGTACTTCCTGA	ATGACAGCACATTTCGCTCTGCACGGTTGGTGGCGTACTTCCTGATCTTCGG GCAGTGTCACGCGTCTGCGAAATGCAAACAGATACCCGGCCCGCTACCCAG CGGGCTTTTTTTT
Salmonella enterica	ATGAAGAGGGTATCTAAAATGGC AGCGACATTTGCATTACACGGTT GGTGGCGCACTTCCT	ATGAAGAGGGTATCTAAAATGGCAGCGACATTTGCATTACACGGTTGGTGG CGCACTTCCTGATAGCGGGCGGTGTATGAACAGCTGTAATCAGCCAAACGA TACCCGGCCCGCCTGTTAAGCGGGCTTTTTTTT
Salmonella bongori	ATGACAACAACATTTGTACTGCA CGGCTGGTGGCGCACTTCCTGA	ATGACAACAACATTTGTACTGCACGGCTGGTGGCGCACTTCCTGATGTCGG GCGGTGTATAACAGCTGTAATAAGCCAAAAAGATACCCGGCCCGCCTGTCA AGCGGGCTTTTTTTT
Shigella sonnei	ATGAAAGCAATTTTCGTACTGAA AGGTTGGTGGCGCACTTCCTGA	ATGAAAGCAATTTTCGTACTGAAAGGTTGGTGGCGCACTTCCTGAAACGGG CAGTGTATTCACCATGCGTAAAGCAATCAGATACCCAGCCCGCCTAATGAG CGGGCTTTTTTTT
Shigella boydii	ATGAAAGCAATTTTCGTACTGAA AGGTTGGTGGCGCACTTCCTGA	ATGAAAGCAATTTTCGTACTGAAAGGTTGGTGGCGCACTTCCTGAAACGGG CAGTGTATTCACCATGCGTAAAGCAATCAGATACCCAGCCCGCCTAATGAG CGGGCTTTTTTTT
Shigella dysenteriae	ATGAAAGCAATTTTCGTACTGAA AGCTTGGTGGCGCACTTCCTGA	ATGAAAGCAATTTTCGTACTGAAAGCTTGGTGGCGCACTTCCTGAAACGGG CAGTGTATTCACCATGCGTAAAGCAATCAGATACCCAGCCCGCCTAATGAG CGGGCTTTTTTTT
Leclercia adecarboxylata	ATGACCGCACATTTCACTCTGCA TGGCTGGTGGCGCACTTCCTGA	ATGACCGCACATTTCACTCTGCATGGCTGGTGGCGCACTTCCTGAAATTCGG GCAGTGTCGTACGTCTGCAATGTGCAAACAGATACCTGGCCCGCCACTGAG CGGGCTTTTTTTT
Cronobacter sakazakii	ATGACTCACATTTTTTCTCTGCAT GGCTGGTGGCGTACCTCCTGA	ATGACTCACATTTTTTCTCTGCATGGCTGGTGGCGTACCTCCTGATCCCGGG CGGTGTCTTCACGTATGCGCAAGCATTCAGATACCAGGCCCGCTCACCGCG GGCTTTTTTTT
Cronobacter malonaticus	ATGGCTCACATTTTTTCTCTGCAT GGCTGGTGGCGTACCTCCTGA	ATGGCTCACATTTTTTCTCTGCATGGCTGGTGGCGTACCTCCTGATCCCGGG CGGTGTCTTCACGTATGCGCAAGCAATCAGATACCAGGCCCGCTCAGTGCG GGCTTTTTTTT
Raoultella terrigena	ATGAAAACGCAAACTATCACTCT GCACGGCTGGTGGCGCACCTCCT GA	ATGAAAACGCAAACTATCACTCTGCACGGCTGGTGGCGCACCTCCTGATAA CGGGCGGCGTGATCGCGTTTTGCACTCAGCATACAGATACCCGGCCCGCCA ATGAGCGGGCTTTTTATT

Supplementary table 2: List of primers

ID #	Primer name	Used for	Sequence
1	prKEIOtrpE	validation of KEIO trpE strain	GATATTATCGAGCAGCAGAATGTC

2	prKEIOtrpR	validation of KEIO trpR strain	CCCGCTAACAATGGCGAC
3	prKEIOKan	validation of KEIO strains	ACGTGTTCCGCTTCCTTTAG
4	PrVal_pVS101_F	Validation of pVS101	GCCAAATGCGGCCATCTG
5	Gibson_pJ251_Lin_R	Gibson liniarization	CCATGACTAAGCTTTTCATTG
6	Gibson_Trp1_F	Gibson reaction	AATGAAAAGCTTAGTCATGGGGCGCACTCCCGTTCTGG
7	Gibson_pJ251_Lin_F	Gibson liniarization	AACAAAATTAGAGAATAACAATGCGTAAAGGCGAAGAG
8	prVal_pVS101_R	Validation of pVS101	CGTCACCTTCACCCTCGC
9	Gibson_pBAC_Lin_R	Gibson liniarization	TCCCCGGGTACCGAGCTC
10	prVal_pVS201_5R	Validation of pVS201	CCGTTATGATGTCGGCGC
11	prValSa14_pVS201_3F	Validation of pVS201	ACCGTTTCGCTTACCCCG
12	Gibson_pBAC_Lin_F	Gibson liniarization	TCCTCTAGAGTCGACCTGCAG
13	prVal_pVS201_3R	Validation of pVS201	CAGCTATGACCATGATTACGCC
14	Gibson_TrpWT_R	Gibson reaction	TGTTATTCTCTAATTTTGTTCAAAAAAAAGCCC
15	Gibson_TrpOp3_F	Gibson reaction	GAACAAAATTAGAGAATAACAATGCAAAC
16	pr_ValSa1_pVS201	Validation of pVS201	GCAAGGCGATTAAGTTGGG
17	pr_Sa2_pVS201	pVS201 sequencing	AATCCGCAGATATCGACAGC
18	pr_Sa3_pVS201	pVS201 sequencing	CTGCTCGCCTGAACGAAC
19	pr_Sa4_pVS201	pVS201 sequencing	ACCGTTATTCCTATGTGATGCAC
20	pr_Sa5_pVS201	pVS201 sequencing	GAGCAATCCGGTGCTGATG
21	pr_Sa6_pVS201	pVS201 sequencing	CGCTGGTGAGCATGAAAATTC
22	pr_Sa7_pVS201	pVS201 sequencing	GATGAAGTTTCATTACACGCGC
23	pr_Sa8_pVS201	pVS201 sequencing	CATTTATAAACATTACGCTTCGGC
24	pr_Sa9_pVS201	pVS201 sequencing	GAATAAAGTATGTGGCCTGACG
25	pr_Sa10_pVS201	pVS201 sequencing	AAAGACGCACGTCTTTTGG
26	pr_Sa11_pVS201	pVS201 sequencing	ACGCTGAAAGATGCCTGTAAC
27	pr_Sa12_pVS201	pVS201 sequencing	TGAAAATGATGCGCGAAAACC
28	pr_Sa13_pVS201	pVS201 sequencing	ATAATGTCGCACCTATCTTCATC
29	Gibson_TrpOP1_F	Gibson reaction	ATTCGAGCTCGGTACCCGGGGGCGCACTCCCGTTCTGG
30	Gibson_TrpOP3_R	Gibson reaction	CCTGCAGGTCGACTCTAGAGGAAAATTATTGATTTTACTGGTGTTATGTTGCGGG
31	Gibson_Trp1_R_BsaI	Gibson reaction	ACCAGGTCTCACCTTTCAGTACGAAAATTGCTTTCATTG
32	Gibson_Trp2_F_BsaI	Gibson reaction	ACTGAAAGGTGAGACCTGGTGGCGCACTTCCTG
33	Gibson_Trp2_R_BsaI	Gibson reaction	TGTTATTCTCTAATTTTGTTCGAGACCAAAAAAAAGCCCGCTCATTAGG
34	RF_Belo_BsaI_F	RF reaction	GCGATAAAATGGTGCTTAACCTGGACAGATCACGTGTTCC

35	RF_Belo_BsaI_R	RF reaction	CCTCAATTTTCTCTATACACTCAGTTGGAA CACGTGATCTGTCC
36	prBeloRF_val_F	RF reaction	GCAGCAGGCATCTAACCTTC
37	prBeloRF_val_R	RF reaction	GGAACAGTGGGCCCAGAG
38	POC_Gib_F_1	POC Gibson	GAAAGCAATTTTCGTACTGAAAGGTTGGTG GGCGTGAAG
39	POC_Gib_F_238	POC Gibson	GAAAGCAATTTTCGTACTGAAAGGTTGGTG GCGCACTTC
40	POC_Gib_F_4	POC Gibson	GAAAGCAATTTTCGTACTGAAAGGTTGGTG GCGGACTTG
41	POC_Gib_F_5	POC Gibson	GAAAGCAATTTTCGTACTGAAAGGTAGGA GGCGCACTTC
42	POC_Gib_F_6	POC Gibson	GAAAGCAATTTTCGTACTGAAAGGTAGGTG ACGCACTTC
43	POC_Gib_F_7	POC Gibson	GAAAGCAATTTTCGTACTGAAAGGTTGGTG GTGGTGGTG
44	POC_Gib_R_1	POC Gibson	GCATTGTTATTCTCTAATTTTGTTCAAAAAA AACGGGCGAG
45	POC_Gib_R_2	POC Gibson	GCATTGTTATTCTCTAATTTTGTTCAAAAAA AAGCGGGCTC
46	POC_Gib_R_3457	POC Gibson	GCATTGTTATTCTCTAATTTTGTTCAAAAAA AAGCCCGCTC
47	POC_Gib_R_6	POC Gibson	GCATTGTTATTCTCTAATTTTGTTCAAAAAA AAGCCCGC
48	POC_Gib_R_8	POC Gibson	GCATTGTTATTCTCTAATTTTGTTCAAAAAA AAGCCCGCG
49	Gibson_TrpAtt_BamHI	Gibson	TGTTATTCTCTAATTTTGTTCGGATCCACCT TTCAGTACGAAAATTGCTTTCATTG
50	Pr_Trp_promoter_gib	Gibson	TGTTATTCTCTAATTTTGTTCGGATCCACCT TTCAGTACGAAAATTGCTTTCATTG
51	Pr_Cm_F	PVS202 validation	GTGTAGAAACTGCCGGAAATCG
52	Pr_Cm_R	PVS202 validation	GGCATCGTAAAGAACATTTTGAGGC
53	Pr_TrpE_F	PVS202 validation	CCAAACTGCCGTGTGCTG
54	Pr_TrpE_R	PVS202 validation	CCATCAGCGTTTCAGCGAG
55	pr_TrpR_val_F	TrpR deletion validation	GTCGTTACTGATCCGCACG
56	pr_TrpR_val_R	TrpR deletion validation	GTAACCACTCCTGGTGACGC
57	pr_TrpR_val_F	TrpR KO validation	CGTTACTGATCCGCACG
58	pr_TrpE_Val_F	TrpE KO validation	CCACGTTATGCCCATTGC
59	K1_KEIO	KEIO Deletion validation	CAGTCATAGCCGAATAGCCT
60	K2_KEIO	KEIO Deletion validation	CGGTGCCCTGAATGAACTGC
61	:TrpE_Del_Val_F	TrpE deletion validation	GCACGGTGCACCAATGCTTC
62	TrpE_Del_Var_R	TrpE deletion validation	GGCAGCGGGTTTGTTAATCC
63	TrpR_Del_Var_F	TrpR deletion validation	GCTGGATCCGGAAACGAATATC
64	TrpR_Del_var_R	TrpR deletion validation	GATTGAAAACGCCAGCCAGC
65	TrpR_Del_var_R2	TrpR deletion validation	GCTACTTTGCCGTTGCCTGC

66	TrpR_Del_var_R3	TrpR deletion validation	GTGATTCTGGAAAAAGTGCGTGAAGG
67	TrpR_Del_Var_F2	TrpR deletion validation	CAACATTGGCACCAGTTACCTGC
68	TrpR_Del_Var_F3	TrpR deletion validation	CCTGCAATATGTTTATCAGCAGTTTGGC

Supplementary table 3: List of PCR reactions.

ID #	PCR reaction	Primers used	PCR program	Used for
1	Phusion High-Fidelity PCR Master Mix: 25 µl F primer: 7.5 µl R primer: 7.5 µl DNA: varied* DDW: complete volume to 50ul *The amount of oligo DNA added was dependent on the percentage of the sub-library out of the entire library. SP – 2ul (dil 1:10) and increased primers to 11.5ul each due to complications with reaction. Ins1,2,3 – 7.5ul (dil 1:10) Control – 1ul (no dil) 3UTR - 4ul (dil 1:10)	1-12	1. 98ºC 30 sec 2. 98ºC 10 sec 3. **X**ºC 20 sec ** 4. 72ºC 30 sec 5. Repeat steps 2-4 for 20 cycles 72ºC 10 min ** SP, control, ins3 - 47ºC 3UTR - 50ºC Ins1,2 - 51ºC	Oligo amplification
2	KAPA HiFi HotStart ReadyMix: 25 µl F primer: 1.5 µl R primer: 1.5 µl Plasmid: ~25ng DDW: complete volume to 50ul	13-14	1. 95ºC 3 min 2. 98ºC 20 sec 3. 56ºC 15 sec 4. 72ºC 5 min 5. Repeat steps 2-4 for 25 cycles 72ºC 9 min	pCfB2223 linearization
3	KAPA HiFi HotStart ReadyMix: 25 µl F primer: 1.5 µl R primer: 1.5 µl Plasmid: ~30ng DDW: complete volume to 50ul	15-26	1. 95ºC 3 min 2. 98ºC 20 sec 3. 50ºC 15 sec 4. 72ºC 5 min 5. Repeat steps 2-4 for 25 cycles 72ºC 9 min ** for primer 15+16 of SP sub-library, steps 3 +4 were changed to: 3. 49ºC 15 sec 4. 72ºC 1 min	pGS2223 linearization

4	KAPA HiFi HotStart ReadyMix: 25 µl F primer: 1.5 µl R primer: 1.5 µl DNA: 2ul of lysate DDW: complete volume to 50ul	27-30	1. 95ºC 3 min 2. 98ºC 20 sec 3. 53ºC 15 sec 4. 72ºC 20 sec 5. Repeat steps 2-4 for 30 cycles 72ºC 3 min	Verification of yeast sub-libraries
5	KAPA HiFi HotStart ReadyMix: 12.5 µl F primer: 0.75 µl R primer: 0.75 µl DNA: ~30ng DDW: complete volume to 25ul	31-38	1. 95ºC 3 min 2. 98ºC 20 sec 3. 55-58ºC 15 sec ** 4. 72ºC 15 sec 5. Repeat steps 2-4 for 20 cycles 72ºC 1 min ** 58 ºC in small-scale competition and 55 ºC in large-scale competition	PCR1 in amplicon sequencing
6	Phusion High-Fidelity PCR Master Mix: 25 µl F primer: 2.5 µl R primer: 2.5µl DNA: 1ul (ranged from 1-10 ng) DDW: 1 9ul	39-41	1. 98ºC 30 sec 2. 98ºC 10 sec 3. 58 - 64ºC 20 sec ** 4. 72ºC 15 sec 5. Repeat steps 2-4 for 15 cycles 72ºC 5 min ** 58 ºC in small-scale competition and 64 ºC in large-scale competition	PCR2 in amplicon sequencing