Supporting Information

Sandoval et al. 10.1073/pnas.1206299109

SI Selection Model with Varying Epistasis

$$P_{tot} = \sum_{n=1}^{5} P_{n0}\left[\sum_{r=1}^{n}\binom{n}{r}\alpha^{r}(1-\alpha)^{n-r}e^{\left(\mu_r f_{antag}^{r}\cdot t\right)}\right]$$

P_{tot} is the total population, P_{n0} is the initial population of those with n mutations (based on laboratory efficiencies), n is the number of mutations in a clone, r is number of beneficial mutations in a clone, α is the fraction of mutations that are beneficial, μ_r is the growth rate with r beneficial mutations, f_{antag} is the antagonistic factor, and t is time.

The calculation of μ_r is as follows:

$$\mu_r = \mu_{WT}\left(f_{increase}\cdot f_{synergy}\cdot r + 1\right)$$

where μ_{WT} is the wild-type growth rate (set to 0.5 1/h), f_{increase} is the fractional increase in growth rate of one beneficial mutation (set to 0.35), f_{synergy} is the synergy factor, r is the number of beneficial mutations in a clone (see here below).

Condition	Synergy factor (f_{synergy})	Antagonistic factor (f_{antag})
Synergistic	1.1	1.0
Additive	1.0	1.0
Less than additive	0.9	1.0
Antagonistic	0.9	0.85 for $n > 1$, 1 for $n = 1$

SI Materials and Methods

Construction of the RMR Acetate Library. For the acetate library, eight targets were chosen from the highest fitness TRMR alleles (Table S1). In this library construction, 11 rounds of recombination were performed with the pool of 8 oligonucleotides. The replacement efficiency was estimated to be 4.4% per round, yielding a population that has 31% single mutants, 7.1% double mutants, and 1.0% triple mutants. The library underwent serial transfer selection in 10 g/L acetate for 110 h, and the growth rate increased by 75%; none of the 23 colonies picked exhibited an increase in growth rate. Four colonies were selected and had each of their 8 target regions genotyped (32 total reactions), but no mutations were found.

TRMR Sequencing and Microarray Analysis. Final and preselection cultures were harvested for microarray analysis and aliquots of final selection cultures were sampled onto solid LB plates for colony picking. One billion cells (10^9) per sample were taken to ensure a representative population. TRMR DNA barcodes were amplified from extracted genomic DNA via PCR as previously described (Promega Wizard Genomic DNA Purification kit or Invitrogen PureLink Genomic DNA kit) (1). The barcode DNA was purified after being run on an agarose gel via a gel extraction kit (Qiagen). From each sample, 600 ng of the sample barcode tags were applied to the Geneflex Tag4 16K V2 array (Affymetrix), and the resulting data were analyzed as described previously (1).

Control tags were supplemented in known quantities so that calculations of the concentration of sample barcode tags might be made from the array signal. The concentration of each allele was divided by the total amount of sample DNA applied to the array to calculate the frequency of each allele. The fitness of each allele (W) was calculated by dividing the postselection frequency of the allele by the preselection frequency of the allele.

Individual clones were genotyped by amplifying the barcode region of the insert via PCR and then subsequently sequenced conventionally.

Strain Construction and Library Construction of Mutated RBS for Low pH Targets. The pSIM5 plasmid was kindly provided by Donald Court. pRETMG was constructed for integration of red recombinase genes into the genome of *E. coli* MG1655. The *recE* and *recT* genes were amplified from *E. coli* MG1655 genome. Red-gam gene was modified using recursive PCR to avoid homologous recombination with the same gene in λ-red. *recE* and *recT* and modified gam gene were cloned into pKD46 to yield pRETMG.

The λ-red recombinase genes for chromosomal integration were amplified from pSIM5 using two 90-mer primers with 60-bp homologous sequence targeting between *ybhB* and *ybhC* as λ phage integration site. *E. coli* MG1655 containing pRETMG was grown at 30 °C to an OD_{600} of ∼0.6 in SOB medium and 10 mM L-arabinose. Cells were made electrocompetent and electroporated with a mixture of 50 μL of cells and 1 μg of PCR product of red recombinase genes. Cells recovered for 2.5 h in SOC medium and plated. Selected colonies were confirmed by PCR amplification and sequencing on integration site of red recombinase gene. Cells were cured of the helper plasmid, pRETMG, by serial cultures without ampicillin, and antibiotic sensitivity was confirmed. This strain was named MG1655LR.

For the construction of *mutS* deletion mutant, MG1655LR was grown at 30 °C to OD_{600} of ∼0.6, and red recombinase was induced by heating at 42 °C for 16 min in a water bath. DNA cassette for deletion of *mutS* was obtained from pKD13 by PCR amplification using primers with 50 bp homologous sequence targeting *mutS* ORF. After electroporation and incubation for 2.5 h at 30 °C, cells were spread on agar with kanamycin (25 μg/mL). Several selected colonies were checked by PCR and sequencing of the deletion of *mutS*. Finally, constructed strain was named as MG1655LRM.

Growth Studies. Stationary phase overnight cultures were used for a 2.5% inoculation of 5 mL morpholinopropane sulfonate (MOPS) minimal medium in a 15-mL centrifuge tube. These cultures were monitored until the optical density at 600 nm (OD_{600}) reached 0.200 ± 0.01. A 4% inoculation was introduced into the growth test medium. The OD was subsequently monitored. For growth studies involving hydrolysate, 1 mL of culture was centrifuged, decanted, and resuspended in water or minimal media before the OD was observed. OD was observed by either a spectrophotometer for flasks and conical tubes or a 96-well microplate reader (Biotek). Strains showing improved tolerance in 96-well microplate reader were analyzed again in flasks.

Design of Oligonucleotides. The oligo pools had an eight-base degenerate sequence five bases upstream of the start codon of the target. The hydrolysate oligos were designed with eight base pairs of degeneracy with the general sequence of 5′-DDDRRDRRD-3′ for up mutations and 5′- NNNYYNYYN-3′ for down mutations (where D represents A, G, or T; Y represents C or T; and R represents A or G). One base was changed in each oligo to create a C-C mismatch, which was been shown to increase recombination efficiency (2). The acetate up allele oligo pools

were designed with the sequence 5′-DDDRDRRD-3′ (where D represents A, G, or T; and R represents A or G). The acetate down allele oligo pools were designed with the completely degenerate sequence of 5′-NNNNNNNN-3′ (3). On either side of the degenerate sequence is a 41-base homology region specific to the target (4). The first four 5′ bases are linked with phosphorothioate bonds to reduce single-stranded exonuclease activity (4). The oligonucleotide was designed so that the sequence was that of the lagging strand, which has been shown to increase recombination efficiency (2, 5, 6).

Oligos were designed for 14 alleles as found in the low pH TRMR library screening under acidic condition (Table S1). Each has 40 bp of homology targeting the lagging strand of genome and total length of 95 bp for up-regulation including artificial RBS sequence (AAGAGGTATATATTA) and 80 bp for down-regulation with removed RBS and start codon. Each oligo was constructed with two phosphothioate bonds at the 5′ end.

The SIMD 70 strain has a $galK_{tyr145UAG}$ mutation that disables the strain from metabolizing galactose. An oligonucleotide that corrects this mutation to code for a functioning galK gene (mutated region 5′-. . .CAACTATATCACCTA. . .-3′; corresponding with oligo 478 in ref. 4) was used in conjunction with MacConkey agar plates with 1% (mass/vol) galactose to test for recombination efficiency.

1. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* 28: 856–862.
2. Costantino N, Court DL (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci USA* 100:15748–15753.
3. Wang HH, et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460:894–898.
4. Sawitzke JA, et al. (2011) Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J Mol Biol* 407:45–59.
5. Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci USA* 98:6742–6746.
6. Li XT, et al. (2003) Identification of factors influencing strand bias in oligonucleotide-mediated recombination in Escherichia coli. *Nucleic Acids Res* 31:6674–6687.

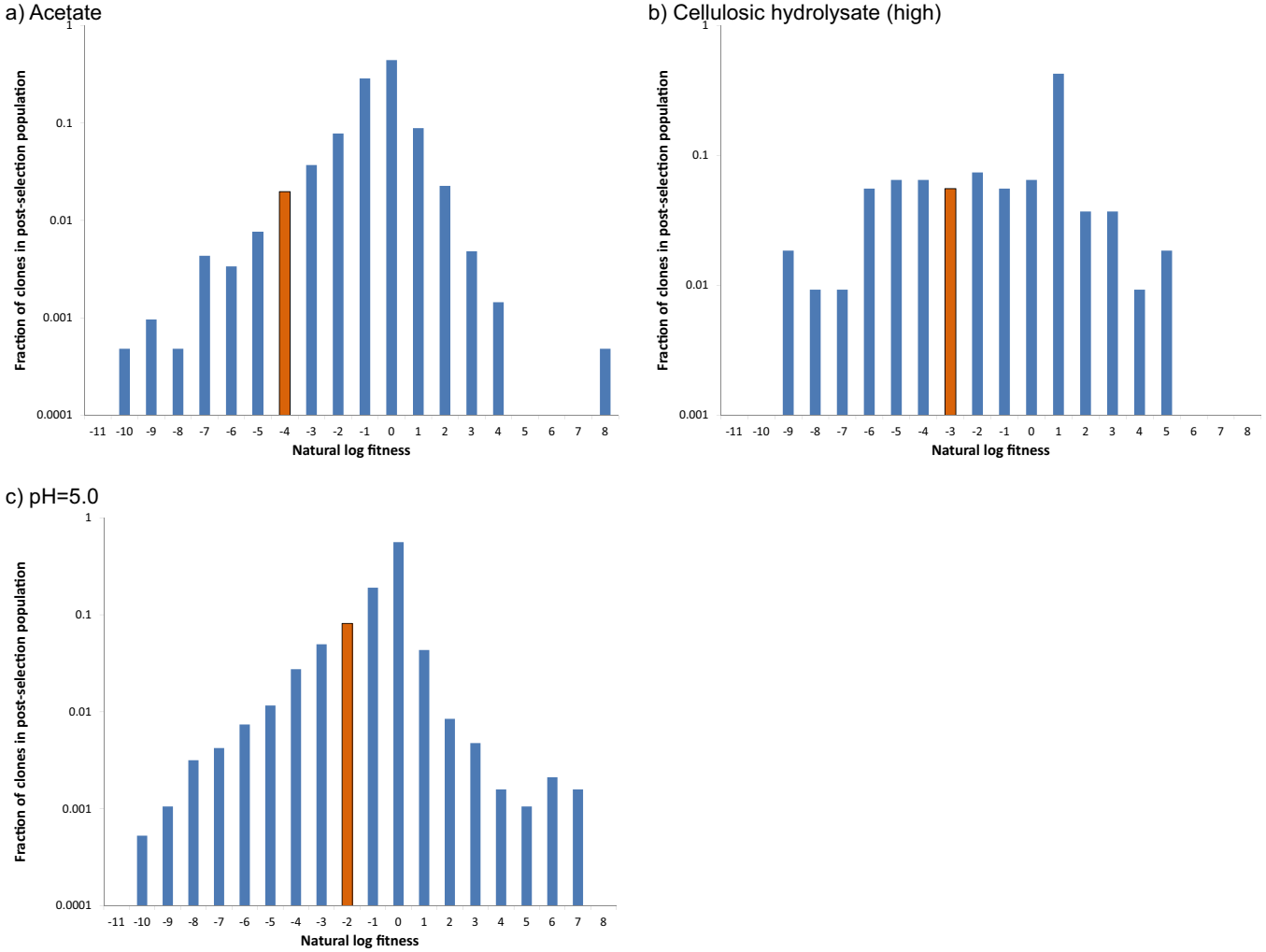


Fig. S1. Histograms of postselection population clones highlighting control strain fitness. Fraction of postselection population is plotted in bins of range 1 natural log fitness (lnW). The bin in which the neutral control strain JWKAN is contained is colored orange and is outlined in black. Note the fraction of population plotted logarithmically. (*A*) Acetate TRMR selection (2,078 clones). (*B*) High cellulosic hydrolysate TRMR selection (108 clones). (*C*) Low pH TRMR selection (1,893 clones).

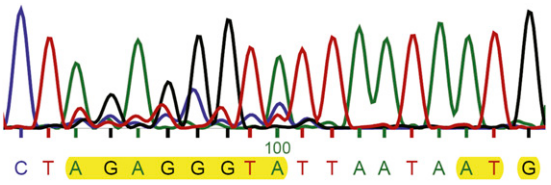


Fig. S2. Chromatograph of the *lpp* RBS region of the final hydrolysate selection population. Diversity of nucleotide sequence can be observed in the region targeted by the ssDNA recombination oligonucleotide with a degenerate sequence.

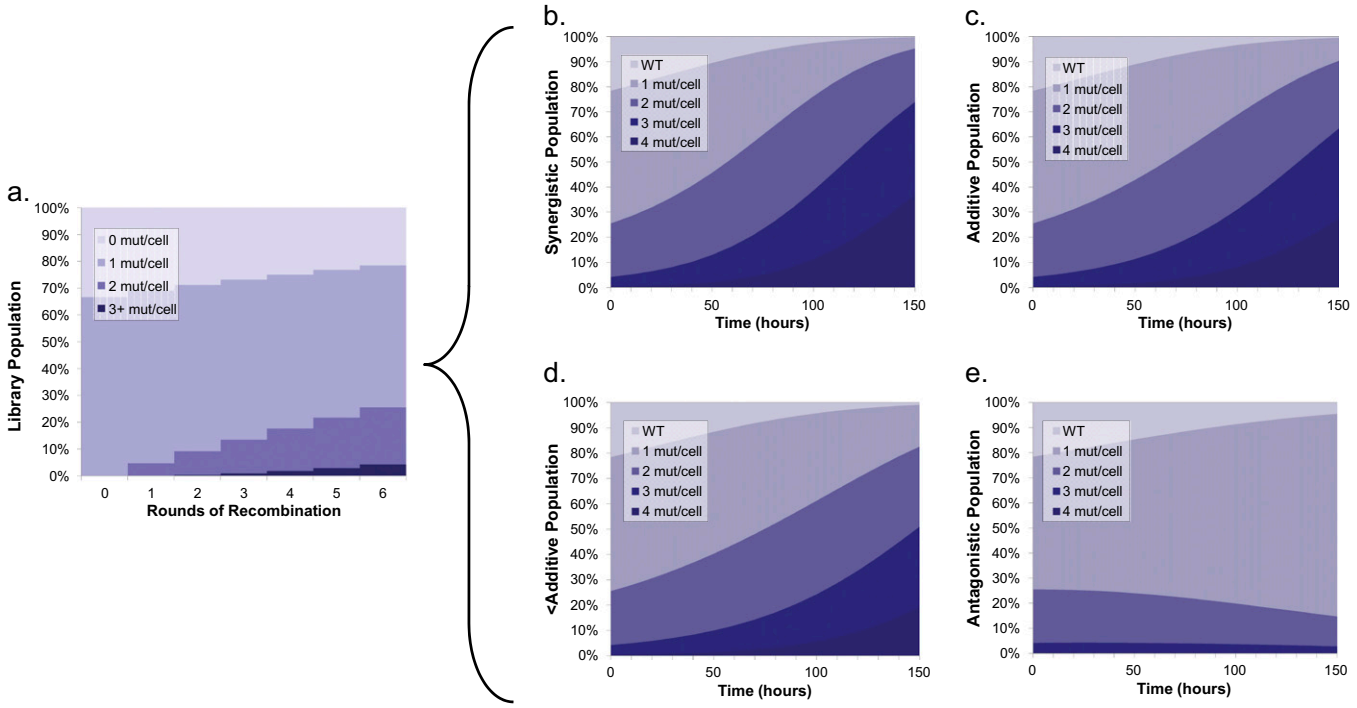
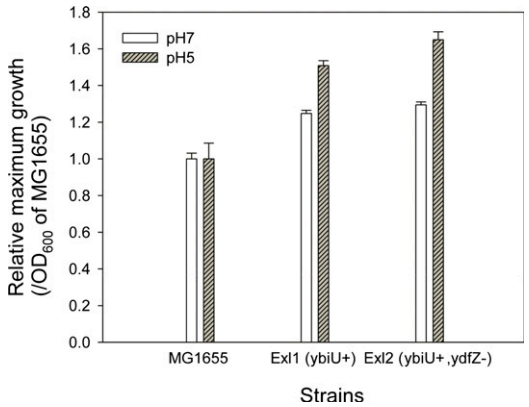


Fig. S3. Model of second recursive multiplex recombineering library growth. (*A*) Construction of the library. Shown is the theoretical population distribution of the library where recombination efficiency is 7.0%. After six rounds of recombination, single, double, and triple mutants represent 52.9%, 21.3%, and 3.9%, of the total library populations, respectively. (*B*–*E*) Four cases of varying epistasis in a growth selection. WT growth rate was set to 0.05 1/h (typical for 40% hydrolysate growth). Mutations were modeled to be either beneficial (35% increase in growth rate over control; 10% of mutations) or neutral (no change in growth rate; 90% of mutations). (*B*) Synergistic: combinations of mutations increase in growth 10% more than additive. (*C*) Additive: benefits of individual mutations are additive. (*D*) Less-than-additive: combinations 10% less than additive. (*E*) Antagonistic: combinations of mutations reduce growth rate by 15% compared with individual mutations.




```
ybiU up      CGAATGAAAGTAAGAGGTATATATTAATGGCT
                       |  |  |  |  |  ||||||
Wild Type    CGAATGAAAGTTAATGAGAGGTGGTTATGGCT

ydfZ down    TGTAAATCAGA----------ACG
                       |          |||
Wild Type    TGTAAATCAGAAAGGAGAAAACACCTATGACG
```

Fig. S4. Recursive multiplex recombineering low pH selection isolates. (*A*) Relative growth over 16 h. (*B*) Mutations in RBS of tolerant clones.

Table S1. Targets for RMR

Condition and gene	TRMR lnW	Gene function
18–20% hydrolysate		
cyaA down	5.65	Adenylate cyclase
yciV down	5.17	Conserved protein
ahpC up	4.61	Alkylhydroperoxide reductase component
eutL up	3.77	Predicted structural protein, ethanolamine utilization microcompartment
ygjQ up	3.48	Conserved protein
ydjG down	3.14	Methylglyoxal reductase (NADH-dependent)
rseP down	2.85	Zinc protease, signal peptide peptidase
lpp down	2.45	Murein lipoprotein
talA up	2.26	Transaldolase A in pentose phosphate pathway
motA up	2.02	Proton conductor component of motor
lacZ down	1.89	β-Galactosidase
tonB up	1.66	Energy transducer; uptake of iron, cyanocobalimin
nanC up	1.52	*N*-acetylneuraminic acid outer membrane channel
ygaZ up	1.51	L-valine LIV-E exporter
yedQ up	1.5	Predicted diguanylate cyclase
ybaB down	1.37	Conserved DNA-binding protein
ugpE down	1.29	Inner membrane polypeptide
15–17% hydrolysate		
moeA up	6.94	Molybdopterin biosynthesis
ptsI up	6.45	Phosphotransferase system enzyme I
cyaA down	6.27	Adenylate cyclase
ilvM up	5.08	Acetohydroxyacid synthase II component
ybjN down	4.52	Predicted oxidoreductase
lolC up	4.16	Predicted membrane component of the LolCDE lipoprotein transporter
ybaB up	4.16	Conserved DNA-binding protein
agaS up	4.12	Putative tagatose-6-phosphate aldose/ketose isomerase
ptsI down	3.24	Phosphotransferase system enzyme I
puuE down	2.91	4-Aminobutyrate aminotransferase
lpcA up	—	D-sedoheptulose 7-phosphate isomerase
talB down	—	Transaldolase B in pentose phosphate pathway
ydaG down	—	Rac prophage; small protein
Low pH (5)		
yiiQ down	1.02	Conserved protein
fliE down	0.99	Flagellar basal-body protein FliE
rcsD up	0.90	RcsD phosphotransferase
paaX up	0.80	DNA-binding transcriptional repressor
ybiU up	0.76	Predicted protein
elaD up	0.49	Deubiquitinase
envR up	0.48	DNA-binding transcriptional repressor
ydfZ down	0.47	Conserved protein
yciS down	0.34	Conserved inner membrane protein
nsrR down	0.20	DNA-binding transcriptional repressor
cspE down	0.18	Transcription antiterminator and regulator of RNA stability
fecI up	−0.01	RNA polymerase, sigma 19 factor
adiC up	−0.10	Arginine:agmatine antiporter
fepB up	−0.69	Component of: ferric enterobactin ABC transporter
16 g/L Acetate		
serS down	8.29	Seryl-tRNA synthetase
tap up	4.63	Inner membrane polypeptide
yahF down	4.36	Predicted acyl-CoA synthetase with NAD(P)-binding domain and succinyl-CoA synthetase domain
deoA down	3.83	Thymidine phosphorylase/uracil phosphorylase
tqsA down	3.18	Quorum signal AI-2 exporter
ddpA up	2.69	Periplasmic binding component of a predicted peptide ABC transporter
prpR down	2.43	DNA-binding transcriptional dual regulator
clcB down	1.12	Chloride ion ClC transporter

Alleles identified through TRMR selection to be targets for recursive multiplex recombineering. From left to right, columns state selection in which the target was identified, the target and allele orientation, the allele fitness in its respective selection, and a brief description of gene function (Ecocyc; www.ecocyc.org). lnW, natural log fitness. —, alleles where fitness cannot be calculated due to lack of frequency in either the initial or final population.