**Supplemental Material**

# Cumate-inducible gene expression system for sphingomonads and other Alphaproteobacteria

**Andreas Kaczmarczyk, Julia A. Vorholt, Anne Francez-Charlot[#]**

Institute of Microbiology, ETH Zurich, Zurich, Switzerland

## Supplemental Material and Methods

### Construction of plasmid pTOPO-TERM193

Tandem copies of a putative bidirectional transcriptional terminator encoded in the Sphingomonas Fr1 genome (between genes Sphme2DRAFT_0228 and Sphme2DRAFT_0229; Fig S1A) were constructed by oligo annealing and extension with Phusion DNA polymerase using oligos oJVZ793 and oJVZ794. Oligo oJVZ793 contains the full terminator sequence and, when double-stranded after the first round of PCR with help of oVZ794, can self-anneal through identical 3'- and 5'-ends in a way that allows the formation of larger concatemers in each PCR cycle (Fig. S1B). Cycling conditions were as follows: 2 min initial denaturation at 98°C; 10 s 98°C, 10 s 57°C, 40 s 72°C (20 cycles). The PCR reaction was subjected to agarose gel electrophoresis (4% agarose) and bands between 200 and 400 bps were purified, T-tailed by incubation with MasterTaq DNA polymerase (5Prime), and cloned into pCR2.1-TOPO (Invitrogen). Sequencing revealed that one clone carried 2 repeats of the terminator, with the last terminator corresponding to the wild-type terminator, and the first terminator carrying a 1-bp deletion at the base of the stem loop (plasmid pTOPO-TERM193; Fig. S1C).

### Construction of plasmid pQF

Plasmid pQF is the parent for all other PQ5-based plasmids and was assembled in multiple steps (an outline is given in Fig. S2). First, a fragment containing a high-GC codon-optimized version of the repressor-encoding gene *cymR* (*cymR**) driven by the *neo* promoter, the $P_{syn2}$-derivative $P_{syn2a}$ flanked by two CuO operator sites followed by a MCS (including N- and C-terminal triple-FLAG-tags [3xFLAG-tags]) and the putative transcriptional terminator T193 was synthesized by GeneArt (Invitrogen). This fragment was amplified using oligos oJVZ1286 and oJVZ1287, digested with EcoRI/PscI and cloned between the same sites of pCM62, generating pQ0 (Fig. S2A). Sequencing revealed that T193 carried a point mutation in the stem. Several attempts to revert this mutation by PCR-based site-specific mutagenesis failed for unknown reasons and we thus proceeded with this mutant terminator, termed T193*. Next $P_{neo}$, $P_{syn2a}$ and the CuO

sequences of pQ0 were replaced by a synthesized DNA fragment (MWG Operon Eurofins) containing the *bla* promoter (driving *cymR** expression) and $P_{syn2}$ with a single CuO operator sequence downstream of the $P_{syn2}$ -10 box (Fig. S2B); the fragment had been delivered on a standard cloning plasmid and was directly subcloned using NheI/SpeI restriction sites, generating pQ0new. *lacZ* on the XbaI/EcoRI fragment of pAK127lacZ(MCS) was then subcloned into the BcuI/EcoRI sites of pQ0new, resulting in pQ0newlacZ (Fig. S2C). The NheI/AgeI-fragment of pQ0newlacZ harboring $P_{bla}$ was replaced by a fragment harboring a modified *bla* promoter core sequence ($P_{bla\text{-mut1}}$), in which the -35 hexamer has been changed from TTCAAA to TTGACA and the -10 hexamer from GACAAT to NACAAT, generated by annealing of oligos oJVZ1098 and oJVZ1099, giving a plasmid library called pQ0newlacZ-Pblamut1 (note that the first base in the -10 box is randomized). Next, the plasmid library was digested with AgeI/NcoI and a DNA fragment encoding the CuO operator sequence generated by annealing oligos oJVZ1100 and oJVZ1101 was cloned in between, giving plasmid library pQ0lacZ-Pblamut1-CuO. This library was transformed into *Sphingomonas* Fr1 and several transformants were tested for β-galactosidase activity with and without cumate and the plasmid of one transformant, which showed the highest induction ratio (clone 5), was extracted and sequenced. This plasmid, which encoded the -10 hexamer TACAAT in the modified *bla* promoter (referred to as $P_{bla\text{-mut1T}}$), was designated pQF-lacZ (Fig. S1D). To restore the original multiple cloning site, a fragment harboring *cymR** and $P_{Q5}$ was PCR-amplified from pQF-lacZ (oligos oJVZ1286/oJVZ1288) and cloned in between the PscI/BcuI sites of pQ0new, yielding pQF.

**Table S1.** Oligonucleotides used in this work

| Primer name | Primer sequence (5' to 3') |
|---|---|
| oJVZ534 | ATTTTAAGCTTAGCAGGGGGCATGGCCTC |
| oJVZ739 | ATTTGGTACCTCTAGCTAGCGATATCACC |
| oJVZ740 | ATTTTCTAGAGACAAGTTTGTACAAAAAAGC |
| oJVZ743 | ATTTGGTACCAGAAGGGAGAGACCCCGAATGG |
| oJVZ744 | ATTTGAATTCTTACTTGTACAGCTCGTCCATG |
| oJVZ745 | ATTTTACATGTGATGGATATCTGCAGAATCCGCCC |
| oJVZ746 | ATTTTAAGCTTGCCGCCAGTGTGCTGGGATTCGCCC |
| oJVZ793 | TTTTGTTTGCGGGGGTTCGCCGGTCACGCAAAAAAGGCCGGGGGAGCGATCC CCCGGCCTTTTTGTTTGCGGG |
| oJVZ794 | CCCGCAAACAAAAAGGCC |
| oJVZ1082 | ATTTGGTACCCACAGGAAACAGCTATGACC |
| oJVZ1083 | ATTTGAATTCAATACGGGCAGACATGGCC |
| oJVZ1086 | AGCTTTTTCTAGAGGGCCGGATCCCG |
| oJVZ1087 | GTACCGGGATCCGGCCCTCTAGAAAA |
| oJVZ1088 | AGCTTCCGCATGTACCCGTACGACGTCCCGGACTACGCCAT |
| oJVZ1089 | TAATGGCGTAGTCCGGGACGTCGTACGGGTACATGCGGA |
| oJVZ1090 | AGCTTCCGCATGGAGCAGAAGCTGATCTCGGAAGAGGACCTGGGCTCGAT |
| oJVZ1091 | TAATCGAGCCCAGGTCCTCTTCCGAGATCAGCTTCTGCTCCATGCGGA |
| oJVZ1093 | ATTTTGGTACCGCATCGCGGCCCATCACGGC |
| oJVZ1098 | CCGGTACATTGACATATGTATCCGCTCATGANACAATAACCCTGATG |
| oJVZ1099 | CTAGCATCAGGGTTATTGTNTCATGAGCGGATACATATGTCAATGTA |
| oJVZ1100 | CCGGAACAAACAGACAATCTGGTCTGTTTGTA |
| oJVZ1101 | CATGTACAAACAGACCAGATTGTCTGTTTGTT |
| oJVZ1102 | ATTTAGATCTTCCCCGAAAAGTGCCACC |
| oJVZ1103 | ATTTACTAGTAGGCCTAGCACCAGGCGTTTAAGG |
| oJVZ1104 | ATTTGGATCCGCTGATGTCCGGCGGTGC |
| oJVZ1105 | ATTTCCATGGATGGATATCTGCAGAATC |
| oJVZ1106 | ATTTGCTAGCACGGGCAGACATGGCCTG |
| oJVZ1107 | ATTTAAGCTTCCGCATGGTGAGCAAGGGCGAGG |
| oJVZ1108 | ATTTATTAATCTTGTACAGCTCGTCCATG |
| oJVZ1286 | ATTTACATGTTCAGCGCTTGAACTTGGC |
| oJVZ1287 | ATTGAATTCAAAAAAAAGGCCGGGGGAGCG |

| | |
|---|---|
| oJVZ1288 | TGCAAACTAGTTACAAACAGACC |
| oJVZ1291 | AGCTGCTGCAATGGAACCAACGAACGCGACTTTGGTTTCATCGCC |
| oJVZ1292 | GATCGGCGATGAAACCAAAGTCGCGTTCGTTGGTTCCATTGCAGC |
| oJVZ1434 | GTACAGCATTGACGGGCCCGGCCACCCCGCGTAGAAGCGCGCCAC |
| oJVZ1435 | GTGGCGCGCTTCTACGCGGGGTGGCCGGGCCCGTCAATGCT |
| oJVZ1436 | ATTTAGGCCTACGGGTTGCTGCGCAACC |
| oJVZ1437 | ATTTACATGTTGAAGACGAAAGGGCCTCG |
| PL28pNTGACN_s | ATTTAAGCTTCCGGCGTGCGACGGTGCGCGGATCGCCCCATGGNTGACNGATCGCCGCGAAGCCGC |
| PL28pTANNNGC_as | ATGGATCCGGCTTTAAATCCTGAATTCGGTTGCCCGGAAAGCCGGCGCNNNTAGCGGCTTCGCGGCGATC |

**Table S2.** Putative housekeeping genes used for MEME motif search

| Gene (IMG locus tag) | Annotation (according to the IMG system) | Upstream region used for motif search (in bp) |
|---|---|---|
| Sphme2DRAFT_0192 | ribosomal protein L21 | 100 |
| Sphme2DRAFT_0653 | ribosomal protein L36 | 100 |
| Sphme2DRAFT_0682 | ribosomal protein L28 | 100 |
| Sphme2DRAFT_0799 | DNA-directed RNA polymerase, omega subunit | 100 |
| Sphme2DRAFT_1509 | ribosomal protein S12 | 100 |
| Sphme2DRAFT_1511 | translation elongation factor EF-G | 100 |
| Sphme2DRAFT_1512 | translation elongation factor TU | 100 |
| Sphme2DRAFT_1514 | ribosomal protein L3 | 100 |
| Sphme2DRAFT_1524 | ribosomal protein L14 | 100 |
| Sphme2DRAFT_1543 | ribosomal protein S13 | 120 |
| Sphme2DRAFT_1545 | DNA-directed RNA polymerase, alpha subunit | 100 |
| Sphme2DRAFT_1546 | ribosomal protein L17 | 100 |
| Sphme2DRAFT_1816 | ribosomal protein L33 | 100 |
| Sphme2DRAFT_2298 | ribosomal protein L13 | 100 |
| Sphme2DRAFT_2439 | DNA primase, catalytic core | 100 |
| Sphme2DRAFT_2588 | ribosomal protein L10 | 100 |
| Sphme2DRAFT_2592 | ribosomal protein L11 | 99 |
| Sphme2DRAFT_2614 | signal recognition particle protein | 100 |
| Sphme2DRAFT_2637 | ribosomal protein S21 | 100 |
| Sphme2DRAFT_2679 | ribosomal protein L25 | 100 |
| Sphme2DRAFT_2712 | ribosomal protein S1 | 100 |
| Sphme2DRAFT_2792 | ribosomal protein L31 | 100 |
| Sphme2DRAFT_3405 | ribosomal protein S6 | 100 |

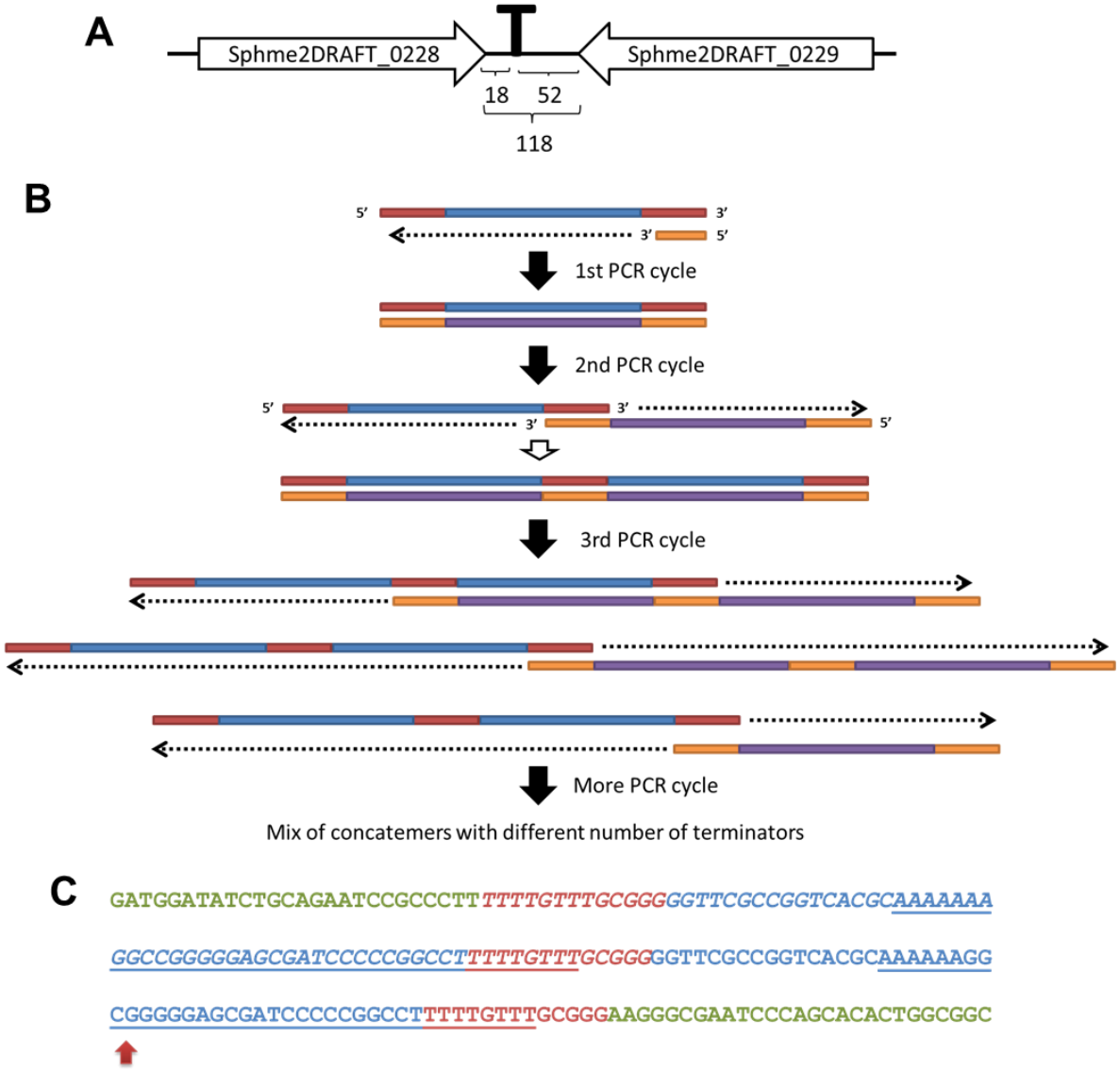


**Fig. S1. Genomic context, construction and sequence of TERM193. (A)** Location of the original terminator on the genome (not to scale). Numbers indicate the distance of genes to each other and the beginning/end of the terminator in base pairs. **(B)** Construction of TERM193 using oligonucleotide annealing and extension with DNAP. The red regions in oligonucleotide oJVZ793 represent identical sequences at the 5' and 3' ends, the blue region encodes the terminator. Oligo oJVZ794 is depicted in orange. Sequences complementary to the red and blue regions in oJVZ793 are depicted in orange and purple, respectively. Only a selection of productive annealing/elongation events is shown. **(C)** Sequence of the fragment encoding TERM193. Terminators are underlined. A single copy of oJVZ793 is indicated in italics and the repetetive elements are shown with the same color code as in **(B)**. The green sequences are derived from the backbone of pCR2.1-TOPO. The red arrow indicates the location of the 1-bp deletion (see the text for details).

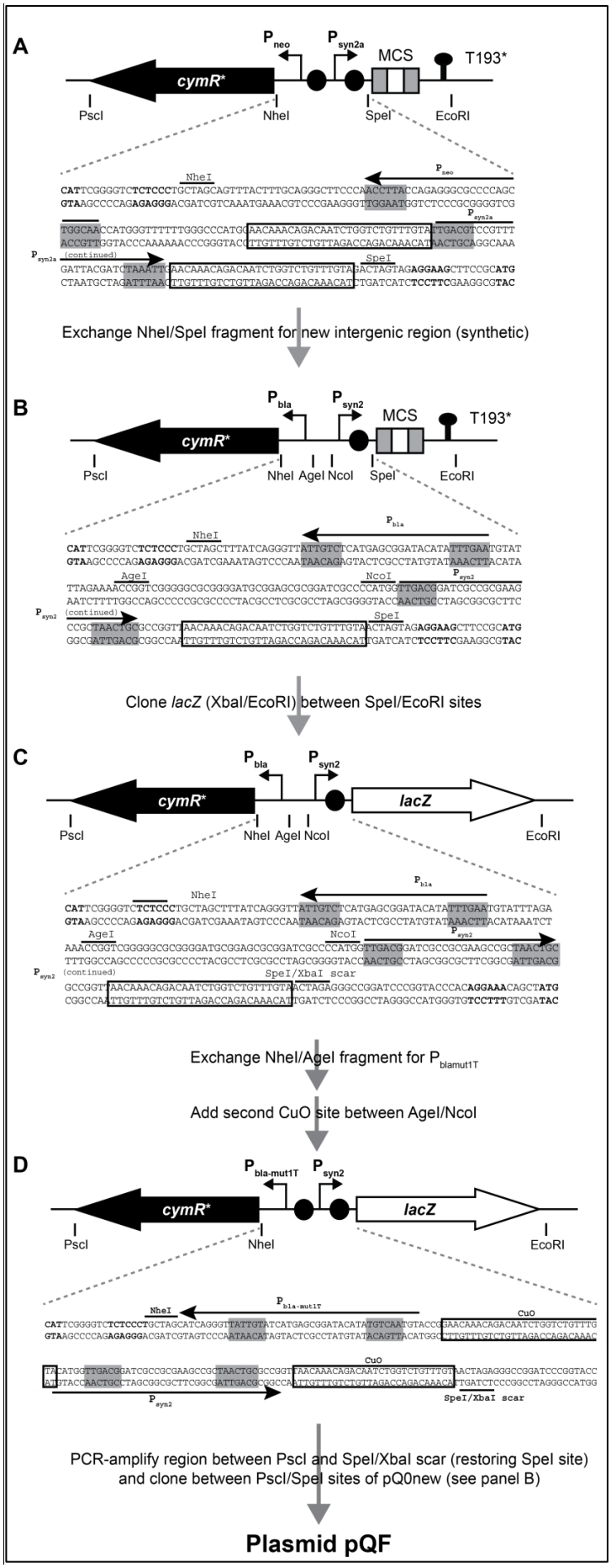


**Fig. S2. Construction of plasmid pQF. (A)** Plasmid pQ0. **(B)** Plasmid pQ0new. **(C)** Plasmid pQ0new-lacZ. **(D)** Plasmid pQF-lacZ. In the sequences, promoter regions are indicated by arrows and their -35 and -10 boxes are highlighted in grey. *cymR** and MCS/*lacZ* start codons and ribosome bind sites are in bold and CuO operator sites are boxed. The schematic drawings show genes (large arrows) and promoters (bended arrows), the CuO operator sites (black circles), the MCS including the 3xFLAG tags (grey boxes) and relevant restriction sites. Note that cloning of *lacZ* destroyes the SpeI site (indicated as 'SpeI/XbaI scar' in panels **(C)** and **(D)**). For details see the text.

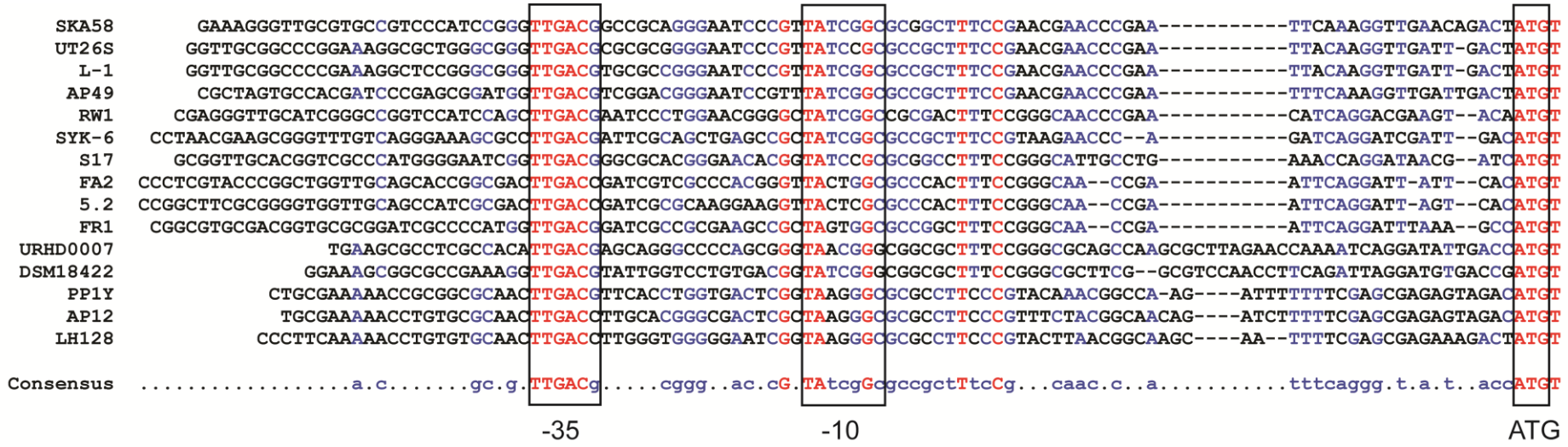


**Fig. S3.** Alignment of the 100 base pairs upstream of the *rpmB* open reading frame from different sphingomonads. Highly conserved nucleotides (>90%) are displayed in red and modestly conserved nucleotides (>50%) in blue. The putative -35 and -10 boxes and the *rpmB* start codon (ATG) are boxes. FR1, *Sphingomonas* sp. strain Fr1; S17, *Sphingomonas* sp. strain S17; SKA58, *Sphingomonas* sp. strain SKA58; RW1, *Sphingomonas wittichii* strain RW1; UT26S, *Sphingobium japonicum* UT26S; L-1, *Sphingobium chlorophenolicum* L-1; SYK-6, *Sphingobium* sp. SYK-6; AP49, *Sphingobium* sp. AP49; FA2, *Sphingomonas phyllosphere* FA2; 5.2, *Sphingomonas phyllosphere* 5.2; URHD0007, *Sphingomonas* sp. URHD0007; DSM18422, *Sphingomonas jaspsi* DSM 18422; PP1Y, *Novosphingobium* sp. PP1Y; AP12, *Novosphingobium* sp. AP12; LH128, *Sphingomonas* sp. LH128.


**Fig. S4.** Sequence alignment of the 14 identified strong promoters (core regions only). The -35 and -10 boxes are highlighted in grey and bases that had been randomized are in bold. Identical promoters that served as the basis for $P_{syn2}$ (core sequence TTGACG-$N_{17}$-TAACTGC) are highlighted in bold.