1

Multiple-site diversification of regulatory sequences enables inter-species operability of genetic devices

by

Angeles Hueso-Gil[1], Ákos Nyerges[2], Csaba Pál[2], Belén Calles[1] and Víctor de Lorenzo[1,*]

[1]Systems Biology Program, Centro Nacional de Biotecnología, Cantoblanco-Madrid, Spain.

[2]Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre, Szeged H-6726, Hungary

Keywords: Pseudomonas, optogenetics, CcaSR, biofilms,

Running title: Enabling *E. coli-P. putida* device interoperability

*Correspondence to: *Víctor de Lorenzo* (vdlorenzo@cnb.csic.es)

Centro Nacional de Biotecnología (CNB-CSIC)

Campus de Cantoblanco, 28049 Madrid (Spain)

Tel: (+34 91) 585 45 73; Fax: (+34 91) 585 45 06

Abstract

The features of the light-responsive cyanobacterial CcaSR regulatory node that determine interoperability of this optogenetic device between *Escherichia coli* and *Pseudomonas putida* have been examined. For this, all structural parts (i.e. *ho1* and *pcyA* genes for synthesis of phycobilin, the *ccaS/ccaR* system from *Synechocystis* and its cognate downstream promoter) were maintained but their expression levels and stoichiometry diversified by [i] reassembling them together in a single broad host range, standardized vector and [ii] subjecting the non-coding regulatory sequences to multiple cycles of directed evolution with random genomic mutations (DIvERGE), a recombineering method that intensifies mutation rates within discrete DNA segments. Once passed to *P. putida*, various clones displayed a wide dynamic range, insignificant leakiness and excellent capacity in response to green light. Inspection of the evolutionary intermediates pinpointed translational control as the main bottleneck for interoperability and suggested a general approach for easing the exchange of genetic cargoes between different species i.e. optimization of relative expression levels and upturning of subcomplex stoichiometry.

_____

Introduction

Many naturally-occurring biosystems respond to the presence of light by activating or repressing distinct gene sets. Typically, light-reacting proteins and regulatory nodes are found in photosynthetic organisms, such as plants, *algae* and several bacterial groups like cyanobacteria and other marine microorganisms that use light to pump protons for obtaining an extra energy supply[1]. The control of bacterial behavior using light has become a most useful asset in synthetic biology, as it allows an instant and non-disruptive on/off switching of genetic constructs. Light as an inducer is cheap, precise, fast and clean, avoiding problems of chemical residues in the culture media when the circuit needs to be turned off. One useful source of parts for engineering light-responsive switches is the two-component CcaSR system

1   of cyanobacterial *Synechocystis* species[2]. This green light-responsive setup controls

2   expression of the phycobilisome linker protein CpcG2. When 520 nm light hits the membrane

3   sensor protein CcaS bound to cofactor phycocianobilin (PCB), the protein autophosphorylates

4   and transfers the phosphoryl group to the CcaR transcription factor. This in turn binds the so-

5   called G box of P_{cpcG2} promoter and activates its transcription by σ^{70}-RNA polymerase[2]. PCB

6   can be easily produced in non-cyanobacterial hosts from the *heme* group through the action of

7   enzymes Ho1 and PcyA. On this basis, the system was adapted, reshaped and implemented

8   for controlling gene expression in *E. coli* upon exposure to the appropriate wavelength[3]. Since

9   then, a collection of growingly improved light switches has been developed to the same end

10   derived from the CcaSR node, using as a host both *E. coli* and *Bacillus subtilis*[4-6].

11   Unfortunately, while *E. coli* is a good chassis for prototyping of genetic devices, its

12   applicability to biotechnological endeavors in industry and the environment is limited[7, 8] and

13   other hosts of engineered circuits need to be considered. Yet, as the parameters and transfer

14   functions that operate the constructs most often change with the biological context,

15   optogenetic tools optimized for *E. coli* do not perform well—or at all— when transferred to

16   other bacteria.

17

18   For adapting existing light switches—and any other genetic devices—to new hosts, the

19   correct balance of the functional components for *nesting* in the new regulatory and

20   biochemical network of the new recipient is crucial. This requires the exploration of an ample

21   solution space for finding the best combination of expression parameters of multiple

22   constituents in a given host[9]. Along this line Schmidl et al[5] adopted a combinatorial assembly

23   of ribosomal binding sites (RBSs) and a shortened version of the promoter P_{cpcG2} (called

24    $P_{cpcG2-172}$) for improving the dynamic range of the CcaSR-based light switch in *E. coli,* which

25   brought about a more balanced expression of the components. But things become more

26   challenging when the issue is interoperability of the same device between two different,

27   distant bacterial species. One appealing recipient of such light-responsive modules is the soil

28   bacterium *Pseudomonas putida*, an emerging chassis for Synthetic Biology-based metabolic

29   engineering because of its excellent pre-evolved properties of stress resistance and high

30   capacity to host strong redox reactions[10, 11]. But if intra-species optimization is demanding

4

1 enough, inter-species portability adds a further screw turn to the problem, as the new context

2 alters the multiple parameters involved for a usable performance.

3

4 In this work, we have adopted a novel strategy for making the popular CcaSR-based light

5 switch usable in *P. putida*. This workflow relies on keeping intact the structural parts of the

6 device but letting its regulatory DNA (i.e. their promoters and RBS) to navigate through a

7 very large solution space. This was enabled by DIvERGE (directed evolution with random

8 genomic mutations[12]), a recombineering method that causes hypermutation *in vivo* of multiple,

9 pre-set DNA segments of a target replicon. Following diversification and a two-step screening

10 process, expression levels of each of the CcaSR system component (*ccaS, pcyA, ho1, ccaR,*

11 $P_{cpcG2-172}$)[5] re-balanced themselves for an optimal nesting in *P. putida*. This outcome not only

12 solved a practical problem, but it also showcased a general stratagem for handling

13 interoperability of SynBio devices between otherwise distant hosts.

14

15 **Results and Discussion**

16

17 **Gross implantation of the CcaS-CcaR system into *P. putida*.** The CcaSR light-sensitive,

18 two-component system was discovered in cyanobacteria *Synechocystis* species[2] and later

19 adapted and used to control gene expression in *E. coli* under illumination with an adequate

20 wavelength[4, 5]. In these works, the CcaSR system components and enzymes for PCB

21 production were expressed independently from constitutive promoters in isolated

22 transcriptional units, with the exceptions of *ho1* and *pcyA*, which together formed a two-gene

23 operon. These parts were carried in two different narrow host range plasmids: pSR43.6

24 harbouring the cassette *ccaS-ho1-pcyA* and pSR58.6 bearing the segment *ccaR-msfGFP*

25 (Supplementary Table S5). These plasmids encode resistance to antibiotics and origins of

26 replication suitable for manipulation in *E. coli* but not adequate for the destination

27 microorganism, *P. putida*. For overcoming this problem, we moved the two relevant cassettes

28 from their original plasmids to broad-host range vectors pSEVA631 and pSEVA241,

29 respectively[13]. These approximately preserved the copy number of every functional DNA

30 segment of the original system and thus maintained the dose of each component (Fig. 1A).

5

1 The resulting constructs, pCcaSPCB (harbouring the DNA for *ccaS-ho1-pcyA*) and

2 pCcaRGFP (cassette *ccaR-msfGFP*), were co-transformed into *P. putida* KT2440. This new

3 strain was separately incubated under green light and under darkness conditions to check its

4 functionality. After 8 h of incubation, the GFP fluorescent output was measured with flow

5 cytometry, as this technology—unlike plate readersyields detailed information at single cell

6 level. As shown in Fig. 1B, GFP expression indeed showed a response to the inducer (green

7 light) when *P. putida* was transformed with both pCcaSPCB and pCcaRGFP plasmids.

8 However, the mere adaptation of the antibiotic resistances and origins of replications to the

9 new host was not enough to build an efficient system. Although basal levels were relatively

10 low, fold induction was not as high as reported for *E. coli*[5]. Another interesting observation

11 was a bimodal distribution of fluorescence reflected in two separate peaks both in darkness

12 and in green light. In the non-induced state, there is a subpopulation displaying relative high

13 fluorescence values while under inducing conditions a considerable amount of cells did not

14 respond to illumination. In view of this clearly suboptimal performance we set out to achieve

15 an homogeneous behavior, less leakiness, a higher dynamic range and capacity to better

16 differentiate ON and OFF states of regulation.

17

18 In previous works with this system in *E. coli*, the key to obtaining a successful performance

19 of the system involved balancing the production of each part of the optogenetic device[5]. For

20 example, an excess of CcaR could provoke an unspecific phosphorylation of the regulator

21 through other kinases and, therefore, unspecific transcription activation. On the other hand,

22 low CcaR levels would not be enough for a differential GFP expression between light and

23 darkness states. In addition, enzymes Ho1 and PcyA that produced the PCB cofactor could be

24 poorly transcribed and/or translated. Moreover, there could be a bottleneck for the correct

25 utilization of light by CcaS[5]. In any case, it was clear that due to the genetic context, parts and

26 functional modules behaved unpredictably once interacting with the new cellular environment

27 of *P. putida.* With these considerations in mind, we reasoned that it would be important to

28 optimize the balance of CcaS/CasR for *P. putida* along with an adequate level of PCB

29 biosynthesis.

30

1   In order to ease the tuning of the stoichiometry of the component of the device in *P. putida,*

2   we first reassembled the whole set of parts (*ccaS*, *ho1*, *pcyA*, *ccaR* and $P_{cpcG2-172}$→ *msfGFP*)

3   in the same DNA segment and cloned it in two orientations in vector pSEVA621.

4   Constructing two plasmids with opposite orientation was necessary because efficient

5   diversification with DIvERGE (see below) requires the precise targeting of the replicating

6   lagging strand with mutagenic oligonucleotides[14, 15]. As data on the structure of the replicore

7   of RK2 origin-based expression vectors were not available, we overcome this issue by

8   constructing and subsequently targeting two plasmids in opposing orientation. The resulting

9   plasmids (Fig. 2A) carried such a cargo in either the same strand as the replication protein of

10 RK2 origin, *trfA* (pGSFw) or in the other strand (pGSRv). As a consequence, both plasmids

11 had the same functional segments, but oriented differently. The two constructs were then

12 passed to *P. putida* and their responsiveness to green light tested.

13

14 As shown in Fig. 2BC, cells bearing pGSFw and pGSRv could sense and respond to green

15 light but they showed a very different behavior, both between them and in comparison with

16 the 2-plasmid precursor system in which the relevant features of the device were split into two

17 plasmids (Fig. 2B-D). Specifically, cells with pGSFw showed a high GFP production in the

18 light-induced state, but also a high basal level without any illumination, meaning that its

19 dynamic range and capacity was very low. In contrast, pGSRv showed a comparatively low

20 basal level and a noticeable induction by light. These data exposed one more time the

21 challenge of context dependency, specifically that the DNA strand carrying the coding

22 information was making a considerable difference in the behavior of the thereby reassembled

23 device. Although the ultimate basis of the dissimilarity between pGSFw and pGSRv remained

24 speculative, the data of Fig. 2B-D suggested that the position of the first gene of our cassette,

25 *ccaS,* could influence its transcription levels due to the upstream context, affecting the

26 stoichiometry and behavior of the system. This advised to focus on the intergenic regulatory

27 regions between genes as a way to optimize the behavior of the light switch in the new host.

28

29 **Diversification of regulatory regions of the CcaSR-based optogenetic device.** Motivated

30 by the results above, we set out to rebalance the expression level of each protein of the

1 optogenetic device [*ccaS•pcyA/ho1•cca*R• $P_{cpcG2-172}$ →] borne by pGSRv and pGSFw by

2 introducing random mutations in their intergenic regulatory regions. Once such DNA

3 segments were identified (*ccaS, pcyA/ho1, ccaR* and $P_{cpcG2-175}$ RBSs andpromoters, see Fig. 3),

4 we subjected them to directed evolution with random genomic mutations (DIvERGE;

5 Nyerges *et al*., 2018), an ssDNA recombineering-based method that uses mutagenic

6 oligonucleotides to integrate randomly distributed, random mutations to user-defined target

7 regions, without inducing off-target mutagenesis. This approach involves the addition of

8 mixtures of mutagenic oligonucleotides carrying randomly distributed point mutations that

9 are incorporated into the target sequence upon invasion of the replication fork and integration

10 as Okazaki fragments[16, 17], facilitated by the β-recombinase of the Red system of phage λ[18, 19].

11 Such oligo-integration is easier in the lagging DNA strand which, as a result, has a higher

12 mutation rate[18, 20]. In contrast, the leading strand has a more processive replication and

13 integration of the oligonucleotides in the fork is less likely leading to a lower mutation rate.

14 The primers used in our case were designed to target just one of the strands (Fig. 3A) and

15 carried just 1 to 10 single point mutations per oligonucleotide molecule (i.e. they were soft-

16 randomized[12]). Although we did not know upfront which one was the leading and which one

17 was the lagging in the RK2 replicon of the SEVA plasmid, having pGSFw and pGSRv at

18 hand covered both possibilities of hitting the right orientation of the DNA targets.

19

20 Diversification of the targeted DNA sequences of pGSFw and pGSRv (Fig. 3) was carried out

21 with *E. coli* MG1655 as a surrogate host for the procedure, bearing the pORTMAGE3[19]

22 plasmid with all the necessary machinery for the recombineering method. This includes a

23 thermoinducible multicistronic operon encoding the Redβ recombinase and the negative

24 dominant E32K variant of MutL that transiently suppresses the mismatch repair system and

25 thus enables unbiased introduction of all possible nucleotide replacements[21].

26

27 The result of 5 cycles of DIvERGE were two plasmid libraries per construct (A and B), each

28 of which stemming from a different combination of mutagenic oligonucleotides. The

29 dissimilarity between libraries A and B relies on oligo DIvERGE5, as it came in two versions.

30 DIvERGE5a did not target the G-Box of the $P_{cpcG2-172}$ promoter while DIvERGE5b did. The

1   G-Box is a sequence motif (CACGTG) present in a large number of plant and cyanobacterial

2   promoters that are recognized by a cognate transcription factor for expression control. In the

3   device assembled in pGSFw and pGSRv this motif is bound by CcaR, what is followed by the

4   recruitment of σ^{70}-RNAP polymerase[2, 22]. By using two variants of DIvERGE5 we could

5   either spare the G-box or try to improve it. On these basis, the target sequences of pGSFw

6   plasmid were diversified using pool A of ssDNA oligonucleotides (DIvERGE1, DIvERGE2,

7   DIvERGE3, DIvERGE4 and DIvERGE 5a), which resulted in the library Fw-A. The same

8   plasmid subjected to pool B of oligos (DIvERGE1, DIvERGE2, DIvERGE3, DIvERGE4 and

9   DIvERGE5b) produced library Fw-B (Fig. 3). By the same token, treatment of cells bearing

10   pGSRv with oligo mixtures A or B generated library Rv-A and Rv-B, respectively (Fig. 3).

11   Mutagenesis thus simultaneously reached out each promoter and RBS of every gene of the

12   CcaSR-based system. Note that—as mentioned above—DIvERGE oligos were soft-

13   randomized such that every ssDNA molecule was calculated to bear 1 to 25 indiscriminately

14   distributed mutations compared to the parental, wild-type sequence (see Supplementary Table

15   S7). This ensured efficient oligo-integration at the replication fork during the course of

16   DIvERGE procedure. Based on the number of targeted cells and the average efficiency of

17   pORTMAGE-based oligo integration (i.e., 30%), we have estimated that the effective library

18   size approached approximately10^9 variants per each library[12]. At this point, the population of

19   *E. coli* bearing libraries Fw-A, Fw-B, Rv-A and Rv-B were separately pooled, lysed to extract

20   the plasmid contents and the cccDNA fraction transferred to *P. putida* by transformation (see

21   Materials and Methods) for securing a minimal loss of variability. The mere visual inspection

22   of the resulting *P. putida* transformants (Supplementary Fig. S5) suggested that Rv libraries

23   had a higher variability in their response to illumination, in contrast with the more

24   homogeneous reaction to green light of the clones with the Fw. We interpreted this as an

25   indication that the lagging strand of the RK2 origin of replication has the orientation of the

26   antisense strand encoding the *repA* protein. It thus seemed that diversification was more

27   efficient in the libraries generated with pGSRv plasmid, something surely related to the

28   location of the lagging strand in vector pSEVA621.

29

30   **Isolation and enrichment of improved light-responsive devices in *P. putida*.** In order to

1   identify improved variants of the [$ccaS•pcyA/ho1•cca$R• $P_{cpcG2-172}$ →] system with a lower

2   basal expression level and a superior output in *P. putida*, we started by sorting the 4 libraries

3   to isolate clones with the highest fluorescence levels in response to light induction. The

4   sorting process involved three positive rounds where the best GFP emitters were selected after

5   induction. This was then followed by a fourth negative cycle to collect clones with lower

6   fluorescent emissions under darkness conditions, as detailed in the Material and Methods

7   section. Total events selected per round of sorting and their relative percentages of the total

8   events per population are shown in Supplementary Table S3. During the sorting process, we

9   noticed that the 4 libraries evolved differently through the enrichment steps. As shown in

10   Figure 4A, none of the Fw libraries (whether A or B versions) failed to display any

11   meaningful shift in their cell cytometry profiles because of an improvement of their

12   fluorescence. In contrast both Rv-A and Rv-B libraries were enriched in high-GFP variants

13   after every sorting cycle. As a consequence, we chose only clones bearing Rv libraries for the

14   rest of the screening, as the probability to find an improved device was higher because of its

15   greater variability. On this basis, we isolated 2610 clones (1305 from the Rv-A library and

16   1305 from the Rv-B library) in 96-well plates with the help of a colony picker. By comparing

17   with samples before sorting (Fig. 4A), we noticed that the enrichment process had

18   significantly increased the number of clones that responded differentially to the presence of

19   light (Fig. 4A, Supplementary Fig. S5). After GFP measurements under light induction and in

20   darkness conditions of the 30 plates with the clones, we picked what appeared to be the best

21   84 clones. The two main parameters of these (basal level fluorescence and fold induction)

22   were separately measured again in three biological replicates and plotted (Fig. 4B) for

23   selecting the 5 best performers for further characterization.

24

25   **Phenotypic characterization of evolved light-responsive devices.** The 5 clones mentioned

26   above were taken through a suite of Flow Cytometry experiments. This technology not only

27   allows the monitoring of ON/OFF states but also exposes homogeneity of gene expression

28   throughout the population. The plasmids carried in those clones were named after the well

29   where they were found in the cognate 96-well plate: pB10, pC8, pA8, pB11 and pB9 (Figure

30   4B). Comparing the results of the five clones among them and with those bearing the non-

1   mutated construct of pGSRv, we noticed that pB9 originated the lowest fluorescence when

2   cultures were not induced and had the widest dynamic range when exposed to green light. In

3   order to confirm that these results were only due to the relevant parts of pB9, the DNA

4   segment containing the complete evolved CcaSR system was reassembled again in the same

5   vector backbone, pSEVA621, thereby yielding plasmid pGreenL (Fig. 5A-C), which had the

6   same behavior as the parental one. *P. putida* cells carrying pGreenL were then subjected to

7   further analyses. First, we inspected whether the system was specifically responding to green

8   light and not to illumination by any other wavelength. In particular, we wished to check

9   whether red light inhibited activity of the CcaSR system at levels below those in darkness,

10 non-induction levels, as had been previously reported[2, 5]. To answer this question, we

11 illuminated cultures of *P. putida* (pGreenL) with red light for 8 hours. The results shown in

12 Supplementary Fig. 6B indicated that [i] the CcaSR sensor was blind to red light and [ii] the

13 same red light had no effect on basal expression levels of GFP.

14

15 **Mapping mutations enabling inter-operativity of the CcaSR system**. The next logic step

16 was figuring out the evolutionary landscape of changes through the intergenic regions of the

17 CcaSR system that had allowed a well-regulated performance in *P. putida*. To this end, the

18 entire DNA constructs were subject to Pacific Biosciences SMRT (Single Molecule Real-

19 Time) PCR amplicon sequencing, based on a modified version of Nyerges *et al.*[12, 23, 24], as

20 explained in Materials and Methods. The 5 analyzed samples included libraries Rv-A (sample

21 1) and Rv-B (sample 2), sorted libraries Rv-A (sample 3) and Rv-B (sample 4) and the 86

22 best performing clones (sample 5; Fig. 6). As expected, we found that mutations gathered just

23 in the tagged regulatory regions. Non-sorted libraries RvA and RvB accumulated mutations in

24 the sites that control expression of the two main proteins CcaS and CcaR along with others in

25 the regulatory region of the *ho1-pcyA* genes. In contrast $P_{cpcG2-172}$ included a lower number of

26 changes. However, after the sorting process, we observed a different pattern of mutations in

27 both libraries comparing them with the previous non-selected populations. The amount of

28 mutations spotted in the regulatory region of CcaS increased notably —unlike the number of

29 mutations included upstream of the rest of the genes. This finding suggested that regulation of

30 CcaS expression was one of the key factors that influenced operation of the system. In the last

11

1 sample (composed by the 86 best performing clones), we observed a similar distribution of

2 mutations, confirming the relevance of CcaS expression levels for optimal reusability of the

3 device. In a subsequent step, we sequenced the entire CcaSR system from pGreenL by

4 classical Sanger sequencing in order to identify the changes that influenced the observed

5 modifications of the basal level and the wider dynamic range of the sensor. We found 9

6 alterations located in only two regions: the RBS region of the CcaS protein had accumulated 7

7 mutations, while the RBS region of CcaR had 2 changes (the complete sequences can be

8 found in the Supplementary files). The fact that the selected clone had only mutations in those

9 regions strengthened our earlier hypothesis that the key bottleneck for re-nesting the

10 optogenetic device from *E. coli* to *P. putida* was the imbalance of expression of main

11 components of the CcaSR system. This evidence was also supported by the results of the RBS

12 strength calculation given by Salis software[25] (Supplementary Table S9), as the translation

13 rates for *ccaS* and *ccaR* genes were re-balanced in the pGreenL clone.

14

15 **Conclusion.** The functioning of any given genetic device depends not only on the intrinsic

16 qualities of the biological parts involved or the relational logic among them. What ultimately

17 makes them to work is the embodiment of a particular set of parameters that fit every transfer

18 function to the preceding one and the next one for an optimal signal processing flow through

19 the complete system[26, 27]. While design of electric circuits can be managed by adopting digital

20 0/1 abstractions, biological counterparts are complicated by the large number of parameters

21 that affect the behavior of genetic constructs. Even when one engineered device is optimized

22 for a desired performance in a certain host and fixed conditions, its conduct can typically

23 change when moved to a different context—whether genomic, nutritional or

24 physicochemical—let alone to other hosts[28-30]. The number of parameters to consider for such

25 *interoperability* is so high that relocation of a genetic device from one biological recipient to

26 another cannot be intended from first principles. Instead, the most useful approach is to set the

27 system out to explore an ample solution space in the new environment and let it fluctuate

28 while artificially rewarding displacement towards a preset optimum. The most popular

29 approaches to this end involve combinatorial assembly of the biological parts involved[9], with

30 a focus on regulatory sequences (e.g. promoters and RBS sequences) and then enrichment or

1 selection of the best combinations. While these stratagems have generally worked well, they

2 ask for intensive labor (each diversification round requires extraction and manipulation of

3 DNA) and the outcome is limited by the competence of the hosts to acquire exogenous DNA.

4 As shown in the work above, most of these drawbacks can be overcome by adoption of

5 DIvERGE[12] a variant of MAGE (*Multiplex Automated Genome Engineering*[31]) technology

6 that focuses the generation of hyper-diversity in specific sites of a target DNA segment *in*

7 *vivo*. In the instance described in this work, we have exploited such a technique for

8 simultaneously increasing mutation rates of the 5 regulatory regions of the light-responsive

9 device [*ccaS•pcyA/ho1•cca*R•$P_{cpcG2-172}$→] of Fig. 3 with the objective of setting parameters

10 suitable for its use in the soil bacterium *P. putida*. In our case, we have applied the technique

11 to a device assembled in a low copy number plasmid in order to create libraries with a wide

12 landscape of values and containing workable combinations for their use in the new host. As

13 shown above, by first reducing the population with cell sorting and then isolating specific

14 clones with a robot we ultimately found a new variant of the module with a highly improved

15 performance.   To   the   best   of   our   knowledge,   the   resulting   module

16 [*ccaS•pcyA/ho1•cca*R•$P_{cpcG2-172}$ →] adapted for use in *P. putida* constitutes the first

17 optogenetic device available for this biotechnologically important bacterium[10, 32, 33].

18 Furthermore, the soft randomization-sorting-isolation-selection workflow hereby described

19 should enable the portability and reusability of many other interesting devices initially

20 developed for *E. coli* in other bacteria but the biotechnological value of which may be

21 deployed only when passed to other species.

22

23 **Materials and Methods**

24

25 **Culture conditions.** For general cultures, *E. coli* and *P. putida* strains were grown in Luria

26 Bertani (LB) medium, which had been filtered when performing Flow Cytometry experiments.

27 Solid plates contained LB solidified with agar at 1.5% (w/v). Terrific Broth (TB) was used for

28 the recovering of bacteria after oligo transformation during diversification. When necessary,

29 antibiotics were added at the proper concentrations: kanamycin (Km, 50 µg mL^{-1}),

30 streptomycin (Sm, 50 µg mL^{-1}) and gentamycin (Gm, 10 µg mL^{-1}). For *E. coli*, the

13

1   temperature used for incubation was 37ºC, except with *E. coli* MG1655 strain bearing the

2   mutagenesis machinery encoded in the pORTMAGE3 (Addgene Plasmid Number 72678,

3   http://n2t.net/addgene:72678 ; RRID: Addgene_72678). In that case, incubation at 30ºC was

4   used for periods of non-induction and 42ºC temperature was used for the induction of the

5   mutagenesis apparatus. *P. putida* was incubated at 30 ºC.

6

7   **Strains.** Bacterial strains used in this study are described in Supplementary Table S4. For the

8   construction of new plasmids, *E. coli* strains DH5α, JM109 and DH10B[+] were used. For

9   DIvERGE diversification, *E. coli* K12 MG1655 strain was used. *P. putida* KT2440[34] was the

10   strain selected for the final sorting and testing of the clones.

11

12   **Green light switch constructs.** For most of the constructs, classical protocols of digestion,

13   ligation and transformation in *E. coli* were used[35]. PCRs for new constructs were performed

14   using Q5 High Fidelity polymerase (New England Biolabs, Ipswich, Massachusetts, USA),

15   and analytic PCRs were performed using DNA AmpliTools Green Master Mix (BioTools,

16   Madrid, Spain). Restriction enzymes were purchased from New England Biolabs (Ipswich,

17   MA, USA) and T4 ligase (Roche, Basilea, Switzerland) was used for ligation purposes.

18   Plasmid DNA was extracted from bacteria using QIAprep Spin Miniprep Kit (Qiagen, Venlo,

19   Netherlands). Isolation of DNA from agarose bands was carried out using the NucleoSpin

20   Extract II kit (Macherey Nagel, Düren, Germany). For the transformation of *E. coli* strains,

21   either heat shock (90 seconds at 42 ºC) or electrical shock (cuvettes 0.2 cm, 2.4 kv) was

22   performed. For Gibson assembly reactions, both home-made mix[36] and Gibson Assembly

23   Master Mix were used (New England Biolabs, Ipswich, Massachusetts, USA). The complete

24   list of constructed and used plasmids and primers can be found in Supplementary Tables S5

25   and S6. Plasmids pSR43.6 and pSR58.6 were kindly provided by J. Tabor and S. Schmidl[5].

26   They were used as templates for the amplification of the cassettes *CcaS-ho1-pcyA* and *CcaR*-

27   P*cpc*G2-172-*msfGFP*, respectively. Gibson assembly reaction was performed for one side using

28   the *CcaS-ho1-pcyA* cassette and pSEVA631 (digested with EcoRI and HindIII). Another

29   Gibson assembly reaction was performed to connect the *CcaR*-P$_{cpcG2-172}$-*msfGFP* cassette and

30   pSEVA241 (digested with EcoRI and HindIII again). The resulting respective plasmids,

1   pCcaSPCB and pCcaRGFP, were altogether transformed into *E. coli* JM109 by

2   electroporation. Accuracy of the cloned DNA fragment was confirmed by DNA sequencing.

3   After that, they were transformed into *P. putida* KT2440, wherein their function was tested.

4   In a second batch of constructs, pGSFw and pGSRv plasmids were built assembling the

5   *CcaS-ho1-pcyA* and *CcaR*-P$_{cpcG2-172}$-*msfGFP* cassettes, previously amplified using oligos

6   listed in the Supplementary files, with a PacI/HindIII digested pSEVA621 backbone. Gibson

7   reaction was performed for the assembly. pGSFw and pGSRv were transformed into *E. coli*

8   MG1655 for diversification. The outcome libraries and the parental non mutated plasmids

9   were later transformed into *P. putida* as described previously. After diversification,

10 characterization and sequencing of the selected clone, the optimized light switch cassette was

11 re-cloned in a pSEVA621 using AscI and SwaI enzymes. The final plasmid was called

12 pGreenL and it was characterized as explained above.

13

14 **DIY Arduino-controlled light panel.** For illumination with green light, we used a computer

15 that controlled an RGB LED panel matrix connected to an Arduino device. Components used

16 for the assembly of the LED panel and Arduino are listed in Supplementary Table S1. For the

17 Arduino UNO board, a firmware was developed in the Arduino IDE v1.0.5r2. The code also

18 needed two libraries provided by the manufacturer of the LED panel: <Adafruit_GFX.h>

19 found at https://github.com/adafruit/Adafruit-GFX-Library and <RGBmatrixPanel.h> found

20 at https://github.com/adafruit/RGB-matrix-Panel. Arduino board firmware waits for user

21 commands from the PC through a USB-CDC interface. When commands are received, it

22 checks if they have a correct format and then proceeds to the colour and shape update in the

23 LED matrix panel. Although the Arduino-LedMatrix composite can show different shapes

24 and colours, green and red were the only configuration used in all the LEDs of the panel.

25 Once the LED panel matrix is setup, the user can send a command to save the current colour

26 and shape. Therefore, the next time the Arduino device is powered up there is no need to

27 reprogram it again. The firmware code was developed using a subset of C/C++ language with

28 the Arduino Native libraries. The Arduino UNO device was connected to the LED panel.

29 Power supply connectors were also joined between Arduino and the panel. A specific

30 software was designed to let the user write commands about the light colour and the number

1   of switched-on LEDs. Then, commands are sent to the Arduino board through the USB

2   connector. Once the Arduino board processes them, the colours and shapes in the LED matrix

3   panel are updated. This code is a Python 2.7.8 script for Windows 7, and it also uses the

4   python libraries pyserial-2.7, pywinusb-0.3.3 and, setuptools-5.4.2. Once the Arduino board is

5   programmed, the connection to the PC can be removed so that both the Arduino board and the

6   LED matrix panel composite are portable. Disposition of LED panel, Arduino board and

7   computer for the illumination of cultures is shown in Supplementary Fig. S2.

8

9   **Diversification.** Directed Evolution with Random Genomic Mutations (DIvERGE)

10   experiments were performed using the procedure described by Nyerges and his colleagues[12].

11   Oligos, designed for DIvERGE, were synthesised at the Nucleic Acid Synthesis Laboratory of

12   the Biological Research Centre, Hungarian Academy of Sciences (Szeged, Hungary). Soft-

13   randomized oligos (listed in the Supplementary Table S7) typically included 1 to 25

14   mutations in the desired region that can later be introduced to the targeted region of the

15   genome (or a plasmid) of the microorganism of interest. *E. coli* MG1655 strain was

16   cotransformed with pORTMAGE3 and pGSFw plasmids for one side, and the same strain

17   was cotransformed with pORTMAGE3 and pGSRv for the other side. As a result, two new

18   MG1655 strains (named Fw and Rv) were constructed carrying the machinery for DIvERGE

19   and the cassette *CcaS-ho1-pcyA-CcaR*-P$_{cpcG2-172}$-*msfGFP* in the forward orientation (Fw strain)

20   and in the reverse orientation (Rv strain). Both cultures, carrying either the pGSFw or the

21   pGSRv plasmid, were in turn exposed to 5 rounds of DIvERGE mutagenesis using two

22   combinations of oligos per each culture: 1+2+3+4+5a which produced a library A, and

23   1+2+3+4+5b which produced a library B (Supplementary Table S2). The difference between

24   oligo 5a and 5b is that the former does not target the G box of the p$_{cpcG2-172}$ regulatory region

25   where CcaR binds[4, 5] and, in consequence, does not introduce mutations, while oligo 5b does

26   target this region. As a result, we had four different libraries coming from the four steps of

27   diversification of two plasmids.

28

29   **Transformation of *P. putida* and cell sorting enrichment.** *E. coli* MG1655 bearing the

30   libraries Fw-A, Fw-B, Rv-A and Rv-B were cultured separately overnight (O/N) in 20 mL of

1 LB medium. Afterwards, DNA plasmid from the total 20 mL was extracted and transferred

2 into *P. putida* via 4 different transformations per library. Cultures were incubated O/N in

3 liquid media with the appropriate antibiotic, and then subject to the first screening. First, 100

4 µL of the saturated cultures were taken to inoculate 10 mL of LB in a flask, and they were

5 incubated under green light conditions during 16 h. The rest of the *P. putida* library was

6 preserved in a glycerol stock. After the light incubation, 1 mL of the culture was taken, and

7 the OD_{600} was adjusted to 0.5. Then, the different libraries were sorted using a cell sorter

8 FACSVantage SE with an installed BD FACSDiVa Software (Becton Dickinson, Franklin

9 Lakes, New Jersey, USA). The top GFP fluorescent population was selected in every sorting

10 event (Supplementary Fig. S3), while the rest of the clones were discarded. Final volume of

11 the outcome was supplemented to reach a total of 10 mL and was then grown O/N. Next, 100

12 µL were used to inoculate another 10 mL LB media for a new cycle of green light exposition

13 and cell sorting, and the rest of the culture was saved in glycerol stock at -80ºC. After three

14 sorting cycles that included incubation under green light during 16 h followed by selection of

15 the most promising GFP positive clones and the preservation of the outcome, the population

16 containing the most green fluorescent cells was incubated in complete darkness for 16 h and

17 passed through a final sorting with the opposite selection criteria, that being selection of the

18 cells with a lower fluorescence. The final outcome was incubated O/N at 30 ºC and preserved

19 in glycerol at -80 ºC.

20

21 **Isolation of sorted clones.** For the following steps, only the RvA and RvB libraries were

22 used, as they showed a higher variability in terms of response to light. For the colony picker

23 screening, we used 30x30 cm plates with 300 mL LB agar. O/N cultures of the RvA and RvB

24 libraries obtained from the sorting cycles described above were diluted 1 to 50 in LB, and 100

25 µL of the outcome were spread in each square plate, to plate the maximum amount of

26 colonies with an adequate separation amongst them. Plates grown O/N at 30ºC were placed in

27 a colony picker Getenetix QPix (Molecular Devices, San José, California) that took and

28 inoculated clones in 96-well microtiter plates containing 200 µL of LB in each well. In total,

29 15 microtiter plates harboring 1305 variants from the RvA library and another 15 with the

30 same amount of clones from the RvB library were obtained by this procedure and preserved

1 by addition of 15% glycerol (Supplementary Fig. S4). Each of those 30 plates keeping the

2 2610 selected clones was organized as indicated in Supplementary Fig. S4A. Supplementary

3 Fig. S4B shows the workflow from the sorted libraries, to the colony picker isolation and

4 finally to the selected clone.

5

6 **Selection of best performers.** Each plate containing the isolated clones was used to inoculate

7 200-µl aliquots placed on two replica 96-well plates for fluorescence measurement (Costar

8 black plates with clear bottom; Thermo Fisher Scientific Inc., Pittsburgh, PA, USA;

9 Supplementary Fig. S4). One of the replica plates was placed under darkness conditions

10 (covered with aluminium foil) while the other under light conditions during 16 h without

11 shaking. Both plates were measured for OD_{600} and fluorescence using a SpectraMax Plate

12 reader (Molecular Devices, San José, CA, USA). After normalisation of GFP values using the

13 OD_{600}, X-fold induction was calculated for each clone. Mean, standard error (SD) and

14 covariance (CV) were calculated for the X-Fold of the 8 replicates of the parental strain. CV

15 illustrates the reproducibility of the parental strain results; this variability probably extends as

16 well to the results of each clone. Each calculated X-Fold for the isolated clones was divided

17 by the X-Fold mean of the parental strain. If the resulting number, deemed the relative X-Fold,

18 was superior to 1+ CV/100, it was considered to be a significant value. The higher this value

19 was, the better the clone in regards to its induction capacity. After analyzing all the isolated

20 variants, the best 86 performers were further examined in three biological replicates using the

21 same method. Finally, the five clones that showed the best behavior were studied by single

22 cell flow cytometry experiments in more detail.

23

24 **Flow cytometry.** Overnight cultures of each of the five best performers were used to

25 inoculate fresh filtered LB (1:5000) in 10 mL tubes and incubated with 190 rpm shaking at

26 30ºC during 8 hours in duplicate. One of the replica was grown in darkness (covered with

27 aluminium foil) and the other in a light-induced state. Next, cells were sedimented and

28 washed with filtered 1X PBS and resuspended in the same buffer at the adequate OD_{600} to

29 pass them through a MACSQuant Flow Citometer (Miltenyi, Bergisch Galdbach, Germany).

30 An Ar laser, diode-pumped solid state, was used to excite msfGFP at 488 nm and the

1   fluorescence signal was recovered with a 525/40 nm band-pass filter. For the experiments,

2   ≥120,000 events were counted per sample. Results were analysed using FlowJo software

3   (FlowJo LLC, Ashland, Oregon).

4

5   **Deep amplicon sequencing of DIvERGE libraries and individual clones.** The allelic

6   composition of the CcaSR libraries and the best-performer ten clones were determined by a

7   Pacific Biosciences SMRT (Single Molecule Real-Time) PCR amplicon sequencing

8   protocol[12]. To create sequencing libraries, a 6595 bp long region, encoding *ccaS, pcyA, ho1,*

9   *ccaR*, P*cpcG2-172-msfgfp*, was PCR amplified from the previously isolated, pooled plasmid

10   DNA samples by using the corresponding barcoded primer pairs (see Supplementary Table).

11   To multiplex sequencing samples, symmetrically barcoded PCR primers were designed[12].

12   Each primer-pair consisted of a symmetrical 16 nucleotide-long Pacific Biosciences Sequel

13   barcode sequence, a 3 nucleotide-long 5' terminal sequence, besides the 3' terminal plasmid-

14   specific primer sequences. All PCR primers were synthesized by the Nucleic Acid Synthesis

15   Laboratory of the Biological Research Centre, Hungarian Academy of Sciences, Szeged,

16   Hungary, and purified using high-performance liquid chromatography (HPLC). Following

17   DNA synthesis and purification, primers were suspended in nuclease-free H2O at 100 μM

18   final concentration. Next, the [*ccaS•pcyA/ho1•cca*R• *PcpcG2-172* →] region from each plasmid

19   sample was amplified in 3×25 μl volumes, consisting of 37.5 μl 2× Q5 Hot-Start MasterMix

20   (New England Biolabs), 2.5 μl of the corresponding sample-specific, barcoded primers (10

21   μM), 200 ng template plasmid DNA, and 33 μl nuclease-free H2O. PCRs were performed in

22   thin-wall PCR tubes in a BioRad CFX96 qPCR machine with the following thermal profile:

23   98 °C 3 minutes, 20 cycles of (98 °C 15 seconds, 72 °C 4:30 minutes), and a final extension

24   of 72 °C for 10 minutes. Following PCRs, the 6633 basepair-long amplicons were purified by

25   using a Zymo Research DNA Clean and Concentrator™ Kit according to the manufacturer's

26   protocol (Zymo Research) and eluted in 30 μl 0.5× Tris-EDTA (TE) buffer (pH 8.0). To

27   prepare samples for sequencing, amplicons were quantified using Qubit dsDNA BR assay kit

28   (Thermo Fisher Scientific), mixed, and libraries were sequenced on a Pacific Biosciences

29   Sequel instrument using Sequel Polymerase v2.1 on an SMRT cell v2 with Sequencing

30   chemistry v2.1 (both from Pacific Biosciences). The SMRT cell was loaded by diffusion, and

1 the sequencing movie time was 1200 min. After sequencing, raw sequencing subreads were

2 demultiplexed according to their corresponding barcodes (see Supplementary Table S8) by

3 using Demultiplex Barcodes pipeline on SMRT Link v5.1.0.26412 (SMRT Tools

4 v5.1.0.26366). A minimum barcode score of 26 was used identify high-quality barcodes. The

5 average sequencing read counts were 9000 per library sample. To determine the sequence of

6 individual clones, we relied on Long Amplicon Analysis and determined the phased

7 consensus sequences for each of the corresponding demultiplexed datasets. Following

8 demultiplexing, sequencing reads were mapped to their corresponding reference sequence by

9 using bowtie2 2.3.4 [37] in "--very-sensitive" mode and the nucleotide composition was

10 extracted for each nucleotide position within the CcaSR region. Finally, allelic replacement

11 frequencies at nucleotide positions were quantified by measuring the distribution and ratio of

12 nucleotide substitutions for each reference nucleotide position.

13

14

15 **ASSOCIATED CONTENT**

16

17

18 **Supplementary Figure S1.** Representation of the CcaSR two-component system and its
19   mechanism of action.

20 **Suplementary Figure S2.**   LED matrix panel used to induce the CcaSR system with green
21   light.

22 **Supplementary Figure S3.** Schematic representation of the procedure followed for the
23   libraries' cell sorting.

24 **Figure S4.**   Isolation of clones from libraries.

25 **Supplementary Figure S5.** Outcome plates at different points of the screening.

26 **Supplementary Figure S6:** pGreenL behaviour. **Supplementary Table S1**.   List of
27   components used for the LED plate and the company where they
28   were purchased.

1 **Supplementary Table S2.**   Plasmids and oligos used for DIVERGE diversification and the
2   resulting libraries.

3 **Supplementary Table S3.**   Sorting rounds of each library and their selection events.

4 **Supplementary Table S4.**   List of strains used in this study

5 **Supplementary Table S5.**   List of plasmids used in this study

6 **Supplementary Table S6.**   List of primers used for the constructions performed in this
7   study.

8 **Supplementary Table S7.**   List of primers used for the diversification of the regulatory
9   regions.

10 **Supplementary Table S8.**   List of primers used for the sequencing for the mutations profile.

11 **Supplementary Table S9.**   Results of the RBS strength calculation for the original (WT)
12   and the mutated (pGreenL) RBS regions of *ccaS* and *ccaR* genes
13   using Salis software.

14 **pGreenL sequence**

15 **COMPETING INTERESTS**

16
17 Akos Nyerges and Csaba Pal have filed a patent application (PCT/EP2017/082574) related to
18 DIvERGE.

19

20 **AUTHORS' CONTRIBUTIONS**

21

22 AN and VDL conceived the study and BC, AHG and AN designed the experiments. AHG
23 carried out the experimental parts of the work. AHG, VDL and AN wrote the article. All the
24 authors contributed to the discussion of the research and interpretation of the data.

25

26 **ACKNOWLEDGEMENTS**

27

28 Authors are indebted to Laura Molero from the SIdI service at the Universidad Autónoma de
29 Madrid for the sorting of libraries. We also thank David Gonzalez and Miguel Alcalde for

14
15

16 **REFERENCES**
17

18 1. Beja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P.,
19   Jovanovich, S. B., Gates, C. M., Feldman, R. A., Spudich, J. L., Spudich, E. N., and
20   DeLong, E. F. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in
21   the sea, *Science 289*, 1902-1906.
22 2. Hirose, Y., Shimada, T., Narikawa, R., Katayama, M., and Ikeuchi, M. (2008)
23   Cyanobacteriochrome CcaS is the green light receptor that induces the expression of
24   phycobilisome linker protein, *Proc Natl Acad Sci U S A 105*, 9528-9533.
25 3. Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M.,
26   Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., and Voigt, C. A.
27   (2005) Synthetic biology: engineering Escherichia coli to see light, *Nature 438*, 441-
28   442.
29 4. Tabor, J. J., Levskaya, A., and Voigt, C. A. (2011) Multichromatic control of gene
30   expression in Escherichia coli, *J Mol Biol 405*, 315-324.
31 5. Schmidl, S. R., Sheth, R. U., Wu, A., and Tabor, J. J. (2014) Refactoring and

optimization of light-switchable Escherichia coli two-component systems, *ACS Synth Biol 24*, 24.

6. Castillo-Hair, S. M., Baerman, E. A., Fujita, M., Igoshin, O. A., and Tabor, J. J. (2019) Optogenetic control of Bacillus subtilis gene expression, *Nat Commun 10*, 019-10906.

7. Kim, J., Salvador, M., Saunders, E., González, J., Avignone-Rossa, C., and Jiménez, J. I. (2016) Properties of alternative microbial hosts used in synthetic biology: towards the design of a modular chassis, *Essays Biochem 60*, 303-313.

8. Adams, B. L. (2016) The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field, *ACS Synth Biol 5*, 1328-1330.

9. Yokobayashi, Y., Weiss, R., and Arnold, F. H. (2002) Directed evolution of a genetic circuit, *Proc Natl Acad Sci U S A 99*, 16587-16591.

10. Nikel, P. I., Chavarria, M., Danchin, A., and de Lorenzo, V. (2016) From dirt to industrial applications: Pseudomonas putida as a Synthetic Biology chassis for hosting harsh biochemical reactions, *Curr Opin Chem Biol 34*, 20-29.

11. Martinez-Garcia, E., and de Lorenzo, V. (2017) Molecular tools and emerging strategies for deep genetic/genomic refactoring of Pseudomonas, *Curr Opin Biotechnol 47*, 120-132.

12. Nyerges, A., Csorgo, B., Draskovits, G., Kintses, B., Szili, P., Ferenc, G., Revesz, T., Ari, E., Nagy, I., Balint, B., Vasarhelyi, B. M., Bihari, P., Szamel, M., Balogh, D., Papp, H., Kalapis, D., Papp, B., and Pal, C. (2018) Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance, *Proc Natl Acad Sci U S A 115*, E5726-E5735.

13. Silva-Rocha, R., Martinez-Garcia, E., Calles, B., Chavarria, M., Arce-Rodriguez, A., de Las Heras, A., Paez-Espino, A. D., Durante-Rodriguez, G., Kim, J., Nikel, P. I., Platero, R., and de Lorenzo, V. (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes, *Nucleic Acids Res 41*, 23.

14. Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009) Programming cells by multiplex genome engineering and accelerated evolution, *Nature 460*, 894-898.

15. Swaminathan, S., Ellis, H. M., Waters, L. S., Yu, D., Lee, E. C., Court, D. L., and Sharan, S. K. (2001) Rapid engineering of bacterial artificial chromosomes using

oligonucleotides, *Genesis 29*, 14-21.

16. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and Sugino, A. (1968) Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains, *Proc Natl Acad Sci U S A 59*, 598-605.

17. Sakabe, K., and Okazaki, R. (1966) A unique property of the replicating region of chromosomal DNA, *Biochim Biophys Acta 129*, 651-654.

18. Ellis, H. M., Yu, D., DiTizio, T., and Court, D. L. (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides, *Proc Natl Acad Sci U S A 98*, 6742-6746.

19. Nyerges, A., Csorgo, B., Nagy, I., Balint, B., Bihari, P., Lazar, V., Apjok, G., Umenhoffer, K., Bogos, B., Posfai, G., and Pal, C. (2016) A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species, *Proc Natl Acad Sci U S A 113*, 2502-2507.

20. Costantino, N., and Court, D. L. (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants, *Proc Natl Acad Sci U S A 100*, 15748-15753.

21. Nyerges, A., Csorgo, B., Nagy, I., Latinovics, D., Szamecz, B., Posfai, G., and Pal, C. (2014) Conditional DNA repair mutants enable highly precise genome engineering, *Nucleic Acids Res 42*, 5.

22. Ezer, D., Shepherd, S. J. K., Brestovitsky, A., Dickinson, P., Cortijo, S., Charoensawan, V., Box, M. S., Biswas, S., Jaeger, K. E., and Wigge, P. A. (2017) The G-Box Transcriptional Regulatory Code in Arabidopsis, *Plant Physiol 175*, 628-640.

23. Hebert, P. D. N., Braukmann, T. W. A., Prosser, S. W. J., Ratnasingham, S., deWaard, J. R., Ivanova, N. V., Janzen, D. H., Hallwachs, W., Naik, S., Sones, J. E., and Zakharov, E. V. (2018) A Sequel to Sanger: amplicon sequencing that scales, *BMC Genomics 19*, 018-4611.

24. Rhoads, A., and Au, K. F. (2015) PacBio Sequencing and Its Applications, *Genomics Proteomics Bioinformatics 13*, 278-289.

25. Salis, H. M. (2011) The ribosome binding site calculator, *Methods Enzymol 498*, 19-42.

26. Endy, D. (2005) Foundations for engineering biology, *Nature 438*, 449-453.

27. Canton, B., Labno, A., and Endy, D. (2008) Refinement and standardization of

synthetic biological parts and devices, *Nat Biotechnol 26*, 787-793.

28. Cardinale, S., and Arkin, A. P. (2012) Contextualizing context for synthetic biology-- identifying causes of failure of synthetic biological systems, *Biotechnol J 7*, 856-866.

29. Boo, A., Ellis, T., and Stan, G.-B. (2019) Host-aware synthetic biology, *Current Opinion in Systems Biology 14*, 66-72.

30. Kittleson, J. T., Wu, G. C., and Anderson, J. C. (2012) Successes and failures in modular genetic engineering, *Curr Opin Chem Biol 16*, 329-336.

31. Carr, P. A., Wang, H. H., Sterling, B., Isaacs, F. J., Lajoie, M. J., Xu, G., Church, G. M., and Jacobson, J. M. (2012) Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection, *Nucleic acids research 40*, e132-e132.

32. Nikel, P. I., and de Lorenzo, V. (2018) Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism, *Metabolic Engineering 50*, 142-155.

33. Martinez-Garcia, E., and de Lorenzo, V. (2016) The quest for the minimal bacterial genome, *Curr Opin Biotechnol 42*, 216-224.

34. Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A., Fouts, D. E., Gill, S. R., Pop, M., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R. T., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Chris Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzez, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Stjepandic, D., Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J. A., Timmis, K. N., Dusterhoft, A., Tummler, B., and Fraser, C. M. (2002) Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440, *Environ Microbiol 4*, 799-808.

35. Maniatis, T., F Fritsch, E., and Sambrook, J. (2018) *Molecular cloning : a laboratory manual / J. Sambrook, E.F. Fritsch, T. Maniatis*.

36. Aparicio, T., de Lorenzo, V., and Martínez-García, E. (2017) Broadening the SEVA Plasmid Repertoire to Facilitate Genomic Editing of Gram-Negative Bacteria, In *Hydrocarbon and Lipid Microbiology Protocols: Genetic, Genomic and System Analyses of Pure Cultures* (McGenity, T. J., Timmis, K. N., and Nogales, B., Eds.), pp 9-27, Springer Berlin Heidelberg, Berlin, Heidelberg.

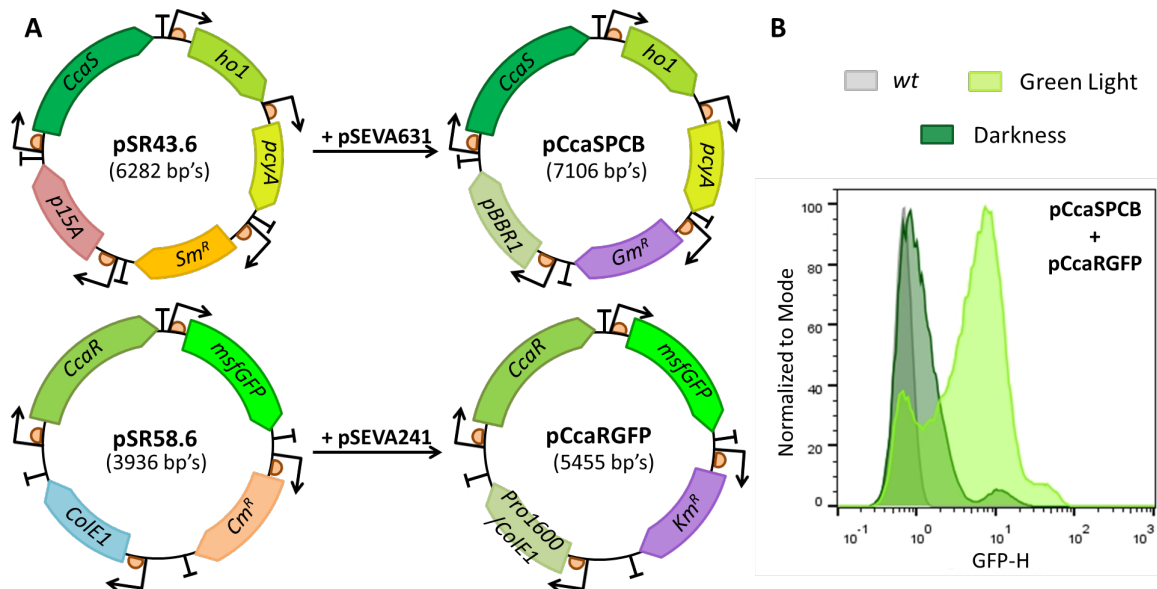37. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2,

25

1   *Nat Methods* 9, 357-359.

2

3

1 ## FIGURES

2

3 **Figure 1.** CcaSR two-component system original distribution and its behaviour.

4

5

6
7
8
9 **A.** The functional DNA segments of the original vectors were adapted to the broad-host range
10 pSEVA plasmids for their transfer to *P. putida*. Notice that the gene distribution was
11 maintained and that the origins of replication of the destination backbone were selected to
12 keep a similar plasmid copy number. **B.** Cell cytometry of *P. putida* bearing CcaSR-encoding
13 plasmids after 8 h of induction. *P. putida* KT2440 *wt* with no GFP is plotted as a negative
14 control (grey), together with *P. putida* harbouring pCcaSPCB + pCcaRGFP induced by green
15 light (light green) and incubated in darkness (dark green).
16
17
18

**Figure 2.** Behavior of a merged CcaSR-derived light-responsive device in *P. putida*
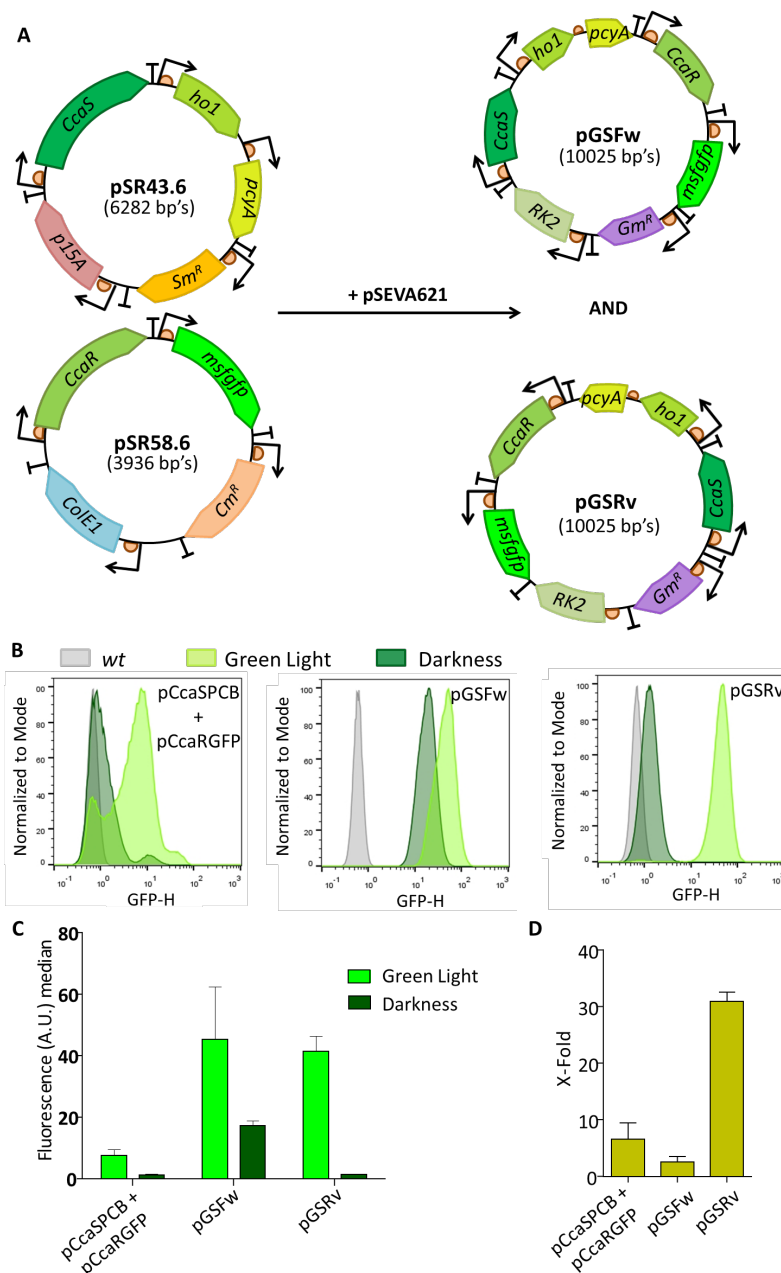
A. Second adaptation of the CcaSR system to *P. putida*. This time, both gene packs were merged in one single construct. The fused protein-coding cassette was placed in different orientations for the two final vectors: in the forward direction (pGSFw) and in its complementary counterpart (pGSRv). **B.** GFP expression profile of *P. putida* harbouring the CcaSR system in three plasmid formats. Cell cytometry for *P. putida* harbouring the two adapted plasmids, pCcaSPCB and pCcaRGFP, the single pGSFw plasmid with all the necessary genes in the forward orientation and the single pGSRv plasmid with same construct but in reverse orientation. In the three cases, fluorescence of *P. putida* KT2440 *wt* (grey

1   curve), green light-induced cultures (light green) or darkness/non-induced cultures (dark
2   green) are shown. **C.** Plotted medians for three flow cytometry biological replicates of light-
3   induced and non-induced *P. putida* cultures with the same constructs. **D.** X-fold induction of
4   the CcaSR system encoded in the three combinations of plasmids.

1   **Figure 3.** Strategy for diversification of regulatory sites
2
3
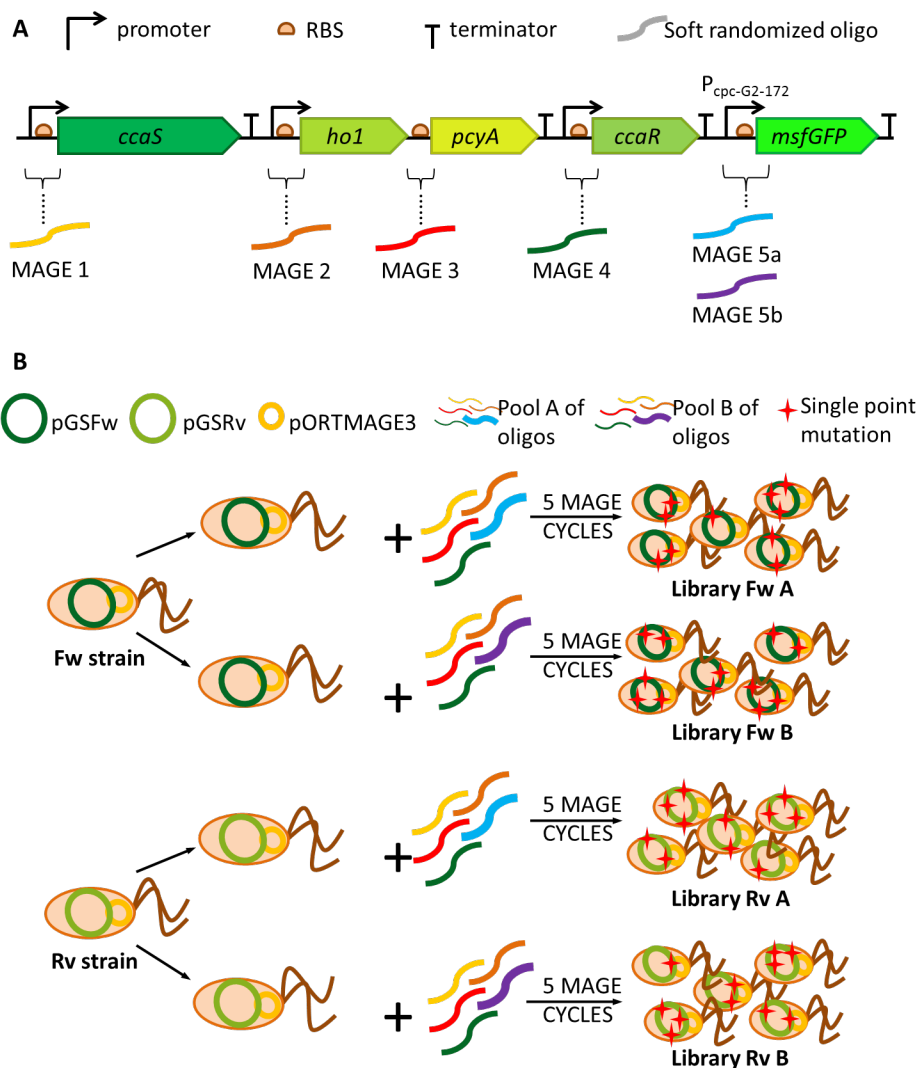
4
5
6
7   **A.** CcaSR cassette and DIVERGE oligo targets. Disposition of components as separate
8   transcriptional units (except the two-gene operon *ho1-pcyA*) is shown. Each soft-randomized
9   oligo mixture (including 1 or 2 mutations per molecule) was designed for the regulatory
10   region of every gene including the promoters and the RBSs. For the $p_{cpc-G2-172}$ sequence
11   controlling the reporter gene, two different oligos were designed (DIVERGE 5a and
12   DIVERGE 5b), targeting the G-Box recognized by CcaR or not including this sequence in the
13   mutagenic process, respectively. **B.** Generation of CcaSR libraries using DIvERGE approach.
14   Sketch of DIvERGE diversification over pGSFw and pGSRv plasmids using 2 pools of soft
15   randomized oligos, carried out with *E. coli* MG1655 as host. The use of either mix of primers
16   per each construct gave as a result 4 different libraries with random mutations in the
17   regulatory regions of the CcaSR components. The four libraries are Fw A, Fw B, Rv A and
18   Rv B.
19

1   **Figure 4.** Analysis of libraries FwA, FwB, RvA and RvB

2
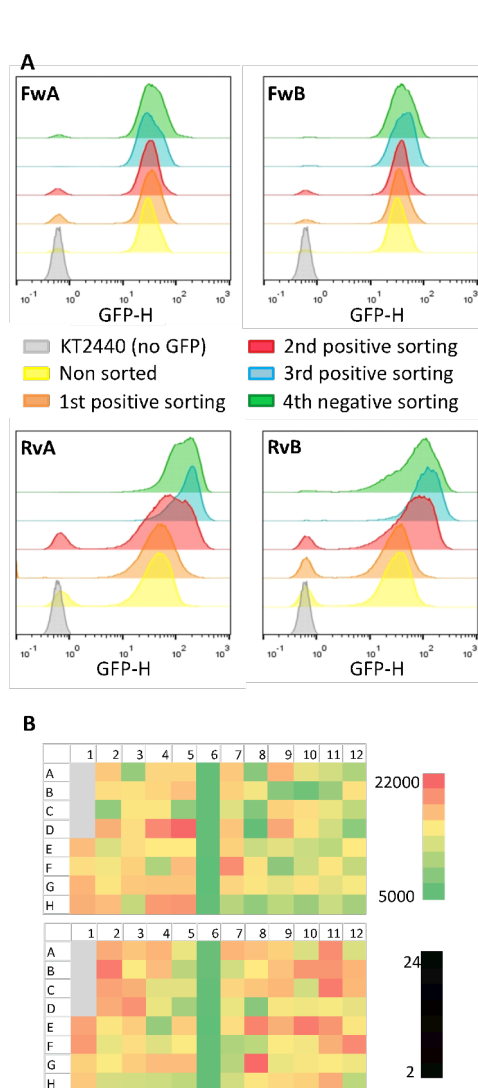
3
4   **A.** Population fluorescence of the four libraries
5   before being sorted using FACS (yellow curve),
6   after three sortings for selection of the induced
7   population (first cycle, orange curve; second cycle,
8   red curve; third cycle, blue curve) and after a
9   fourth sorting event selecting the most GFP-
10   negative non-induced population (green curve).
11
12   **B.** Heatmaps for the final selected 84 clones
13   representing the two criteria for selecting the best
14   performers. On top, the basal values of
15   fluorescence were plotted. The lower values are
16   represented in green, while higher values
17   correspond to red squares. In the bottom, X-fold
18   induction values between the non-induced state
19   and the light-induced state are shown. Higher X-
20   fold values are in green while lower counterparts
21   are plotted in red. In either heatmap, grey squares
22   correspond to blank measurements, while column
23   6 was inoculated with the pGSRv-containing
24   strain, thereby showing a uniform colour in both
25   cases. Represented values are the calculated mean
26   of three biological replicates. To select the 5 best
27   clones, we chose wells with lower basal levels
28   combined with higher X-fold induction.

1 **Figure 5:** Light induction of the 5 best clones after the screening process.
2
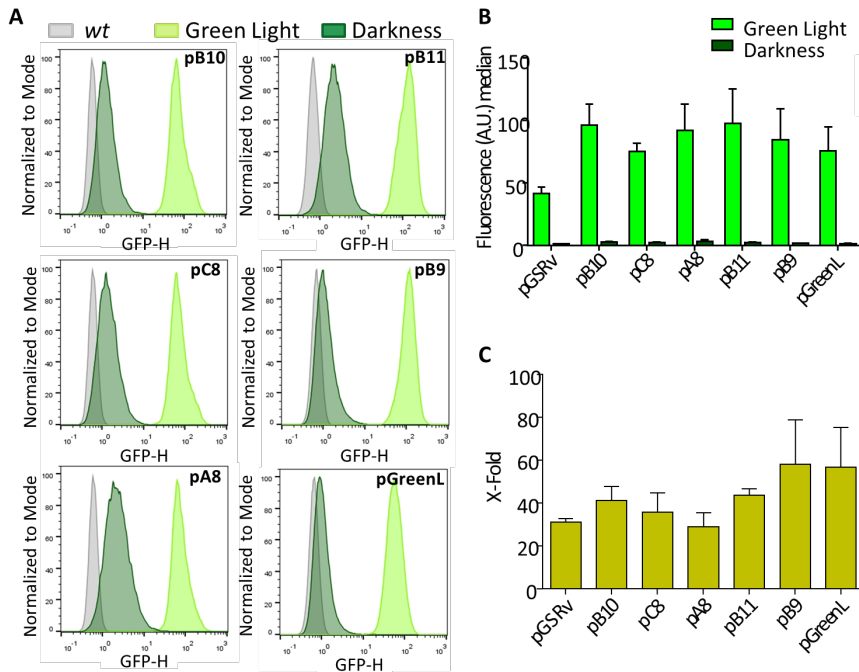
3
4
5 **A,** Cell cytometry for the 5 pGSRv derivatives carried by *P. putida*. Measurements include
6 light-induced cultures (light green), darkness cultures (dark green) and the control *P. putida*
7 KT2440 *wt* (grey). **B.** Median averages from three biological replicates of cell cytometry with
8 induced and non-induced cultures, including the non-mutated plasmid pGSRv. **C.** Average X-
9 Fold induction for each of the clones compared to pGSRv.
10
11
12
13

Figure 6: Deep sequencing of resulting libraries from the diversification process.
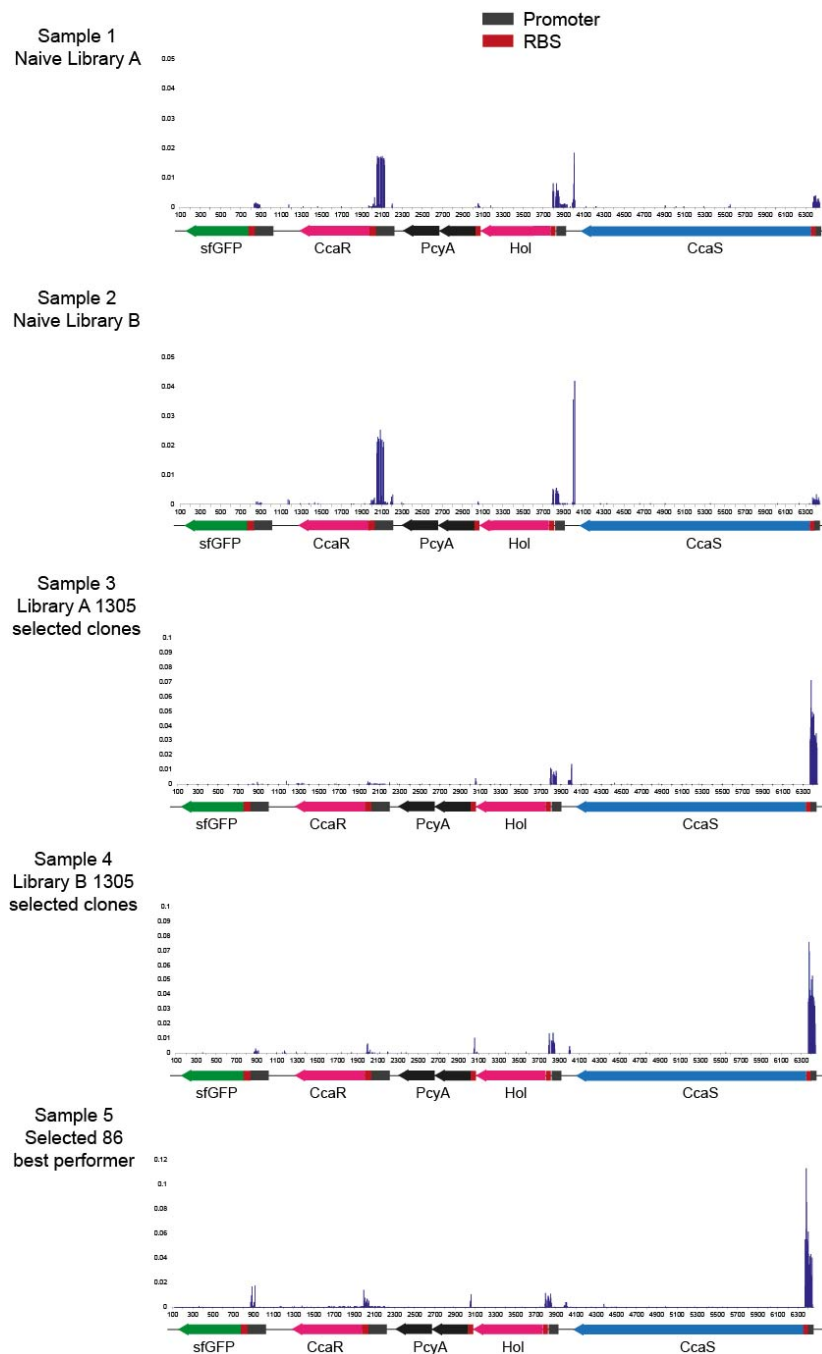
Samples 1 and 2 correspond to naïve libraries RvA and RvB, both result of the mutagenesis process before sorting. Sample 3 and 4 correspond to the 1305 clones of library RvA and RvB (respectively) that were sorted and isolated for its characterization in terms of light responsiveness. Sample 5 represents the sequences of the 86 best performers that were selected from the 2610 (1305 + 1305) clones from samples 3 and 4.