Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population

Aisha I. Khan,[1]*† Duy M. Dinh,[1]* Dominique Schneider,[2,3] Richard E. Lenski,[4] Tim F. Cooper[1]‡

Epistatic interactions between mutations play a prominent role in evolutionary theories. Many studies have found that epistasis is widespread, but they have rarely considered beneficial mutations. We analyzed the effects of epistasis on fitness for the first five mutations to fix in an experimental population of *Escherichia coli.* Epistasis depended on the effects of the combined mutations—the larger the expected benefit, the more negative the epistatic effect. Epistasis thus tended to produce diminishing returns with genotype fitness, although interactions involving one particular mutation had the opposite effect. These data support models in which negative epistasis contributes to declining rates of adaptation over time. Sign epistasis was rare in this genome-wide study, in contrast to its prevalence in an earlier study of mutations in a single gene.

Evolutionary theory predicts that epistatic interactions between mutations can play an important role in determining patterns of adaptation (*1*–*5*). Nevertheless, despite many studies of mutational interactions, little is known about the distribution of their fitness effects. Do most mutations combine additively, or are epistatic interactions widespread? When mutations interact epistatically, are their combined fitness effects usually greater or less than expected from their separate effects? Do the prevalence and form of epistasis depend on how the mutations are generated—for example, whether the mutations are random and therefore usually deleterious or, alternatively, ones that have been fixed by selection? Recent studies have examined epistatic interactions by measuring the relative fitness of genotypes with known numbers and combinations of mutations (*6*–*14*). These studies have typically found that epistasis is common, although there is often no consistent directional effect. Nevertheless, some general patterns have been suggested. For example, synergistic epistasis may be more common in organisms with complex genomes (*15*). However, most studies have examined deleterious mutations, and few data exist on interactions among beneficial mutations that arose during a population's evolution.

The difficulty of identifying and precisely manipulating beneficial mutations has meant that the role of epistasis in adaptive evolution has usually been inferred indirectly from its effects on evolutionary dynamics and outcomes (*16*–*21*). A common observation in microbial evolution experiments is that the rate of fitness increase tends to decelerate over time (*21*–*25*). Negative epistasis, in which the combined effect of beneficial mutations is smaller than would be expected from their separate effects, could explain that tendency. However, such deceleration might instead occur simply because beneficial mutations of large effect will tend to be incorporated earlier owing to their faster spread and greater success in the face of competing beneficial mutations (*26*), and this explanation does not require epistasis. The capacity to sequence experimentally evolved genomes, as well as to enumerate beneficial mutations over time, adds another dimension to evolutionary dynamics that can inform efforts to understand the role of epistasis in adaptive evolution (*24*). Kryazhimskiy *et al.* (*4*) recently proposed that trajectories for fitness and accumulated beneficial mutations could be jointly analyzed to infer the nature of epistasis among the beneficial mutations. By analyzing these combined data for the same experimental population that is the focus of our study, they inferred that interactions between beneficial mutations were predominantly negative in form. Here we will directly test that inference.

To that end, we constructed all possible combinations of the first five beneficial mutations that were fixed in one population of *Escherichia coli* from a long-term evolution experiment. In that experiment, 12 replicate populations were started from a common ancestor and independently evolved in a defined environment supplemented with glucose as the limiting resource (*22*). Fitness increased in all populations, relative to the ancestor, in that same environment (*22*, *27*). This experiment was designed, in part, to examine the repeatability of adaptive evolution, and subsequent studies have documented many examples of both phenotypic and genetic parallelism (*24*, *27*–*30*). Nevertheless, replicate populations have diverged in other phenotypic and genetic traits (*24*, *31*, *32*). A striking example of divergence is the ability to grow on citrate that evolved in only 1 of the 12 populations (*32*). Aside from that case, all of the populations show a strong tendency toward decelerating rates of fitness increase (*27*, *28*).

Whole-genome sequencing of a clone that was isolated after 20,000 generations from one of the populations (designated Ara-1) identified 45 mutational differences from the ancestor (*24*). Many other mutations appeared in the population, of course, but most were eliminated by random drift or negative selection. Other beneficial mutations also arose, including some that reached detectable frequencies, but these were lost by interference from superior beneficial mutations (*24*, *26*) and, in at least one case, because they were less able to evolve than the eventual winners (*33*). Here we focus on the first five mutations that fixed in this population and whose spread coincided with the period of fastest adaptation. These mutations together produced

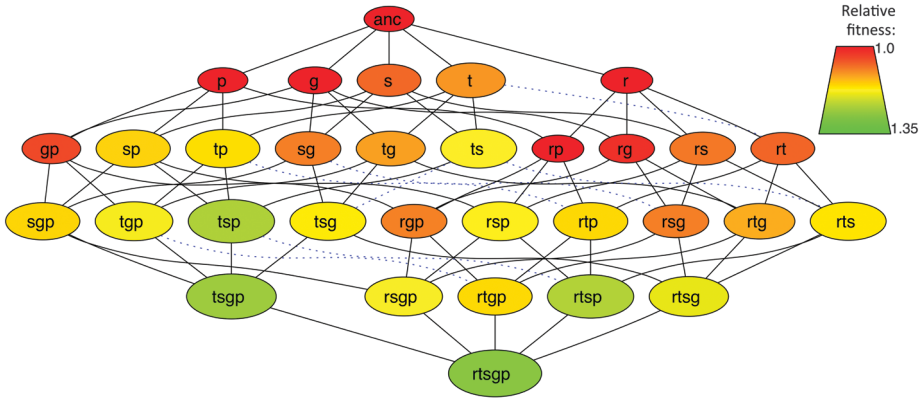

Fig. 1. Mutational network connecting constructed genotypes. Each node represents one of 32 possible combinations of five mutations. Anc indicates the ancestral strain. Other labels indicate mutations affecting these genes: r, *rbs*; t, *topA*; s, *spoT*; g, *glmUS*; and p, *pykF*. Node colors and sizes reflect the genotype's fitness relative to the ancestor. Edges show all one-step mutations. Solid and dotted lines indicate whether the additional mutation caused an increase or decrease in fitness, respectively.

[1]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA. [2]Laboratoire Adaptation et Pathogénie des Microorganismes, Université Joseph Fourier, Institut Jean Roget, F-38041 Grenoble, France. [3]CNRS UMR5163, F-38041 Grenoble, France. [4]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA.

*These authors contributed equally to this work.
†Present address: University of Texas Medical Branch, Galveston, TX 77555, USA.
‡To whom correspondence should be addressed. E-mail: tfcooper@uh.edu

a fitness increase of ~30% relative to the ancestor, and thus they account for much of the ~80% increase in mean fitness that this population achieved by 20,000 generations (*24*, *34*). These mutations, listed in the order they arose, are in these genes or gene regions: *rbs* operon, *topA*, *spoT*, *glmUS* promoter, and *pykF*.

The five beneficial mutations were moved into the ancestral chromosome to produce a set of 32 (= 2^5) genotypes representing all possible combinations of those mutations (table S1). This approach follows previous studies that examined combinations of mutations underlying focal traits in single genes or gene regions (*7*, *9*, *12*). The fitness of each constructed genotype was then estimated in direct competition against a marked variant of the ancestor, which enabled us to describe the local fitness landscape for this full set of genotypes (Fig. 1 and fig. S1). Note that each mutational step that was followed in the evolving population produced an increase in fitness relative to the immediate progenitor. Therefore, each of these mutations was beneficial on the genetic background in which it arose.

To test for epistasis between the focal beneficial mutations, we first used a multiplicative null model that gave the expected fitness of a genotype as the product of the individual fitness effects of the relevant mutations (*35*). The absolute epistatic deviation was then calculated as the difference between observed and expected fitness values (fig. S2). We also calculated the relative epistatic deviation as the log ratio of observed and predicted fitness values (*35*). In both cases, positive (or negative) deviations indicated that the observed fitness of a genotype was greater (or less) than expected under the null model.

We found no overall trend in the contribution of either absolute or relative deviations to the fitness of the 26 constructed genotypes with two or more mutations {mean absolute epistatic deviation = 0.006 ± 0.021 [95% confidence interval (CI)], t_{25} = 0.553, P = 0.585; mean relative deviation = 0.002 ± 0.008 (95% CI), t_{25} = 0.670, P = 0.510} (Fig. 2). The nonsignificant average deviations might mean that epistatic effects were small, rare, or both; alternatively, nonsignificance may indicate that positive and negative effects tended to offset one another (*6*). Consistent with the latter possibility, epistasis was common in the constructed genotypes. The absolute epistatic deviation was significantly positive in seven genotypes and negative in five ($P < 0.05$) (table S2).

There were seven cases in which the addition of a mutation appeared to reduce fitness (Fig. 1), but only one was significant ($P < 0.05$) when analyzed in isolation, and it became nonsignificant with a Bonferroni correction for multiple tests. In any case, even if all seven cases of apparent marginal declines were real, the local adaptive landscape has only one fitness peak, because it is possible to reach the most-fit genotype by uphill trajectories from any starting point. Moreover, 86 of the 120 possible mutational trajectories connecting the ancestor and the most-fit genotype are monotonically increasing for fitness (fig. S1 and table S3). Although not all of the step-wise changes are significant, the paucity of even nonsignificant declines supports the inference that most steps are individually positive. The preponderance of monotonically increasing fitness trajectories in our study stands in sharp contrast to a previous study of five mutations in a β-lactamase gene, where only 18 of the 120 trajectories gave strictly increasing levels of antibiotic resistance (*9*).

What factors might explain the observed variation in the direction and strength of epistasis in our study? One possibility is that epistasis depends on the magnitude of the fitness effects of the contributing mutations. To test this hypothesis, we examined the relation between epistasis and the expected fitness of each genotype in the absence of any epistatic effect. We observed an overall negative relation, indicating that epistatic effects became more negative as the expected fitness rose [absolute epistasis: correlation coefficient (r) = −0.578; relative epistasis: r = −0.586], although this relation appeared to obscure some underlying genetic complexity (Fig. 3). Henceforth, we will focus on the relation between expected fitness and relative epistatic deviation because it provides a measure of the influence of epistasis on the strength of selection for a new beneficial mutation.

The negative relation between epistasis and expected fitness could reflect a consistent effect, such as saturating the potential improvement in some underlying physiological process. Alternatively, the presence or absence of specific mutations might produce groups of genotypes that have different average fitness and epistasis values. For example, if one particular mutation conferred a large benefit in the ancestor but interacted negatively with the other mutations, then it might drive an overall correlation that would not be typical of most mutational interactions. To distinguish between these possibilities, we used analysis of covariance to test whether the presence of any single mutation, or any combination of

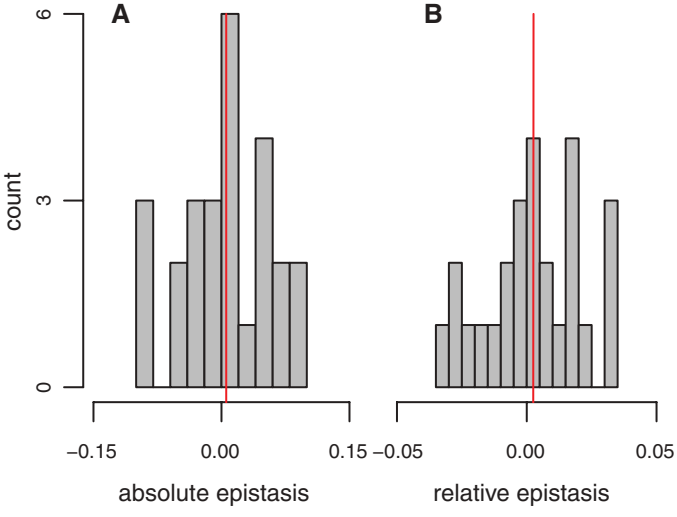

Fig. 2. Distributions of estimated epistatic fitness deviations for 26 genotypes with at least two mutations. (**A**) Absolute epistatic deviations. (**B**) Relative epistatic deviations. Red lines indicate mean values. Note that the *pykF* mutation drives almost all of the positive values of epistasis (Fig. 3 and table S2).

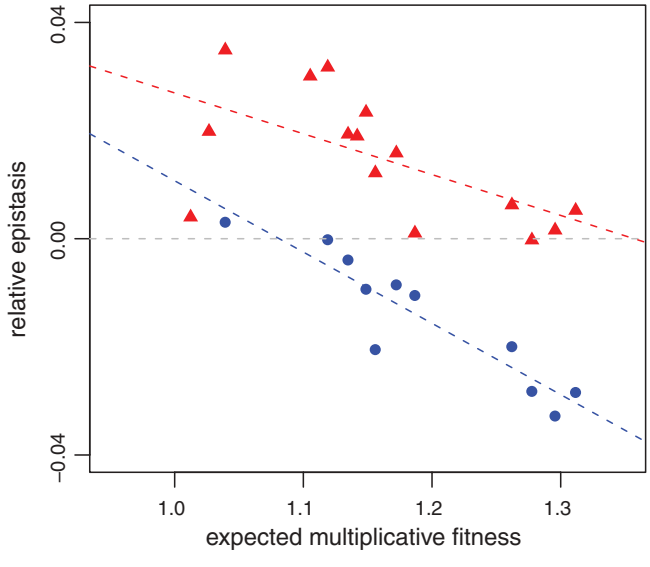

Fig. 3. Relation between relative epistasis and expected fitness under a null model without epistasis. Blue circles and red triangles show genotypes with the ancestral and evolved *pykF* alleles, respectively.

two mutations, explained a significant fraction of the variation in the relation between epistasis and expected fitness. The only gene that showed a significant effect was *pykF* (Fig. 3 and fig. S3) ($F_{1,23}$ = 61.93, P < 0.001) (table S4). The mean epistatic deviation was significantly negative among genotypes with the ancestral allele (mean = –0.014, t_{10} = –3.942, P = 0.003), whereas it was significantly positive among those with the evolved allele (mean = 0.015, t_{14} = 4.913, P < 0.001) (fig. S4). Thus, the *pykF* mutation enhances fitness through its epistatic interactions with the other evolved mutations. However, the sets of genotypes with ancestral and evolved *pykF* alleles both exhibit negative correlations between relative epistasis and expected fitness values (ancestral *pykF*: r = –0.923; evolved *pykF*: r = –0.610) (Fig. 3 and fig. S5). As a consequence, the slope of the major axis regression (*35*) of the observed and expected fitness values is less than unity in both cases, although only marginally so for those genotypes with the evolved *pykF* allele (ancestral *pykF*: slope = 0.628, 95% CI of 0.519 to 0.748; evolved *pykF*: slope = 0.850, 95% CI of 0.698 to 1.029) (fig. S6). These data thus indicate a strong negative relation between expected fitness and epistatic deviations, although the details of this relation also clearly depend on the particular beneficial mutations involved.

We also examined the relation between fitness and epistasis by arranging the 32 genotypes into 16 pairs, such that each pair differed only by the presence or absence of a particular mutation. This pairing allowed us to quantify how the marginal fitness effect of each mutation varied with the fitness of the progenitor background in which the mutation was placed. The benefits conferred by the first four mutations that fixed in the population (Δ*rbs*, *topA*, *spoT*, and *glmUS*) tended to decline as the fitness of the progenitor increased, and that correlation was significant (P < 0.05) in three of those cases (Fig. 4). Notably, the relation was reversed for the *pykF* mutation, such that it tended to confer a greater benefit in the more-fit genetic backgrounds. In fact, the *pykF* mutation was selectively neutral in the ancestral background, unlike the earlier beneficial mutations. Thus, the effect of background fitness on the strength of epistasis depended on the specific mutation, which suggests that these mutations conferred their beneficial effects through different underlying physiological processes.

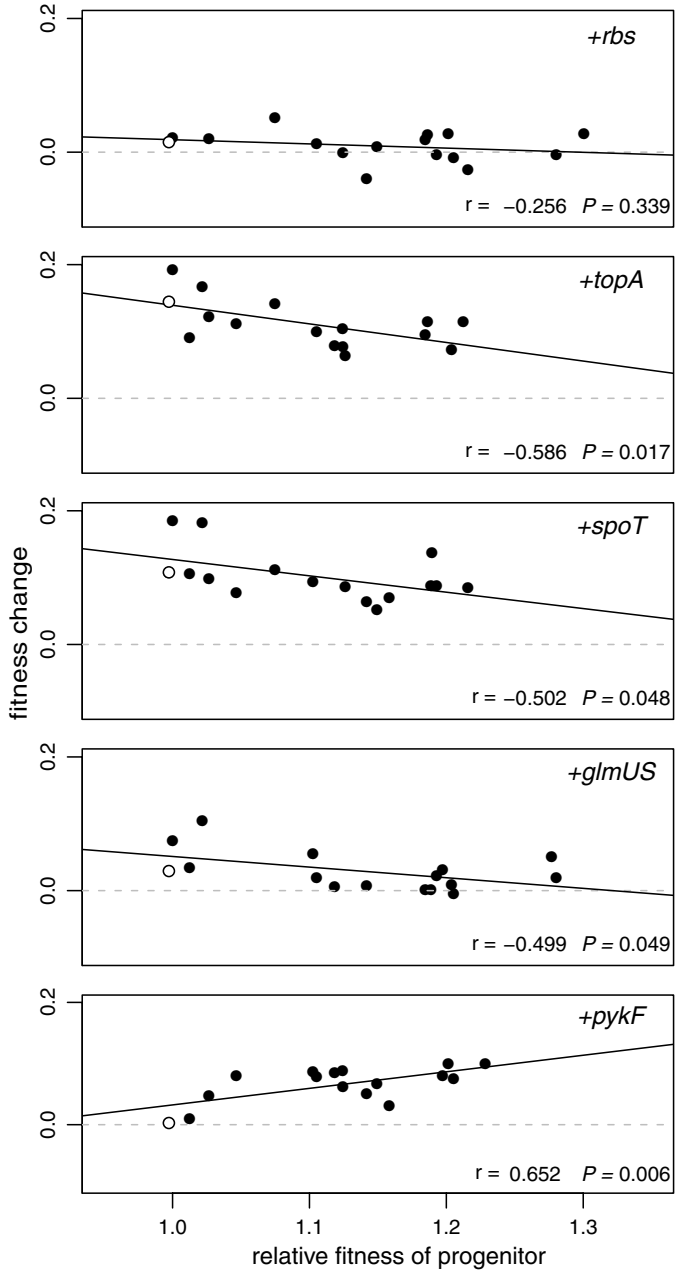

Fig. 4. Relation between the marginal fitness effect of adding a particular mutation and the fitness of the progenitor background to which it was added for each one of five focal mutations. Each panel includes the Pearson correlation coefficient and its significance. The open symbols show the effects of adding each focal mutation to the ancestral strain.

A conspicuous feature of the mean-fitness trajectory for this population—and indeed for most experimental populations evolving in a constant environment—is that the rate of adaptation declined over time (*21*–*25*). Mechanisms that may explain this deceleration include reductions in the number and effect-size of beneficial mutations as a population becomes better adapted to its environment (*21*, *22*, *36*). The strong negative relation between epistasis and the expected fitness of a genotype in the absence of epistasis (Fig. 3) suggests that epistatic interactions contribute greatly to this deceleration by reducing the effect-size of the remaining beneficial mutations as a population approaches a fitness peak. In other words, epistasis acts as a drag that reduces the contribution of later beneficial mutations. Note that similar trends were seen by Chou *et al.* (*37*), who examined the fitness interactions among five beneficial mutations that occurred during adaptation of an engineered strain of *Methylobacterium extorquens* AM1. That study, like ours, found that four mutations interacted to yield diminishing fitness returns, whereas one mutation had the opposite effect.

Our results are also consistent with a recent theoretical study that used population-genetic models to infer that negative epistasis between beneficial mutations could explain the fitness trajectory in the same experimental population that we have studied (*4*). Although we observed widespread epistatic interactions among beneficial mutations in this study, and there was strong epistasis with another beneficial mutation that arose but did not fix in the same population (*33*), we did not observe the extreme constraints on possible mutational paths to higher fitness seen in an earlier study of the β-lactamase gene (*9*). More generally, our results suggest that a relatively simple epistasis function might be incorporated into models that seek to predict the dynamics of adaptation, at least for asexual haploid populations evolving under constant conditions. However, our results also caution that there will be exceptions to any simple function, as evidenced by the finding that the *pykF* mutation greatly affected the magnitude, although not the trend, of the relation between epistasis and fitness.

References and Notes

1. A. S. Kondrashov, *Nature* **369**, 99 (1994).
2. H. A. Orr, M. Turelli, *Evolution* **55**, 1085 (2001).
3. D. M. Weinreich, R. A. Watson, L. Chao, *Evolution* **59**, 1165 (2005).
4. S. Kryazhimskiy, G. Tkacik, J. B. Plotkin, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18638 (2009).
5. J. A. Draghi, T. L. Parsons, G. P. Wagner, J. B. Plotkin, *Nature* **463**, 353 (2010).
6. S. F. Elena, R. E. Lenski, *Nature* **390**, 395 (1997).
7. M. Lunzer, S. P. Miller, R. Felsheim, A. M. Dean, *Science* **310**, 499 (2005).
8. R. Sanjuán, J. M. Cuevas, A. Moya, S. F. Elena, *Genetics* **170**, 1001 (2005).
9. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science* **312**, 111 (2006).
10. L. Jasnos, R. Korona, *Nat. Genet.* **39**, 550 (2007).
11. J. A. G. M. de Visser, S. C. Park, J. Krug, *Am. Nat.* **174** (suppl. 1), S15 (2009).
12. J. da Silva, M. Coetzer, R. Nedellec, C. Pastore, D. E. Mosier, *Genetics* **185**, 293 (2010).
13. K. M. Pepin, H. A. Wichman, *Evolution* **61**, 1710 (2007).
14. L. Jasnos, K. Tomala, D. Paczesniak, R. Korona, *Genetics* **178**, 2105 (2008).
15. R. Sanjuán, S. F. Elena, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14402 (2006).
16. R. Korona, C. H. Nakatsu, L. J. Forney, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9037 (1994).
17. C. L. Burch, L. Chao, *Nature* **406**, 625 (2000).
18. F. B.-G. Moore, D. E. Rozen, R. E. Lenski, *Proc. Biol. Sci.* **267**, 515 (2000).
19. R. Montville, R. Froissart, S. K. Remold, O. Tenaillon, P. E. Turner, *PLoS Biol.* **3**, e381 (2005).
20. R. Sanjuán, J. M. Cuevas, V. Furió, E. C. Holmes, A. Moya, *PLoS Genet.* **3**, e93 (2007).
21. O. K. Silander, O. Tenaillon, L. Chao, *PLoS Biol.* **5**, e94 (2007).
22. R. E. Lenski, M. R. Rose, S. C. Simpson, S. C. Tadler, *Am. Nat.* **138**, 1315 (1991).
23. M. R. Goddard, H. C. J. Godfray, A. Burt, *Nature* **434**, 636 (2005).
24. J. E. Barrick *et al.*, *Nature* **461**, 1243 (2009).
25. S. E. Schoustra, T. Bataillon, D. R. Gifford, R. Kassen, *PLoS Biol.* **7**, e1000250 (2009).
26. P. J. Gerrish, R. E. Lenski, *Genetica* **102-103**, 127 (1998).
27. V. S. Cooper, R. E. Lenski, *Nature* **407**, 736 (2000).
28. R. E. Lenski, M. Travisano, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6808 (1994).
29. T. F. Cooper, D. E. Rozen, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1072 (2003).
30. R. Woods, D. Schneider, C. L. Winkworth, M. A. Riley, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9107 (2006).
31. M. Travisano, F. Vasi, R. E. Lenski, *Evolution* **49**, 189 (1995).
32. Z. D. Blount, C. Z. Borland, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7899 (2008).
33. R. J. Woods *et al.*, *Science* **331**, 1433 (2011).
34. J. A. G. M. de Visser, R. E. Lenski, *BMC Evol. Biol.* **2**, 19 (2002).
35. Materials and methods are available as supporting material on *Science* Online.
36. H. A. Orr, *Genetics* **163**, 1519 (2003).
37. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, *Science* **332**, 1190 (2011).

Acknowledgments: This work was supported by grants from the NSF (DEB-1019989 to R.E.L. and DEB-0844355 to T.F.C.), the James S. McDonnell Foundation (220020174 to T.F.C.), the Agence Nationale de la Recherche (Program Génomique, Grant ANR-08-GENM-023-001 to D.S.) and the Defense Advanced Research Projects Agency "Fun Bio" Program (HR0011-09-1-0055 to R.E.L. and T.F.C.). We thank R. Azevedo, T. Paixão, and D. Stoebel for valuable discussions and helpful comments on the manuscript. R.E.L. will make the ancestral and evolved strains used in this study available to qualified recipients, subject to completion of a material transfer agreement that can be found at http://technologies.msu.edu/forms.html. T.F.C. will make the strains constructed in this study available to qualified recipients. Competition experiment counts, summary input data, and analysis scripts that pertain to the experiments and analyses reported in this paper have been deposited at http://dx.doi.org/10.5061/dryad.5rv40.

Supporting Online Material

www.sciencemag.org/cgi/content/full/332/6034/1193/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S4
References and Notes

3 February 2011; accepted 25 April 2011
10.1126/science.1203801

Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades

Lulu Qian[1] and Erik Winfree[1,2,3]*

To construct sophisticated biochemical circuits from scratch, one needs to understand how simple the building blocks can be and how robustly such circuits can scale up. Using a simple DNA reaction mechanism based on a reversible strand displacement process, we experimentally demonstrated several digital logic circuits, culminating in a four-bit square-root circuit that comprises 130 DNA strands. These multilayer circuits include thresholding and catalysis within every logical operation to perform digital signal restoration, which enables fast and reliable function in large circuits with roughly constant switching time and linear signal propagation delays. The design naturally incorporates other crucial elements for large-scale circuitry, such as general debugging tools, parallel circuit preparation, and an abstraction hierarchy supported by an automated circuit compiler.

The power and mystery of life is entangled within the information processing at the heart of all cellular machinery. Engineering molecular information processing systems may allow us to tap into that power and elucidate principles that will help us to understand and appreciate the mystery. DNA is an excellent engineering material for biochemical circuits because its biological nature supports technological applications in vivo, its easy chemical synthesis facilitates practical experiments in vitro, its combinatorial structure provides sufficient sequence design space, and the Watson-Crick complementarity principle enables predictable molecular behavior.

DNA has been used as a computing substrate since the first demonstration of solving a seven-city Hamiltonian path problem in 1994 (*1*) and has evolved away from competing with silicon to embedding control within molecular systems. Although DNA automata can be built with deoxyribozymes (*2*, *3*) or with restriction enzymes (*4*), the introduction of toehold-mediated DNA strand displacement enabled enzyme-free DNA machinery that is automated by hybridization alone (*5*–*8*). A DNA strand can serve as a signal when it is free, but is inhibited when it is bound to a complementary strand. A single-stranded DNA signal can first bind to a partially double-stranded complex by a single-stranded domain called a toehold, then release the originally bound strand after branch migration has occurred. Thus, an output signal can be activated upon the arrival of an input signal, and the reaction rate can be controlled by the length of the toehold. This principle has inspired the development of a rich theory (*9*, *10*) and practice (*11*–*13*) of DNA strand displacement circuits, resulting in a wide range of applications such as medical therapeutics in vivo (*14*), molecular instruments in situ (*15*), and biomedical diagnostics in vitro (*16*). To date, the largest digital circuit built with DNA strand displacement cascades involved 12 initial DNA species (*11*). However, their logic gates were constructed with multistranded DNA complexes, challenging sequence design constraints were required, and signal restoration occurred only at the circuit output, perhaps explaining why the performance decayed surprisingly with scale.

To create a scalable DNA circuit architecture, we proposed (*17*) a simple DNA gate motif—a "seesaw" gate—that makes use of a reversible strand displacement reaction based on the principle of toehold exchange (*8*, *12*). In this context, seesawing is the reversible reaction that exchanges

[1]Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA. [2]Computer Science, California Institute of Technology, Pasadena, CA 91125, USA. [3]Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: winfree@caltech.edu

Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population

Aisha I. Khan, Duy M. Dinh, Dominique Schneider, Richard E. Lenski and Tim F. Cooper

Science **332** (6034), 1193-1196.
DOI: 10.1126/science.1203801