Supplementary Information

Experimental procedures

Identifying the viable circular permutation sites of *E. coli* EPSPS

Bioinformatics prediction and *E. coli* complementation assay were performed together to identify the viable circular permutation sites of EPSPS. First, Cpred, a web server (http://sarst.life.nthu.edu.tw/CPred)[1] was used to predict viable circular permutation sites of EPSPS (PDB ID: 2AA9). Only circular permutation sites with high scores (> 0.7) were taken into consideration. For several adjacent sites with scores all > 0.7, only the site with the highest score was selected. In addition, sites which were close to N termini (< 50) or C termini (> 400) were not considered given the resulted variants might be similar to wild type EPSPS. According to this principle, 18 circular permutation sites (Table S1) were finally selected for further study. A tandem fusion template of *aroA* gene was created for PCR cloning of all the predicted cpEPSPS variants. The first copy of *aroA* gene was PCR amplified by forward primer *aroA*-P1 (CTACGAA-TCATGGAATCCCTGACGTTAC) and reverse primer *aroA*-P2 (TAGTGGATCCGCCGCC-ACCGGCTGCCTGGCTAATCCG) with *E. coli* MG1655 strain genome as the template. PCR product was inserted into the EcoRI site and BamHI sites of pUC19 to get pUC19-*aroA*-1. The second copy of *aroA* gene was amplified with forward primer *aroA*-P3 (GATC-TAGGATCCGGCGGTGGAGGCTCAGGTGGTGGAGGCTCAATGGAATCCCTGACGTTAC) and reverse primer *aroA*-P4 (CTTAGAAGCTTGGCTGCCTGGCTAATCCG) using the same template of the first copy. The resulted PCR product was inserted into the BamHI site and HindIII sites of pUC19-*aroA*-1 to obtain pUC19-*aroA*-2. Thus pUC19-*aroA*-2 contained two copies of *aroA* ORF with a sequence (GGTGGCGGCGGATCCGGCGGTGGAGGCTCAGGTGGTGGA-GGCTCA) which codes for a $(GGGGS)_3$ linker connecting them. ORFs of all predicted cpEPSPS variants were generated by PCR amplification of the pUC19-*aroA*-2 with forward primers and reverse primers corresponding to their new termini. A start codon and a stop codon were introduced into the forward primers and the reverse primers, respectively. The PCR products of cpEPSPS variants were cloned between the NcoI site and SalI sites of a modified pACYC184 derivative vector pKU8001 (pACYC184 with a NcoI site introduced at the translation start site of the *tet* gene, the former NcoI site inside *cat* gene of pACYC184 was eliminated by PCR point mutation). All the cpEPSPS variants were controlled by the constitutive *tet* promoter and were transformed into the *E. coli aroA* deleted strain BD2100[2] to test whether they could restore its growth in M63 minimal medium.

Measurement of growth rate and enzyme specific activity

To measure the growth rate of BD2100 harboring different constructs in M63 minimal medium, cells were first grown in LB broth at 37 °C to stationary phase as seed culture and collected by centrifugation. Cell pellets were then diluted 1:100 into fresh M63 liquid medium and grew to stationary phase as pre-cultures. The pre-cultures were harvested and retransferred to fresh M63 liquid medium with initial $OD_{600} \approx 0.03$. Cell samples were taken at least four times to measure OD_{600} during exponential phase (typically at OD_{600} between 0.1 and 0.5). Doubling times of cells were obtained by plotting the OD_{600} against growing time.

For measurement of crude enzyme activity, BD2100 harboring different constructs were grown at 37 °C in 80 mL LB or M63 minimal medium to stationary phase. Cells were centrifuged at 12000 rpm for 5 min and the cell pellets were resuspended at 20 mL 50 mM Tris-HCl buffer (pH 7.5, with 0.1 mM DTT). Cells were then subjected to ultrasonication and centrifuged at 12000 rpm for 10min to obtain the supernatant. Ammonia sulfate was added to precipitate the supernatant and the fraction between 35% and 70% saturation of ammonia sulfate was harvested. The precipitate was dissolved in 50 mM Tris-HCl buffer (pH 7.5, with 0.1 mM DTT) and dialyzed overnight against the same buffer to get crude enzyme sample. EPSPS activity was determined by measuring phosphate release rate using the malachite green dye assay method as previously described.[2,3]

Protein expression and purification

ORFs of WT EPSPS and the seven selected cpEPSPS variants were cloned between the NcoI site and XhoI sites of pET28a. All the constructs were transformed into *E. coli* BL21(DE3) pLyS strain (Invitrogen) for protein purification. For details: Cells were grown overnight at LB broth containing 25 μg/mL kanamycin as pre-cultures. Pre-cultures were then transferred into (1:100 dilution) 500 mL LB broth containing 25 μg/mL kanamycin and incubated at 37 °C (220 rpm) until $OD_{600} \approx 0.6$; 0.5 mM IPTG was added and protein expression was induced at 18 °C (220 rpm) for 20 h. Cultures were collected by centrifugation at 8000 rpm for 15min at 4 °C. The cell pellets were resuspended at Buffer A (pH 8.0, HEPES 20 mM, NaCl 500 mM, Glycerol 5%, 0.1 mM DTT) and lysed by ultrasonication. Cell debris was removed at 13000 rpm for 1 h at 4 °C. Proteins were first purified by a 5 mL size HisTrap Ni^{+} affinity column (GE-healthcare) using Buffer B (pH 8.0, HEPES 20 mM, NaCl 500 mM, Glycerol 5%, Imidazole 500 mM, 0.1 mM DTT). The resulted proteins were further purified through a Superdex-200 gel filtration chromatography column (GE-healthcare) with elution of Buffer C (pH 8.0, HEPES 20 mM, NaCl 200 mM, Glycerol 5%, 0.1 mM DTT). Purified proteins were observed in 12% SDS-PAGE.

Kinetic analysis of enzymes

For kinetic analysis, enzyme concentration was fixed at 20 nM. Enzyme substrate affinity K_m(PEP) and K_m(S3P) were determined by Lineweaver-Burk method as previously described.[3] For measuring K_m(PEP), S3P concentration was fixed at 1 mM and PEP concentration was varied from 30 to 300 μM. For measuring K_m(S3P), PEP concentration was fixed at 1 mM and S3P concentration was varied from 30 to 300 μM. Enzyme catalytic constant k_{cat} was obtained through division of maximal reaction rate (V_{max}) by enzyme concentration. Glyphosate inhibition constant K_i was determined by Dixon plot as previously described[4] with little modifications. In our assay, S3P concentration was fixed at 1 mM. PEP concentration was varied from 50 to 200 μM and glyphosate concentration was varied from 2 to 50 μM.

Circular dichroism (CD) spectroscopy of enzymes

UV-CD spectra were obtained using a MOS-450/AF-CD spectrometer（Bio-Logic corporate, France）. Protein samples were dialyzed overnight at 4 °C against 20 mM potassium phosphate buffer (pH 8.0). For far-UV CD measurement, the protein samples (200 μg/mL) were loaded into a 0.1 cm path-length cell. Ellipticities were measured from 200 nm to 250 nm. For near-UV CD measurement, the protein samples (2 mg/mL) were loaded into a 1 cm path-length cell. Elipticities were measured from 260 nm to 320 nm. Other parameters included: Time constant: 1 s; Scan rate: 30 nm/min; Numbers of Scan: 3 times, which were common for both far-UV CD and near-UV CD. CD raw data were converted to mean residue ellipticity $[\theta]_{mrw}$ given by: $[\theta]_{mrw} = MRW \times \theta_\lambda / 10 \times$

d × c, where θ_λ is the observed ellipticity (degrees) at wavelength λ, d is the path-length (cm), and c is the concentration (g/ml).[5]

Enzyme thermodynamic analysis with DSC

DSC data were obtained on a CSC model 6300 Nano III differential scanning calorimeter (Calorimetry Sciences Corp, Lindon, UT). Protein samples (1.5 mg/mL) were dissolved in 20 mM potassium phosphate buffer. Sample volume was kept at 584 μL for all the proteins. DSC experiments were performed from 30 °C to 90 °C at a heating rate of 0.5 °C /min. A constant pressure of 3 atm was exerted to prevent bubble formation during the whole process. The DSC raw data were analyzed by NanoAnalyze software.

Establishment of protein fragment complementation (PFC) for EPSPS

To establish PFC system for EPSPS, GCN4 leucine zipper was fused with the EPSPS fragments split at seven selected circular permutation sites. For details: coding sequence for *Saccharomyces cerevisiae* GCN4 basic leucine zipper domain (residue 236 to 281 of GCN4 protein) was chemically synthesized and cloned between the NcoI and BamHI sites of both a pBR322 derivatives vector pKU8000 (pBR322 with a NcoI site introduced at the translation start site of the *tet* gene by point mutation) and the pACYC184 derivative vector pKU8001. The resulted vectors were inserted by EPSPS fragments between BamHI and SalI sites with N-terminal fragments at pKU8000 vector and C-terminal fragments at pKU8001 vector. A 15aa $(GGGGS)_3$ linker coding sequence was in frame added to the forward primer of EPSPS fragments. Thus a leucine zipper was fused to the N terminus of all the EPSPS fragments. The N- and C-terminal fragments without leucine zippers were cloned directly into the NcoI and SalI sites of pKU8000 and pKU8001 respectively. The Reverse primers for N-terminal fragments and the forward primers of C-terminal fragments were added with a stop codon and a start codon respectively. All the fragments were thus driven by the constitutive *tet* promoter of pKU8000 and pKU8001. The N- and C-terminal fragment constructs at seven split sites were co-transformed into BD2100 strain for survival assay and enzyme activity measurement. To express leucine zipper-fused fragment pairs by T7 promoter, all the N-terminal fragments fused with leucine zipper were cloned between NcoI and SalI sites of the pET28a vector. To create a T7 expression vector which is compatible with pET28a, the T7 expression cassette between BglII site and SalI site of pET28a were inserted into the BamHI site (BamHI and BglII is a pair of isocaudamers) and SalI site of the pACYC184 derivative vector pKU8001 to get vector pKU8002. All the C-terminal fragments fused with leucine zipper were cloned between NcoI and SalI sites of the pKU8002 vector. Thus all the seven leucine zipper-fused fragment pairs were driven by T7 promoter. They were transformed into BD2100 strain (It is the BL21(DE3) *aroA* deleted strain). As a positive control, wild type *aroA* gene was also inserted between NcoI and SalI sites of the pKU8002 vector and co-transformed with pET28a vector into BD2100 strain. All the T7 constructs were induced with 100 μM IPTG.

Reference:

1 W. C. Lo, L. F. Wang, Y. Y. Liu, T. Dai, J. K. Hwang, P. C. Lyu, *Nucleic Acids Res*, 2012, 40, W232–7.

2 Y. Li, Y. C. Sun, H. Q. Yan, Y. P. Wang, *Chinese Sci Bull*, 2001, 56, 514–520.

3 Y. C. Sun, Y. C. Chen, Z. X. Tian, F. M. Li, X.Y. Wang, J. Zhang, Z. L. Xiao, M. Lin, N. Gilmartin, D. N. Dowling and Y. P. Wang, *Appl. Environ. Microbiol.*, 2005, 71, 4771–4776.

4 Y. S. Tian, A. S. Xiong, J. Xu, W. Zhao, F. Gao, X. Y. Fu, H. Xu, J. L. Zheng, R. H. Peng and Q. H. Yao, *Appl. Environ. Microbiol.*, 2010, 76, 6001–6005.

5 S. M. Kelly, T. J. Jess, N. C. Price, *Biochim. Biophys. Acta.*, 2005, 1751, 119–139.

Figures and Tables

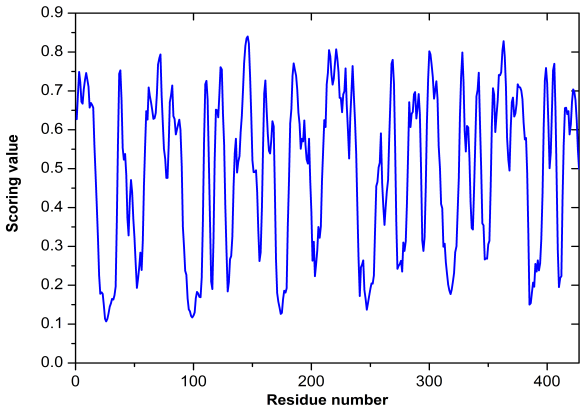

Fig. S1 Prediction of circular permutation sites for *E. coli* EPSPS through CPred. The structure information of *E. coli* EPSPS (PDB ID: 2AA9) was submitted to Cpred server. X axis denotes each residue of *E. coli* EPSPS. Y axis denotes the corresponding scoring value of each residue. The scoring values range from 0 to 1. The higher the value, with greater likelihood the residue is a viable circular permutation site.

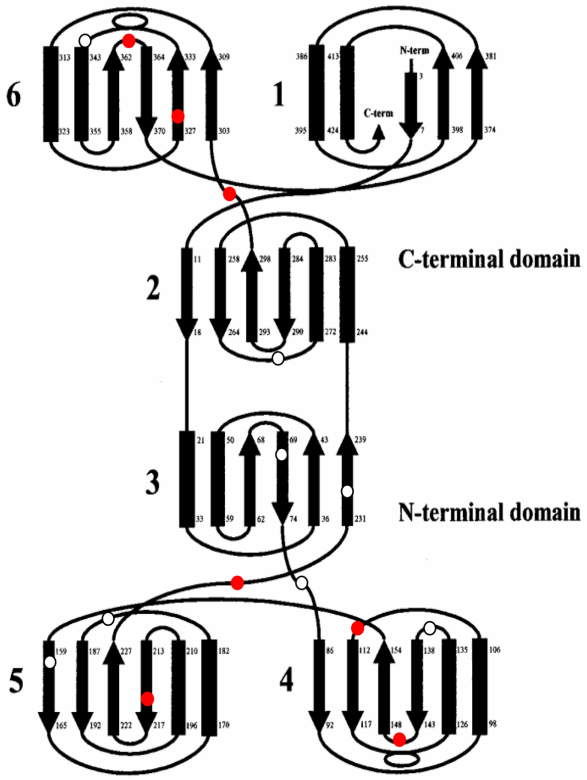

Fig. S2 Viable circular permutation sites in *E. coli* EPSPS. The new N termini of viable cpEPSPS variants were labeled in the schematic diagram of *E. coli* EPSPS. Red close circles represent new N termini of seven selected variants that subjected to detailed analysis and white open circles represent N termini of the rest variants.

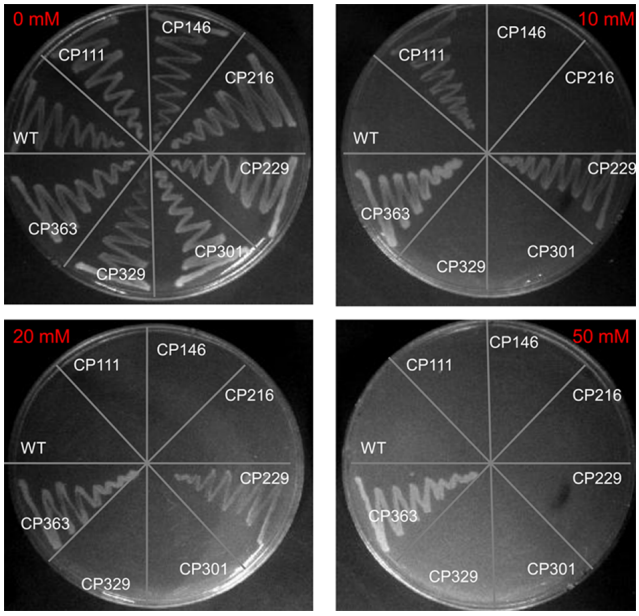

Fig. S3 Survival assay of BD2100 harboring WT EPSPS or the seven selected cpEPSPS variants in M63 minimal medium supplemented with varied concentrations of glyphosate. Cells were grown at 37 °C for 48 h.

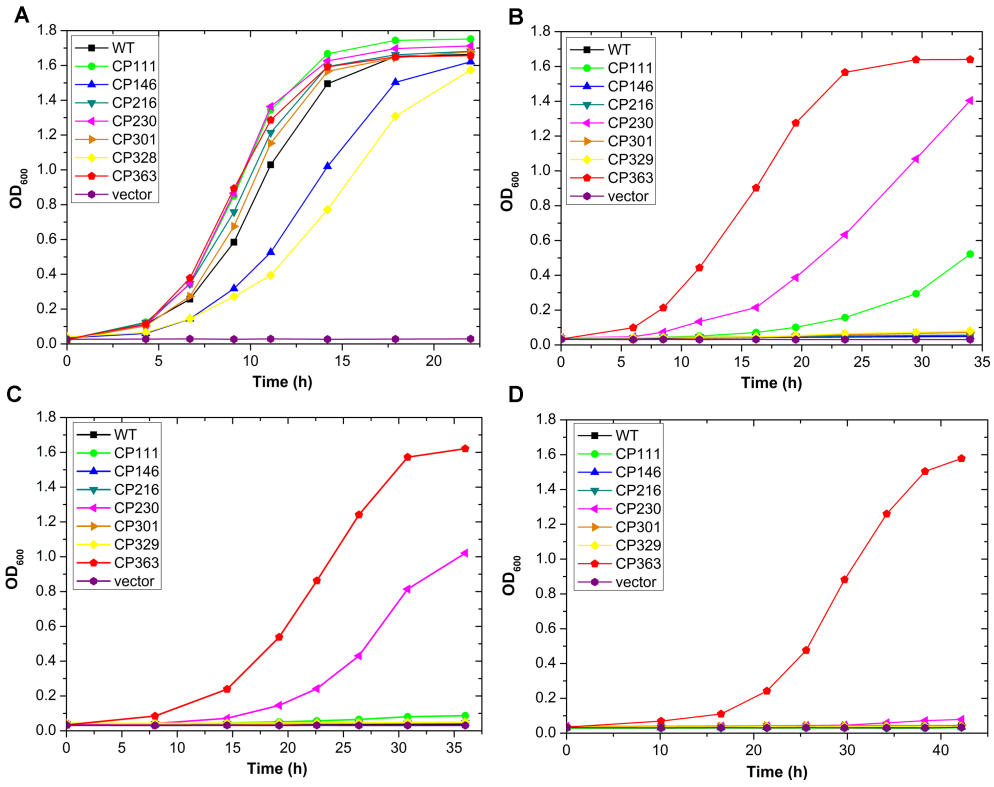

Fig. S4 Growth curve of BD2100 harboring WT EPSPS or the seven selected cpEPSPS variants in liquid M63 minimal medium added with varied concentrations of glyphosate, (A) 0 mM, (B) 10 mM, (C) 20 mM, (D) 50 mM. Cells were grown at 37 °C (220 rpm).

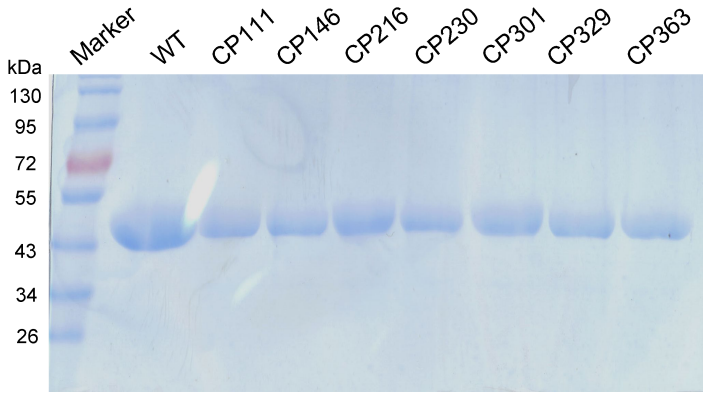

Fig. S5 Purified WT EPSPS and the seven selected cpEPSPS variants. Proteins were observed in 12% SDS-PAGE. The molecular size should be about 46.8 kDa for WT EPSPS and 47.9 kDa for the seven cpEPSPS variants due to a extra $(GGGGS)_3$ linker.

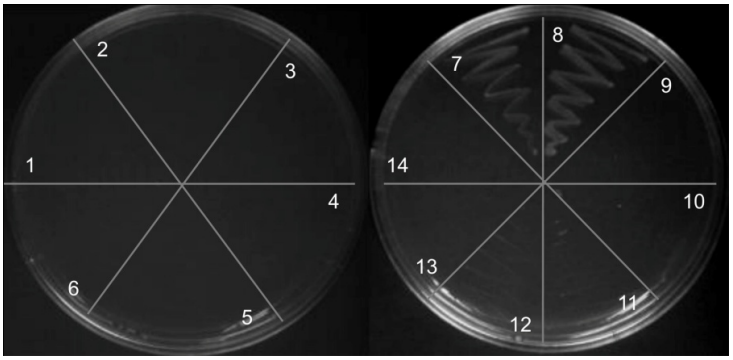

Fig. S6 Survival assay of *E.coli* BD2100 harboring N- and C-terminal fragments with only one carried leucine zipper in M63 minimal medium: (1) Z-N110/C111; (2) N110/Z-C111; (3) Z-N145/C146; (4) N145/Z-C146; (5) Z-N215/C216; (6) N215/Z-C216; (7) Z-N229/C230; (8) N229/Z-C230; (9) Z-N300/C301; (10) N300/Z-C301; (11) Z-N328/C329; (12) N328/Z-C329; (13) Z-N362/C363; (14) N362 /Z-C363. "Z" denotes leucine zipper, e.g. Z-N110 means leucine zipper fused to the N terminus of N110 fragment.

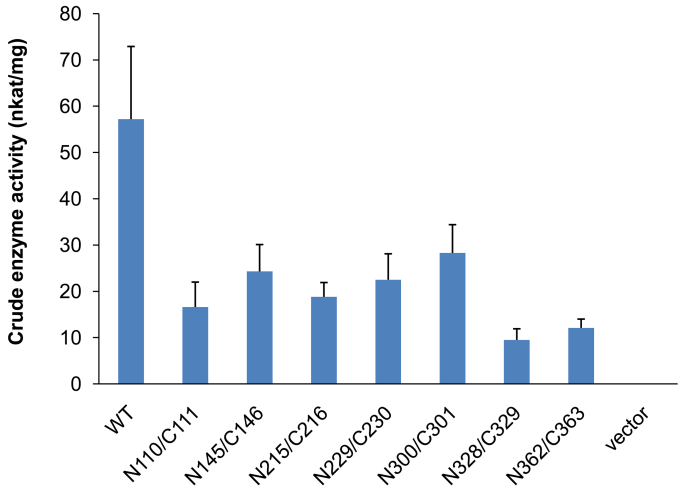

Fig. S7 Crude enzyme activity of BD2100 harboring the WT EPSPS and the seven leucine zipper-fused fragment pairs expressed by the inducible T7 promoter. Cells were grown in LB broth supplemented with 100 μM IPTG at 37 °C (220 rpm).

Enzyme variants	Circular permutation site [a]	New termini region [b]	Doubling time [c] (min)
CP72	T71-72R	Inside β strand	495
CP83	L82-83H	loop between folding units	378
CP111	S110-N111	loop between α helix and β strand	88
CP124	E123-124R	loop between α helix and β strand	/
CP137	G136-137G	loop between α helix and β strand	593
CP146	Q145-E146	loop between β strands	122
CP161	G160-N161	Inside β strand	210
CP186	E185-D186	loop between α helix and β strand	224
CP216	I215-E216	Inside β strand	91
CP230	Q229-S230	loop between folding units	75
CP235	P234-G235	Inside β strand	321
CP269	N268-269S	loop between α helix and β strand	352
CP289	C288-289W	Inside β strand	/
CP301	E300-L301	loop between folding units	79
CP306	D305-W306	Inside β strand	/
CP329	T328-T329	Inside β strand	169
CP343	T342-D343	loop between α helix and β strand	226
CP363	G362-H363	loop between β strands	94
WT			97

Table S1 All the viable cpEPSPS variants identified in our study. a. Circular permutation sites refer to the new termini of the circular permutants (e.g. S110-N111 means the new N termini and C termini of CP111 variants is N111 and S110 of the wild type enzyme respectively). b. New termini region refers to the locations of new N termini at the structure of WT enzyme. **c.** Doubling time of BD2100 harboring corresponding cpEPSPS variants was measured in M63 liquid minimal medium. "/" denotes no growth.

Enzymes	T_m (°C)	ΔH_{cal} (kcal/mol)	ΔS_m (cal/mol•k)
WT	56.9	118.8	360
CP111	56.3	118.3	359
CP146	55.1	123.6	376
CP216	50.6	96.2	297
CP230	44.7	125.7	395
CP301	50.4	116.8	361
CP329	47.2	91.2	285
CP363	52.9	106.5	328

Table S2 DSC thermodynamic parameter of the eight purified proteins. T_m, transition midpoint temperature; ΔH_{cal}, calorimetric enthalpy; ΔS_m, entropy of unfolding.

	with leucine zippers	without leucine zippers
N110/C111	114	/
N145/C146	83	/
N215/C216	231	/
N229/C230	77	237
N300/C301	71	/
N328/C329	133	/
N362/C363	288	/

Table S3 Doubling time (min) of BD2100 harboring seven fragment pairs either with or without leucine zippers in liquid M63 minimal medium at 37 °C (220 rpm). " / " denotes no growth.