Accelerated evolution using the phage T7 and its application to evolve riboswitches

Authors

Eduardo Goicoechea[1] and Alfonso Jaramillo[1,2*]

Affiliations

[1] Warwick Integrative Synthetic Biology Centre and School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK.

[2] *De novo* Synthetic Biology Lab, I2SysBio, CSIC-University of Valencia. Parc Científic Universitat de València. Calle Catedrático Agustín Escardino, 9. 46980 Paterna. Spain.

* To whom correspondence should be addressed:

Prof. Alfonso Jaramillo. *De novo* Synthetic Biology Lab. I2SysBio, CSIC-University of Valencia. Parc Científic Universitat de València. Calle Catedrático Agustín Escardino, 9. 46980 Paterna. Spain. https://orcid.org/0000-0002-6313-9689, Alfonso.Jaramillo@csic.es.

Abstract

Phages can be used to accelerate the evolution of proteins by developing selection systems based on the conditional gene complementation. The evolution is accelerated because of the mutation load that the phages can tolerate. It is also possible to evolve gene switches by designing dual selection systems, consisting of positive and negative selection steps. Until now only the filamentous phage M13 had been used for this purpose and never for gene switches made of non-coding RNA. Riboswitches are a type of gene switches consisting of small fragments of non-coding mRNA, usually working as genetic sensors to regulate gene expression. Much smaller and faster than protein-based regulation, they can act at the translation or transcriptional level. Here we propose a genetic selection system (T7AE) for the phage-assisted evolution using the T7 phage, which we exemplify by evolving new riboswitches. We have engineered the genome of the phage T7 and its propagation host *Escherichia coli* to implement a conditional error-prone phage replication when an evolving gene of interest is active or inactive. For this, we encoded in the phage T7 the cI transcription factor from λ phage regulated by a theophylline riboswitch. We then randomised the riboswitches to create a library of phages, which were subjected to positive and negative selections in succession. We found that riboswitch variability within the phage population progressively decreased over selection cycles, leading to novel riboswitches showing significant changes in activation between different selection stages.

Keywords: Riboswitch, Genetic selection, Directed Evolution, Phage-assisted evolution

Introduction

Phages can be used to accelerate the directed evolution of genetic systems because of they have the fastest replication rates in nature and can cope with large mutagenesis loads. The PACE (Phage Assisted Continuous Evolution) technology (1) uses the M13 phage to evolve proteins in continuous cultures. PACE employs modified M13 phages, where a specific essential gene has been removed from their genome to be then placed in the propagation host but in a transcriptionally inactivated form. PACE selection can be used to evolve any gene of interest (GOI) that could be able to transcriptionally activate the essential gene. More recently, the PACEmid (2) technology modified PACE by moving almost all genes from the phage to the host, which allowed to freeze the adaptation of these genes. This amounted to perform the evolution of satellite phages, where plasmids get packed as functional virions after adding to them the phage's packaging signal (these plasmids are then called phagemids). PACEmid is useful in applications where the initial GOI activity is very small like in their case, where the authors evolved a transcription factor activator from the *cro* repressor, creating the smallest known transcription factor activator of only 63 amino acids. PACEmid also allows increasing tenfold the non-specific mutation rate. But the use of the phage M13 poses some serious challenges during evolution in bioreactors because bacteria don't die after infection, allowing the biofilm-producing cells to secrete non-evolved phages. This can even be the dominant production of phages when the selection is very stringent, when the initial GOI activity is very small, or when the GOI activity is very low or very high depending on the chemical context ("gene switch"). Gene switches also demand designing different, ideally modular selection circuits for different chemical contexts (3,4).

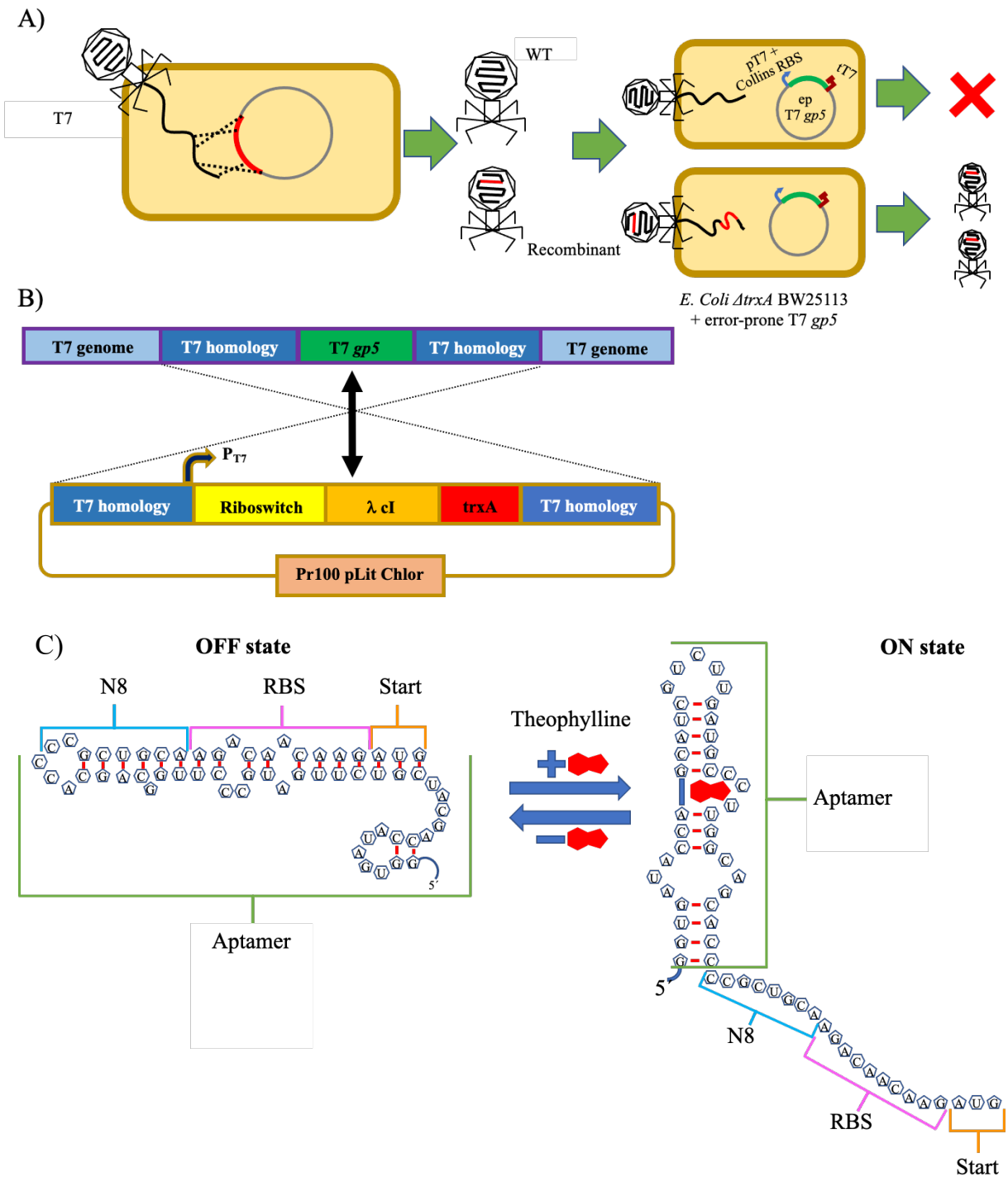
Here we propose a methodology to evolve genetic switches by extending PACE to use a lytic phage. We chose a well-known lytic phage of *Escherichia coli*, phage T7, with a short

replication time and large burst size (it produces approximately 100 virions per cell) (5). Moreover, its mutation rate can be further increased specifically by using an error-prone variant of its DNA polymerase (encoded in gene gp5) (5). As a proof-of-concept, we focused on the evolution of a riboswitch, the smallest of all gene switches. Phage-assisted evolution techniques had been used mainly with proteins, and non-coding RNA has relied on the use of computational design methods (6–9).

Riboswitches are structured regions of mRNA able to form a specific tertiary structure and influence downstream gene expression (10). This tertiary structure is capable of binding small molecules, and upon binding, a structural rearrangement occurs. This causes the mRNA structure of the sequence immediately downstream of the riboswitch to change its secondary structure, which can generate a new transcriptional terminator or open up a previously occluded ribosome binding site (RBS) (11). Riboswitches are naturally found in all forms of life, but have been mostly observed in archaea and prokaryotes, with some in eukaryotes (12) and even prophages (13). Riboswitches have also been engineered to work in viruses (14). This ability of riboswitches to be functional across different organisms makes them of great interest for further study with broad potential applications. They are amenable to mutations, generating novel riboswitches that no longer recognise their original ligand (15) and can also be combined in novel ways to flip gene regulation, such as a ligand-binding gene-repressor riboswitch converted to a ligand-binding gene-activating riboswitch (16). Due to their small size and understood mechanism of function they are one of the simplest genetic switches.

This broad spectrum of possibilities has motivated efforts to discover novel riboswitches (17), their improvement (18) and their *de novo* design (19). As a result, there are currently three main ways to develop new riboswitches: computer modelling (20), directed evolution(21), and

101 *de novo* design (22). Although most experimental methods have been focusing on *in vitro*
102 selection approaches, which later require individual validation in living cells, there are also
103 examples of *in vivo* methods (23). An *in vivo* selection methodology allows screening for viable
104 riboswitches during the sequence search phase, increasing the sequence space and, therefore,
105 the final activity. Dual selection methods are required for the directed evolution of genetic
106 switches, where positive and negative selections are used in environments stabilizing the
107 switch in its "ON" and "OFF" states respectively (24).
108
109 Here we aim at developing a methodology to use lytic phages for accelerated evolution, which
110 we will exemplify by engineering new riboswitches. In the following, we will develop a dual
111 selection strategy consisting of positive and negative selections to evolve genes using the phage
112 T7 that we name T7AE. As gene of interest, we integrated a theophylline riboswitch on the T7
113 phage, which is later randomised using phage genome engineering methods (25). This is the
114 first report of a functional riboswitch in a phage, as there are no known examples of
115 riboswitches integrated in the genome of phages, although phages are known to encode
116 ribozymes (26). We will show how we passed a heterogeneous phage population carrying
117 randomized theophylline riboswitches through several cycles of T7AE. We will then
118 characterize the evolving phage population by measuring the variability of riboswitch
119 sequences and the phage activity under different inducer conditions.
120
121 **Results**
122 **T7AE: Dual selection method for T7-assisted directed evolution**
123 The strategy we have followed to develop our T7 phage-based screening method to identify
124 new riboswitches sequences is shown in Figure 1.


A)
T7
WT
Recombinant
pT7 +
Collins RBS
tT7
ep
T7 gp5
E. Coli ΔtrxA BW25113
+ error-prone T7 gp5
B)
T7 genome
T7 homology
T7 gp5
T7 homology
T7 genome
P T7
T7 homology
Riboswitch
λ cI
trxA
T7 homology
Pr100 pLit Chlor
C)
OFF state
N8
RBS
Start
Aptamer
Theophylline
ON state
Aptamer
5'
N8
RBS
Start

D)

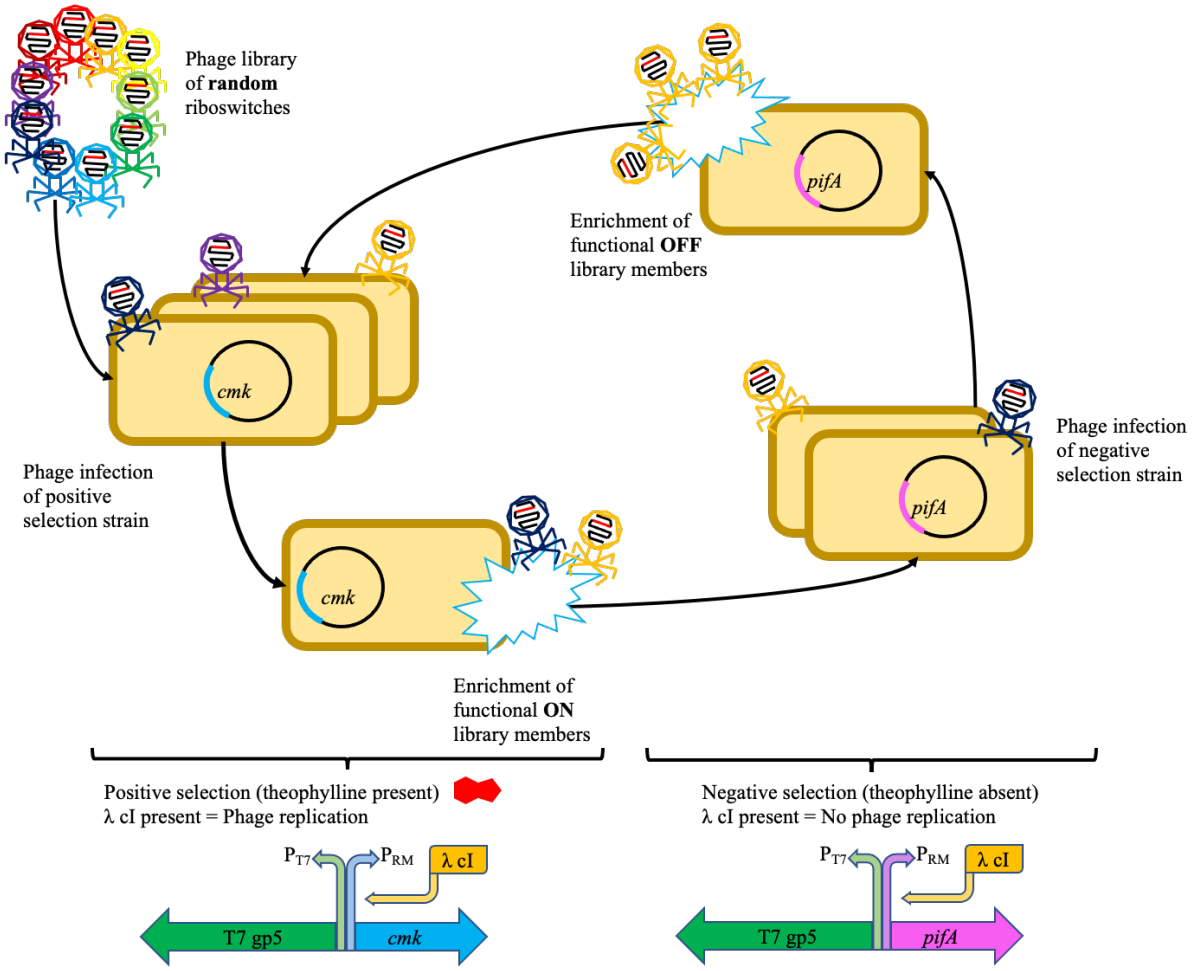
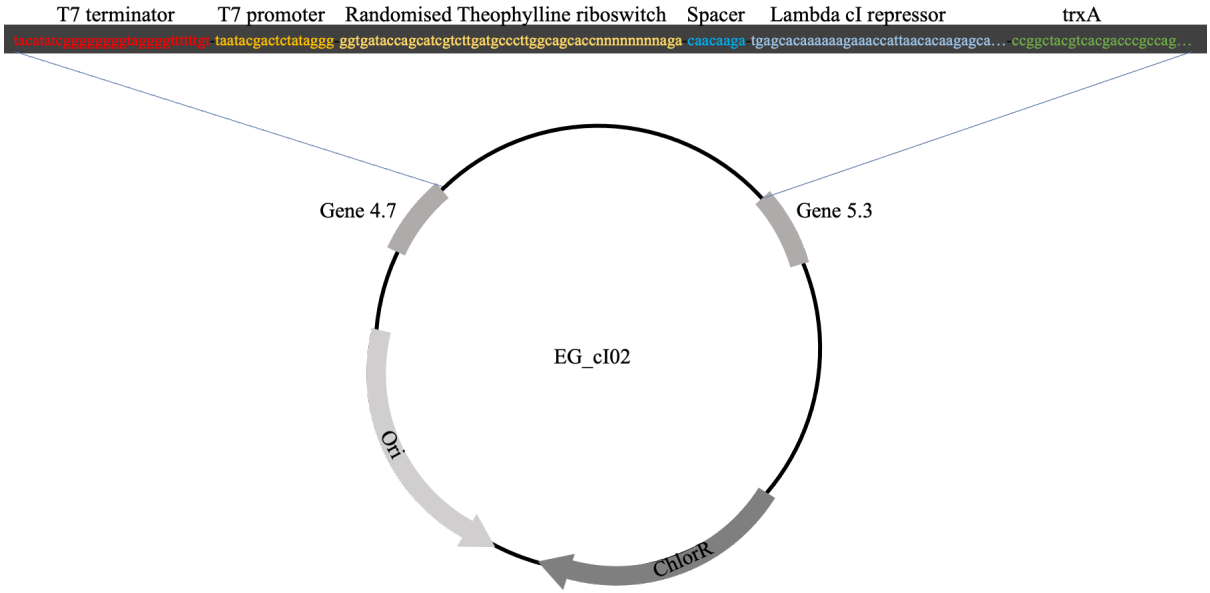


Figure 1. Schema showing the design and use of our T7AE system. A) Obtaining recombinant phages. Schematic version of the homologous recombination process undergone by the phage. As a result, some of the offspring will be recombinant, and be selected by infecting a specific replication strain **B) Detail of genetic constructs.** Representation of the T7 genome and the recombination plasmid, indicating homology regions and the elements they flank. Through homologous recombination, these 2 fragments are swapped. **C) Versions of the theophylline riboswitch.** "OFF" state of the riboswitch (left), "ON" state of the riboswitch (right), **D) General structure of the full phage evolutionary procedure.** 1) A phage library of switches is used to infect the positive strain, containing a positive selection plasmid. 2) In the presence of theophylline, only phages containing working variants of the riboswitch will be able to replicate properly. 3) These surviving phages will then go on to the negative selection step. 4) In this case, the absence of theophylline means that riboswitches should stay inactive, but to eliminate constitutively active phages, the *pifA* gene is transcribed and those phages are unable to replicate properly. The resulting phages will then go back to infecting the positive selection strain, and then the cycle will start once again, progressively forcing the evolutionary pressure towards functional and improved riboswitches over time.

141 First we inserted our DNA fragment of interest in the genome of T7 phages via homologous
142 recombination (25,27) (Fig 1A). Different plasmids were designed, containing regions
143 homologous to the phage's genome, aimed at eliminating the gp5 gene from the phage. The
144 plasmids contained either the sequence of a positive control riboswitch (EG_cI01), or the
145 riboswitch library (EG_cI02). A schematic version of the latter can be found in Figure 2. More
146 information on the constructs can be found in the Supplementary materials. In this case, the
147 recombined gene fragment contains a T7 promoter, a version of the theophylline riboswitch
148 (28) controlling the expression of the cI transcription factor from phage λ (29); and a gene
149 essential for phage replication (*trxA*) (5), which was used as a cloning marker to select for
150 recombinant T7 genomes (25) (Fig 1B). During selection, the λ cI repressor will activate the
151 P_{RM} promoter (30) and thus regulate the expression of genes downstream from it.
152



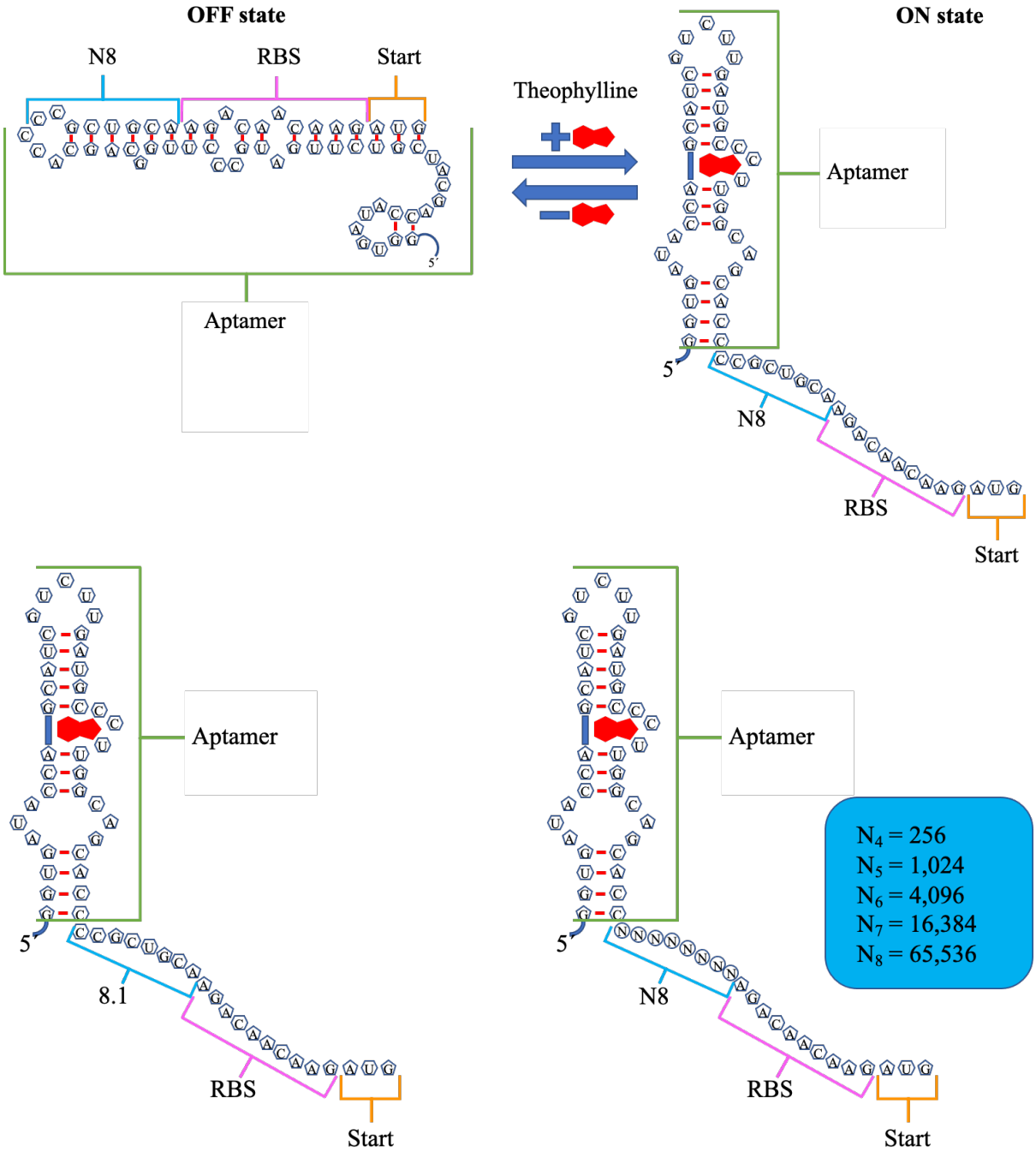
153 **Figure 2. Schematic representation of HR plasmid EG_cI02.** This is the version of the homology
154 plasmid that was used for the creation of the library, carrying the fragment of interest flanked by two T7 genes to
155 act as homology regions.
156

157 The directed evolution of a genetic device to behave conditionally according to an input signal
158 requires implementing a dual selection (31,32), where selection is enforced in two steps:
159 positive and negative (see Fig 1D). In the positive step of our phage-assisted riboswitch
160 evolution, theophylline is added at high concentration during phage replication, where the
161 riboswitch controls (via λ cI) the replication rate. The selection is done by infecting an
162 engineered *E. coli* positive selection strain with our T7 phage carrying the riboswitch. The
163 strain contains a knockout of the *cmk* gene, essential gene for T7 phage replication (5) but not
164 for *E. coli* growth. Moreover, we transformed the strain with a plasmid encoding the P_{RM}
165 promoter regulating the *cmk* gene. Enforcing a theophylline dilution in the cultures selects for
166 the amplified phages that encoded a riboswitch in the active state, because only they produced
167 the λ cI required to activate the essential *cmk* gene. In the negative selection step, we used a
168 different *E. coli* strain for infection. The strain had the original copy of the *cmk* gene, but it is
169 transformed with a plasmid containing the *pifA* gene under the control of a P_{RM} promoter. The
170 gene *pifA* belongs to the F-plasmid and it was found to block T7 replication (33). The
171 theophylline is absent in the negative step and a good riboswitch should be inactive in this case.
172 Any activity would increase the translation rate of λ cI, which would activate the *pifA* gene
173 producing an abortive infection. This way, phages encoding a riboswitch mutation unable to
174 fully repress the λ cI translation in the absence of theophylline will not be able to replicate
175 properly. By cycling between positive and negative selections in presence and absence of
176 theophylline respectively, we can sieve through the initial pool of random riboswitches (65,536
177 or 4^8) and obtain more efficient riboswitches.
178
179 **Library design**
180 We created a library of phages by randomising an 8-nucleotide region of the riboswitch. The
181 region was chosen based on previous results by (28) (Fig. 3). The theoretical diversity of the

182 library is 65,536 different variants and considering that our transformation efficiency was 1.11

183 *10^6 cfu/µg, we could obtain a 17-fold coverage of the library with the cultures used. (Fig 3).

184



185

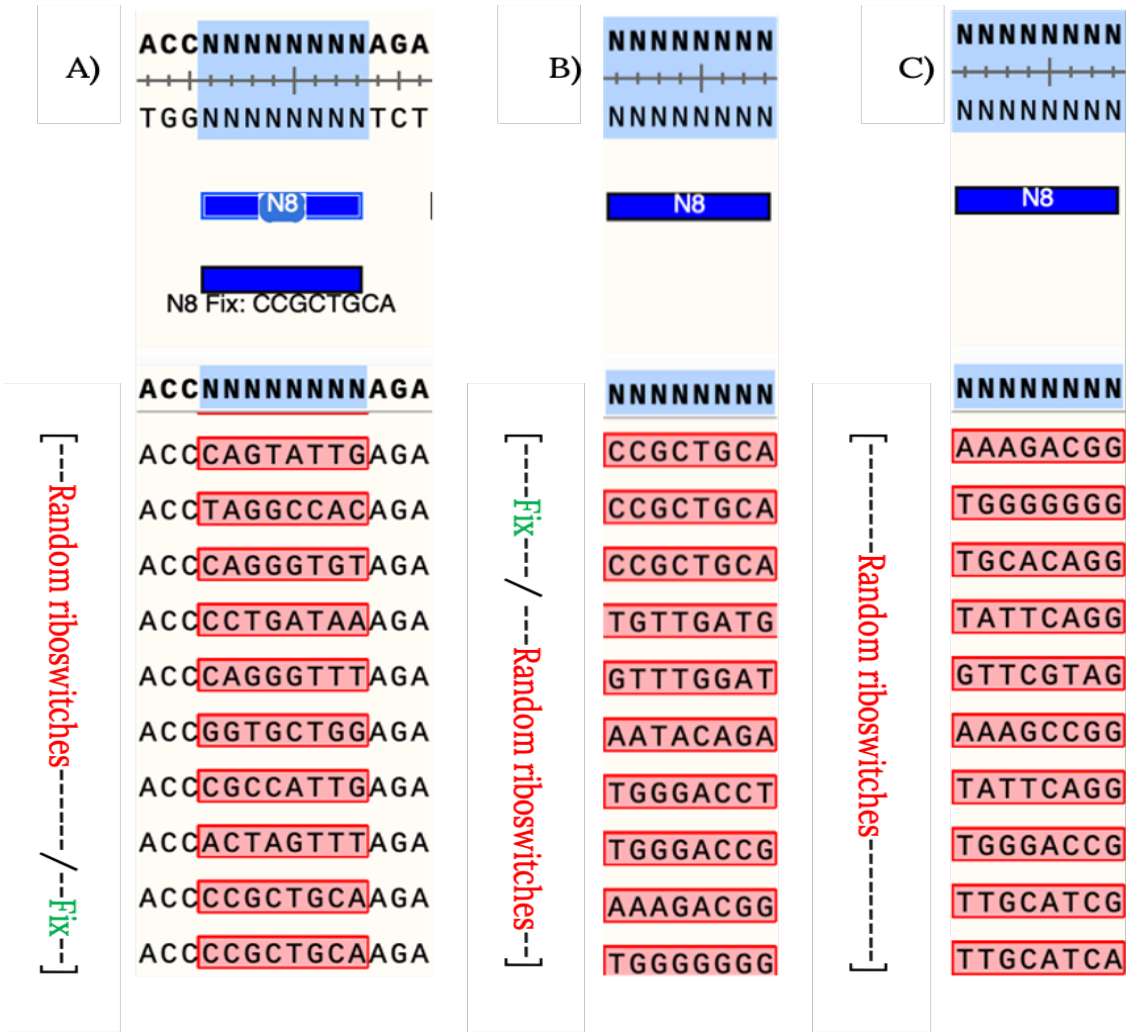
186 **Figure 3. Design of riboswitch library.** "OFF" state of the riboswitch (top left), "ON" state of the

187 riboswitch (top right), 8.1 version of the riboswitch, which showed a 36-fold activation in Lynch, et al. 2007

188 (bottom left) and structure of the theoretical library obtained after randomising the 8 nucleotides following the
189 stem of the riboswitch (bottom right).
190
191 **Plasmid and Phage library production**
192 In order to confirm if our procedure could be used to select and eventually evolve riboswitches,
193 we developed via homologous recombination a phage library of switches with the above
194 indicated 8 randomised nucleotides (see Fig 3), which is then used to infect two alternating
195 bacterial selection strains as described above.
196
197 To assess whether enough variability was present in the initial libraries, samples from the
198 different plasmid constructs were sequenced. After the homologous recombination process,
199 some of the phages that incorporated the fragment were also sequenced before undergoing any
200 selection. The results from the sequencing of elements from EG_cI01 and EG_cI02, as well as
201 that of phages, showed different versions of the riboswitch in both the plasmid and phage

202 libraries (see Figure 4). Both the plasmid and the phage populations containing the randomised

203 libraries showed no repetition of sequences.



204

205 **Figure 4. Sequences obtained from initial plasmid and phage libraries.** Collection of sequences

206 obtained throughout the development of the different libraries. **A)** Riboswitch sequences obtained by Sanger

207 sequencing of the PCR of single colonies carrying plasmids EG_cI01 (Riboswitch used as a positive control,

208 marked as "Fix" in figure) and EG_cI02 (Random library of switches). Phages carrying the control were used to

209 initially set up the procedure.

210
211
212
213

214 **Dual selection cycles show enrichment with an engineered T7-phage library of**
215 **randomized riboswitches**

216 We performed 16 cycles of positive and negative selections and sequenced 24 libraries (see
217 Material and Methods). Analyses were carried out with smaller library samples (10^5 reads)
218 across the 16 selection cycles and the sequence variation in each set assessed. This was
219 considered appropriate, as it was large enough to encompass the sequence library at its
220 maximum variety (65,536 riboswitch variants). Despite this, further changes allowing for more
221 specific riboswitches to be obtained (such as concentrations of the activator) can be considered
222 within the strategy. All phage populations before generation 10 were selected with 1.5 mM
223 theophylline, whilst every generation afterwards was selected with 1 mM in order to have a
224 more stringent selection process.

225

226 The results are shown in the form of diagrams representing the 8 random nucleotides flanked
227 by 5 fixed nucleotides on each side (Figure 5). The larger size of a particular nucleotide denotes
228 overrepresentation of that nucleotide in that position. This analysis was performed for multiple
229 libraries (shown in **Error! Reference source not found.**) throughout the evolutionary process,
230 with each step marked as "cmk" or "pifA" depending on whether it corresponded to a positive
231 or negative selection step respectively.

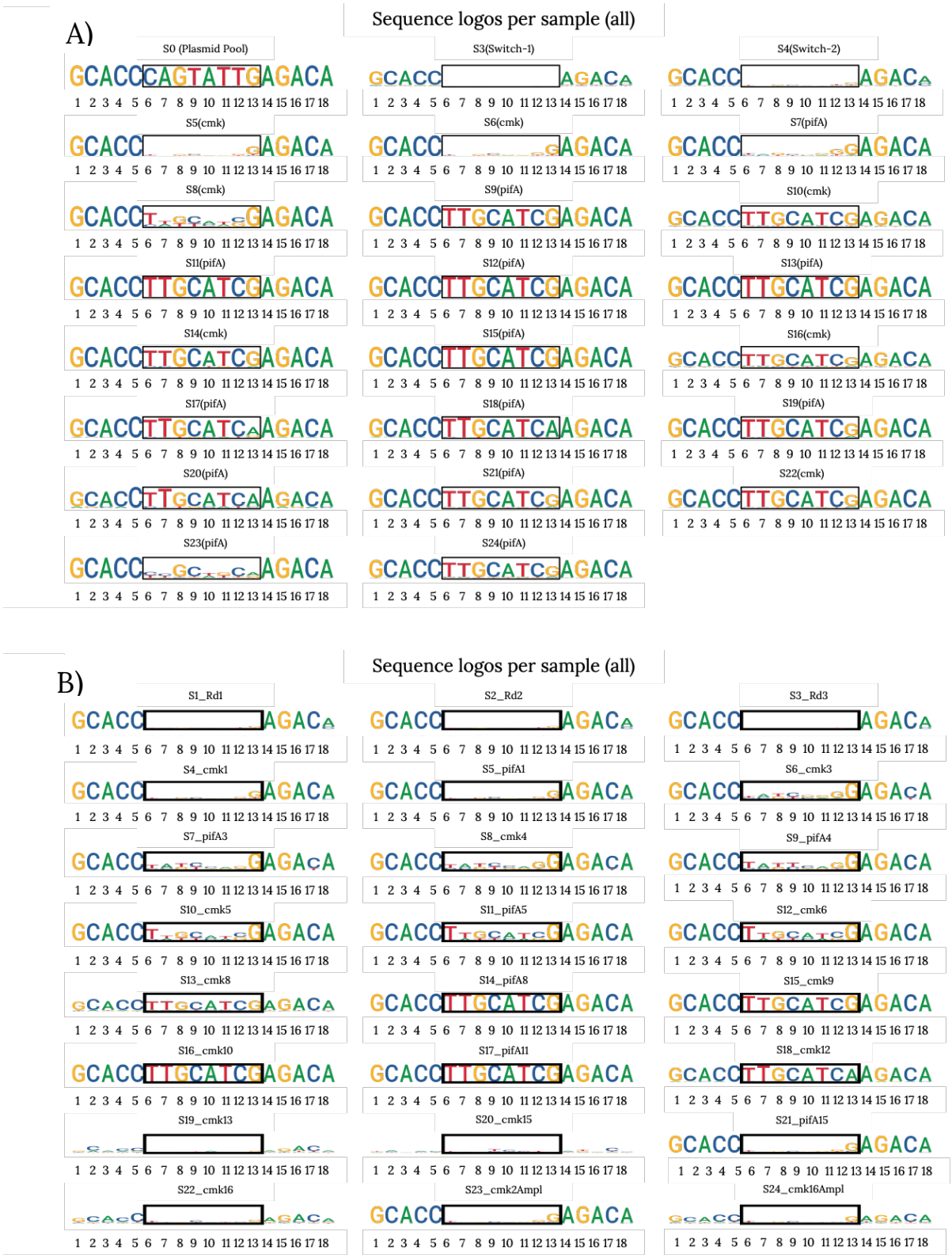


Figure 5. Evolution of predominant riboswitch sequences throughout dual selection cycles. Diagram representing the sequence variation for each of the libraries, starting from the earliest (top left) to the most recent (bottom right). Each set represents the predominant sequence present in a phage population

after each T7AE selection cycle. Bases included within the black square correspond to the randomised riboswitch sequences. The size of a nucleotide in the diagram denotes its representation percentage in the sample population, and the nomenclature "S+number" indicates the generation number since the beginning of the project. **A)** Representation of the first set of libraries that were assessed. At the beginning of the selection process (Switch-1 and 2) no prominence of any particular sequence was observed. However, a progressive increase over time towards one dominant sequence was observed, which helped us pinpoint the approximate point where it took hold. **B)** Representation of a second set of libraries, focusing more on the early generations, to fill the gaps present in A) as well as assess new phage generations. A more thorough description of the process can be found in Materials and Methods, and the relevant library information for these experiments can be found in Table S1. The last set of logos possibly were affected by human error when being prepared, which explain the sequences observed.

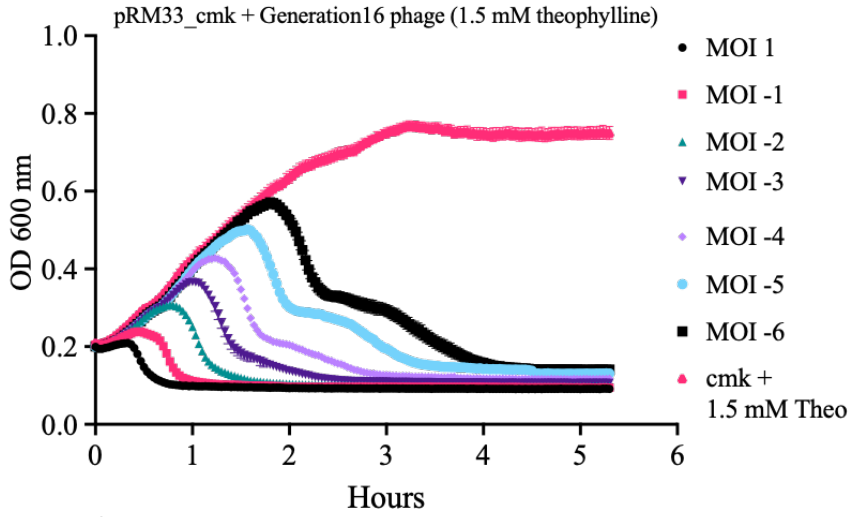
To fully assess this, Next Generation Sequencing (NGS) of different libraries was carried out: this time within a smaller range of generations (Figure 5B). Based on the initial set of sequences analysed, no sequence is overrepresented, however the variation within each phage population decreases over the course of the selection steps, with a an early preference for G in position 8, with one single sequence (TTGCATCG) having a progressively prominent presence by generation 6, in a manner similar to previous literature (12, 30). Most of the sequences analysed in the second round of NGS show that as well; however, the reason why S23 in A, and S19, 20, 21 22 and 24 in B shows a much higher variation is unclear.

Estimation of riboswitch activity from phage phenotype shows riboswitch adaptation during evolution

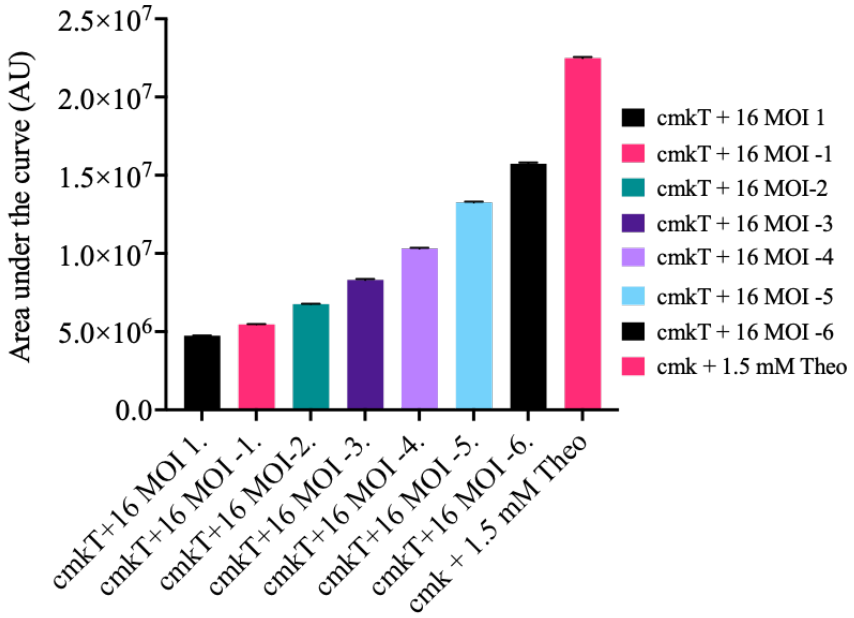
To validate the results of the selection a subsequent virulence index (VI) program of assays was carried out, as described in Materials and Methods. To do so, different phage populations obtained throughout the process were compared. The selected phages included the recombinant phage with a known, working riboswitch, used as a control for the selection process (positive control); phages from generations 2, 5, and 16; and a Wild Type T7 phage population, lacking

264 a riboswitch in its genome (negative control, since its virulence index would remain constant
265 in the presence or absence of theophylline). All populations were tested in the presence and
266 absence of theophylline, and in positive and negative selection strains. An example of a set of
267 results for one of the calculations, showing all the steps from the OD_{600} absorbance curves in
268 the presence and absence of phages, to the regional virulence values, can be seen in Figure 6.
269 This set of data was obtained after infecting the positive selection strain with phages that had
270 undergone each selection 16 times in total.

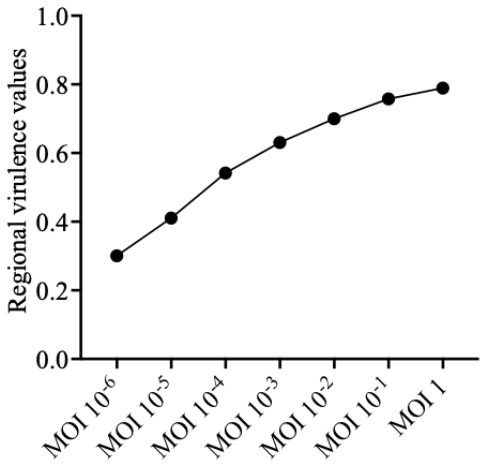
A)


pRM33_cmk + Generation16 phage (1.5 mM theophylline)
OD 600 nm
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
Hours
MOI 1
MOI -1
MOI -2
MOI -3
MOI -4
MOI -5
MOI -6
cmk +
1.5 mM Theo

B)


Area under the curve (AU)
2.5×10^7
2.0×10^7
1.5×10^7
1.0×10^7
5.0×10^6
0.0
cmkT+16 MOI 1.
cmkT+16 MOI -1.
cmkT+16 MOI-2.
cmkT+16 MOI -3.
cmkT+16 MOI -4.
cmkT+16 MOI -5.
cmkT+16 MOI -6.
cmk + 1.5 mM Theo
cmkT + 16 MOI 1
cmkT + 16 MOI -1
cmkT + 16 MOI-2
cmkT + 16 MOI -3
cmkT + 16 MOI -4
cmkT + 16 MOI -5
cmkT + 16 MOI -6
cmk + 1.5 mM Theo

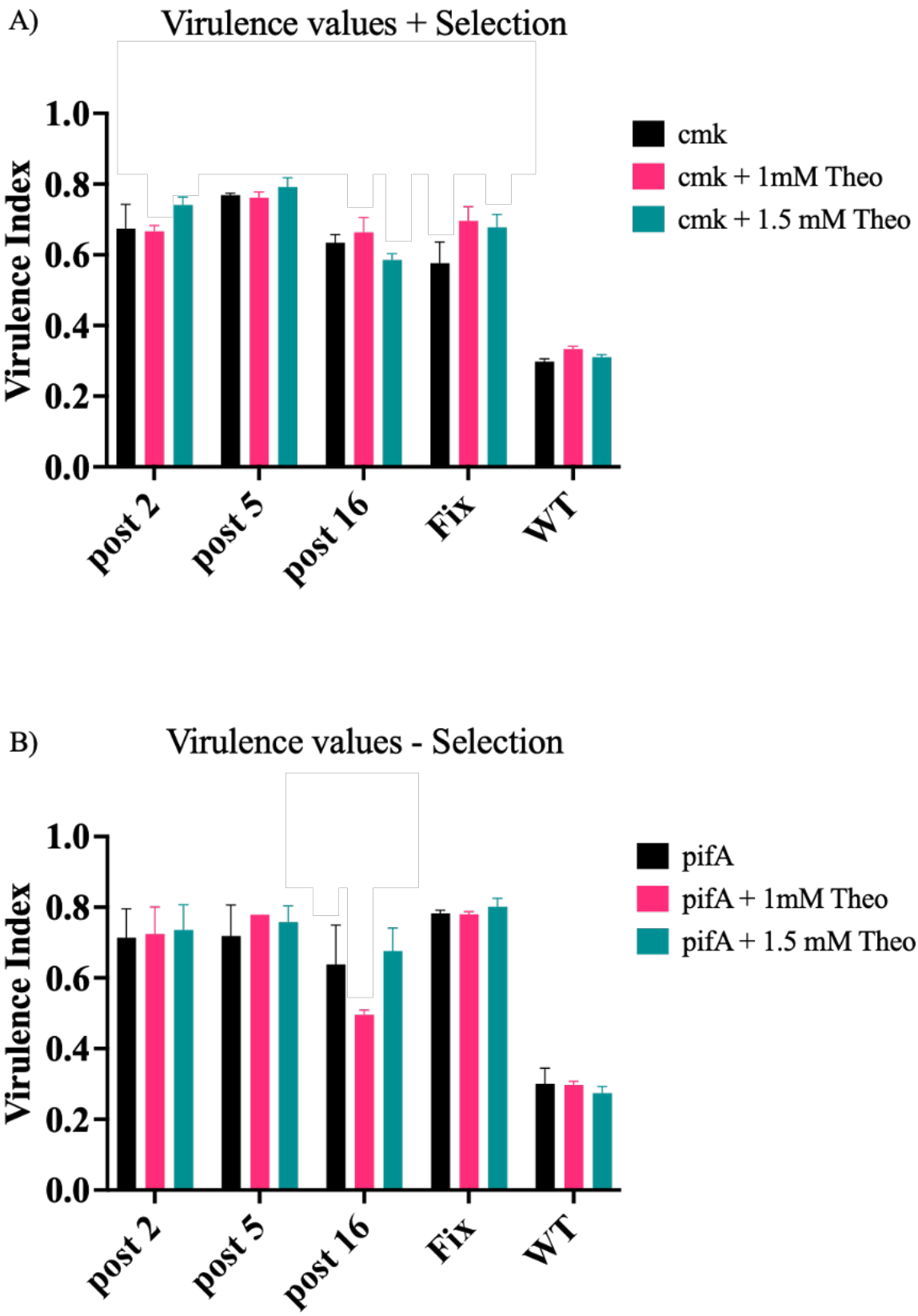
C)


Regional virulence values
1.0
0.8
0.6
0.4
0.2
0.0
MOI 10^-6
MOI 10^-5
MOI 10^-4
MOI 10^-3
MOI 10^-2
MOI 10^-1
MOI 1

271

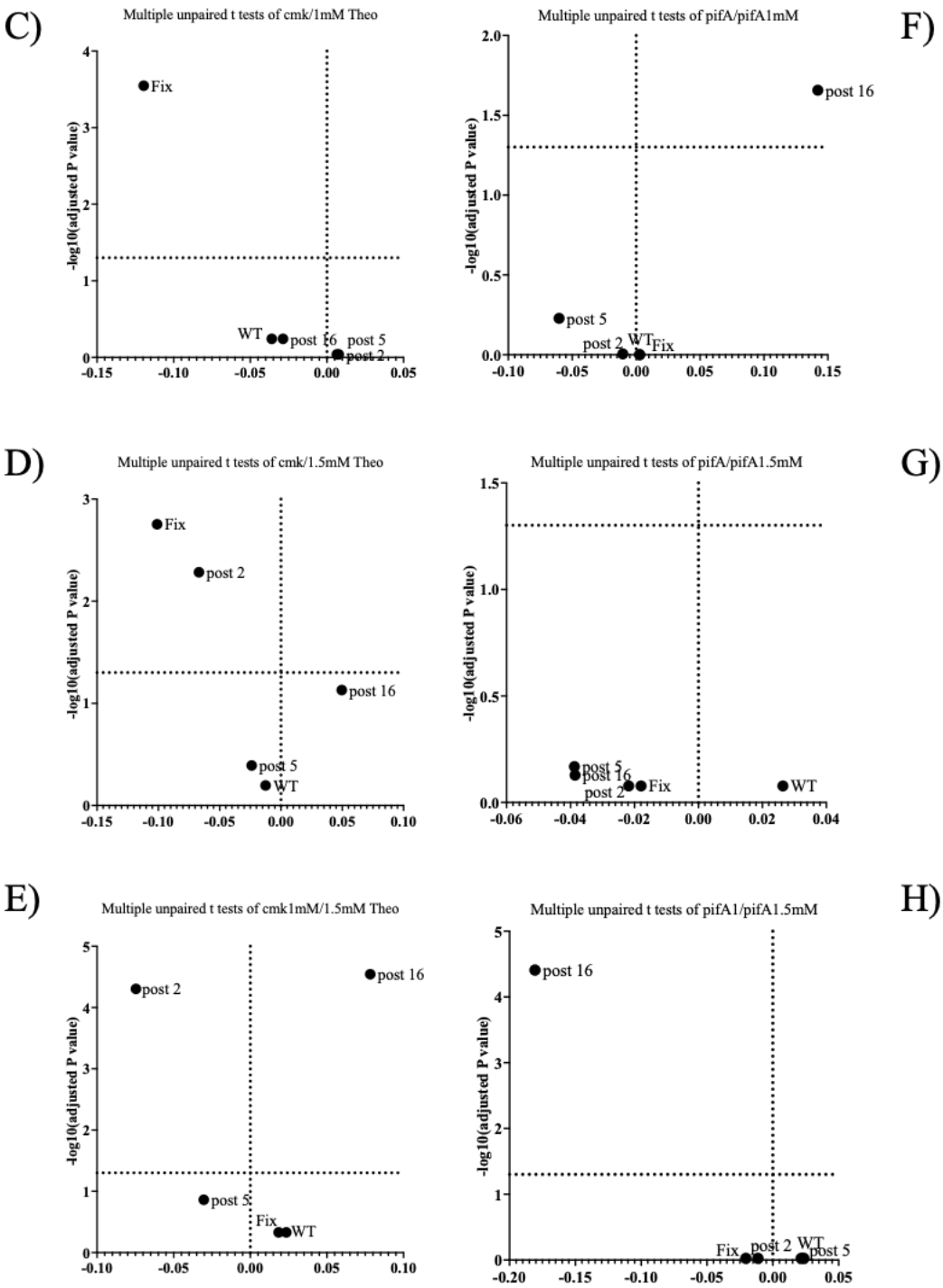
(Previous page) Figure 6. Standard set of data used to obtain the virulence index for a phage population infecting a specific strain. **A)** Killing curves of the cell cultures where phages at different MOIs (Phage to cell ratios) have been added, as well as growth curve of the cell strain without phages. Notice the point where it enters stationary phase, which marks the last timepoint included to calculate the area under the curves. **B)** Values resulting from calculating the area under each of the previous curves. The Y axis represents Arbitrary Units (AU). cmkT represents cultures of cells with theophylline present in the media. **C)** Representation of the values obtained after using the data from B) in Equation 1, as seen in Materials and Methods. The value for the area under this curve corresponds to the virulence index for that phage in that strain.

Once these calculations have been made for all the different combinations of phage, strains, and inducers the resulting values were grouped and represented as columns (Figure 7). Each of the tested phage populations (Wild Type T7, Positive control Riboswitch, and cycles 2, 5 and 16) was represented for both positive and negative selections, in the absence and presence of different concentrations of theophylline. These values were then used to calculate the riboswitch's activation efficiency, as shown in Materials and Methods. WT T7´s virulence index showed no change, and the positive control riboswitch sequence showed a general higher activation fold-change. The test samples, which included phages from cycles 2, 5, and 16, showed virulence changes between them. Furthermore, the results from the negative selection show that phage populations that were selected at 1.5 mM (2 and 5) have a better activation rate than the positive control, while the one selected at 1mM (16) performed worse than all other populations. However, at 1 mM, while the populations 2 and 5 were once again less efficient, population 16 shows the highest activation rate with an increase of 28% when the riboswitch changes conformation.


A)
Virulence values + Selection
Virulence Index
1.0
0.8
0.6
0.4
0.2
0.0
post 2
post 5
post 16
Fix
WT
cmk
cmk + 1mM Theo
cmk + 1.5 mM Theo
B)
Virulence values - Selection
Virulence Index
1.0
0.8
0.6
0.4
0.2
0.0
post 2
post 5
post 16
Fix
WT
pifA
pifA + 1mM Theo
pifA + 1.5 mM Theo

297
298

299


C)
Multiple unpaired t tests of cmk/1mM Theo
-log10(adjusted P value)
Fix
WT
post 16
post 5
post 2
F)
Multiple unpaired t tests of pifA/pifA1mM
-log10(adjusted P value)
post 16
post 5
post 2
WT
Fix
D)
Multiple unpaired t tests of cmk/1.5mM Theo
-log10(adjusted P value)
Fix
post 2
post 16
post 5
WT
G)
Multiple unpaired t tests of pifA/pifA1.5mM
-log10(adjusted P value)
post 5
post 16
post 2
Fix
WT
E)
Multiple unpaired t tests of cmk1mM/1.5mM Theo
-log10(adjusted P value)
post 2
post 16
post 5
Fix
WT
H)
Multiple unpaired t tests of pifA1/pifA1.5mM
-log10(adjusted P value)
post 16
Fix
post 2
WT
post 5

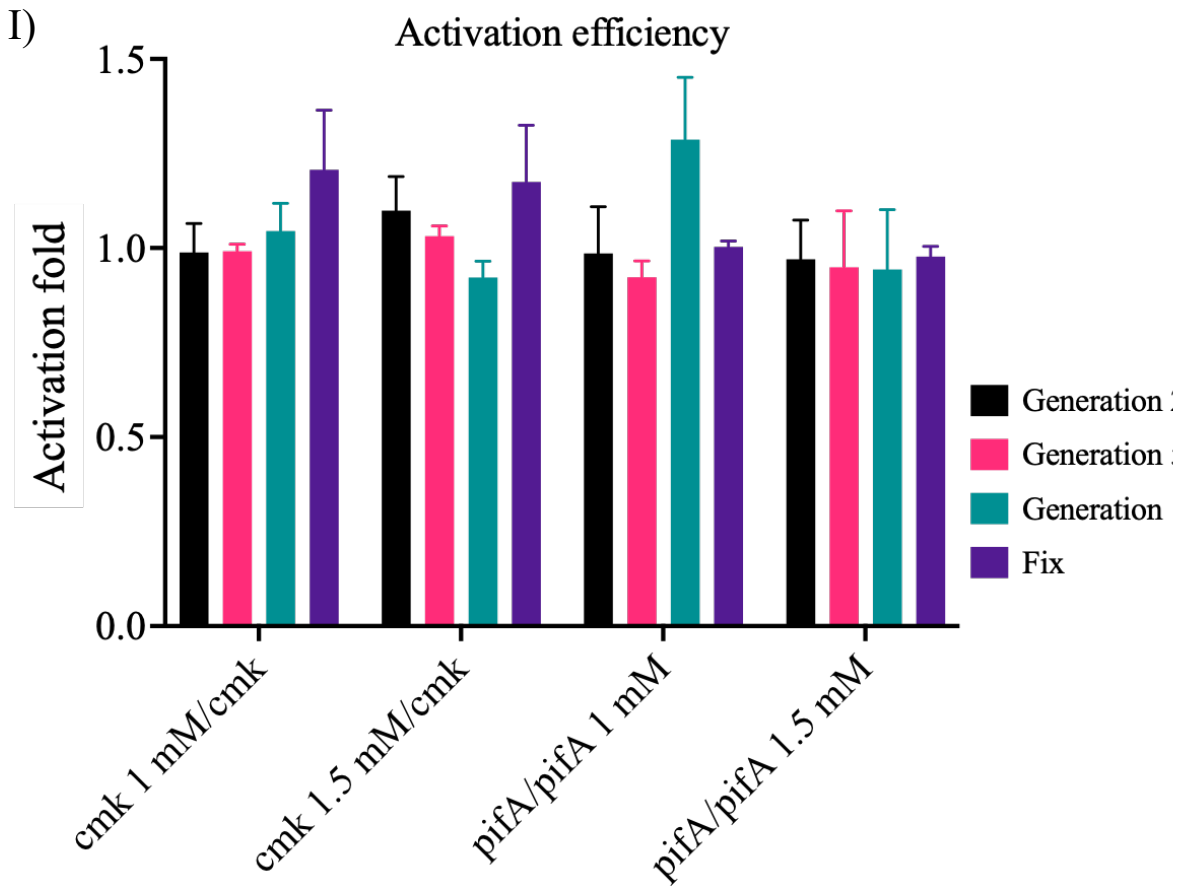


Figure 7. Virulence comparisons between phage populations (Wild Type/Negative control, Control riboswitch/Positive control, and after 2, 5, or 16 cycles of selection) when infecting the positive and negative selection strains in the presence and absence of different concentrations of theophylline. Higher virulence index indicates higher virulence of that phage population, with a range between 0 and 1. A ¨T¨ in the X axis represents the presence of theophylline in those cultures. **A)** Virulence indexes for the 5 different phage populations after infecting the positive selection strain in the absence and presence of theophylline. In this case, virulence should be higher in the presence of theophylline, as the riboswitch changes to an "ON" conformation in its presence and allows for the expression of cmk. **B)** Virulence index for the different phage populations when infecting the negative selection strain in the absence and presence of theophylline. In this case, virulence should be higher in the absence of theophylline, as the riboswitch stays in an "OFF" state. Once theophylline is being added and the switch turned "ON", pifA is expressed, blocking efficient T7 replication. **C to H)** Representation of the p-values corresponding to the pair comparisons between the different conditions, in both positive and negative selections. Statistics can be found in Table S4. **I)** Representation of the fold increase in virulence in both positive and negative selections for the different concentrations of theophylline used in each. In the positive

316 selection, the virulence values in theophylline^{+} cultures were divided by the value in theophylline^{-} cultures, with
317 the opposite being done for the negative selection, as indicated in the X axis. Significance in figures C-H was
318 calculated using unpaired t-tests. Error bars for I) were obtained using error propagation.
319
320 To further illustrate these changes in the available sequences, as well as virulence, a
321 comparison between the predicted structures of different riboswitch sequences is presented in
322 Figure S1(34). It includes random sequences from the initial steps, as well as the Positive
323 control riboswitch and the dominating sequences present by the final steps of selection.

Discussion

We have developed T7AE, a novel genetic system for the selection of gene switches using T7 phages, which we applied to the evolution of a theophylline riboswitch. The sequencing of the evolved populations provided us with an approximate timeline of the evolution of the phage population. This gave us the generation at which certain nucleotides get fixed in a particular position, and the point where a sequence is overrepresented in the population (Figure 5). Together with Table S1, this shows that the phages that survived the first selection step already show a preference for G as the 8th nucleotide of the sequence. By cycle 6, there is a sequence being overly represented in the population, which eventually takes over and remains dominant throughout (TTGCATCG, or the slightly improved TTGCATCA) The sequence shares some similarities with the control sequence used (CCGCTGCA, from (28)), with 6 out of 8 purines and pyrimidines in the same position. Virulence index calculations are a robust way of comparatively measuring the different efficiencies for each of the phage populations, in a much better way than directly using growth curves (35). To fully characterise the timeline of virulence, the virulence values for different populations were assessed (Figure 6). These values represent the efficiency of infection by those particular phages in a specific strain. These experiments included two different controls: the Wild Type T7, and the Control version of the

riboswitch (28) described in Figure 2B. Wild Type T7 was used as the purpose of the experiment was measuring the activation of the switch, and how that affected the virulence, while the Control riboswitch was used as a standard working switch. That explains why in the first case the value remains at around 30% (Figure 6A and B), since there is no switch to be activated. The Positive control phages that were used to set up the procedure show an increase in virulence of 20% and 17% for the positive selection, and no difference in the negative selection.

However, statistical analysis shows that despite what we observe in A and B, differences between ON and OFF states of the riboswitch are not significant in all cases. In the case of the positive selection, the positive control shows differences when comparing between both, regardless of the concentration; while the negative control shows no difference whatsoever in any case, as expected. As for the different populations, generation 5 shows no difference at all, while results from generations 2 and 16 indicate differences when comparing ON and OFF states (2) and when comparing activation between both concentrations (2 and 16).
In the negative selection, no population other than generation 16 shows any significant difference when compared between different states. Both sets of results further support the riboswitch population changes over subsequent generations.

We believe these results are directly related to the lack of knowledge on the pifA molecular cascade (36,37), as well as potential leakage in the system. Currently, the mechanism of the pifA protein and the phage exclusion process are still not fully understood (36–38), which could mean some effects may be overlooked in the selection and cause a larger variation between results than in the case of the positive selection.

Regarding the actual test populations, two out of three have higher VI values than the phage carrying the established control riboswitch sequence. However, as evidenced by the activation efficiency in both cases, that higher infectivity does not correlate with better riboswitch sequences (Figure 7). The best example are the results from population 5, which show the highest virulence values, but with no increase or even some decrease in activation efficiency, at least in the positive selection. There are several possible explanations for this phenomenon, the simplest being that a particular sequence that was constitutively "ON" was present in the population in such numbers that it was able to overcome the competitors during the growth assays. When that generation is checked via NGS (Figure 5, S8 in A and S10 in B), no sequence has already been completely fixed within the population, which would explain why virulence indexes from then onwards improved with further selection steps.

Conclusion

To our knowledge, our use of T7AE with riboswitches represents the first instance of a phage-assisted evolution of a non-coding RNA. Even though the technical design may be improved upon in future iterations, we are confident in the positive impact that such a strategy could have in the field of riboswitch and biosensor development. When taken together, virulence index and NGS data seem to confirm that our strategy allows for the selection of efficient riboswitch sequences from a pool of random riboswitches, as evidenced in Figures 4 and 6. This gives us confidence in the ability of the process to directly evolve riboswitches. We have also concluded that phage virulence does not directly correlate to riboswitch efficiency, as some phages have increased the virulence over subsequent generations, despite carrying seemingly less efficient riboswitch sequences. This could be due to some phages carrying constitutive sequences, that give them an advantage in a specific selection step, or specific sequences affecting regulation in unplanned ways. Furthermore, our data shows that the obtained riboswitches have improved

characteristics when compared to the switch used as positive control for the process. This switch was ratified by the literature (28), and the comparison can be seen in Figures 6 and 7. Further developments include modifying T7AE to follow the PACEmid approach, where we would evolve satellite T7 phages. Cosmids of the T7 phage have already been used in evolution (39), where only the packaging signal is required in the evolving DNA. Our T7AE selection could be applied to obtain riboswitches with novel ligand specificities and, more generally, our dual selection provides a methodology for the phage-assisted accelerated evolution of gene switches.

Materials and methods

LB media: All bacterial strains were grown in LB media for the different experiments, with any alternative media used only for specific experiments, and described accordingly. The formulation for 1L includes 10 g Peptone 140, 5 g yeast extract, and 5 g sodium chloride, diluted in ddH2O.

Saline-Magnesium buffer: Solution used to store phage samples, avoiding the risk of contamination by microorganisms (40). For 1L, 50 mM Tris-HCl 7.5, 100 mM NaCl, 100 mM $MgSO_4$ diluted in ddH20.

Cell strains: Cells from the Keio Collection (41), namely the parental Strain BW25113, and KO for genes *cm* and *trxA* (CGSC, http://cgsc2.biology.yale.edu). The Keio Collection cells were used to transform the confirmed plasmids and to use the resulting cells for the positive and negative selections.

Plasmids

Plasmid specifications and assembly details are described in the Supporting information.

Primers and gBlocks: All gene sequence constructs produced by Integrated DNA Technologies and can be found in Table S2. Melting temperatures (T_m) of the primers were calculated using the NEB calculator. PCR reactions were carried out using Phusion Master Mix from NEB in a final volume of 25 µL for diagnostic PCRs, and 50 µL for purification PCRs.

Digestion enzymes: All used enzymes were part of the FastDigest line from Thermo Scientific. All digestions were done at 37 ºC, using 200 ng of DNA template, 1 µl of FastDigest enzymes, and 10x buffer solutions from ThermoFisher in a final volume of 20 µL.

Polymerase Chain Reaction: All PCR reactions used for plasmid assembly were done using the 2x Phusion mix from NEB, in a 50 µL volume, according to manufacturer specifications. In cases where the secondary structure of the sequences interfered with the amplification, different percentages of DMSO were used (3, 6 %) to help the correct amplification. Diagnostic PCRs from colonies or phage samples were done using a 2x Taq or GreenTaq Mix (Thermo Scientific) under manufacturer's specifications.

Goldengate assembly: Based on Engler, Kandzia and Marillonnet (42). The desired plasmids were amplified via PCR. The primers used contain an overhang not present in the original template, which introduces a BsaI/Eco31I cutting site. Compatible ends were included into the tails of primers used to amplify the desired fragment introduced in the backbone. Using 100 pmol of each DNA template, 1 µL of BsaI (New England Biolabs/NEB), 1 µL of T4 DNA (Fermentas), 2 µL of 10x ligase buffer and H2O up to 20 µL, each reaction tube underwent the following

series of cycles: 2 minutes at 37°C followed by 5 minutes at 16°C, 35 cycles, 5 minutes at 50°C and finally 5 minutes at 80°C.

Agarose electrophoresis: Amplified samples were run in 1% agarose gels (43), to confirm their size and approximate quantity, at 95V for 30 minutes for most experiments. Smaller samples were run in 2% agarose gels.

Purifications: All PCR reactions were purified using the Thermo GeneJET PCR purification kit. In cases where several bands appeared when running the sample in a gel, the GeneJET Gel extraction kit was used.

Transformations: Plasmids were transformed into both electrically and chemically competent cells. For transformations in electrically competent bacteria, a minimum of 100 ng of DNA were added to 50 µL of competent cells and left on ice for 15 minutes, along with an electroporator cuvette. The volume was then transferred to said cuvette, the cells electroporated and then grown in LB media for 1 hour before being plated. For chemically competent cells, a minimum of 100 ng of DNA was added to the 50 µL of cells, which were left on ice for 30 minutes, then heat shocked at 42ºC for 90 seconds, and returned to ice for 10 minutes before recovery in LB media for an hour at 37ºC. Cells were then plated in an LB agar plate with their corresponding antibiotic resistances.

Growth: To obtain cell cultures for the various assays, colonies picked from grown plates were grown overnight in 5 ml of LB with the appropriate antibiotic, using Falcon 15 ml round bottom tubes. In the cases of positive selection strains, to prepare them for the assays, 300 µL of 25 mM Theophylline were added to the 5 ml of LB. Overnight cultures were refreshed the next

day and used when an OD between 0.2 and 0.3 was reached. The remaining overnight culture was used to create glycerol stocks.

Phage killing and homologous recombination: To obtain phage populations or test their killing efficiencies, phages were added to cultures within the growth phase (OD 0.2-0.3), making sure the phages would be able to properly replicate. In cases where the objective was to produce homologous recombination, a rate of 1 phage per 10^5 cells was used. These tests required phages to infect Δ*trxA* cells containing gp5/T7 DNA polymerase. This ensured that only recombinant phages would be able to replicate, as trxA is an essential protein for the process (5). All lysates were then centrifuged for 10 min at 3000 rpm and filtered with 2 µm filters, obtaining clean phage samples.

Selection procedure: 8 different strains were tested for each of the selection steps, using the control riboswitch to find the most efficient strain to be used for evolution. This was done by infecting each of the strains of interest with the recombined T7 phages in the following ways:

- Growth assays: In a 96-well plate, each well is filled with 180 µL of cells at an OD of 0.2-0.3, and 20 µL of the tested phages in a decreasing order along the length of the plate. In the case of cultures with theophylline, two ways were tested: 180 µL of a culture already containing theophylline, or 163.5 µL of culture plus 12.5 µL of 25 mM theophylline, to which the phages were added. As a way to keep all records the same, and have a consistent concentration in the cultures, the first method was preferred. The plate reader (model, make) took OD measurements every 2 minutes for a duration of 5 hours. To compare the differences between strains and select the best candidates for the

evolution procedure, the growth curves for each cell strain were plotted using Microsoft Excel and Graphpad Prism 9.

- Plaque assays: These assays were developed to check the replication efficiency of the phages (44). By infecting 300 µL of cells at an OD of 0.2-0.3 with 100 µL of phage at different dilutions, mixing it with 3 ml of soft LB agar (LBA 50% in LB), and then plating it, clear spots or "plaques" can be observed in the bacterial lawn of the plate. A higher number of plaques indicates a higher number of phages, indicating which strains were better for the replication of the phages. These assays were made both in the presence and absence of 1.5 mM Theophylline, to assess the "leakiness" of the system and the activation of the riboswitch.

- One-Step assay: These assays (Figure S2) follow the same procedure as a plaque assay, but phage samples were obtained from infecting a culture (45). Samples were taken every 4 minutes over a period of 40 minutes, showing the rise in the number of phages as the infection progresses. The assays were also made in the presence or absence of 1.5 mM theophylline to determine the differences due not only to the two types of sample treatment, but to the addition of the activating molecule to the media, thus affecting the phages' replication process. These phage samples were then used for plaque assays as described previously.

Next-Generation Sequence Analysis: This strategy was designed following the 16S amplicon sequencing protocol from Illumina, using its Nextera XT DNA Library Kit (Illumina #FC-131-001) to prepare the samples for sequencing in an Illumina MiSeq (Figure S2). A first PCR was carried out to amplify a 175 bp fragment from the phage genome containing the closest

elements present upstream and downstream from the region containing the 8 randomized nucleotides of the riboswitch (Figure S3A). The primers used in this PCR also incorporated sequences necessary to carry out the indexing PCR for the sequencing (Figure S3B). Combinations of primers for the Index PCR are shown in Table S2.

Riboswitch sequence analysis: Due to the lack of available computing power, the initial set of analyses were done in a fraction of the total number of sequences, approximately 10^5 per sample.

Sequence quality control was performed using the FastQC software (46). After trimming, all reads corresponding to each sample were aligned to the initial "parent" sequence (as designed and cloned into the initial vector prior to selection) using a recursive Needleman-Wunsch pairwise alignment algorithm (47). Once aligned, all reads were trimmed from both ends until only the 8 nt-long riboswitch sequence plus 5 bases on either side remained. Trimming was performed using base R, by finding the 8 nucleotide long "NNNNNNNN" motif in the parent sequence and extracting this position (±5bp) for all reads in each sequence. Once extracted, nucleotide proportions were calculated by stacking each position and counting the number of each nucleotide that was informative (not ambiguous), which is considered a Bit. The program does not only count which percentage of each nucleotide is present in every position, as there are many reads showing an ambiguous read or just a gap in these positions. To account for this, it creates a list and only counts the number of times every base is clear, represented as an informative Bit, according to information theory. The entire process of nucleotide counting and plotting was performed using the R package *ggseqlogo* (48).

Evolution assessment via virulence index calculation

The selected procedure shown previously (35,49), was used, by which a value known as the virulence index was ascribed to different phages to compare them to one another. The virulence index was obtained via a two-step process:

I. Calculating the local virulence (υ1) via the following formula:

Equation 1: $\upsilon 1 = 1 - \frac{Ai}{A0}$

where A_0 represents the area under the growth curve of a cell culture without phages until reaching stationary phase, and A_i the area under the curve of a culture at a certain MOI until the same timepoint. υ1 is measured in a range of 0 to 1, with 0 being no virulence whatsoever, and 1 being the maximum theoretical virulence, instantaneous death of the infected cells.

.

II. Once the local virulences over a range of MOIs were assessed, they were represented on a curve. The area under that curve (A_p) was calculated following the formula

Equation 2: $A_p = \int_i^0 \upsilon 1 \mathrm{d}(\log MOI)$

with *i* being the base10 log of the lowest MOI, which in this case was -6.

III. That value was then divided by the theoretical maximum value (A_{max}) for the same curve, giving the virulence value for a specific phage on a specific strain. In this case, this maximum value corresponded to 6.

IV. These values were readily comparable with each other, and were taken both as they were, and to calculate the "activation efficiency" for the riboswitches inside the phages, indicating the fold increase between states of the switch in those cases.

The virulence index measurements were made for 5 different phage populations: 3 at different stages of selection, 1 for the original riboswitch-carrying phage, and one for WT T7. Each of these phages was used to infect the positive and negative selection strains under three conditions: 0 mM, 1 mM and 1.5 mM Theophylline; to assess their virulence under each condition. All calculations were made using Microsoft Excel 2019 and Graphpad Prism 9. Statistical significance between the different virulence indexes was done via Two-way ANOVA, with Tukey´s multiple comparison test.

Riboswitch Activation metric: This is a value representing the activation fold of the riboswitch in each of the populations. This was obtained by dividing the virulence values for a particular strain in cases where the switch should be "ON", by the value in the case where it should be "OFF". This allows to see the fold increase in virulence between the presence/absence of theophylline and thus, the activation/inactivation fold of the riboswitch.

Supporting information

The supporting information includes a list and description of used plasmids and their assembly methods, and a list of relevant primers and their sequences, as well as diagrams of several experimental procedures and nomenclatures.

Abbreviations

cmk: CMP kinase *trxA*: Thioredoxin A MFE: Minimum free energy

MOI: Multiplicity of infection PACE: Phage Assisted Continuous Evolution

SELEX: Systematic Evolution of Ligands by Exponential enrichment

Acknowledgements

We thank WISB for the use of their facilities during this work. This work was supported by the European Union's Horizon 2020 research and innovation programme under ITN Marie Sklodowska-Curie (grant agreement 642738 [MetaRNA]), the Biotechnology and Biological Sciences Research Council (grant BB/P020615/1 [EVO- ENGINE], the Spanish Ministerio de Ciencia e Innovación´s (grant PID2020-118436GB-I00 [GeneCircuits++]) and the European Innovation Council (grant 101070948 [PhotoSynH2]) to AJ.

References

1. Esvelt KM, Carlson JC, Liu DR. A System for the Continuous Directed Evolution of Biomolecules. Nature. 2011;472(7344):499–503.
2. Brödel AK, Rodrigues R, Jaramillo A, Isalan1 M. Engineering the smallest transcription factor: accelerated evolution of a 63-amino acid peptide dual activator-repressor Andreas. Sci Adv. 2019;44(0):1–29.
3. Bernard E, Wang B. Synthetic Cell-Based Sensors with Programmed Selectivity and Sensitivity. Methods Mol Biol. 2017;1572:349–63.
4. Rugbjerg P, Genee HJ, Jensen K, Sarup-Lytzen K, Sommer MOA. Molecular Buffers Permit Sensitivity Tuning and Inversion of Riboswitch Signals. ACS Synth Biol. 2016;5(7):632–8.

5. Qimron U, Marintcheva B, Tabor S, Richardson CC. Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage. Proc Natl Acad Sci USA. 2006;103(50):19039–44.

6. Rodrigo G, Landrain TE, Jaramillo A. De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. Proced Natl Acad Sci United States Am. 2012;109(38):15271–6.

7. Shen S, Rodrigo G, Prakash S, Majer E, Landrain TE, Jaramillo A, et al. Dynamic signal processing by ribozyme-mediated RNA circuits to control gene expression. Nucleic Acids Res. 2015;43(10):5158–70.

8. Rostain W, Shen S, Cordero T, Rodrigo G, Jaramillo A. Engineering a Circular Riboregulator in Escherichia coli . BioDesign Res. 2020;2020:1–9.

9. Rodrigo G, Jaramillo A. RiboMaker : computational design of conformation-based riboregulation. Bioinformatics. 2014;30(17):2508–10.

10. Tucker BJ, Breaker RR. Riboswitches as versatile gene control elements. Curr Opin Struct Biol. 2005;15(3 SPEC. ISS.):342–8.

11. Omotajo D, Tate T, Cho H, Choudhary M. Distribution and diversity of ribosome binding sites in prokaryotic genomes. BMC Genomics. 2015;16(1):1–8.

12. Sudarsan N, Barrick JE, Breaker RR. Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA. 2003;(203):644–7.

13. Fortier LC. The Contribution of Bacteriophages to the Biology and Virulence of Pathogenic Clostridia. Adv Appl Microbiol. 2017;101:169–200.

14. Wang S, White KA. Riboswitching on RNA virus replication. Proc Natl Acad Sci U S A. 2007;104(25):10406–11.

15. Dixon N, Duncan JN, Geerlings T, Dunstan MS, McCarthy JEG, Leys D, et al. Reengineering orthogonally selective riboswitches. Proc Natl Acad Sci U S A. 2010;107(7):2830–5.

16. Link KH, Breaker RR. Engineering ligand-responsive gene-control elements : lessons learned from natural riboswitches. Gene Ther. 2009;16(10):1189–201.

17. Bengert P, Dandekar T. Riboswitch finder — a tool for identification of riboswitch RNAs. Nucleic Acids Res. 2004;32:154–9.

18. Berens C, Suess B. Riboswitch engineering - making the all-important second and third steps. Curr Opin Biotechnol. 2015;31:10–5.

19. Groher F, Suess B. Synthetic riboswitches - A tool comes of age. Biochim Biophys Acta - Gene Regul Mech. 2014;1839(10):964–73.

20. Gong S, Wang Y, Wang Z, Zhang W. Computational methods for modeling aptamers and designing riboswitches. Int J Mol Sci. 2017;18(11):1–19.

21. Lynch SA, Gallivan JP. A flow cytometry-based screen for synthetic riboswitches. Nucleic Acids Res. 2009;37(1):184–92.

22. Wachsmuth M, Findeiß S, Weissheimer N, Stadler PF, Mo M. De novo design of a synthetic riboswitch that regulates transcription termination. 2013;41(4):2541–51.

23. Topp S, Gallivan JP. Random Walks to Synthetic Riboswitches — A High-Throughput Selection Based on Cell Motility. ChemBioChem. 2008;25(9(2)):210–3.

24. Brödel AK, Jaramillo A, Isalan M. Engineering orthogonal dual transcription factors for multi-input synthetic promoters. Nat Commun. 2016;7:13858.

25. Grigonyte AM, Harrison C, MacDonald PR, Montero-Blay A, Tridgett M, Duncan J, et al. Comparison of CRISPR and marker-based methods for the engineering of phage T7. Viruses. 2020;12(2).

26. Ford E, Ares M. Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozimes derived from a group I intron of phage T4. Proced Natl Acad Sci United States Am. 1994;91(April):3117–21.

27. Hendrix RW. Bacteriophage genomics. Curr Opin Microbiol. 2003;6(5):506–11.

28. Lynch SA, Desai SK, Sajja HK, Gallivan JP. A High Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights into their Function. Chem Biol. 2007;14(2):173–84.

29. Fehér T, Karcagi I, Blattner FR, Pósfai G. Bacteriophage recombineering in the lytic state using the lambda red recombinases. Microb Biotechnol. 2012;5(4):466–76.

30. Huang D, Holtz WJ, Maharbiz MM. A genetic bistable switch utilizing nonlinear protein degradation. J Biol Eng. 2012;6:1–13.

31. Yokobayashi Y, Arnold FH. A dual selection module for directed evolution of genetic circuits.

Nat Comput. 2005;4(3):245–54.

32. Brödel AK, Isalan M, Jaramillo A. Engineering of biomolecules by bacteriophage directed evolution. Curr Opin Biotechnol. 2018;51:32–8.

33. Garcia LR, Molineux I a NJ. Incomplete entry of bacteriophage T7 DNA into F plasmid-containing Escherichia coli. J Bacteriol. 1995;177(14):4077–83.

34. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, et al. Software News and Updates NUPACK: Analysis and Design of Nucleic Acid Systems. J Comput Chem. 2011;32:170–3.

35. Storms ZJ, Teel MR, Mercurio K, Sauvageau D. The Virulence Index : A Metric for Quantitative Analysis of Phage Virulence. Phage. 2019;1(1):17–26.

36. Schmitt CK, Molineux IJ. Expression of gene 1.2 and gene 10 of bacteriophage T7 is lethal to F plasmid-containing Escherichia coli. J Bacteriol. 1991;173(4):1536–43.

37. Cheng X, Wang W, Molineux IJ. F exclusion of bacteriophage T7 occurs at the cell membrane. Virology. 2004;326(2):340–52.

38. Wang WF, Cheng X, Molineux IJ. Isolation and identification of fxsA, an Escherichia coli gene that can suppress F exclusion of bacteriophage T7. J Mol Biol. 1999;292(3):485–99.

39. Yosef I, Goren MG, Globus R, Molshanski-Mor S, Qimron U. Extending the Host Range of Bacteriophage Particles for DNA Transduction. Mol Cell. 2017;66(5):721-728.e3.

40. Lee WJ, Billington C, Hudson JA, Heinemann JA. Isolation and characterization of phages infecting Bacillus cereus. Lett Appl Microbiol. 2011;52(5):456–64.

41. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. Mol Syst Biol. 2006;2.

42. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PLoS One. 2008;3(11):1–7.

43. Serwer P. Agarose gels: Properties and use for electrophoresis. Electrophoresis. 1983;4(6):375–82.

44. Kropinski AM, Mazzoco A, Waddell TE, Linghor E, Johnson RP. Bacteriophages : Methods in Molecular Biology. In: Clokie MRJ, Kropinski AM, editors. Methods in molecular biology.

Humana Press; 2009. p. 69–76.

45. Yu YP, Gong T, Jost G, Liu WH, Ye DZ, Luo ZH. Isolation and characterization of five lytic bacteriophages infecting a Vibrio strain closely related to Vibrio owensii. FEMS Microbiol Lett. 2013;348(2):112–9.

46. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

47. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suit. Trends Genet. 2000;16(6):276–7.

48. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.

49. Xie Y, Wahab L, Gill JJ. Development and validation of a microtiter plate-based assay for determination of bacteriophage host range and virulence. Viruses. 2018;10(4):1–16.

Supporting Information

Accelerated evolution using the phage T7 and its application to evolve riboswitches

Authors

Eduardo Goicoechea[1] and Alfonso Jaramillo[1,2]

Affiliations

[1] Warwick Integrative Synthetic Biology Centre and School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK.

[2] *De novo* Synthetic Biology Lab, I2SysBio, CSIC-University of Valencia. Parc Científic Universitat de València. Calle Catedrático Agustín Escardino, 9. 46980 Paterna. Spain.

* To whom correspondence should be addressed:

Prof. Alfonso Jaramillo. *De novo* Synthetic Biology Lab. I2SysBio, CSIC-University of Valencia. Parc Científic Universitat de València. Calle Catedrático Agustín Escardino, 9. 46980 Paterna. Spain. https://orcid.org/0000-0002-6313-9689, Alfonso.Jaramillo@synth-bio.org.

Supplementary Materials and Methods

Plasmid List:

Below is a list of the different plasmids utilized throughout the project, with their official denomination, their name throughout the project, and the serial number in the lab's records. Their specific assembly is discussed in the next section.

- Pr100 pLit Chlor/EG_001/PAJ343:_3494 bp, with a pMB1 origin of replication and Chloramphenicol resistance. Genes 4.7 and 5.3 from the T7 phage's genome; both flanking a T7 promoter followed by the gene for thioredoxin A to incorporate them via homologous recombination.

- pSEVA631/EG_002: 3001 bp, containing the default multiple cloning site flanked by an rrnB T1 terminator and a lambda t0 terminator, a pBBR1 origin of replication, and a gentamicin resistance. This plasmid was used as the backbone for all selection plasmids.

  Parental positive selection plasmid (EG002+): 7135 bp

  Parental negative selection plasmid (EG002-): 8677 bp

- HRF plasmid/EG_cI01/PAJ341: 3569 bp, based upon the Δ*cmk* version of Pr100 pLit Chlor. Via the use of GoldenGate assembly, a gene fragment containing a T7 terminator, a T7 promoter, and a fixed version of the theophylline riboswitch with a high-fold activation, taken from (2), as seen in Figure 1B. This plasmid was used as a positive control for the efficiency of the method.

- HRR plasmid/EG_cI02/PAJ342: 3569 bp. Exactly the same elements as EG_c01, but with the riboswitch contains a randomised region of 8 nucleotides after the stem of the riboswitch and before the RBS (Figure 1C). Thus, a plasmid library of riboswitches containing theoretically 65,536 variants. These will be used for experiments once the method has been properly set up for the fixed version of the riboswitch.

- Positive selection variants: Based upon the positive selection plasmid, different combinations of promoter and RBS were tested to select for the ideal one to use in the current experimental design. This meant a combination of promoter and RBS that gave very low production of *cmk* for a riboswitch in an "OFF" state (without theophylline), and a high one for an "ON" state riboswitch (with theophylline). 8 variants were used, 6 of them test strains, and 2 of them used as controls.

    - The positive control/EG_cI03/SAJ782 (6093 bp) contained a T7 promoter and a strong synthetic RBS in control of the *cmk* gene, ensuring its expression and the proper replication of T7 phages, regardless of the presence of theophylline.

    - The negative control/EG_cI10/SAJ789 (6054 bp) contained neither promoter nor RBS in front of *cmk,* so phages would not be able to replicate under any circumstances, as the transcription system could not be initiated in this case

    - The test variants/EG_cI04-09/SAJ783-8 (6155 bp) contained the previously described pRM promoter followed by 1 of 6 different RBSs of different

strengths. These RBSs were known as B0030, B0031, B0032, B0033, B0064, and B00Syn (1).

- Negative selection variants: As in the positive selection, 8 variants were tested, 2 of them being controls and 6 of them being tests. All the variations are as previously described, except for the plasmids containing *pifA* instead of *cmk.* Positive control/EG_cI11/SAJ700: 7635 bp. Negative control/EG_cI18/SAJ707: 7600 bp. Test variants/EG_cI12-17/SAJ701-6: 7700 bp.

Plasmid assembly specifications:

All the constructs assembled were based on previously available plasmids present in the laboratory's stock. The assemblies were made either by enzymatic digestion at 37 ºC and overnight ligation at 20ºC, or using the process known as GoldenGate ligation.

- EG001: The Homologous Recombination plasmid was based on a pre-existing plasmid present in the lab, Pr100 pLit Chlor, itself based on pLITMUS 28 (https://www.neb.com/products/n3528-litmus-28i-vector); and containing the genes for CMP/dCMP kinase and thioredoxin A, making it 4226 bp. The plasmid was modified further before assembling the final version, eliminating the *cmk* gene via enzymatic digestion, making it 3490 bp.

- EG_002: Based upon the European standard of SEVA plasmids (2), allowing for easy to reproduce modularity. This original plasmid underwent further modifications to assemble the positive and negative selection plasmids. Such modifications included the addition of an error-prone version of T7 phage's DNA polymerase controlled by a T7 promoter in the 5' direction, and BsaI sites flanking an R0010 promoter and Ribosome

Binding Site (From here onwards RBS) controlling RFP in the 3' direction, followed by B0010 and B0012 terminators; followed by the selection gene (*cmk* or *pifA* for the positive and negative selections, respectively).

- Selection plasmids: All different cargos were introduced into EG002 via Goldengate assembly, using the previously mentioned GoldenGate cutting sites

Supplementary Figures

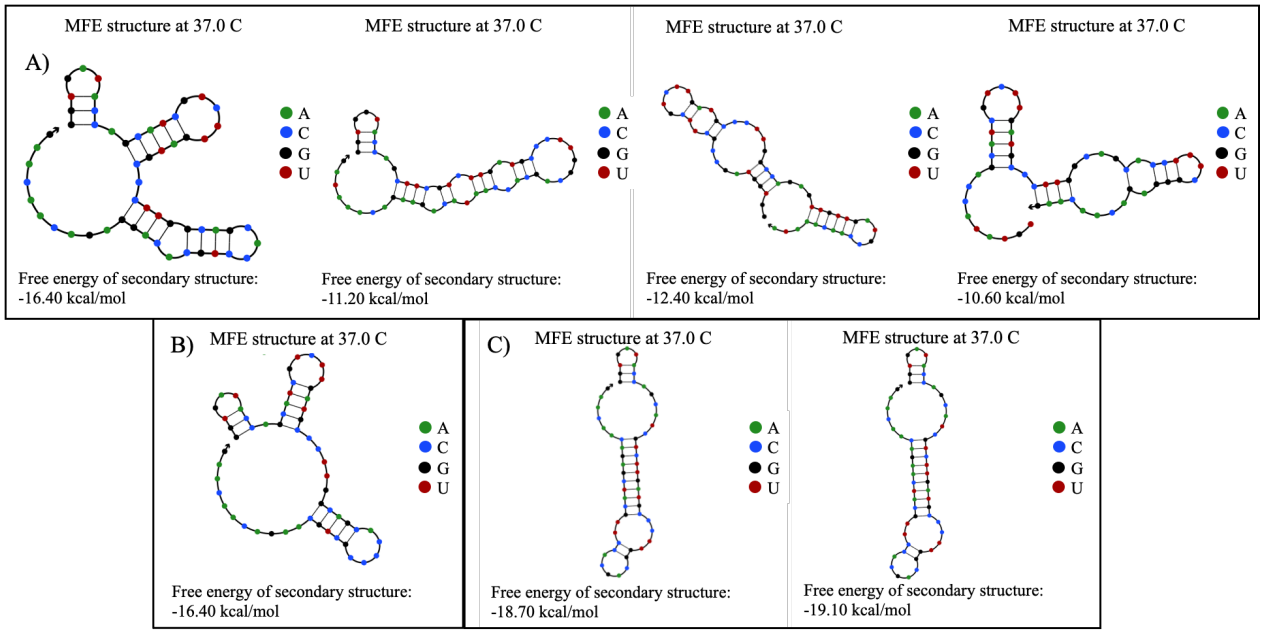


Figure S1. Structural comparison between riboswitch sequences. Using NUPACK (3), the secondary structures corresponding to different sequences throughout the selection process were predicted, as well as their MFE (minimum free energy). **A)** Sequences from the random pool of riboswitches from the early stages in the selection. **B)** Riboswitch used as a standard and positive control, with a fixed sequence (CCGCUGCA). **C)** Sequences that were overly represented by the final steps of the selection, as seen in Figure 3B. Left structure: TTGCATCG, Right structure: TTGCATCA. This shows that starting from a random pool of switches, we have been able to obtain sequences which managed to rival and even surpass the stability of the positive control switch.

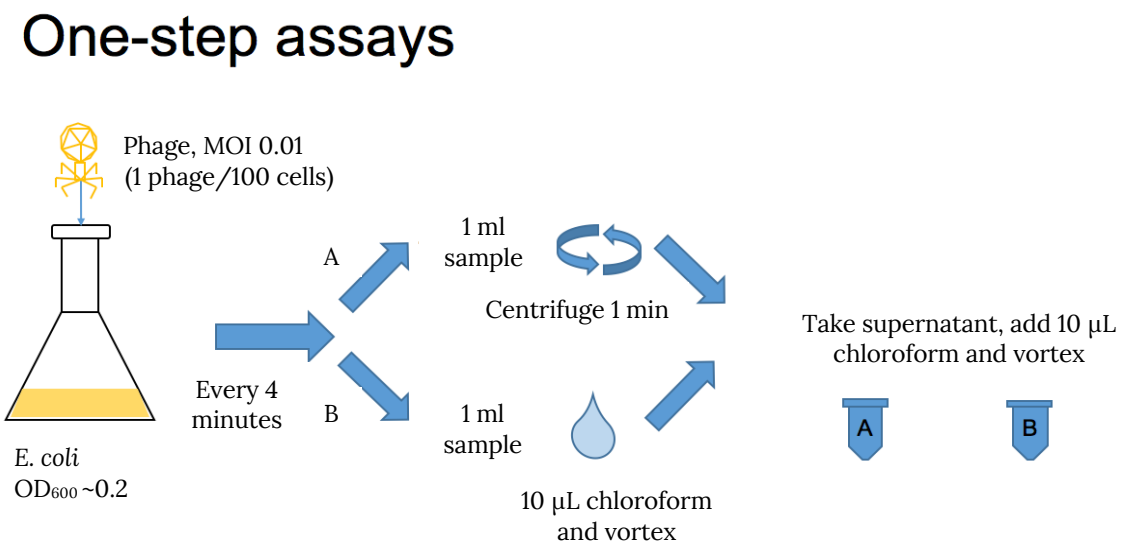


Figure S2. General diagram of a one-step assay. Once the phages, at a MOI (Multiplicity of Infection) of 10^{-2} (1 phage/100 cells) have been added to the culture, samples are taken and undergo different treatments. "A" are centrifuged to pellet the cells and obtain the phages present in the supernatant, while "B" samples have Chloroform added to them, causing bacteria to burst and release phages that had not yet exited the cells. To eliminate any remaining bacteria present in the final samples, Chloroform is added again, and those phage samples are then used for a plaque assay, where a difference should be observable between different samples and time points.

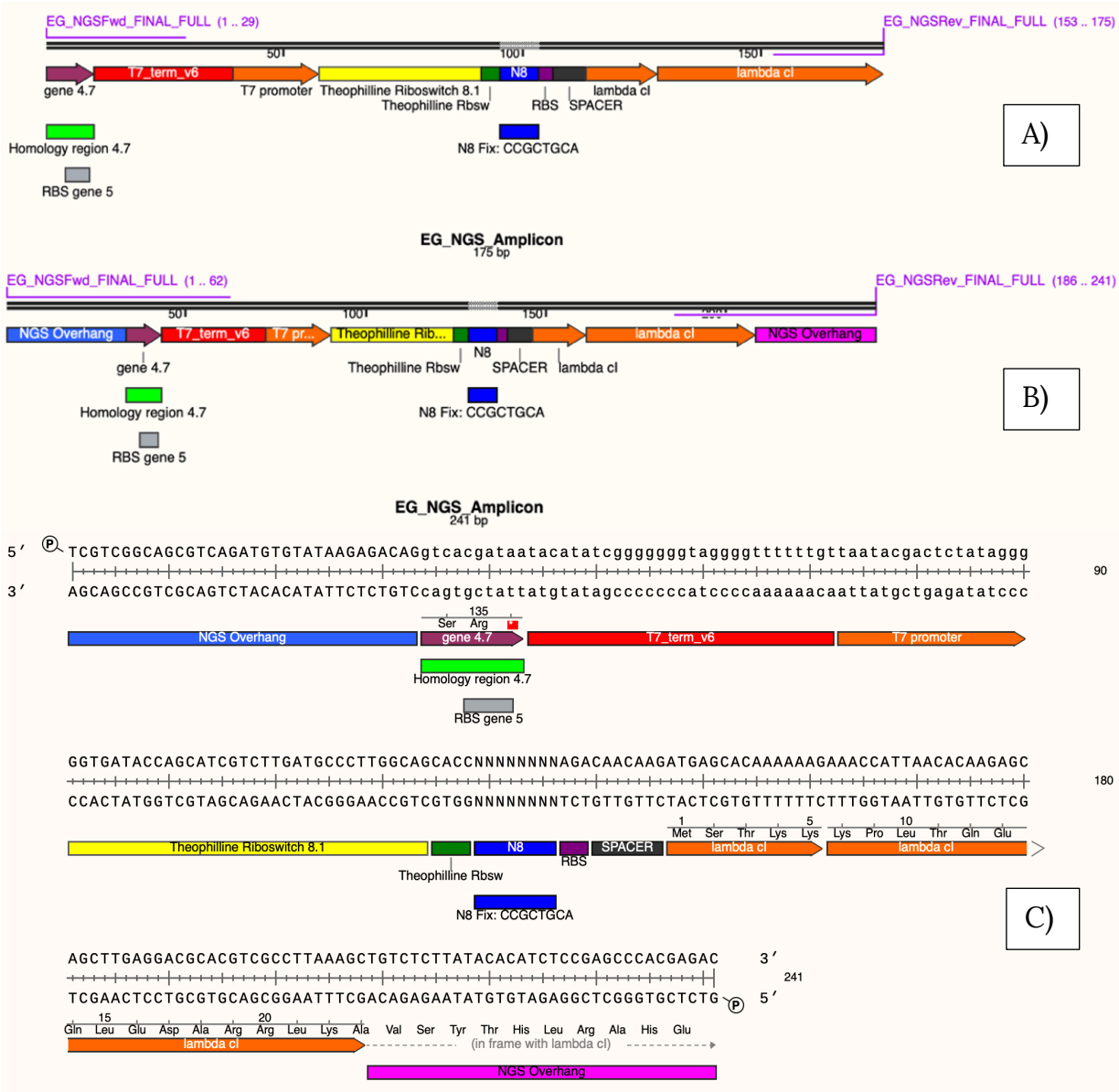


Figure S3. Amplification steps for NGS preparation. Selected amplicon to be used for the NGS procedures. A) Original amplicon as it is present in the recombined phages' genome, 175 base pairs in size, containing fragments upstream and downstream from the random riboswitch sequence. B) Amplicon obtained after the PCR, with the desired primers having added the required Illumina overhangs for the Indexing PCR. C) Detailed sequence of the final amplicon, indicating the original sequences and the overhangs.

SAMPLES A	NUMBER	SAMPLES B	NUMBER
PLASMID Pool	S0	**Rd Switch 1**	S1
Fix Switch	S2	**Rd Switch 2**	S2
Rd Switch 1	S3	**Rd Switch 3**	S3
Rd Switch 2	S4	**cmk1**	S4
Cmk1	S5	**pifA1**	S5
cmk2	S6	**cmk3**	S6
pifA2	S7	**pifA3**	S7
cmk5	S8	**cmk4**	S8
pifA6	S9	**pifA4**	S9
cmk7	S10	**cmk5**	S10
pifA7	S11	**pifA5**	S11
pifA9	S12	**cmk6**	S12
pifA10	S13	**cmk8**	S13
cmk11	S14	**pifA8**	S14
pifA11	S15	**cmk9**	S15
cmk12	S16	**cmk10**	S16
pifA12	S17	**pifA11**	S17
pifA12B	S18	**cmk12**	S18
pifA13	S19	**cmk13**	S19
pifA13D	S20	**cmk15**	S20
pifA13B_8-11	S21	**pifA15**	S21
cmk14	S22	**cmk16**	S22
pifA14_6-11	S23	**cmk2 Amplif.**	S23
pifA14	S24	**cmk16 Amplif.**	S24

Table S1. NGS sample nomenclature. Correspondence between the generation number of the different analysed libraries and the nomenclature used in the sequence variation diagram. cmk or pifA indicated whether the sample corresponded to a positive or negative step in the selection, and the number indicated the specific generation.

.

PRIMER NAME	**USE**	**SEQUENCE**
		Designed primers
EG_GG01	HR Goldengate fragment addition Fw	cttcgaagacgataccagcatcgtcttgatgcccttggcagcacctgagacca tagaatgagcacaaaaaaggtctcgaaaccattaaca
EG_GG02	HR Goldengate fragment addition Rev	cttcgaagacctggtatcacccccctatagagtcgtattaacaaaaaaccccctac ccccccgatatgtattatcgtgacttaacaatctct
EG_GG03	cI insert + Rd Rbsw Fwd	aaggtctcccaccnnnnnnnnagacaacaagatgagcacaaaaaagaaa ccgagaccaa
EG_GG04	cI insert + Fix Rbsw Fwd	aaggtctcccaccccgctgcaagacaacaagatgagcacaaaaaagaaac cgagaccaa
EG_GG05	cI insert+ Rbsw Rev	ttggtctcggtttctttttgtgctcatcttgttgtc
EG_064	Goldengate presence test Fw	tgccgataagaccaacatga
EG_065	Goldengate presence test Rv	gcatcaagacgatgctggta
EG_068	Eliminate *cmk* from HR plasmid Fw	ataaacagaggagatatcacatgaccgcaattgcaccggt
EG_069	Eliminate *cmk* from HR plasmid Rev	agtagcctcctgtagtcgacttattaggccagcgccagtt
EG_096	Detect *pifA* presence	cggaagaaagagcgcttcaaccggactgccg
EG_097	*Gene 5* Fwd	ctcaaggtcgccaaagtgg
EG_098	*Gene 5* Rev	gaaggacactttaaatttctttgtcatagcc
EG_099	Change RBS for *cmk* in *mmyb* plasmid (Fw)	aggtctcaataagctcatccgcggsaataattgagraggwcagataatgacc gcaattgcaccggtaattacg
EG_100	Change RBS for *cmk* or *pifA* in *mmyb* plasmid (Rev)	aggtctcattattacgctgcaagggcgtaattttcgtcg
EG_101	Change RBS for *pifA* in *mmyb* plasmid (Fw)	aggtctcaataacccgacaaacccgactmcataagswggtagcagtatga agatactgaggcagctgtggaacc
EG_102	Change origin of replication in *mmyb* plasmid (Fw)	aggtctcagcacgatatacaggattttgccaaagggttcg
EG_103	Change origin of replication in *mmyb* plasmid (Rev)	aggtctcagtgcgaaaaaggatggatataccgaaaaaatcgc
EG_104	Assembly of control for negative selection (Fw)	gtaaggcgccgtcgtcat
EG_105	Assembly of control for negative selection (Fw)	tcatatgacgacggcgcc
EG_106	*Gene 5* base addition for insertion in phagemid plasmid Fw	aggtctcatcgagataaacagaggagatatcacatgattgtg
EG_107	Phagemid base addition for *gene 5* insertion Rev	aggtctcatcgagtcagccaaacgtctcttc
EG_108	*Gene 5* base addition for insertion in phagemid plasmid Rev	aggtctcagtcgacttagtggcaaatcgcc
EG_109	Phagemid base addition for *gene 5* insertion Fw	aggtctcacgacctgcagtaaagcccgcttcg

EG_112	Bind to middle of *pifA* for detection	atgatctggaccgactggagcc
EG_115	Confirmation of plasmid presence 1 (Fw)	ggtttttttgtttgcaagcagcagattacg
EG_116	Confirmation of plasmid presence 1 (Rev)	cggggagtcaggcaactatgg
EG_117	Confirmation of plasmid presence 2 (Fw)	gtatttagaaaaataaacaaataggggttccgcg
EG_118	Confirmation of plasmid presence 2 (Rev)	ttgtgagcggataacaatttcacacagg
EG_119	Confirmation of plasmid presence 3 (Fw)	ctggggtgcctaatgagtgagc
EG_120	Confirmation of plasmid presence 3 (Rev)	gtcgatttttgtgatgctcgtcagg
EG_123	Mid λ cI to check for riboswitch (Rev)	gttgtgcttacccatctctccg
NGS procedure		
EG_NGS _Fwd	Initial NGS amplification primer, 2 bases missing, Fw	tcgtcggcagcgtcagatgtgtataagagacgtcacgataatacatatcgggggggtagg
EG_NGS _Rev	Initial NGS amplification primer, 2 bases missing, Rev	gtctcgtgggctcggagatgtgtataagagacctttaaggcgacgtgcgtcctc
EG_NGSFw_ Final_Full	NGS amplification primer Fw	tcgtcggcagcgtcagatgtgtataagagacaggtcacgataatacatatcgggggggtagg
EG_NGSRev_Final _Full	NGS amplification primer Rev	gtctcgtgggctcggagatgtgtataagagacagctttaaggcgacgtgcgtcctc
		LABORATORY STOCK PRIMERS
Negative Selection Assembly		
pifA_T7te _Fw	Addition of sequence and Eco31I site upstream of *pifA* (Fw)	ttcgggtgggcctttctgcgtttataactagtggtctccatgaagatactgaggcagctg
re_prm30	Addition of sequence containing RBS B0030 and Eco31I site (Rev)	ttatttggtctcttcatctagtatttctcctctttaatctctagacaagatacattgtaaa
re_prm31	Addition of sequence containing RBS B0031 and Eco31I site (Rev)	ttatttggtctcttcatctagtaggtttcctgtgtgactctagacaagatacattgtaaa
re_prm32	Addition of sequence containing RBS B0032 and Eco31I site (Rev)	ttatttggtctcttcatctagtactttcctgtgtgactctagacaagatacattgtaaa
re_prm33	Addition of sequence containing RBS B0033 and Eco31I site (Rev)	ttatttggtctcttcatctagtagtcctgtgtgactctagacaagatacattgtaaa
re_prm64	Addition of sequence containing RBS B0064 and Eco31I site (Rev)	ttatttggtctcttcatctagtatttcccctctttctctagacaagatacattgtaaa
re_prmSyn	Addition of sequence containing Synth RBS and Eco31I site (Rev)	ttatttggtctcttcattctctacctccttaatgtgttctattgacgactcacaagatacattgtaaa
T7rbsFWD	Addition of T7 RBS sequence and Eco31I cut site upstream of *pifA* (Fw)	aataaaggtctctgtaataatacgactcactatagggagtcgtcaatagaacacatt
T7rbsREV	Addition of sequence containing T7 RBS and Eco31I site (Rev)	ttatttggtctcttcattctctacctccttaatgtgttctattgacgactc

Diagnostic	Binding to:	
seq_pLitGG_Fwd	PmB1 Origin of replication	gtcgatttttgtgatgctcgtcag
WR_PR4	Homology region 5.3	ttggtctctgccatactagcatttcagaagcagcacc
WR_PR6	Homology region 5.3	ttggtctctgccacacctgttaagtgtctttcgttc
WR_PR8	Homology region 5.3	ttggtctctgccattcatgagtagcctcctgtagtcg
PS2_T0Rev	T0 region in selection plasmids	gcggcaaccgagcgttc
VF2	G00100_VF2 annealing site in selection plasmids	tgccacctgacgtctaagaa
gp5EP_Rev	Error prone T7 gene 5	cacggttgagttgcagctta
Gp5_mtplx_Fwd	Error prone T7 gene 5	tcaccgacgttggctatactac
Gp5_mtplx_Rev	Error prone T7 gene 5	ccatcctgcttcttgcagtttt
PS1_T1Fw	T1 region in selection plasmids	agggcggcggatttgtcc

Table S2. List of primers used throughout the different plasmid assemblies.

A)	**N701**	**N702**	**N703**	**N704**	**N705**	**N706**
S502	PLASMID Pool	Fix Riboswitch	Rd Switch 1	Rd Switch 2	Cmk1	cmk2
S503	pifA2	cmk5	pifA6	cmk7	pifA7	pifA9
S504	pifA10	cmk11	pifA11	cmk12	pifA12	pifA12B
S517	pifA13	pifA13D	pifA13B_8-11	cmk14	pifA14_6-11	pifA14
B)	N701	N702	N703	N704	N705	N706
S502	Rd1	Rd2	Rd3	Cmk1	pifA1	cmk3
S503	pifA3	cmk4	pifA4	cmk5	pifA5	cmk6
S504	cmk8	pifA8	cmk9	cmk10	pifA11	cmk12
S517	cmk13	cmk15	pifA15	cmk 16	cmk2 Ampl	cmk16 Ampl

Table S3. Primer combinations for each NGS sample. Matrix indicating which primers correspond to each specific phage library in each of the two repetitions. The 6x4 matrix gives the possibility of separating between 24 samples at once, each of them with their own specific signal, thanks to the specific indexes present in each of the 2 primers used for the amplification. The different denominations for each library are based on the selection step they were obtained from, followed by the number of the generation. If there are multiple libraries corresponding to the same generation, a letter or date is included after the denomination. E.g. Sample cmk11 corresponds to the phage library obtained after the 11th positive selection step in the evolutionary process. A and B correspond to the 1st and 2nd set of sequenced libraries, respectively.

Test details									
Test name	Unpaired t test								
Variance assumption	Single pooled variance								
Multiple comparisons	Set P value threshold								
Method	Holm-Šídák method								
Alpha	0,05								
Number of tests performed	5								
Number of rows omitted	0								
A)	Below threshold?	P value	Mean of cmk	Mean of cmk + 1mM Theo	Difference	SE of difference	t ratio	df	Adjusted P Value
post 2	No	0,705422	0,6738	0,6661	0,007694	0,02024	0,3802	50	0,913224
post 5	No	0,751494	0,7684	0,762	0,006444	0,02024	0,3184	50	0,913224
post 16	No	0,201953	0,6344	0,6631	-0,02867	0,02217	1,293	50	0,571508
Fix	Yes	0,000057	0,5767	0,6962	-0,1195	0,02715	4,401	50	0,000283
WT	No	0,190931	0,2978	0,3338	-0,036	0,02715	1,326	50	0,571508
	Below threshold?	P value	Mean of cmk	Mean of cmk + 1.5mM Theo	Difference	SE of difference	t ratio	df	Adjusted P Value
post 2	Yes	0,001303	0,6738	0,7407	-0,06689	0,01963	3,407	50	0,005202
post 5	No	0,229343	0,7684	0,7923	-0,02389	0,01963	1,217	50	0,406087
post 16	No	0,025309	0,6344	0,5848	0,04958	0,0215	2,306	50	0,074021
Fix	Yes	0,000353	0,5767	0,6777	-0,101	0,02634	3,835	50	0,001763
WT	No	0,637125	0,2978	0,3103	-0,0125	0,02634	0,4746	50	0,637125
	Below threshold?	P value	Mean of cmk + 1mM Theo	Mean of cmk + 1.5mM Theo	Difference	SE of difference	t ratio	df	Adjusted P Value
post 2	Yes	0,000012	0,6661	0,7407	-0,07458	0,01485	5,022	38	0,00005
post 5	No	0,048098	0,762	0,7923	-0,03033	0,01485	2,042	38	0,137466
post 16	Yes	0,000006	0,6631	0,5848	0,07825	0,01485	5,269	38	0,000029
Fix	No	0,38397	0,6962	0,6777	0,0185	0,021	0,8808	38	0,467438
WT	No	0,270232	0,3338	0,3103	0,0235	0,021	1,119	38	0,467438

B)	Below threshold?	P value	Mean of pifA	Mean of pifA + 1mM Theo	Difference	SE of difference	t ratio	df	Adjusted P Value
post 2	No	0,765117	0,714	0,7244	-0,01046	0,03487	0,3001	62	0,987041
post 5	No	0,200169	0,7188	0,7792	-0,06033	0,04659	1,295	62	0,590746
post 16	Yes	0,004439	0,6379	0,4958	0,1421	0,04812	2,953	62	0,022
Fix	No	0,957819	0,7831	0,7805	0,002556	0,04812	0,05311	62	0,997813
WT	No	0,953239	0,3003	0,2975	0,002833	0,04812	0,05888	62	0,997813
	Below threshold?	P value	Mean of pifA	Mean of pifA + 1.5 mM Theo	Difference	SE of difference	t ratio	df	Adjusted P Value
post 2	No	0,453006	0,714	0,7357	-0,02179	0,0289	0,7541	80	0,836338
post 5	No	0,203151	0,7188	0,7576	-0,03878	0,03022	1,283	80	0,678723
post 16	No	0,289006	0,6379	0,6765	-0,03856	0,03612	1,067	80	0,744457
Fix	No	0,620705	0,7831	0,801	-0,01794	0,03612	0,4968	80	0,836338
WT	No	0,466713	0,3003	0,2739	0,02642	0,03612	0,7313	80	0,836338
	Below threshold?	P value	Mean of pifA + 1mM Theo	Mean of pifA + 1.5 mM Theo	Difference	SE of difference	t ratio	df	Adjusted P Value
post 2	No	0,672187	0,7244	0,7357	-0,01133	0,02659	0,426	44	0,943301
post 5	No	0,524939	0,7792	0,7576	0,02156	0,03364	0,6409	44	0,943301
post 16	Yes	0,000008	0,4958	0,6765	-0,1807	0,03568	5,064	44	0,000039
Fix	No	0,568476	0,7805	0,801	-0,0205	0,03568	0,5746	44	0,943301
WT	No	0,51203	0,2975	0,2739	0,02358	0,03568	0,661	44	0,943301

Table S4. Full statistical data for virulence index significance. Explanation of the type of statistical analysis and corrections made to determine whether or not the differences between different populations were significant. **A)** Positive selection. Showing the pair comparison between all different populations: cells grown in the absence of theophylline (cmk), cells grown in 1mM theophylline (cmk + 1mM Theo), and cells grown in 1.5 mM theophylline (cmk + 1.5mM Theo). **B)** Negative selection. Showing the pair comparison between all different populations: cells grown in the absence of theophylline (pifA), cells grown in 1mM theophylline (pifA + 1mM Theo) and cells grown in 1.5 mM theophylline (pifA + 1.5 mM Theo). All calculations were done with Graphpad Prism 9.

1. Weiss R. Cellular computation and communications using engineered genetic regulatory networks [Internet]. 2001. Available from: http://dspace.mit.edu/handle/1721.1/8228

2. Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, De Las Heras A, et al. The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res. 2013;41(D1):666–75.

3. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, et al. Software News and Updates NUPACK: Analysis and Design of Nucleic Acid Systems. J Comput Chem. 2011;32:170–3.