Cell-free prediction of protein expression costs for growing cells

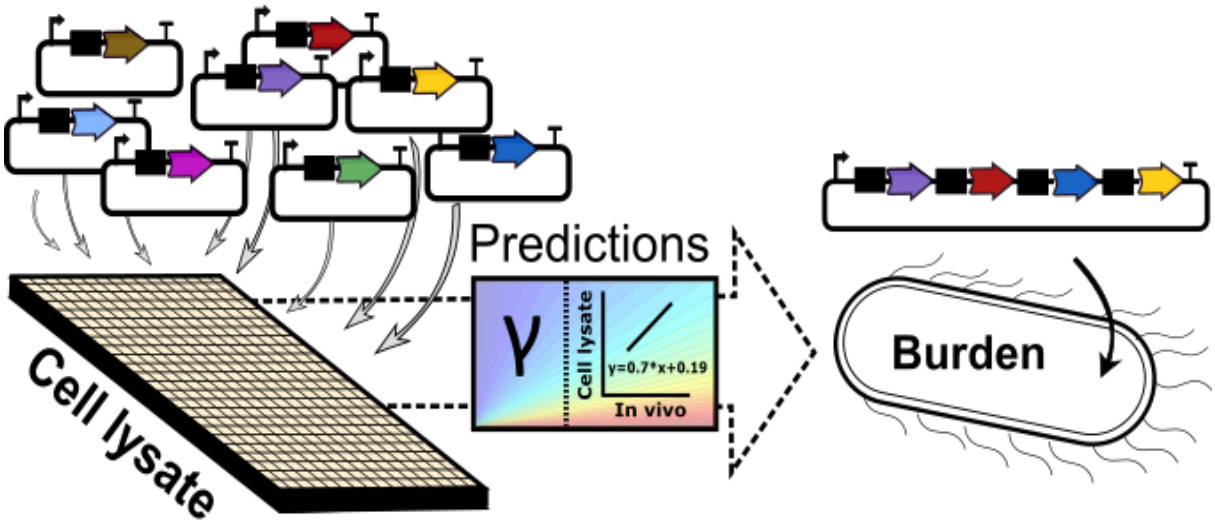
Olivier Borkowski[1,2], Carlos Bricio[1,2], Michaela Murgiano[1,2], Guy-Bart Stan[1,2] and Tom Ellis[1,2]

[1] Centre for Synthetic Biology and Innovation, Imperial College London, London SW7 2AZ, UK.

[2] Department of Bioengineering, Imperial College London, London SW7 2AZ, UK.

Correspondence and requests for materials should be addressed to T.E. email: t.ellis@imperial.ac.uk

Translating heterologous proteins places significant burden on host cells, consuming expression resources leading to slower cell growth and productivity. Yet predicting the cost of protein production for any gene is a major challenge, as multiple processes and factors determine translation efficiency. Here, to enable prediction of the cost of gene expression in bacteria, we describe a standard cell-free lysate assay that determines the relationship between *in vivo* and cell-free measurements and γ, a relative measure of the resource consumption when a given protein is expressed. When combined with a computational model of translation, this enables prediction of the *in vivo* burden placed on growing *E. coli* cells for a variety of proteins of different functions and lengths. Using this approach, we can predict the burden of expressing multigene operons of different designs and differentiate between the fraction of burden related to gene expression compared to action of a metabolic pathway.



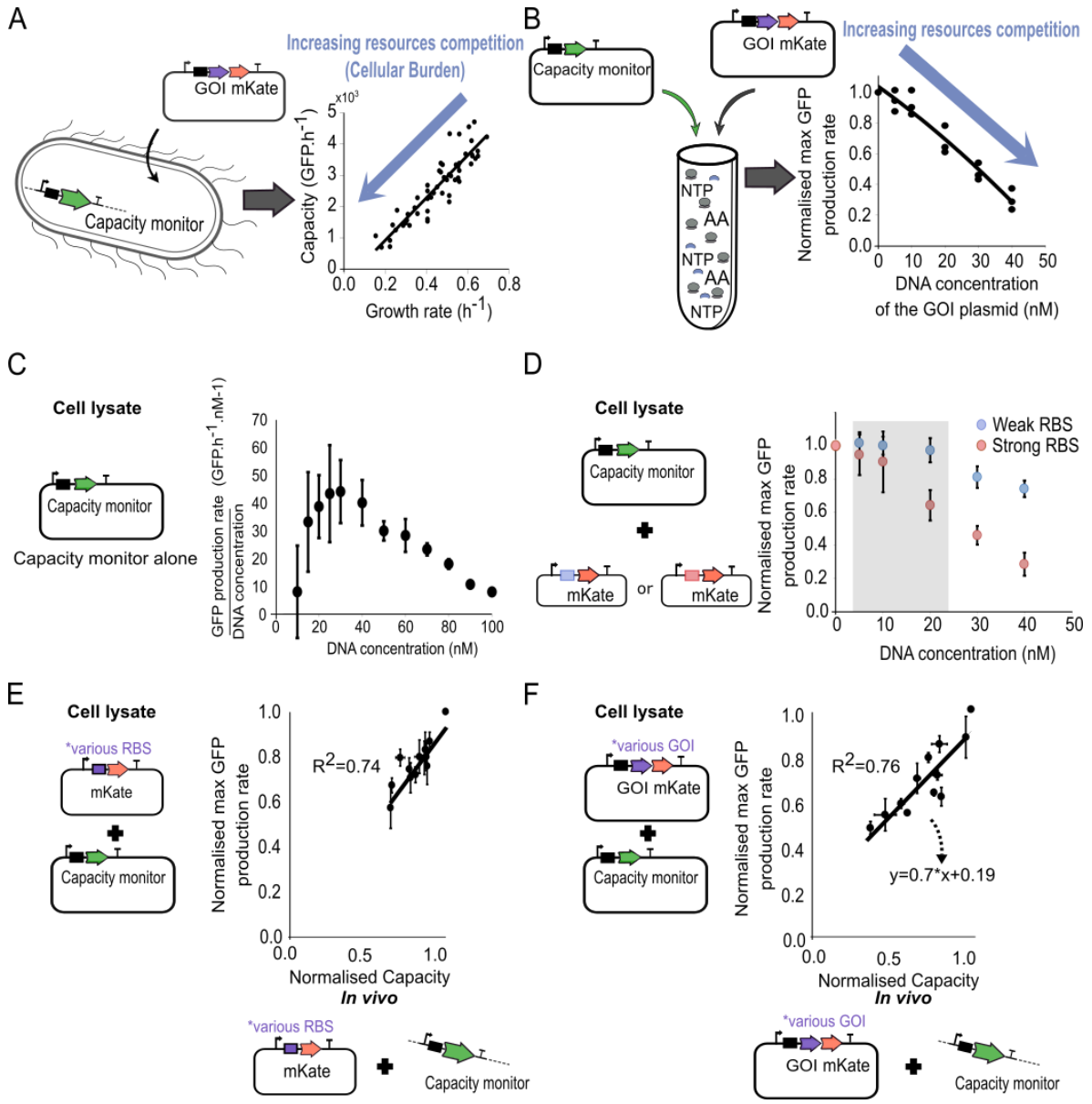


Figure 1: A method to measure resource competition using a capacity monitor in cell lysate. (**A**) *In vivo* resource competition in DH10B *E. coli* cells between a genome-integrated GFP capacity monitor gene and a plasmid-based gene of interest (GoI) tagged with mKate. The graph illustrates the correlation between the inferred expression capacity (measured as GFP production rate per cell) and the cell's growth rate. (**B**) Resource competition in cell lysates expressing the capacity monitor from a plasmid and the GoI from another plasmid. The graph illustrates the correlation between the normalized max GFP production rate and the concentration of the GoI plasmid added to the lysate. (**C**) Measured max GFP production rate divided by DNA concentration with the capacity monitor plasmid added at different concentrations in a 10 µl cell lysate mix. Error bars show standard error of three independent repeats. (**D**) The normalized max GFP production rate measured in 10 µl cell lysate containing 30 nM of the capacity monitor plasmid and different concentration plasmids with *mkate* and either a strong (red) or a weak (blue) RBS sequence. The grey area represents the concentration of plasmid that leads to competition for translational resource. Error bars show standard error of three independent repeats. (**E**) Correlation between normalized max GFP production rates measured in cell lysate and normalized capacity measured *in vivo* with DH10B-GFP cells. The constructs in this experiment all express *mkate* with different strength RBS sequences. Error bars show standard error of three independent repeats. (**F**) Correlation between normalized max GFP production rate measured in cell lysate and normalized capacity measured *in vivo* with DH10B-GFP using constructs with various genes of different sizes paired with RBS BCD2. Error bars show standard error of three independent repeats.

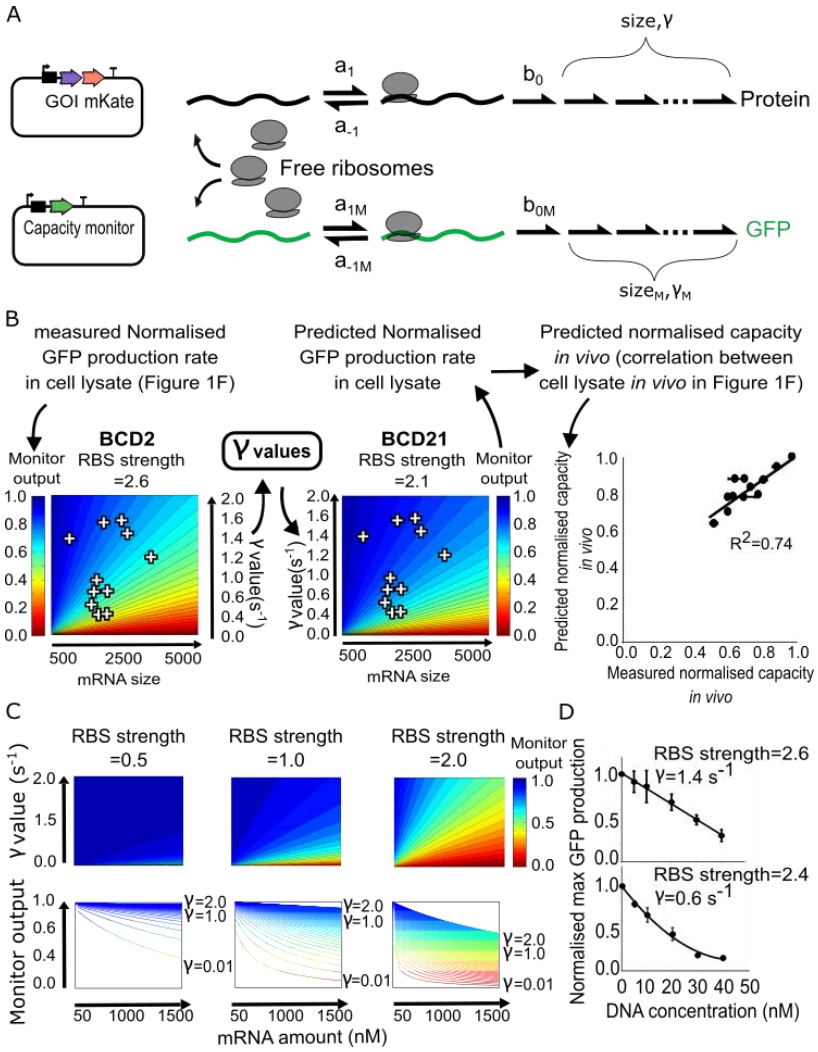


Figure 2: Predictive model for the competition of resources between a synthetic construct and the capacity monitor. (**A**) Overview of a mathematical model of competitive translation with a finite ribosome pool. The model represents the competition for ribosomes between the capacity monitor construct and a construct expressing a GoI. Free ribosomes bind to an unoccupied RBS at a rate a_1 (a_1=a_{1M}*RBS strength) and either unbinds and returns to the free ribosome pool at a rate a_{-1} (a_{-1}=a_{-1M}/RBS strength), or initiates translation at a rate b_0 (b_0=b_{0M}*RBS strength). Once elongation has initiated, the ribosome moves along the transcript at a rate γ. The number of elongation steps depends on "size" (mRNA size / 30 bp as it is the footprint of each ribosome on an mRNA and better represents how many can be queued on a transcript). Each elongation step is considered to proceed at the same rate γ. Once the ribosome reaches the final position on the mRNA it returns to the free ribosome pool. **(B)** Heat maps of simulated capacity monitor (monitor output) when mRNA size and the γ value of a synthetic construct are varied while RBS strength is fixed. The first heat map is used to determine the γ value of each construct used in Figure 1F. The second heat map is used to deduce the monitor output using these calculated γ values. As the prediction is done for cell lysate, the *in vivo* predictions are deduced from the correlation between cell lysate and *in vivo* measurements as per Figure 1F. Error bars show standard error of three independent repeats. (**C**) First row: heat maps of simulated capacity (monitor output) when mRNA amounts and γ values of a synthetic circuit are varied. Each heat map is a construct with a different RBS strength as denoted by the corresponding b_0 values. Second row: capacity (monitor output) as a function of the mRNA amount. Each line corresponds to a different γ value. (**D**) Measurements in cell lysate of the normalised max GFP production rate of the capacity monitor when increasing DNA amounts of synthetic construct are added to the cell lysate assay. Top: *mkate* with a strong RBS. Bottom: *viob-mkate* with a strong RBS. The calculated γ and RBS strength values (from Supplementary Figure 3 B) are shown for each construct.

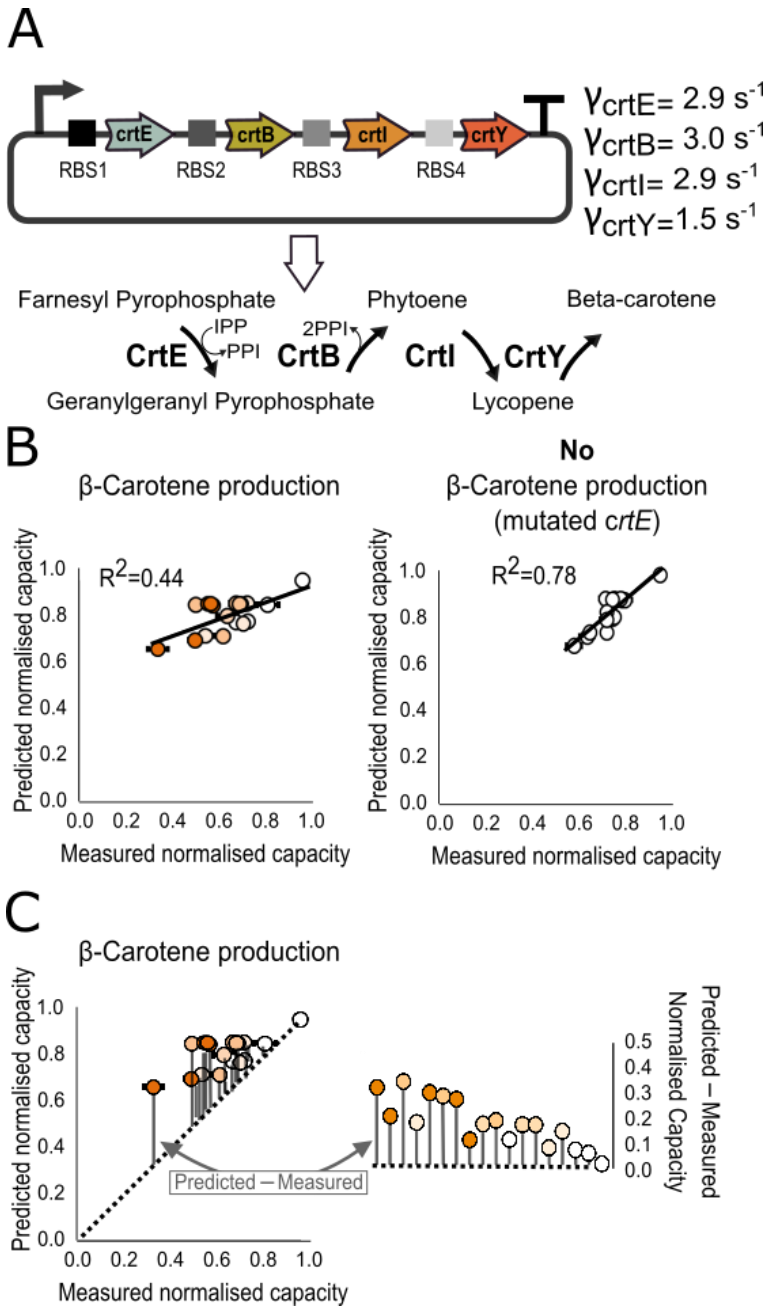
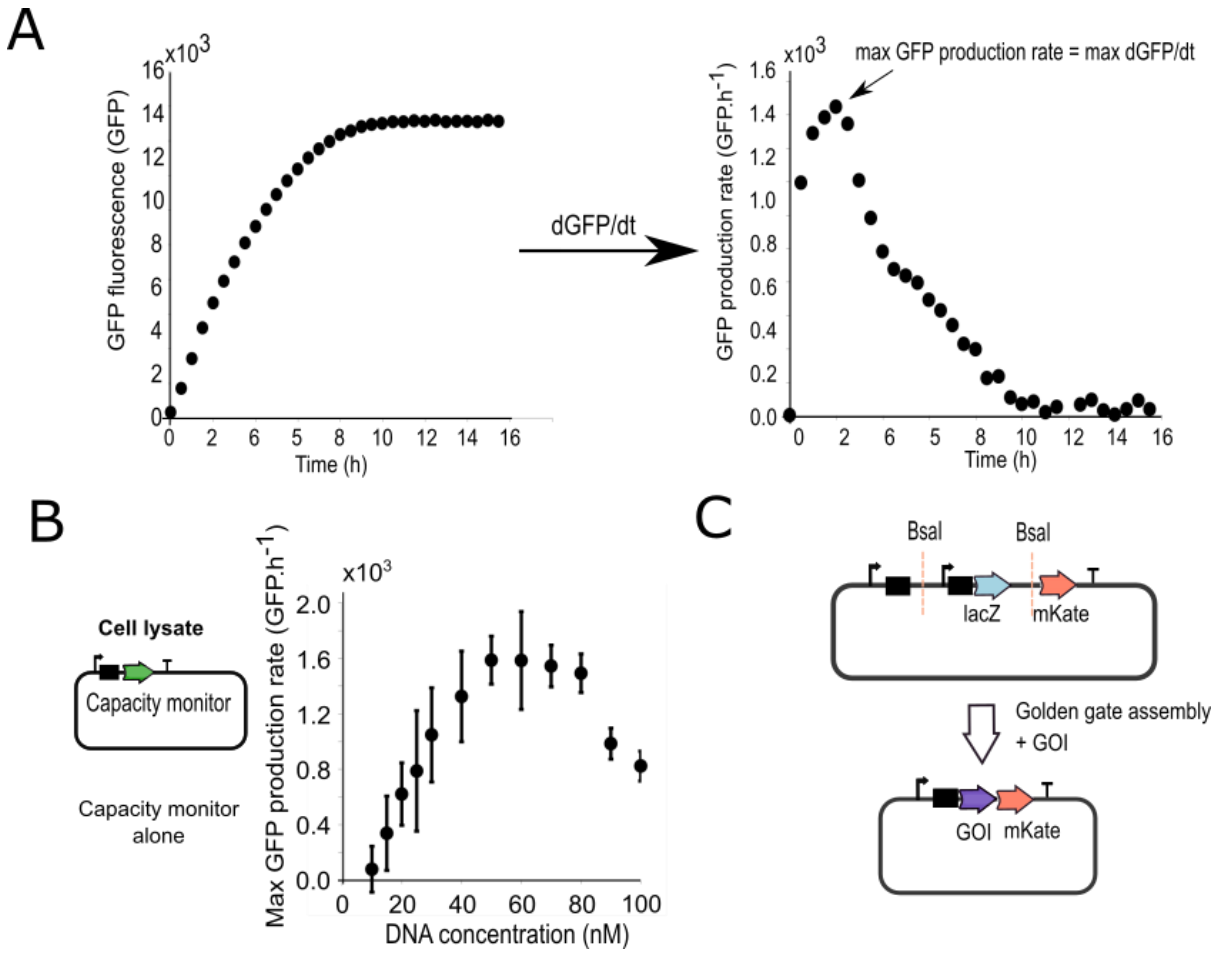
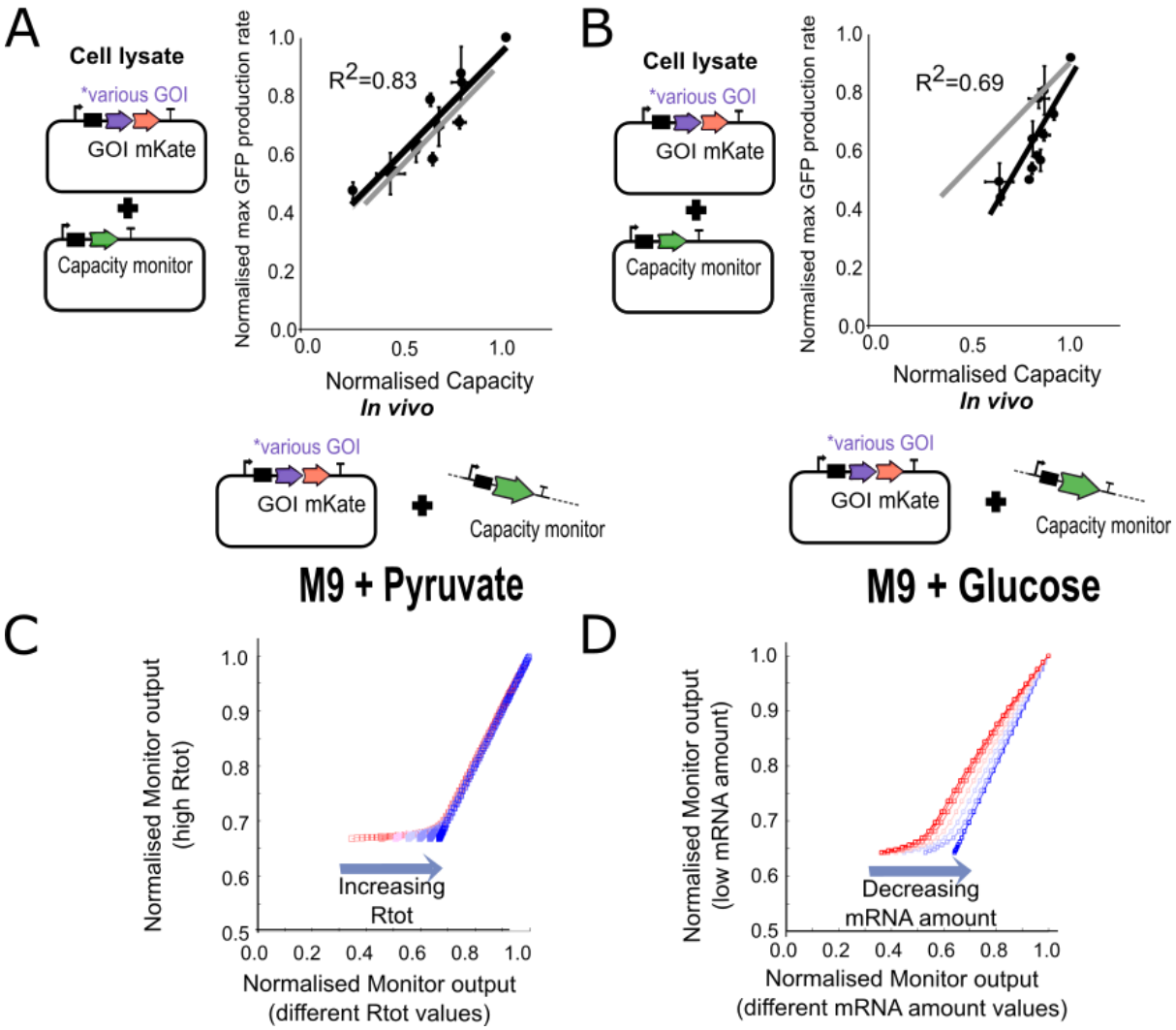


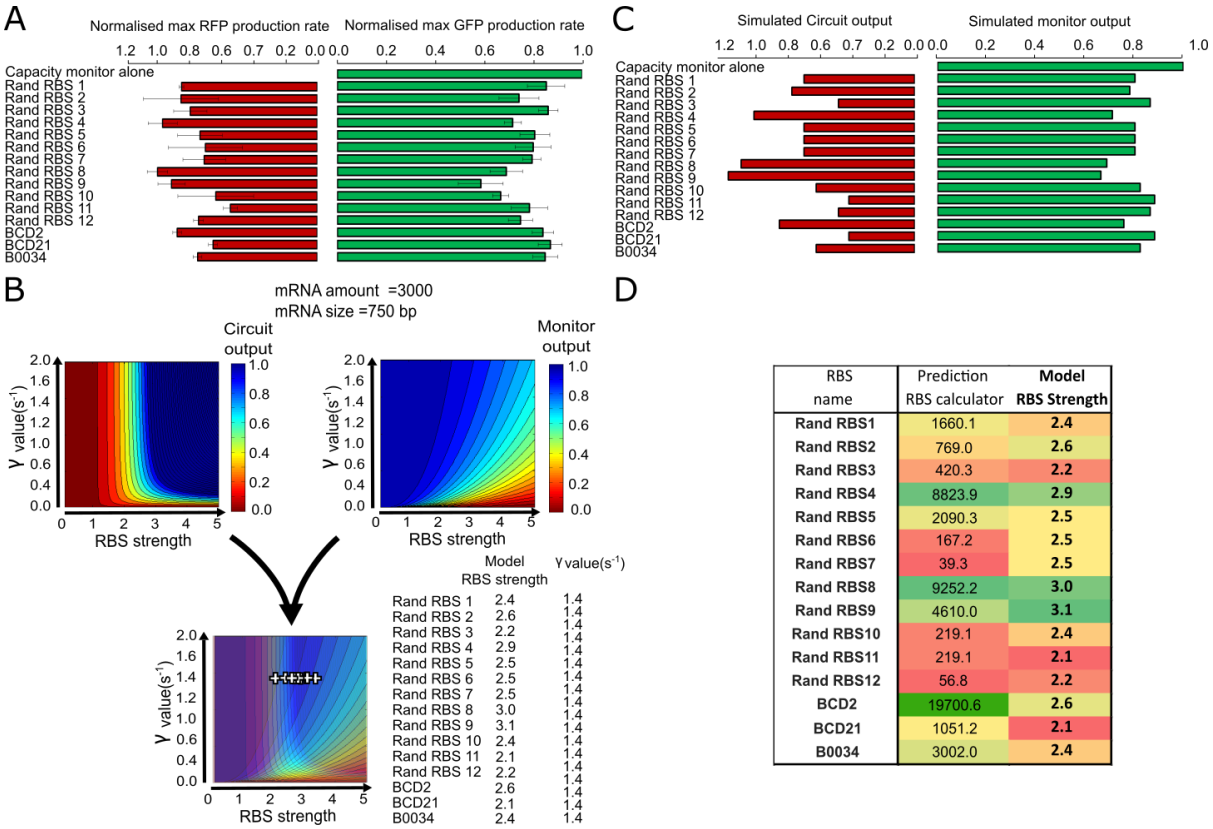
Figure 3: Predicting the burden of operon designs for the beta-carotene biosynthesis pathway. (**A**) Diagram of the beta-carotene pathway and the γ values for the four enzyme-encoding sequences as measured by the cell lysate capacity assay (Supplementary Figure 3). The operon is designed with partially-randomised RBS sequences and one of three different promoters: J23113 (weak), J23106 (medium) or J23100 (strong). (**B**) Model-predicted burden of each operon design compared to the measured capacity of *E. coli* expressing the operons with or without an inactivating mutation in the *crtE* gene. The orange intensity in each circle represents the beta-carotene production as level measured for each strain (see in Supplementary Figure 5). Error bars show standard error of three independent repeats (**C**) Model-predicted burden of each operon design compared to the measured capacity of *E. coli* expressing the working pathway operons (same data as panel B). The diagonal dot line represents equality between the predicted and measured normalised capacity. Grey bars indicate the difference between the predicted and measured normalised capacity of the 17 operons. The righthand plot compares the relative differences between the predicted and measured normalised capacity for the 17 operons and the strain-only control.



Supplementary Figure 1: Cell lysate measurements and backbone design (**A**) The graph on the left is an example of the GFP fluorescence measurements in cell lysate as a function of time using 40 nM of the capacity monitor plasmid cell lysate mix. The graph on the right is the derivate of the GFP fluorescence, i.e. the GFP production rate. GFP production rates were calculated with: GFP production rate at t2 = [total GFP(t3) – total GFP(t1)]/(t3 – t1) with t2 = time of the measurement, t3 = t2 + 0.25 hr and t1= t1 – 0.25 hr. The maximum GFP production rate is the maximum value of GFP production rate. (**B**) Measured max GFP production rate with the capacity monitor plasmid added at different concentrations in a 10 µl cell lysate mix. Error bars show standard error of three independent repeats. (**C**) The entry vector design for this work consists of a promoter J23106, the BCD 2 or BCD 21 or RBS B0034, a *bsa*I restriction site (`ggtctcccatt`), promoter pLac, lacZ, a *bsa*I restriction site, linker sequence (`aggtctcagctt`), codon optimised mkate and terminator. Golden Gate assembly was used to integrate different GoIs leading to GoI-mKate fusion proteins.

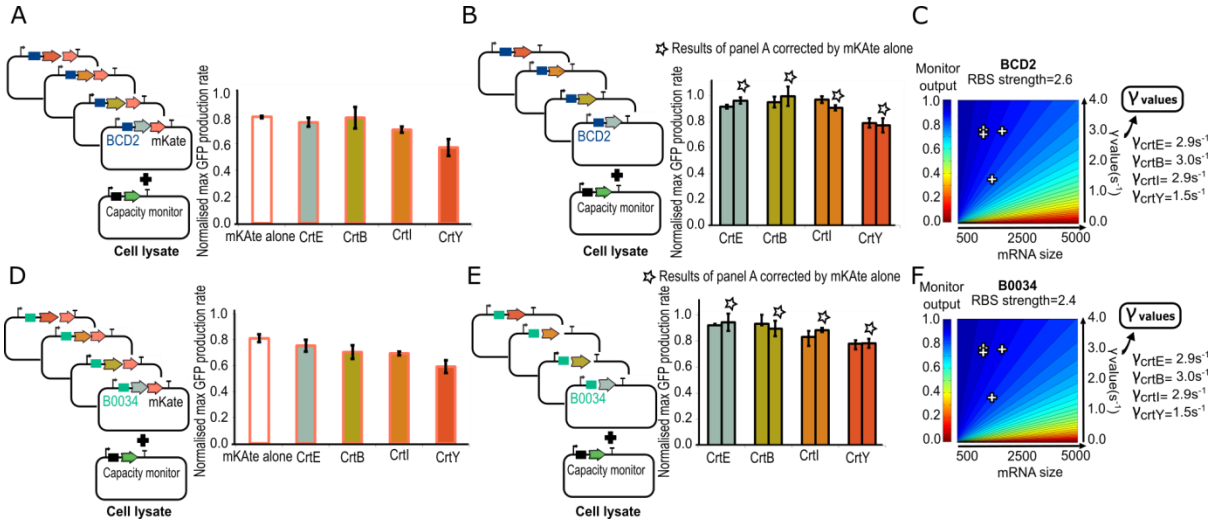


Supplementary Figure 2: Medium-dependent burden. (**A**) Correlation between normalized max GFP production rates measured in cell lysate (same value as Figure 1E) and normalized capacity measured *in vivo* with DH10B-GFP using constructs with various genes of different sizes paired with RBS BCD2. All *in vivo* measurements have been done with the same strain as Figure 1E but grown in minimal media + 0.5% pyruvate. Error bars show standard error of three independent repeats (**B**) Correlation between normalized max GFP production rate measured in cell lysate (same value as Figure 1E) and normalized capacity measured *in vivo* with DH10B-GFP using constructs with various genes of different sizes paired with RBS BCD2. All *in vivo* measurements are with the same strain as Figure 1E but grown in minimal media + 0.5% glucose. Error bars show standard error of three independent repeats. (**C**) Simulations of the Monitor output in competition with circuits with different γ values with Ribosome fixed at 50000 (y axis). Simulations of the Monitor output in competition with circuits with different γ values with Ribosome fixed at 50000, 10000, 5000, 2000 or 1000 nM (x axis) (**D**) Simulations of the Monitor output in competition with constructs with different γ values with mRNA fixed at 100 (y axis). Simulations of the Monitor output in competition with constructs with different γ values with mRNA fixed at 5000, 4000, 3000, 2000, 1000, 500 or 100 nM (x axis).



RBS name	Prediction RBS calculator	Model RBS Strength
Rand RBS1	1660.1	2.4
Rand RBS2	769.0	2.6
Rand RBS3	420.3	2.2
Rand RBS4	8823.9	2.9
Rand RBS5	2090.3	2.5
Rand RBS6	167.2	2.5
Rand RBS7	39.3	2.5
Rand RBS8	9252.2	3.0
Rand RBS9	4610.0	3.1
Rand RBS10	219.1	2.4
Rand RBS11	219.1	2.1
Rand RBS12	56.8	2.2
BCD2	19700.6	2.6
BCD21	1051.2	2.1
B0034	3002.0	2.4

Supplementary Figure 3: Cell lysate measurements and simulated output and capacity at steady state determined by a translational resource model. (**A**) Normalised max RFP production rate and normalised max GFP production rate measurements of constructs with mkate alone under control of different RBS/BCDs (library from Figure 1E). Normalised max RFP production rates are the max RFP production divided by the mean max RFP production of the Rand RBS 8 construct (strongest of the collection). Normalised max GFP production rates are the max GFP production divided by the max GFP production of capacity monitor expressed alone in the cell mix as described in Methods. Error bars show standard error of three independent repeats. (**B**) Simulation of the construct output (corresponding to the normalised max RFP production rate measurement) and monitor output (corresponding to the Normalised max GFP production rate measurement). The construct and monitor outputs graphs have been merged to deduce RBS strengths and γ value of the mkate construct library. All the constructs produce the same mkate protein and thus have the same γ value. (**C**) Results of the simulation of panel B. (**D**) Comparison between the RBS calculator predictions and our model simulations. The order of RBSs strength is respected when compared between RBS Calculator predictions and our model simulations. Our simulations reveal that the strongest RBS of our collection has a RBS strength=3.0, medium RBSs have a RBS strength=2.5 and the weakest RBS has a RBS strength=2.0.

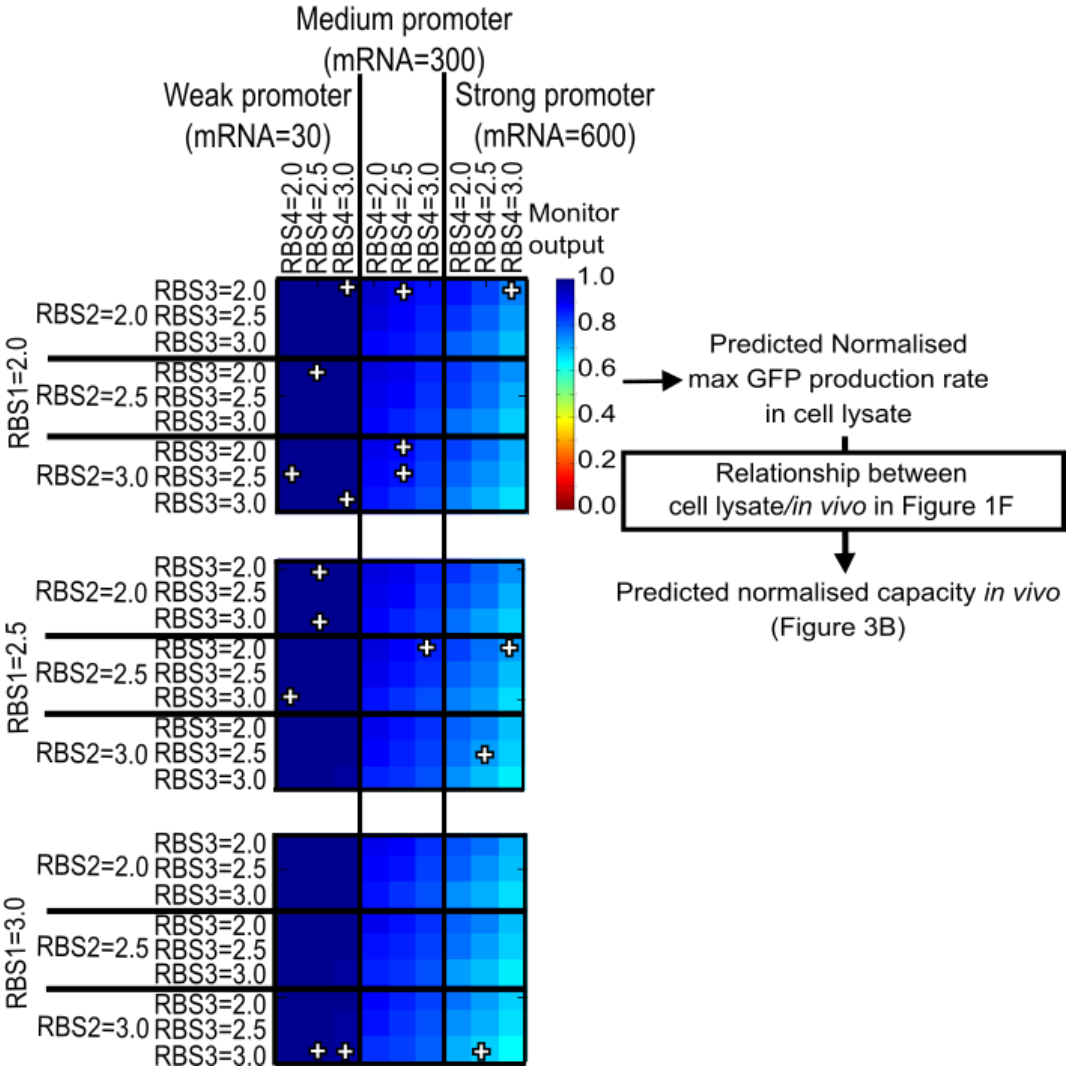


Supplementary Figure 4: Competition for the resources in cell lysate between the enzymes of the beta-carotene pathway and the capacity monitor. (**A**) Normalised max GFP production rate measured with capacity monitor and the crt-mkate fusion added in the cell lysate mix. Crt are the enzymes of the beta-carotene pathway. The production of mKate (mKate alone) under control of BCD2 leads to a competition for the resources as the normalised max GFP production rate is lower than 1. (**B**) Normalised max GFP production rate measured with capacity monitor and the crt enzymes in the cell lysate mix. The mKate CDS has been removed to obtain the competition for resource due to the crt enzyme alone. The normalised max GFP production rate of the crt-mkate fusion corrected by removing the cost of mkate calculating using the data from panel A. The normalised max GFP production rates of the crt enzymes and of the corrected crt-mkate exhibit similar values showing that both backbone with mkate fusion and without mkate can be used for γ value calculation. (**C**) Heat maps of simulated capacity (monitor output) when mRNA size and the γ value of a synthetic construct are varied as RBS strength is fixed (the BCD2 RBS strength has been obtained as explained in Supplementary Figure 2B). (**D, E, F**) Similar experiment as panels A, B, C but with RBS B0034.

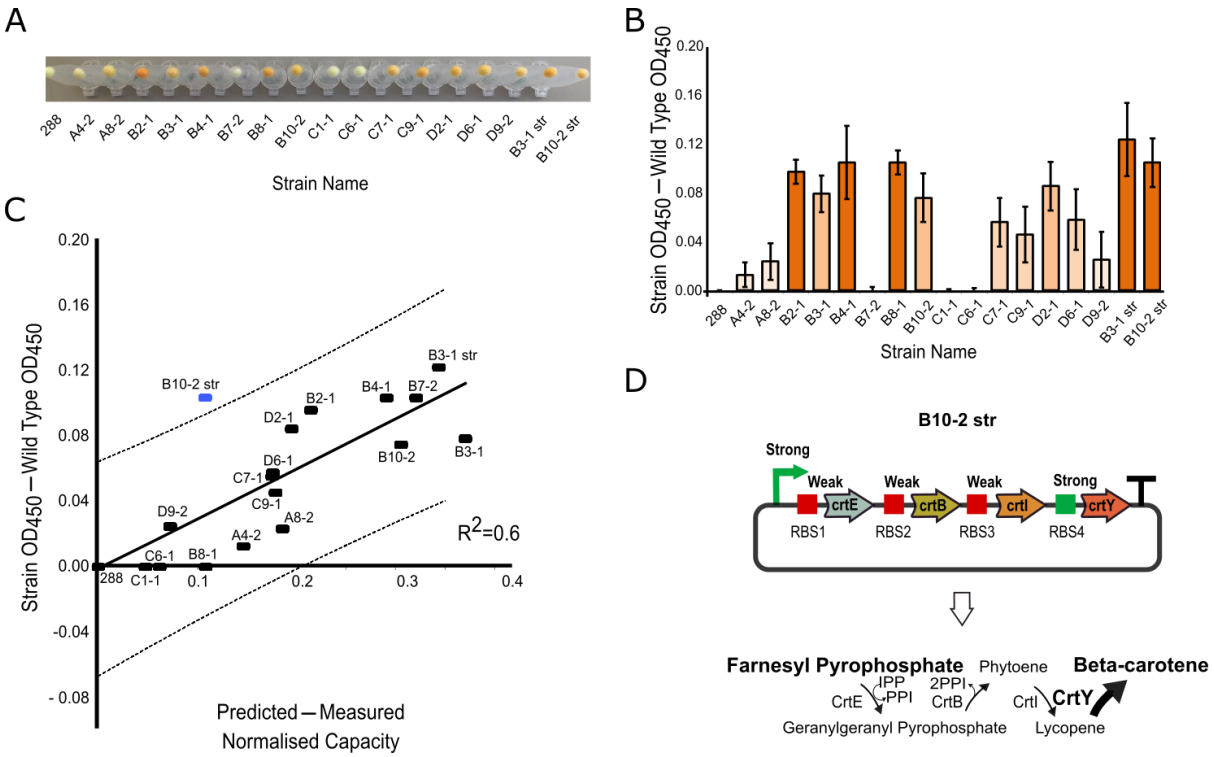
A

Construct name	Promoter name	Model mRNA Strength	Prediction Operon calculator	Model RBS Strength	Prediction Operon calculator	Model RBS Strength	Prediction Operon calculator	Model RBS Strength	Prediction Operon calculator	Model RBS Strength
A4-2	P7-J23113	60	2.6	2.0	1891	2.5	1040	2.0	839	2.5
B3-1	P7-J23113	60	232.3	3.0	2457	3.0	47664	3.0	850	2.5
B4-1	P7-J23113	60	345.2	3.0	3498	3.0	32075	3.0	911460	3.0
B8-1	P7-J23113	60	7.9	2.5	1579	2.5	90441	3.0	44	2.0
B10-2	P7-J23113	60	0.3	2.0	671	2.0	390	2.0	6783	3.0
C1-1	P7-J23113	60	1.8	2.0	2911	3.0	270383	3.0	4027	3.0
C7-1	P7-J23113	60	0.0	2.0	3082	3.0	7264	2.5	154	2.0
D2-1	P7-J23113	60	13.6	2.5	932	2.0	187	2.0	1025	2.5
D6-1	P7-J23113	60	16.3	2.5	771	2.0	63725	3.0	808	2.5
B7-2	P6_J23106	300	0.2	2.0	3096	3.0	3540	2.0	1267	2.5
C6-1	P6_J23106	300	1.4	2.0	2474	3.0	6994	2.5	1117	2.5
C9-1	P6_J23106	300	0.1	2.0	814	2.0	1802	2.0	704	2.5
D9-2	P6_J23106	300	16.3	2.5	1378	2.5	194	2.0	4177	3.0
A8-2	P5-J23100	600	15.8	2.5	4418	3.0	13957	2.5	532	2.5
B2-1	P5-J23100	600	5.1	2.5	1103	2.5	194	2.0	2332	3.0
B3-1-str	P5-J23100	600	232.3	3.0	2457	3.0	47664	3.0	850	2.5
B10-2-str	P5-J23100	600	0.3	2.0	671	2.0	390	2.0	6783	3.0

B



Supplementary Figure 5: Burden prediction for beta-carotene operons. (**A**) RBS Calculator prediction using the operon calculator. We ranked the model RBS value as described in Supplementary Note 2. (**B**) Predictions from the competition between the capacity monitor and the 4 *crt* enzymes using the γ values obtain in Supplementary Figure 3 and the RBS values obtain in panel A.



Supplementary Figure 6: Beta-carotene production. (**A**) Pellet of the strains producing beta-carotene with the constructs of the beta-carotene operon library (**B**) Beta-carotene values obtain at OD_{450} measurements after acetone treatment (**C**) Comparison between the beta-carotene production (panel B) and the difference between the predicted and measured capacity (Figure 3C). The dotted lines represent the 95% confident interval. B10-2 str appears to be an outlier (blue dash). (**D**) Diagram of the beta-carotene pathway of the construct B10-2 str. B10-2 str is a construct with a strong promoter, weak RBS1/RBS2/RBS3 and a strong RBS4 (Supplementary Figure 5A). This construct should produce a higher amount of CrtY, the last enzyme of the beta-carotene operon, than CrtE,CrtE and CrtI.

Supplementary Table 1: Library of BCDs/RBSs

Name	Sequence
Rand RBS1	GCAAGGGGATAGTG
Rand RBS2	GCAAGGGATTGATA
Rand RBS3	GCAAAGGGAAAGATGG
Rand RBS4	GCAAGGGGGGGGTT
Rand RBS5	GCAAGGGAGGGCGCT
Rand RBS6	GCAAGGGGCGGGGT
Rand RBS7	GCAAGGGCGTATG
Rand RBS8	GCAAAGGGGGTGCGTC
Rand RBS9	GCAAGGGGGCGGGGG
Rand RBS10	GCAAGGGTCGTCT
Rand RBS11	GCAAGGGGGGGGTGT
Rand RBS12	GCAAGGGTCGAGT
BCD2	GCAAGGGCCCAAGTTCACTTAAAAAGGAGATCAACAATGAAAGCAATTTT CGTACTGAAACATCTTAATCATGCTAAGGAGGTTTTCT
BCD21	GCAAGGGCCCAAGTTCACTTAAAAAGGAGATCAACAATGAAAGCAATTTT CGTACTGAAACATCTTAATCATGCGAGGGATGGTTTCT
B0034	AAAGAGGAGAAA