

SUPPLEMENTARY FIGURE LEGENDS:

Figure S1.

The *galK-loxP-INV* substrate for the Cre recombinase. **A.** Wild type *galK* region (strain MG1655). **B.** Insertion of a 36 bp in frame oligonucleotide containing the 34 bp *loxP** site between codons 239 and 240. *loxP** denotes single bp change at center of *loxP* site. **C.** Insertion of *cat-sacB* cassette replacing a DNA segment of *galK* from beyond *loxP* through *galM* to *gpmA*. **D.** Replacement of the *cat-sacB* cassette with the exact DNA segment deleted in step C (*loxP*, *galK* through *galM*), but inverted. **E.** The *tetA* gene, amplified from *Tn10*, was inserted into a position just beyond the *galΔ loxP** site. Strain CC942 is MG1655 carrying this step E construct, which is the substrate used in this study to detect Cre-mediated recombination. Strain CC942 is resistant to tetracycline at 12.5 µg/ml driven by a promoter in the upstream *galΔ loxP** region and induced by presence of galactose in the media. **F.** Active Cre recombinase provided *in trans* inverts the DNA between the two *loxP* sites, restoring a functional *galK* gene and a Gal⁺ phenotype. The second *loxP* site has a junction sequence with *gpmA* that creates a strong promoter (red rectangle) driving the expression of the inverted *tetA* gene and making the strain also resistant to tetracycline at 12.5 µg/ml on LB agar.

The DNA sequence of the construct in CC942 (Step E) is found in Figure S2. The DNA sequence of the Cre-mediated inversion (Step F) is found in Figure S3. Bent arrows represent promoters for the *gal* operon, the *cat* cassette, and the inverted *tetA*. Phenotypes are indicated to the right of the figure.

Figure S2.

Sequence of *galK-loxP-INV* in strain CC942 of Figure S1E. The duplex DNA sequence is shown and where a functional protein is made the amino acid sequence is also shown. The *galK* coding sequence is “blue” with the initiation codon marked. The rest of the *gal* operon upstream of *galK* is not shown here. The distal part of *galK* beyond the *loxP** site has been inverted and transposed (see Fig. S1). The *loxP** site in *galK* inserted beyond GCC (codon 239 underlined) is cyan, and the lower case red t/a base pair is a base change to eliminate Cre-independent recombination between the two *Lox* sites. Beyond *loxP* is the *tetA* gene with an engineered

upstream spacer and ribosome binding site. The *tetA* coding sequence is pink and the upstream region black. Beyond *tetA* is the end of the inverted *galM* gene in green. The blue sequence beyond the *galM* is the end of the inverted and fragmented *galK* gene in blue. The second *loxP* site (cyan) is beyond *galK* (the junction is indicated by an inverted K in Fig. S1). The two green base pairs were added with the 34 bp of *loxP* to maintain frame and are next to the insertion site at *galK* codon 240 (CTG underlined). Beyond *loxP* is the *gpmA* gene as mustard color. The large double headed arrow indicates the region that can be inverted by Cre recombination at inverted Lox sites. The -10 and -35 sequences are underlined for the serendipitous promoter generated between *loxP* and *gpmA*.

Figure S3.

Sequence of the Gal⁺ *galK-loxP-INV* after inversion. Sequence of *galK-loxP-INV* region after Cre-mediated inversion of the construct as shown in Figure S1F. Sequence above large arrow has been inverted by Cre recombinase. The color coding of sequences of genes and sites are the same as for Figure S2. Here the *galK* coding sequence has been reconnected by the Cre-mediated inversion as indicated by the amino acid sequence, and codons 239 and 240 are indicated on each side of *loxP*. The *tetA* gene is transcribed from a promoter generated between *loxP* and *gpmA*, for which the -10 and -35, respectively, are underlined.

Figure S4.

Construction of Cre recombinase gene controlled by the arabinose promoter *P_{BAD}*.

A. Wildtype *araBAD* operon with the *polB* gene of DNA polymerase II downstream. SD indicates Shine-Dalgarno ribosome binding site. **B.** Insertion of *cat-sacB bla* cassette downstream of *araA* and removing part of *araD*. This is strain XTL191 as described in (Li *et al.* 2013). **C.** Replacement of SD *araB*, *araA*, *cat*, *sacB*, *bla* and remainder of *araD* with the Cre recombinase gene and an optimized SD* ribosome binding site. **D.** Mutations of wild type TAT at codon 324 to either TGT or TGC, which cause a Tyr to Cys change at the active site as indicated by asterisk (*). A *cat* cassette replacement of the *yabQ* gene, ~6 kbp downstream of *polB*, is indicated. **E.** Replacement of SD *araB*, *araA*, *cat*, *sacB*, *bla* and remainder of *araD* in Figure S4B above with the firefly luciferase gene *luc* and the optimized SD*.

Figure S5.

Sequence of Cre expression construct shown in Figure S4C. The *cre* gene sequence is shown (blue) with the Cre protein sequence below. The long blue arrows indicate the sequences of the single strand oligo primers used to amplify by PCR the wild type *cre* gene from the DNA of bacteriophage P1 (STERNBERG *et al.* 1986). These primers had at their 5' end homology to the upstream and downstream regions of the *ara* operon generating an amplified *cre* gene flanked by these homologies. The upstream homology sequence (red) included the P_{BAD} promoter and the modified SD*. The downstream homology sequence (black) included the *araD* stop codon (TAG) and region beyond. This PCR product was recombined with strain XTL191 (Li *et al.* 2013) containing the *cat-sacB-bla* (Fig. S5B). The sucrose resistant (Cm^S and Amp^S) recombinants had inserted *cre* under P_{BAD} control. Red arrows (5' to 3') indicate the sequencing primers used to verify the sequences shown here. Black arrow indicates the 5' transcription start for the arabinose operon. The black bar represents the optimized sequence of the Shine-Dalgarno (SD*) to enhance translation of Cre. A partial sequence of the *polB* gene downstream of the *ara* operon is shown.

The *cre* gene TAT codon (red) encodes tyrosine at position 324 of the Cre protein. Only tyrosine can be present at this position for Cre to be active. We used two oligos to change this codon to a TGT and a TGC, respectively, generating a defective Cre with cysteine incorporated at position 325. These changes are indicated as a white star in Figure S4D. Each of these two oligos also had changes in the wobble position of four codons adjacent to the cysteine codon change. The sequence of each oligo is shown in Table S2.

Figure S6.

***In vitro* proofreading of G→A error by GreA and GreB on Cre^{TGT} sequence.** TEC C9 was assembled with 10 pmol *E. coli* RNA polymerase core enzyme, 10 pmol of RNA C9-TDS57 Cre^{TGT} hybrid and 20 pmol NDS57 Cre^{TGT} in transcription buffer (20 mM Tris-HCl, pH 7.9, 5 mM MgCl₂, 40 mM KCl and 2 mM 2-mercaptoethanol). The resulting TEC C9 was diluted to 100 nM final concentration in the same buffer and incubated with 1 mM ATP and UTP (and no GTP) and 0.2 mg/ml GreA or GreB for indicated time. In this experiment, RNA polymerase incorporated the cognate UMP to the 3' end of C9 primer followed by misincorporation of AMP for GMP. The fraction of TEC U10 misincorporating AMP is plotted against the incubation time

(in sec). The RNA sequence encoded immediately downstream 3' end of the C9 RNA primer is UGUAUCC with UGU codon underlined and the site for G→A misincorporation shown in bold. On this scaffold, the transcribed sequence surrounding TGT is identical to that of the Cre^{TGT} fidelity reporter of Figure 1A. The sequences of RNA and DNA oligonucleotides for assembly of TEC C9: RNA primer C9 Cre^{TGT} carrying eight 8-nt C-tail and 56FAM fluorophore at the 5' end: 5'-/56FAM/ CCCCCCCCUGUAUGAAC, Template strand TDS57 Cre^{TGT}: 5'-TGTTTCACTATCCAGGTTACGGATACAGTTCATGACAATATTTACATTGGTCCAGCC, non-template strand NDS57 Cre^{TGT}: 5'-GGCTGGACCAATGTAAATATTGTCATGAACTGTATCCGTAACCTGGATAGTGAAAC

SUPPLEMENTARY REFERENCES:

- Koulich, D., M. Orlova, A. Malhotra, A. Sali, S. A. Darst *et al.*, 1997 Domain organization of Escherichia coli transcript cleavage factors GreA and GreB. *Journal of Biological Chemistry* 272: 7201-7210.
- Li, X. T., L. C. Thomason, J. A. Sawitzke, N. Costantino and D. L. Court, 2013 Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. *Nucleic Acids Research* 41.
- Sternberg, N., B. Sauer, R. Hoess and K. Abremski, 1986 Bacteriophage-P1 Cre Gene and Its Regulatory Region - Evidence for Multiple Promoters and for Regulation by DNA Methylation. *Journal of Molecular Biology* 187: 197-212.
- Thomason, L. C., J. A. Sawitzke, X. Li, N. Costantino and D. L. Court, 2014 Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol* 106: 1 16 11-39.

SUPPLEMENTARY TABLES:

Table S1: Strains and plasmids

Strain	Genotype	Ref
CC930	MG1655 <i>galK-loxP-INV araD<>cre^{TAT}-cat</i>	this work
CC942	MG1655 <i>galK-loxP-INV</i>	this work
CC945	MG1655 <i>galK-loxP-INV araD<>cre^{TGT}-cat</i>	this work
CC947	MG1655 <i>galK-loxP-INV araD<>cre^{TGC}-cat</i>	this work
CC1234	MG1655 <i>galK-loxP-INV araD<>luc-cat</i>	this work
CC1235	MG1655 <i>galK-loxP-INV araD<>luc-cat greA<>spc</i>	this work
CC1236	MG1655 <i>galK-loxP-INV araD<>luc-cat greB<>amp</i>	this work
DY330	W3110 $\Delta lacU169 gal490 pgl\Delta 8 [\lambda cI857 \Delta(cro-bioA)]$	(THOMASON <i>et al.</i> 2014)
NB301	DY330 <i>greA<>spc</i>	this work
NB394	DY330 <i>greB<>amp</i>	this work
NB987	CC942 <i>greB<>amp</i>	this work
NB1001	CC942 <i>greA<>spc</i>	this work
NB1004	CC942 <i>greA<>spc greB<>amp</i>	this work
NB1041	CC942 <i>cre^{TGT}</i>	this work
NB1054	NB1001 <i>cre^{TGT}</i>	this work
NB1055	NB987 <i>cre^{TGT}</i>	this work
NB1056	NB1004 <i>cre^{TGT}</i>	this work
NB1103	NB1001 pGreA	this work
NB1104	NB1001 pGreB	this work
NB1105	NB1001 pBR322	this work
NB1115	NB1103 <i>cre^{TGT}</i>	this work
NB1116	NB1104 <i>cre^{TGT}</i>	this work
NB1117	NB1105 <i>cre^{TGT}</i>	this work
pBR322	ColE1 Amp ^R Tet ^R	
pGreA	Derivative of pTrc88A, ColE1 Amp ^R	((KOULICH <i>et al.</i> 1997)
pGreB	Derivative of pTrc88A, ColE1 Amp ^R	(KOULICH <i>et al.</i> 1997)
pSIM18	pSC101ts <i>lexo bet gam</i> Hyg ^R	(THOMASON <i>et al.</i> 2014)

Table S2: Primers *

Primer name	Sequence
greA-spcF	TGGGTGAAGACTTGCCCTATCAGGAATATTCAAGAGGTATAACAA
greA-spcR	ATGCGCTCACGCAACTGGTCC TTCCTTTCTTTACAATACATCAACATCTTGAGTATTGGGTAATTC
greA checkF	TTATTTGCCGACTACCTTGGT
greA checkR	TCTGGTCCCGGTAAGGAGTTATGC
greB-ampF	GCAAAATGCCACGCTCTGTTCTGTTG
greB-ampR	ATACCCAATGTGCGCATTATCAAACAGACAAAGGGAATCAACGAG ATGAGTATTCAACATTTCCGTGTC
greB checkF	GGCTGGCAAAAATGCCAGCCATCGGCAGGAGTTAAGACTCTTCC
greB checkR	TTACCAATGCTTAATCAGTGAGGC
creF	CGGCGAAACATTATTGATTCTGTTG AAATCAGGGGATAGTTATACGGAC CGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTT GGAT <u>AGGAGGAT</u> GAAACGATGTCGAATTTACTGACCGTA
creR	GGCTTGAGTATAGCCTGGTTTCGTTTGATTGGCTGTGGTTTTATA CAGTCACTAATCGCCATCTTCCAGCAG
catF	CAAATTACTGACAACGCAATTATTGAATGAGCCAGAA _{taa} GCTAAG GTTG <u>TGTGACGGAAGATCACTTCGCAG</u>
catR	TGAAATCATCCCACGCTACTGCTAAGGTTGAAGGGGCGTTCCAGCC CCTT <u>ACCAGCAATAGACATAAGCGGC</u>
cat checkF	<u>TG CTTTTTCTTACCCGTTGACC</u>
cat checkR	<u>TTTTTCGTCTCAGCCAATCCC</u>
lucF	CGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTG GAT <u>AGGAGGAT</u> GAAACGATGGAAGACGCCAAAAACATAA
lucR	GGCTTGAGTATAGCCTGGTTTCGTTTGATTGGCTGTGGTTTTATACA GTCATTACACGGCGATCTTCCG
creTGT mutagenic	GCTGGTGGCTGGACCAATGTAAATATTGTCATGAAt <u>Tg</u> TATtCGcAAtC TGGATAGTGAAACAGGGGCAATGGTG
creTGT antisense	<u>ACTATCCAGATTGCGAATCAA</u>
creTGC mutagenic	GCTGGTGGCTGGACCAATGTAAATATTGTCATGAAt <u>Tgc</u> ATtCGcAAtC TGGATAGTGAAACAGGGGCAATGGTG
creTGC antisense	<u>ACTATCCAGATTGCGAATGCAA</u>

*oligos used to construct the *galK-loxP-INV* reporter are available upon request

Figure S1

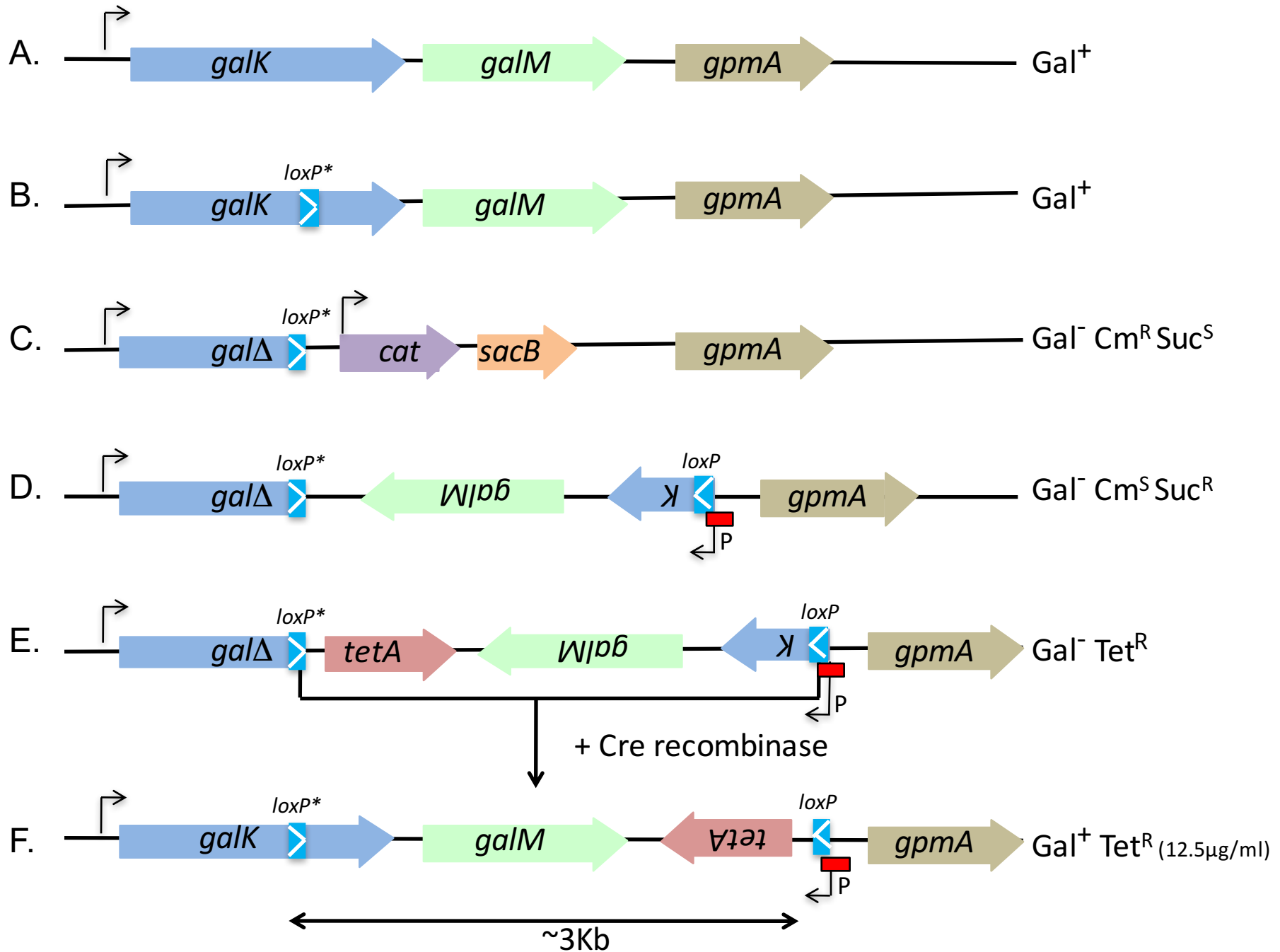


Figure S2

Galk>

CAGAACAGGCAGCAGAGCGTTTGC GCGCAGTCAGCGATATCCATTTTCGCGAATCCGGAGTGTAAGAA ATG
GTCCTTGTCCTCGTCTCGCAAACGCGCGTCAGTCGCTATAGGTAAGCGCTTAGGCCTCACATTCTT TAC

AGT CTG AAA GAA AAA ACA CAA TCT CTG TTT GCC AAC GCA TTT GGC TAC CCT GCC
TCA GAC TTT CTT TTT TGT GTT AGA GAC AAA CGG TTG CGT AAA CCG ATG GGA CGG

ACT CAC ACC ATT CAG GCG CCT GGC CGC GTG AAT TTG ATT GGT GAA CAC ACC GAC
TGA GTG TGG TAA GTC CGC GGA CCG GCG CAC TTA AAC TAA CCA CTT GTG TGG CTG

TAC AAC GAC GGT TTC GTT CTG CCC TGC GCG ATT GAT TAT CAA ACC GTG ATC AGT
ATG TTG CTG CCA AAG CAA GAC GGG ACG CGC TAA CTA ATA GTT TGG CAC TAG TCA

TGT GCA CCA CGC GAT GAC CGT AAA GTT CGC GTG ATG GCA GCC GAT TAT GAA AAT
ACA CGT GGT GCG CTA CTG GCA TTT CAA GCG CAC TAC CGT CGG CTA ATA CTT TTA

CAG CTC GAC GAG TTT TCC CTC GAT GCG CCC ATT GTC GCA CAT GAA AAC TAT CAA
GTC GAG CTG CTC AAA AGG GAG CTA CGC GGG TAA CAG CGT GTA CTT TTG ATA GTT

TGG GCT AAC TAC GTT CGT GGC GTG GTG AAA CAT CTG CAA CTG CGT AAC AAC AGC
ACC CGA TTG ATG CAA GCA CCG CAC CAC TTT GTA GAC GTT GAC GCA TTG TTG TCG

TTC GGC GGC GTG GAC ATG GTG ATC AGC GGC AAT GTG CCG CAG GGT GCC GGG TTA
AAG CCG CCG CAC CTG TAC CAC TAG TCG CCG TTA CAC GGC GTC CCA CGG CCC AAT

AGT TCT TCC GCT TCA CTG GAA GTC GCG GTC GGA ACC GTA TTG CAG CAG CTT TAT
TCA AGA AGG CGA AGT GAC CTT CAG CGC CAG CCT TGG CAT AAC GTC GTC GAA ATA

CAT CTG CCG CTG GAC GGC GCA CAA ATC GCG CTT AAC GGT CAG GAA GCA GAA AAC
GTA GAC GGC GAC CTG CCG CGT GTT TAG CGC GAA TTG CCA GTC CTT CGT CTT TTG

CAG TTT GTA GGC TGT AAC TGC GGG ATC ATG GAT CAG CTA ATT TCC GCG CTC GGC
GTC AAA CAT CCG ACA TTG ACG CCC TAG TAC CTA GTC GAT TAA AGG CGC GAG CCG

AAG AAA GAT CAT GCC TTG CTG ATC GAT TGC CGC TCA CTG GGG ACC AAA GCA GTT
TTC TTT CTA GTA CGG AAC GAC TAG CTA ACG GCG AGT GAC CCC TGG TTT CGT CAA

TCC ATG CCC AAA GGT GTG GCT GTC GTC ATC ATC AAC AGT AAC TTC AAA CGT ACC
AGG TAC GGG TTT CCA CAC CGA CAG CAG TAG TAG TTG TCA TTG AAG TTT GCA TGG

CTG GTT GGC AGC GAA TAC AAC ACC CGT CGT GAA CAG TGC GAA ACC GGT GCG CGT
GAC CAA CCG TCG CTT ATG TTG TGG GCA GCA CTT GTC ACG CTT TGG CCA CGC GCA

loxP

TTC TTC CAG CAG CCA GCC ATA ACT TCG TAT A+t GTA TGC TAT ACG AAG TTA TTG
AAG AAG GTC GTC GGT CGG TAT TGA AGC ATA TaA CAT ACG ATA TGC TTC AAT AAC

TetA-->

CGATGGATCCCACGTATCAGACGGAGCACGCTGGAGCTGGAATGAATAGTTTCGACAAAGATCGCATTGGTAA
GCTACCTAGGGTGCATAGTCTGCCTCGTGCACCTCGACCTACTTATCAAGCTGTTTCTAGCGTAACCATT

▶ M N S S T K I A L V

TTACGTTACTCGATGCCATGGGGATTGGCCTTATCATGCCAGTCTTGCCAACGTTATTACGTGAATTTATTG
AATGCAATGAGCTACGGTACCCCTAACCGAATAGTACGGTCAGAACGTTGCAATAATGCACTTAAATAAC

▶ I T L L D A M G I G L I M P V L P T L L R E F I

CTTCGGAAGATATCGCTAACCACTTTGGCGTATTGCTTGCACCTTATGCGTTAATGCAGGTTATCTTTGCTC
GAAGCCTTCTATAGCGATTGGTGAAACCGCATAACGAACGTGAAATACGCAATTACGTCCAATAGAAAACGAG

▶ A S E D I A N H F G V L L A L Y A L M Q V I F A

CTTGGCTTGAAAAATGTCTGACCGATTTGGTTCGGCGCCAGTGTCTGTTGTTGTCATTAATAGGCGCATCGC
GAACCGAACCTTTTTACAGACTGGCTAAACCAGCCGCGGGTCACGACAACAAGTAATTATCCGCGTAGCG

▶ P W L G K M S D R F G R R P V L L L S L I G A S

TGGATTACTTATTGCTGGCTTTTTCAAGTGCCTTTGGATGCTGTATTTAGGCCGTTTGCTTTCAGGGATCA
ACCTAATGAATAACGACCGAAAAAGTTCACGCGAAACCTACGACATAAATCCGCGAAACGAAAGTCCCTAGT

▶ L D Y L L L A F S S A L W M L Y L G R L L S G I

CAGGAGCTACTGGGGCTGTCGCGGCATCGGTCATTGCCGATACCACCTCAGCTTCTCAACCGGTGAAGTGGT
GTCCTCGATGACCCCGACAGCGCCGTAGCCAGTAACGGCTATGGTGGAGTCGAAGAGTTGCGCACTTCACCA

▶ T G A T G A V A A S V I A D T T S A S Q R V K W

TCGGTTGGTTAGGGGCAAGTTTTGGGCTTGGTTAATAGCGGGCCTATTATTGGTGGTTTTGCAGGAGAGA
AGCCAACCAATCCCCGTTCAAACCCGAACCAAATTATCGCCCCGATAATAACCACCAAACGTCCTCTCT

▶ F G W L G A S F G L G L I A G P I I G G F A G E

TTTCACCGCATAGTCCCTTTTTTATCGCTGCGTTGCTAAATATTGTCGCTTTCCTTGTTGTTATGTTTTGGT
AAAGTGGCGTATCAGGGAAAAAATAGCGACGCAACGATTTATAACAGCGAAAGGAACACCAATACAAAACCA

▶ I S P H S P F F I A A L L N I V A F L V V M F W

TCCGTGAAACCAAAAATACACGTGATAATACAGATACCGAAGTAGGGGTTGAGACGCAATCGAATTCGGTAT
AGGCACTTTGTTTTTATGTGCACTATTATGTCTATGGCTTCATCCCCAACTCTGCGTTAGCTTAAGCCATA

▶ F R E T K N T R D N T D T E V G V E T Q S N S V

ACATCACTTTATTTAAAACGATGCCATTTTGTGATTATTTATTTTTTCAGCGCAATTGATAGGCCAAATTC
TGTAGTGAATAAATTTTGTACGGGTAAAACAATAAATAAAAAGTCGCGTTAACTATCCGTTTAAAG

▶ Y I T L F K T M P I L L I I Y F S A Q L I G Q I

CCGCAACGGTGTGGGTGCTATTTACCGAAAATCGTTTTGGATGGAATAGCATGATGGTTGGCTTTTCATTAG
GGCGTTGCCACACCCACGATAAATGGCTTTTAGCAAAACCTACCTTATCGTACTACCAACCGAAAAGTAATC

▶ P A T V W V L F T E N R F G W N S M M V G F S L

CGGGTCTTGGTCTTTTACTCAGTATTCCAAGCCTTTGTGGCAGGAAGAATAGCCACTAAATGGGGCGAAA
GCCAGAACCAGAAAATGTGAGTCATAAGTTTCGAAACACCGTCCTTCTTATCGGTGATTTACCCCGCTTT

▶ A G L G L L H S V F Q A F V A G R I A T K W G E

AAACGGCAGTACTGCTCGGATTTATTGCAGATAGTAGTGCATTTGCCTTTTATAGCGTTTATATCTGAAGTT
TTTGCCGTCATGACGAGCCTAAATAACGTCTATCATCACGTAACCGAAAAATCGCAAATATAGACTTCCAA

▶ K T A V L L G F I A D S S A F A F L A F I S E G

GGTTAGTTTTCCCTGTTTTAATTTTATTGGCTGGTGGTGGGATCGCTTTACCTGCATTACAGGGAGTGATGT
CCAATCAAAGGGACAAAATTAATAAACCAGACCACCCTAGCGAAATGGACGTAATGTCCCTCACTACA

▶ W L V F P V L I L L A G G G I A L P A L Q G V M

CTATCAAACAAAGAGTCATCAGCAAGGTGCTTTACAGGGATTATTGGTGAGCCTTACCAATGCAACCGGTG
GATAGTTTTGTTTCTCAGTAGTCGTTCCACGAAATGTCCCTAATAACCACTCGGAATGGTTACGTTGGCCAC

▶ S I Q T K S H Q Q G A L Q G L L V S L T N A T G

TTATTGGCCATTACTGTTTGCTGTTATTTATAATCATTCACTACCAATTTGGGATGGCTGGATTGGATTA
AATAACCGGTAATGACAAACGACAATAAATATTAGTAAGTGATGGTTAAACCCTACCGACCTAACCTAAT

▶ V I G P L L F A V I Y N H S L P I W D G W I W I

TTGGTTTAGCGTTTTACTGTATTATTATCCTGCTATCGATGACCTTCATGTTAACCCTCAAGCTCAGGGGA
AACCAAATCGAAAATGACATAAATAAGGACGATAGCTACTGGAAGTACAATTGGGGAGTTCGAGTCCCTT

▶ I G L A F Y C I I I L L S M T F M L T P Q A Q G

GTAACAGGAGACAAGTGCTTAGTGAGGTGTTTGCTGGTGATTTGAACAATATGAGATAAAGCCCTCATGAC
CATTTGCTCTGTTACGAATCACTCCACAAACGACCACTAACTTGTTATACTCTATTTCTGGGAGTACTG

▶ S K Q E T S A

GAGGGCGTAACATTA CTC AGC AAT AAA CTG ATA TTC CGT CAG GCT GGA ATA CTC TTC
CTCCCGCATTGTAAT GAG TCG TTA TTT GAC TAT AAG GCA GTC CGA CCT TAT GAG AAG

▶ E A I F Q Y E T L S S Y E E

GCC AGG ACG CAG GAA GCA GTC CGG TTG CGG CCA TTC AGG GTG GTT CGG GCT GTC
CGG TCC TGC GTC CTT CGT CAG GCC AAC GCC GGT AAG TCC CAC CAA GCC CGA CAG

▶ G P R L F C D P Q P W E P H N P S D

CGG TAG AAA CTC GCT TTC CAG AGC CAG CCC TTG CCA GTC GGC GTA AGG TTC GGT
GCC ATC TTT GAG CGA AAG GTC TCG GTC GGG AAC GGT CAG CCG CAT TCC AAG CCA
P L F E S E L A L G Q W D A Y P E T

TCC CCG CGA CGG TGT GCC GCC GAG GAA GTT GCC GGA GTA GAA TTG CAG AGC CGG
AGG GGC GCT GCC ACA CGG CGG CTC CTT CAA CGG CCT CAT CTT AAC GTC TCG GCC
G R S P T G G L F N G S Y F Q L A P

AGC GGT GGT GTA GAC CTT CAG CTG CAA TTT TTC ATC TGC TGA CCA GAC ATG CGC
TCG CCA CCA CAT CTG GAA GTC GAC GTT AAA AAG TAG ACG ACT GGT CTG TAC GCG
A T T Y V K L Q L K E D A S W V H A

CGC CAC TTT CTT GCC ATC GCC TTT GGC CTG TAA CAA GAA TGC GTG ATC GTA ACC
GCG GTG AAA GAA CGG TAG CGG AAA CCG GAC ATT GTT CTT ACG CAC TAG CAT TGG
A V K K G D G K A Q L L F A H D Y G

TTT CAC TTT GCG CTG ATC GTC GTC GGC AAG AAA CTC ACT GGC GAT GAT TTT GGC
AAA GTG AAA CGC GAC TAG CAG CAG CCG TTC TTT GAG TGA CCG CTA CTA AAA CCG
K V K R Q D D D A L F E S A I I K A

GCT GCG GAA ATC AAA AGA CGT TCC GGC GAC AGA TTT CAG GCC GTC GTG CGG AAT
CGA CGC CTT TAG TTT TCT GCA AGG CCG CTG TCT AAA GTC CGG CAG CAC GCC TTA
S R F D F S T G A V S K L G D H P I

GCC GCC TTC ATC AAC CGG CAG ATA TTC GTC CGC CAG AAT CTG CAA CTT GTG ATT
CGG CGG AAG TAG TTG GCC GTC TAT AAG CAG GCG GTC TTA GAC GTT GAA CAC TAA
G G E D V P L Y E D A L I Q L K H N

GCG CAC GTC AGA CTG CTC GCC GTC AAG ATT GAA ATA GAC GTG ATT AGT CAT ATT
CGC GTG CAG TCT GAC GAG CGG CAG TTC TAA CTT TAT CTG CAC TAA TCA GTA TAA
R V D S Q E G D L N F Y V H N T M N

CAC CGG GCA AGG TTT ATC AAC TGT GGC GCG ATA AGT AAT GGA GAT ACG GTT ATC
GTG GCC CGT TCC AAA TAG TTG ACA CCG CGC TAT TCA TTA CCT CTA TGC CAA TAG
V P C P K D V T A R Y T I S I R N D

GTC GGT CAG ACG ATA TTG CAC CGT CGC GCC GAG ATT ACC CGG GAA GCC CTG ATC
CAG CCA GTC TGC TAT AAC GTG GCA GCG CGG CTC TAA TGG GCC CTT CGG GAC TAG
D T L R Y Q V T A G L N G P F G Q D

ACC ATC ATC TGA ACT CAG GGC AAA CAG CAC CTG ACG ATC GTT CTG GTT CAC AAT
TGG TAG TAG ACT TGA GTC CCG TTT GTC GTG GAC TGC TAG CAA GAC CAA GTG TTA
◀ G D D S S L A F L V Q R D N Q N V I

CTG CCA GCG ACG TTT GTC GAA CCC TTC CGG CCC GCC GTG CAG CTG GTT AAC GCC
GAC GGT CGC TGC AAA CAG CTT GGG AAG GCC GGG CGG CAC GTC GAC CAA TTG CGG
◀ Q W R R K D F G E P G G H L Q N V G

CTG ACT TGG CGA AAG CGT CAC GGT TTC ACC GTC AAA GGT ATA ACG GCT ATT GGC
GAC TGA ACC GCT TTC GCA GTG CCA AAG TGG CAG TTT CCA TAT TGC CGA TAA CCG
◀ Q S P S L T V T E G D F T Y R S N A

GAT ACG GTT GGC ATA ACG ACC AAT AGA GGC CCC CAG AAA CGC GGC CTG ATC CTG
CTA TGC CAA CCG TAT TGC TGG TTA TCT CCG GGG GTC TTT GCG CCG GAC TAG GAC
◀ I R N A Y R G I S A G L F A A Q D Q

ATA GCA TTC CGG GCT GGC ACA GCC GAG CAG CGC CTC GCG GAC GCT GCC ATC GGA
TAT CGT AAG GCC CGA CCG TGT CGG CTC GTC GCG GAG CGC CTG CGA CGG TAG CCT
◀ Y C E P S A C G L L A E R V S G D S

AAG CGG AAT ACG GGC GGA AAG TAA AGT CGC ACC CCA GTC CAT CAG CGT GAC TAC
TTC GCC TTA TGC CCG CCT TTC ATT TCA GCG TGG GGT CAG GTA GTC GCA CTG ATG
◀ L P I R A S L L T A G W D M L T V V

←GalM

CAT CCC TGC GTT GTT ACG CAA AGT TAA CAG TCG GTA CGG CTG ACC ATC GGG TGC
GTA GGG ACG CAA CAA TGC GTT TCA ATT GTC AGC CAT GCC GAC TGG TAG CCC ACG
◀

CAG TGC GGG AGT TTC GTT CA GCA CTG TCC TGC TCC TTG TGA TGG TTT ACA AAC
GTC ACG CCC TCA AAG CAA GT CGT GAC AGG ACG AGG AAC ACT ACC AAA TGT TTG

GTA AAA AGT CTC TTT AAT ACC TGT TTT TGC TTC ATA TTG TTC AGC GAC AGC TTG
CAT TTT TCA GAG AAA TTA TGG ACA AAA ACG AAG TAT AAC AAG TCG CTG TCG AAC

CTG TAC GGC AGG CAC CAG CTC TTC CGG GAT CAG CGC GAC GAT ACA GCC GCC AAA
GAC ATG CCG TCC GTG GTC GAG AAG GCC CTA GTC GCG CTG CTA TGT CGG CGG TTT

ACT TCT GTG CTG AAA CGC GCT ATC CAT ACC CTG TGG AAT GTG CTG GAC GAA CTG
TGA AGA CAC GAC TTT GCG CGA TAG GTA TGG GAC ACC TTA CAC GAC CTG CTT GAC
▶ T S V L K R A I H T L W N V L D E L

GAT CAG GCA TGG CTG CCC GTT GAG AAA TCC TGG AAA CTG AAC GAA CGT CAC TAC
CTA GTC CGT ACC GAC GGG CAA CTC TTT AGG ACC TTT GAC TTG CTT GCA GTG ATG
▶ D Q A W L P V E K S W K L N E R H Y

GGT GCG TTG CAG GGT CTG AAC
CCA CGC AAC GTC CCA GAC TTG
▶ G A L Q G L N

FigureS3

Galk->

CAGAACAGGCAGCAGAGCGTTTGC GCGCAGTCAGCGATATCCATTTTCGCGAATCCGGAGTGTAAGAA ATG
 GTCTTGTCCGTCGTCTCGCAAACGCGCGTCAGTCGCTATAGGTA AAAAGCGCTTAGGCCTCACATTCTT TAC

▶ M

AGT CTG AAA GAA AAA ACA CAA TCT CTG TTT GCC AAC GCA TTT GGC TAC CCT GCC
 TCA GAC TTT CTT TTT TGT GTT AGA GAC AAA CGG TTG CGT AAA CCG ATG GGA CGG
 ▶ S L K E K T Q S L F A N A F G Y P A

ACT CAC ACC ATT CAG GCG CCT GGC CGC GTG AAT TTG ATT GGT GAA CAC ACC GAC
 TGA GTG TGG TAA GTC CGC GGA CCG GCG CAC TTA AAC TAA CCA CTT GTG TGG CTG
 ▶ T H T I Q A P G R V N L I G E H T D

TAC AAC GAC GGT TTC GTT CTG CCC TGC GCG ATT GAT TAT CAA ACC GTG ATC AGT
 ATG TTG CTG CCA AAG CAA GAC GGG ACG CGC TAA CTA ATA GTT TGG CAC TAG TCA
 ▶ Y N D G F V L P C A I D Y Q T V I S

TGT GCA CCA CGC GAT GAC CGT AAA GTT CGC GTG ATG GCA GCC GAT TAT GAA AAT
 ACA CGT GGT GCG CTA CTG GCA TTT CAA GCG CAC TAC CGT CGG CTA ATA CTT TTA
 ▶ C A P R D D R K V R V M A A D Y E N

CAG CTC GAC GAG TTT TCC CTC GAT GCG CCC ATT GTC GCA CAT GAA AAC TAT CAA
 GTC GAG CTG CTC AAA AGG GAG CTA CGC GGG TAA CAG CGT GTA CTT TTG ATA GTT
 ▶ Q L D E F S L D A P I V A H E N Y Q

TGG GCT AAC TAC GTT CGT GGC GTG GTG AAA CAT CTG CAA CTG CGT AAC AAC AGC
 ACC CGA TTG ATG CAA GCA CCG CAC CAC TTT GTA GAC GTT GAC GCA TTG TTG TCG
 ▶ W A N Y V R G V V K H L Q L R N N S

TTC GGC GGC GTG GAC ATG GTG ATC AGC GGC AAT GTG CCG CAG GGT GCC GGG TTA
 AAG CCG CCG CAC CTG TAC CAC TAG TCG CCG TTA CAC GGC GTC CCA CGG CCC AAT
 ▶ F G G V D M V I S G N V P Q G A G L

AGT TCT TCC GCT TCA CTG GAA GTC GCG GTC GGA ACC GTA TTG CAG CAG CTT TAT
 TCA AGA AGG CGA AGT GAC CTT CAG CGC CAG CCT TGG CAT AAC GTC GTC GAA ATA
 ▶ S S S A S L E V A V G T V L Q Q L Y

CAT CTG CCG CTG GAC GGC GCA CAA ATC GCG CTT AAC GGT CAG GAA GCA GAA AAC
 GTA GAC GGC GAC CTG CCG CGT GTT TAG CGC GAA TTG CCA GTC CTT CGT CTT TTG
 ▶ H L P L D G A Q I A L N G Q E A E N

CAG TTT GTA GGC TGT AAC TGC GGG ATC ATG GAT CAG CTA ATT TCC GCG CTC GGC
 GTC AAA CAT CCG ACA TTG ACG CCC TAG TAC CTA GTC GAT TAA AGG CGC GAG CCG
 ▶ Q F V G C N C G I M D Q L I S A L G

AAG AAA GAT CAT GCC TTG CTG ATC GAT TGC CGC TCA CTG GGG ACC AAA GCA GTT
 TTC TTT CTA GTA CGG AAC GAC TAG CTA ACG GCG AGT GAC CCC TGG TTT CGT CAA
 ▶ K K D H A L L I D C R S L G T K A V

TCC ATG CCC AAA GGT GTG GCT GTC GTC ATC ATC AAC AGT AAC TTC AAA CGT ACC
AGG TAC GGG TTT CCA CAC CGA CAG CAG TAG TAG TTG TCA TTG AAG TTT GCA TGG
▶ S M P K G V A V V I I N S N F K R T

CTG GTT GGC AGC GAA TAC AAC ACC CGT CGT GAA CAG TGC GAA ACC GGT GCG CGT
GAC CAA CCG TCG CTT ATG TTG TGG GCA GCA CTT GTC ACG CTT TGG CCA CGC GCA
▶ L V G S E Y N T R R E Q C E T G A R

loxP

TTC TTC CAG CAG CCA GGC ATA ACT TCG TAT A+t GTA TGC TAT ACG AAG TTA TCC
AAG AAG GTC GTC GGT CGG TAT TGA AGC ATA TaA CAT ACG ATA TGC TTC AAT AGG
▶ F F Q Q P A I T S Y I V C Y T K L S

CTG CGT GAT GTC ACC ATT GAA GAG TTC AAC GCT GTT GCG CAT GAA CTG GAC CCG
GAC GCA CTA CAG TGG TAA CTT CTC AAG TTG CGA CAA CGC GTA CTT GAC CTG GGC
▶ L R D V T I E E F N A V A H E L D P

ATC GTG GCA AAA CGC GTG CGT CAT ATA CTG ACT GAA AAC GCC CGC ACC GTT GAA
TAG CAC CGT TTT GCG CAC GCA GTA TAT GAC TGA CTT TTG CGG GCG TGG CAA CTT
▶ I V A K R V R H I L T E N A R T V E

GCT GCC AGC GCG CTG GAG CAA GGC GAC CTG AAA CGT ATG GGC GAG TTG ATG GCG
CGA CGG TCG CGC GAC CTC GTT CCG CTG GAC TTT GCA TAC CCG CTC AAC TAC CGC
▶ A A S A L E Q G D L K R M G E L M A

GAG TCT CAT GCC TCT ATG CGC GAT GAT TTC GAA ATC ACC GTG CCG CAA ATT GAC
CTC AGA GTA CGG AGA TAC GCG CTA CTA AAG CTT TAG TGG CAC GGC GTT TAA CTG
▶ E S H A S M R D D F E I T V P Q I D

ACT CTG GTA GAA ATC GTC AAA GCT GTG ATT GGC GAC AAA GGT GGC GTA CGC ATG
TGA GAC CAT CTT TAG CAG TTT CGA CAC TAA CCG CTG TTT CCA CCG CAT GCG TAC
▶ T L V E I V K A V I G D K G G V R M

ACC GGC GGC GGA TTT GGC GGC TGT ATC GTC GCG CTG ATC CCG GAA GAG CTG GTG
TGG CCG CCG CCT AAA CCG CCG ACA TAG CAG CGC GAC TAG GGC CTT CTC GAC CAC
▶ T G G G F G G C I V A L I P E E L V

CCT GCC GTA CAG CAA GCT GTC GCT GAA CAA TAT GAA GCA AAA ACA GGT ATT AAA
GGA CGG CAT GTC GTT CGA CAG CGA CTT GTT ATA CTT CGT TTT TGT CCA TAA TTT
▶ P A V Q Q A V A E Q Y E A K T G I K

GAG ACT TTT TAC GTT TGT AAA CCA TCA CAA GGA GCA GGA CAG TGC TG AAC GAA
CTC TGA AAA ATG CAA ACA TTT GGT AGT GTT CCT CGT CCT GTC ACG AC TTG CTT
▶ E T F Y V C K P S Q G A G Q C

ACT CCC GCA CTG GCA CCC GAT GGT CAG CCG TAC CGA CTG TTA ACT TTG CGT AAC
TGA GGG CGT GAC CGT GGG CTA CCA GTC GGC ATG GCT GAC AAT TGA AAC GCA TTG

GalM-->

AAC GCA GGG ATG GTA GTC ACG CTG ATG GAC TGG GGT GCG ACT TTA CTT TCC GCC
TTG CGT CCC TAC CAT CAG TGC GAC TAC CTG ACC CCA CGC TGA AAT GAA AGG CGG
▶ M V V T L M D W G A T L L S A

CGT ATT CCG CTT TCC GAT GGC AGC GTC CGC GAG GCG CTG CTC GGC TGT GCC AGC
GCA TAA GGC GAA AGG CTA CCG TCG CAG GCG CTC CGC GAC GAG CCG ACA CGG TCG
▶ R I P L S D G S V R E A L L G C A S

CCG GAA TGC TAT CAG GAT CAG GCC GCG TTT CTG GGG GCC TCT ATT GGT CGT TAT
GGC CTT ACG ATA GTC CTA GTC CGG CGC AAA GAC CCC CGG AGA TAA CCA GCA ATA
▶ P E C Y Q D Q A A F L G A S I G R Y

GCC AAC CGT ATC GCC AAT AGC CGT TAT ACC TTT GAC GGT GAA ACC GTG ACG CTT
CGG TTG GCA TAG CGG TTA TCG GCA ATA TGG AAA CTG CCA CTT TGG CAC TGC GAA
▶ A N R I A N S R Y T F D G E T V T L

TCG CCA AGT CAG GGC GTT AAC CAG CTG CAC GGC GGG CCG GAA GGG TTC GAC AAA
AGC GGT TCA GTC CCG CAA TTG GTC GAC GTG CCG CCC GGC CTT CCC AAG CTG TTT
▶ S P S Q G V N Q L H G G P E G F D K

CGT CGC TGG CAG ATT GTG AAC CAG AAC GAT CGT CAG GTG CTG TTT GCC CTG AGT
GCA GCG ACC GTC TAA CAC TTG GTC TTG CTA GCA GTC CAC GAC AAA CGG GAC TCA
▶ R R W Q I V N Q N D R Q V L F A L S

TCA GAT GAT GGT GAT CAG GGC TTC CCG GGT AAT CTC GGC GCG ACG GTG CAA TAT
AGT CTA CTA CCA CTA GTC CCG AAG GGC CCA TTA GAG CCG CGC TGC CAC GTT ATA
▶ S D D G D Q G F P G N L G A T V Q Y

CGT CTG ACC GAC GAT AAC CGT ATC TCC ATT ACT TAT CGC GCC ACA GTT GAT AAA
GCA GAC TGG CTG CTA TTG GCA TAG AGG TAA TGA ATA GCG CGG TGT CAA CTA TTT
▶ R L T D D N R I S I T Y R A T V D K

CCT TGC CCG GTG AAT ATG ACT AAT CAC GTC TAT TTC AAT CTT GAC GGC GAG CAG
GGA ACG GGC CAC TTA TAC TGA TTA GTG CAG ATA AAG TTA GAA CTG CCG CTC GTC
▶ P C P V N M T N H V Y F N L D G E Q

TCT GAC GTG CGC AAT CAC AAG TTG CAG ATT CTG GCG GAC GAA TAT CTG CCG GTT
AGA CTG CAC GCG TTA GTG TTC AAC GTC TAA GAC CGC CTG CTT ATA GAC GGC CAA
▶ S D V R N H K L Q I L A D E Y L P V

GAT GAA GGC GGC ATT CCG CAC GAC GGC CTG AAA TCT GTC GCC GGA ACG TCT TTT
CTA CTT CCG CCG TAA GGC GTG CTG CCG GAC TTT AGA CAG CGG CCT TGC AGA AAA
▶ D E G G I P H D G L K S V A G T S F

GAT TTC CGC AGC GCC AAA ATC ATC GCC AGT GAG TTT CTT GCC GAC GAC GAT CAG
CTA AAG GCG TCG CGG TTT TAG TAG CGG TCA CTC AAA GAA CGG CTG CTG CTA GTC
▶ D F R S A K I I A S E F L A D D D Q

CGC AAA GTG AAA GGT TAC GAT CAC GCA TTC TTG TTA CAG GCC AAA GGC GAT GGC
GCG TTT CAC TTT CCA ATG CTA GTG CGT AAG AAC AAT GTC CGG TTT CCG CTA CCG
▶ R K V K G Y D H A F L L Q A K G D G

AAG AAA GTG GCG GCG CAT GTC TGG TCA GCA GAT GAA AAA TTG CAG CTG AAG GTC
TTC TTT CAC CGC CGC GTA CAG ACC AGT CGT CTA CTT TTT AAC GTC GAC TTC CAG
▶ K K V A A H V W S A D E K L Q L K V

TAC ACC ACC GCT CCG GCT CTG CAA TTC TAC TCC GGC AAC TTC CTC GGC GGC ACA
ATG TGG TGG CGA GGC CGA GAC GTT AAG ATG AGG CCG TTG AAG GAG CCG CCG TGT
▶ Y T T A P A L Q F Y S G N F L G G T

CCG TCG CGG GGA ACC GAA CCT TAC GCC GAC TGG CAA GGG CTG GCT CTG GAA AGC
GGC AGC GCC CCT TGG CTT GGA ATG CGG CTG ACC GTT CCC GAC CGA GAC CTT TCG
▶ P S R G T E P Y A D W Q G L A L E S

GAG TTT CTA CCG GAC AGC CCG AAC CAC CCT GAA TGG CCG CAA CCG GAC TGC TTC
CTC AAA GAT GGC CTG TCG GGC TTG GTG GGA CTT ACC GGC GTT GGC CTG ACG AAG
▶ E F L P D S P N H P E W P Q P D C F

CTG CGT CCT GGC GAA GAG TAT TCC AGC CTG ACG GAA TAT CAG TTT ATT GCT GAG
GAC GCA GGA CCG CTT CTC ATA AGG TCG GAC TGC CTT ATA GTC AAA TAA CGA CTC
▶ L R P G E E Y S S L T E Y Q F I A E

TAATGTTACGCCCTCGTCATGAGGGCTTTATCTCATATTGTTCAAATCACCAGCAAACACCTCACTAAGCAC
ATTACAATGCGGGAGCAGTACTCCCGAAATAGAGTATAACAAGTTTAGTGGTCGTTTGTGGAGTGATTTCGTG
◀ A S

TTGTCTCCTGTTTACTCCCCTGAGCTTGAGGGGTTAACATGAAGGTCATCGATAGCAGGATAATAATACAGT
AACAGAGGACAAATGAGGGGACTCGAACTCCCAATTGTAATTCCAGTAGCTATCGTCCTATTATTATGTCA
◀ T E Q K S G Q A Q P T L M F T M S L L I I I C Y

AAAACGCTAAACCAATAATCCAAATCCAGCCATCCCAAATTGGTAGTGAATGATTATAAATAACAGCAAACA
TTTTGCGATTTGTTATTAGGTTTAGGTCGGTAGGGTTTAAACCATCACTTACTAATATTTATTGTCGTTTGT
◀ F A L G I I W I W G D W I P L S H N Y I V A F L

GTAATGGGCAATAACACCGGTTGCATTGGTAAGGCTCACCAATAATCCCTGTAAAGCACCTTGCTGATGAC
CATTACCCGGTATTGTGGCCAACGTAACCATTCCGAGTGGTATTAGGGACATTTTCGTGGAACGACTACTG
◀ L P G I V G T A N T L S V L L G Q L A G Q Q H S

TCTTTGTTTGGATAGACATCACTCCCTGTAATGCAGGTAAGCGATCCCACCACCAGCCAATAAAATTA
AGAAACAAACCTATCTGTAGTGAGGGACATTACGTCCATTTTCGCTAGGGTGGTGGTCGTTATTTTAATTTT
◀ K T Q I S M V G Q L A P L A I G G G A L L I L V

CAGGGAAAATAACCAACCTTCAGATATAAACGCTAAAAAGGCAAATGCACTACTATCTGCAATAAATCCGA
GTCCCTTTTGGATTGGTTGGAAGTCTATATTTGCGATTTTTCGTTTACGTGATGATAGACGTTATTTAGGCT
◀ P F V L W G E S I F A L F A F A S S D A I F G L

GCAGTACTGCCGTTTTTTTCGCCCCATTTAGTGGCTATTCTTCCTGCCACAAAGGCTTGAATACTGAGTGTA
CGTCATGACGGCAAAAAAGCGGGGTAATCACCGATAAGAAGGACGGTGTTCGGAACCTTATGACTCATAT
◀ L V A T K E G W K T A I R G A V F A Q F V S H L

AAAGACCAAGACCCGCTAATGAAAAGCCAACCATCATGCTATTCCATCCAAAACGATTTTTCGGTAAATAGCA
TTTCTGGTTCTGGGCGATTACTTTTTCGGTTGGTAGTACGATAAGGTAGTTTTGCTAAAAGCCATTTATCGT
◀ L G L G A L S F G V M M S N W G F R N E T F L V

CCCACACCGTTGCGGGAATTTGGCCTATCAATTGCGCTGAAAAATAAATAATCAACAAAATGGGCATCGTTT
GGGTGTGGCAACGCCCTTAAACCGGATAGTTAACGCGACTTTTTTATTATTAGTTGTTTTACCCGTAGCAAA
◀ W V T A P I Q G I L Q A S F Y I I L L I P M T K

TAAATAAAGTGATGTATACCGAATTCGATTGCGTCTCAACCCTACTTCGGTATCTGTATTATCACGTGTAT
ATTTATTTCACTACATATGGCTTAAGCTAACGCAGAGTTGGGGATGAAGCCATAGACATAATAGTGACATA
◀ F L T I Y V S N S Q T E V G V E T D T N D R T N

TTTTGGTTTCACGGAACCAAAACATAACCACAAGGAAAGCGACAATATTTAGCAACGCAGCGATAAAAAAGG
AAAACCAAAGTGCCTTGGTTTTGTATTGGTGTTCCTTTGCTGTTATAAATCGTTGCGTGCCTATTTTTTCC

◀ K T E R F W F M V V L F A V I N L L A A I F F P

GACTATGCGGTGAAATCTCTCCTGCAAACCACCAATAATAGGCCCGCTATTAACCAAGCCCAAACCTTG
CTGATACGCCACTTTAGAGAGGACGTTTTGGTGGTTATTATCCGGGCGATAATTTGGTTCGGGTTTTGAAC

◀ S H P S I E G A F G G I I P G A I L G L G F S A

CCCCTAACCAACCGAACCCTTCACGCGTTGAGAAGCTGAGGTGGTATCGGCAATGACCGATGCCGCGACAG
GGGGATTGGTTGGCTTGGTGAAGTGCGCAACTCTTCGACTCCACCATAGCCGTTACTGGCTACGGCGCTGTC

◀ G L W G F W K V R Q S A S T T D A I V S A A V A

CCCCAGTAGCTCCTGTGATCCCTGAAAGCAAACGGCCTAAATACAGCATCCAAAGCGCACTTGAAAAAGCCA
GGGGTCATCGAGGACACTAGGGACTTTCGTTTGCCGATTTATGTCGTAGGTTTCGCGTGAACCTTTTTCGTT

◀ G T A G T I G S L L R G L Y L M W L A S S F A L

GCAATAAGTAATCCAGCGATGCGCCTATTAATGACAACAACAGCACTGGGCGCCGACCAAATCGGTCAGACA
CGTTATTCATTAGGTCGCTACGCGGATAATTACTGTTGTTGTCGTGACCCGCGGCTGGTTTAGCCAGTCTGT

◀ L L Y D L S A G I L S L L L V P R R G F R D S M

TTTTTCCAAGCCAAGGAGCAAAGATAACCTGCATTAACGCATAAAGTGCAAGCAATACGCCAAAGTGTTAG
AAAAAGGTTGCGTTCCTCGTTTCTATTGGACGTAATTGCGTATTCACGTTGTTATGCGGTTTCACCAATC

◀ K G L W P A F I V Q M L A Y L A L L V G F H N A

CGATATCTTCCGAAGCAATAAATTCACGTAATAACGTTGGCAAGACTGGCATGATAAGGCCAATCCCCATGG
GCTATAGAAGGCTTCGTTATTTAAGTGCATTATTGCAACCGTTCTGACCGTACTATTCCGGTTAGGGGTACC

◀ I D E S A I F E R L L T P L V P M I L G I G M A

<--TetA

CATCGAGTAACGTAATTACCAATGCGATCTTTGTCGAACTATTCAATCCAGCTCCAGCGTGCTCCGTCTGAT
GTAGCTCATTGCATTAATGGTTACGCTAGAAACAGCTTGATAAGTAAGGTCGAGGTTCGCACGAGGCAGACTA

◀ D L L T I V L A I K T S S N M

loxP

ACGTGGGATCCATCG CAA TAA CTT CGT ATA GCA TAC ATT ATA CGA AGT TAT CTGCAGAT
TGCACCCTAGGTAGC GTT ATT GAA GCA TAT CGT ATG TAA TAT GCT TCA ATA GACGTCTA

TTGCAACTCAATATTCACAACAACCTTACACTGCGCCACTATTTTTGCTATGGTTATGCGTAAGCATTGCTG
AACGTTGAGTTATAAGTGTGTTGGAATGTGACGCGGTGATAAAAGCGATACCAATACGCATTGTAACGAC

GpmA-->

TTGCTTCGTCGCGCAATATAATGAGAATTATTATCATTAAAAGATGATTTGAGGAGTAAGTAT ATG GCT
AACGAAGCAGCGCCGTTATATTACTCTTAATAATAGTAATTTTCTACTAAACTCCTCATTTCATA TAC CGA

▶ M A

GTA ACT AAG CTG GTT CTG GTT CGT CAT GGC GAA AGT CAG TGG AAC AAA GAA AAC
CAT TGA TTC GAC CAA GAC CAA GCA GTA CCG CTT TCA GTC ACC TTG TTT CTT TTG
▶ V T K L V L V R H G E S Q W N K E N

CGT TTC ACC GGT TGG TAC GAC GTG GAT CTG TCT GAG AAA GGC GTA AGC GAA GCA
GCA AAG TGG CCA ACC ATG CTG CAC CTA GAC AGA CTC TTT CCG CAT TCG CTT CGT
▶ R F T G W Y D V D L S E K G V S E A

AAA GCA GCA GGT AAG CTG CTG AAA GAG GAA GGT TAC AGC TTT GAC TTT GCT TAC
TTT CGT CGT CCA TTC GAC GAC TTT CTC CTT CCA ATG TCG AAA CTG AAA CGA ATG
▶ K A A G K L L K E E G Y S F D F A Y

ACT TCT GTG CTG AAA CGC GCT ATC CAT ACC CTG TGG AAT GTG CTG GAC GAA CTG
TGA AGA CAC GAC TTT GCG CGA TAG GTA TGG GAC ACC TTA CAC GAC CTG CTT GAC
▶ T S V L K R A I H T L W N V L D E L

GAT CAG GCA TGG CTG CCC GTT GAG AAA TCC TGG AAA CTG AAC GAA CGT CAC TAC
CTA GTC CGT ACC GAC GGG CAA CTC TTT AGG ACC TTT GAC TTG CTT GCA GTG ATG
▶ D Q A W L P V E K S W K L N E R H Y

GGT GCG TTG CAG GGT CTG AAC
CCA CGC AAC GTC CCA GAC TTG
▶ G A L Q G L N

Figure S4

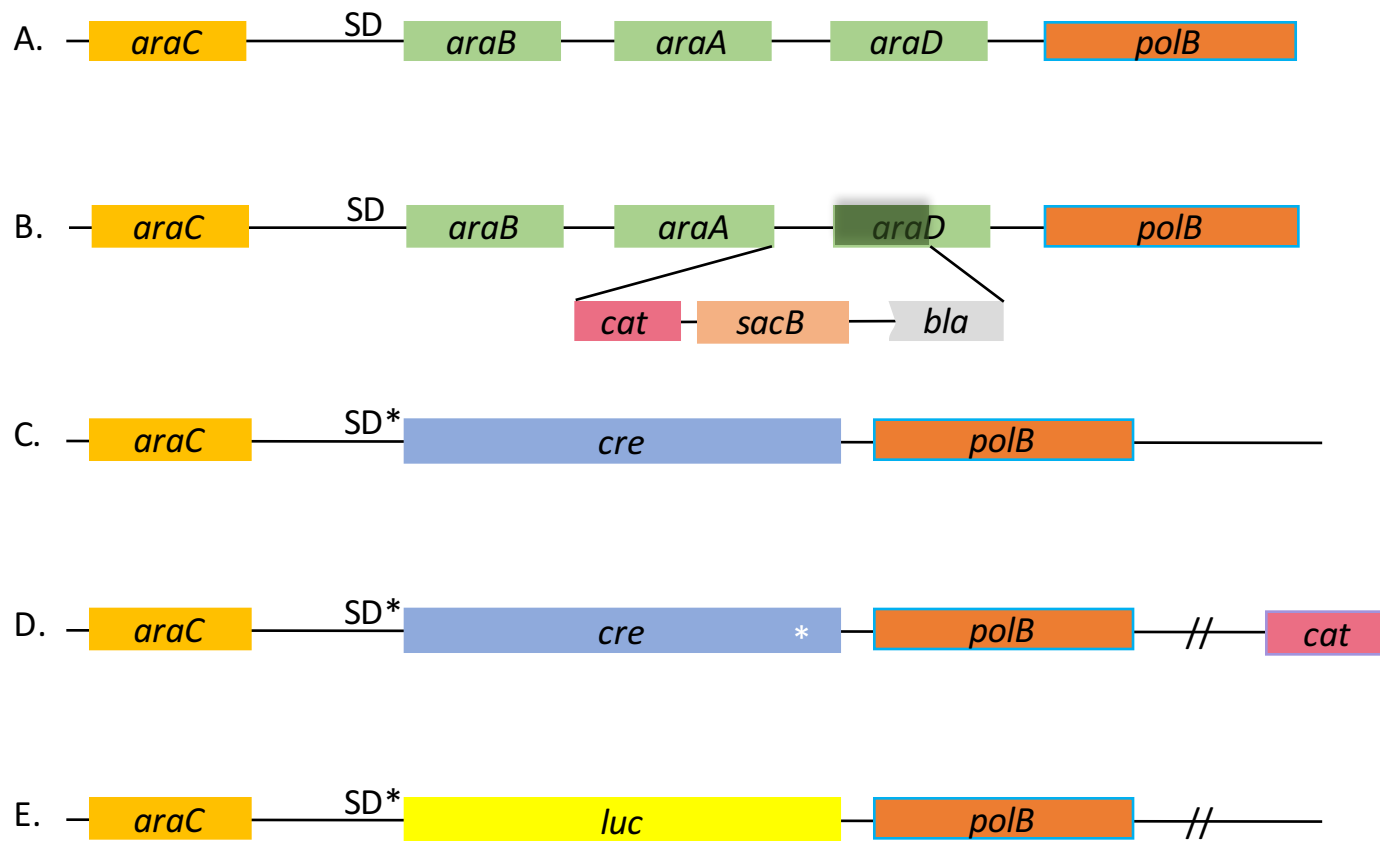


Figure S5

AAGTGTCTATAATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTAC
 TTCACAGATATTAGTCCGCTTTTTAGGTGTAATAAAACGTGCCGAGTGTAAACGATACGGTATCGTAAAAATAGGTATTCTAATCGCCTAGGATG

P_{BAD} *SD** *cre*

CTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGATAGGAGGATGAAACG ATG TCG AAT TTA CTG ACC GTA CAC
 GACTGCGAAAAATAGCGTTGAGAGATGACAAAGAGGTATGGGCAAAAAACCTATCCTCCTACTTTGC TAC AGC TTA AAT GAC TGG CAT GTG

▶ M S N L L T V H

CAA AAT TTG CCT GCA TTA CCG GTC GAT GCA ACG AGT GAT GAG GTT CGC AAG AAC CTG ATG GAC ATG TTC AGG GAT
 GTT TTA AAC GGA CGT AAT GGC CAG CTA CGT TGC TCA CTA CTC CAA GCG TTC TTG GAC TAC CTG TAC AAG TCC CTA
 ▶ Q N L P A L P V D A T S D E V R K N L M D M F R D

CGC CAG GCG TTT TCT GAG CAT ACC TGG AAA ATG CTT CTG TCC GTT TGC CGG TCG TGG GCG GCA TGG TGC AAG TTG
 GCG GTC CGC AAA AGA CTC GTA TGG ACC TTT TAC GAA GAC AGG CAA ACG GCC AGC ACC CGC CGT ACC ACG TTC AAC
 ▶ R Q A F S E H T W K M L L S V C R S W A A W C K L

AAT AAC CGG AAA TGG TTT CCC GCA GAA CCT GAA GAT GTT NGC GAT TAT CTT CTA TAT CTT CAG GCG CGC GGT CTG
 TTA TTG GCC TTT ACC AAA GGG CGT CTT GGA CTT CTA CAA NCG CTA ATA GAA GAT ATA GAA GTC CGC GCG CCA GAC
 ▶ N N R K W F P A E P E D V - D Y L L Y L Q A R G L

GCA GTA AAA ACT ATC CAG CAA CAT TAT TTG GGC CAG CTA AAC ATG CTT CAT CGT CGG TCC GGG CTG CCA CGA CCA
 CGT CAT TTT TGA TAG GTC GTT GTA ATA AAC CCG GTC GAT TTG TAC GAA GTA GCA GCC AGG CCC GAC GGT GCT GGT
 ▶ A V K T I Q Q H Y L G Q L N M L H R R S G L P R P

AGT GAC AGC AAT GCT GTT TCA CTG GTT ATG CGG CGG ATC CGA AAA GAA AAC GTT GAT GCC GGT GAA CGT GCA AAA
 TCA CTG TCG TTA CGA CAA AGT GAC CAA TAC GCC GCC TAG GCT TTT CTT TTG CAA CTA CGG CCA CTT GCA CGT TTT
 ▶ S D S N A V S L V M R R I R K E N V D A G E R A K

CAG GCT CTA GCG TTC GAA CGC ACT GAT TTC GAC CAG GTT CGT TCA CTC ATG GAA AAT AGC GAT CGC TGC CAG GAT
 GTC CGA GAT CGC AAG CTT GCG TGA CTA AAG CTG GTC CAA GCA AGT GAG TAC CTT TTA TCG CTA GCG ACG GTC CTA
 ▶ Q A L A F E R T D F D Q V R S L M E N S D R C Q D

ATA CGT AAT CTG GCA TTT CTG GGG ATT GCT TAT AAC ACC CTG TTA CGT ATA GCC GAA ATT GCC AGG ATC AGG GTT
 TAT GCA TTA GAC CGT AAA GAC CCC TAA CGA ATA TTG TGG GAC AAT GCA TAT CGG CTT TAA CGG TCC TAG TCC CAA
 ▶ I R N L A F L G I A Y N T L L R I A E I A R I R V

AAA GAT ATC TCA CGT ACT GAC GGT GGG AGA ATG TTA ATC CAT ATT GGC AGA ACG AAA ACG CTG GTT AGC ACC GCA
 TTT CTA TAG AGT GCA TGA CTG CCA CCC TCT TAC AAT TAG GTA TAA CCG TCT TGC TTT TGC GAC CAA TCG TGG CGT
 ▶ K D I S R T D G G R M L I H I G R T K T L V S T A

GGT GTA GAG AAG GCA CTT AGC CTG GGG GTA ACT AAA CTG GTC GAG CGA TGG ATT TCC GTC TCT GGT GTA GCT GAT
 CCA CAT CTC TTC CGT GAA TCG GAC CCC CAT TGA TTT GAC CAG CTC GCT ACC TAA AGG CAG AGA CCA CAT CGA CTA
 ▶ G V E K A L S L G V T K L V E R W I S V S G V A D

XT592

GAT CCG AAT AAC TAC CTG TTT TGC CGG GTC AGA AAA AAT GGT GTT GCC GCG CCA TCT GCC ACC AGC CAG CTA TCA
 CTA GGC TTA TTG ATG GAC AAA ACG GCC CAG TCT TTT TTA CCA CAA CGG CGC GGT AGA CGG TGG TCG GTC GAT AGT
 ▶ D P N N Y L F C R V R K N G V A A P S A T S Q L S

ACT CGC GCC CTG GAA GGG ATT TTT GAA GCA ACT CAT CGA TTG ATT TAC GGC GCT AAG GAT GAC TCT GGT CAG AGA
 TGA GCG CGG GAC CTT CCC TAA AAA CTT CGT TGA GTA GCT AAC TAA ATG CCG CGA TTC CTA CTG AGA CCA GTC TCT
 ▶ T R A L E G I F E A T H R L I Y G A K D D S G Q R

TAC CTG GCC TGG TCT GGA CAC AGT GCC CGT GTC GGA GCC GCG CGA GAT ATG GCC CGC GCT GGA GTT TCA ATA CCG
 ATG GAC CGG ACC AGA CCT GTG TCA CGG GCA CAG CCT CGG CGC GCT CTA TAC CGG GCG CGA CCT CAA AGT TAT GGC
 ▶ Y L A W S G H S A R V G A A R D M A R A G V S I P

GAG ATC ATG CAA GCT GGT GGC TGG ACC AAT GTA AAT ATT GTC ATG AAC TAT ATC CGT AAC CTG GAT AGT GAA ACA
CTC TAG TAC GTT CGA CCA CCG ACC TGG TTA CAT TTA TAA CAG TAC TTG ATA TAG GCA TTG GAC CTA TCA CTT TGT
▶ E I M Q A G G W T N V N I V M N Y I R N L D S E T

araD downstream

GGG GCA ATG GTG CGC CTG CTG GAA GAT GGC GAT TAG TGACTGTATAAAACCACAGCCAATCAAACGAAACCAGGCTATACTCAAGCCTGG
CCC CGT TAC CAC GCG GAC GAC CTT CTA CCG CTA ATC ACTGACATATTTTGGTGTGCGTTAGTTTGCTTTGGTCCGATATGAGTTCGGACC
▶ G A M V R L L E D G D



polB

TTTTTTGATGGAATTACAGC GTG GCG CAG GCA GGT TTT ATC TTA ACC CGA CAC TGG CGG GAC ACC CCG CAA GGG ACA GAA
AAAAAACTACCTTAATGTCG CAC CGC GTC CGT CCA AAA TAG AAT TGG GCT GTG ACC GCC CTG TGG GGC GTT CCC TGT CTT
▶ V A Q A G F I L T R H W R D T P Q G T E

GTC TCC TTC TGG CTG GCG ACG GAC AAC GGG CC
CAG AGG AAG ACC GAC CGC TGC CTG TTG CCC GG
▶ V S F W L A T D N G



Figure S6

