

# Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation

Hsin-Hung Chou,<sup>1\*</sup> Hsuan-Chao Chiu,<sup>2</sup> Nigel F. Delaney,<sup>1</sup> Daniel Segrè,<sup>2,3</sup> Christopher J. Marx<sup>1,4†</sup>

Epistasis has substantial impacts on evolution, in particular, the rate of adaptation. We generated combinations of beneficial mutations that arose in a lineage during rapid adaptation of a bacterium whose growth depended on a newly introduced metabolic pathway. The proportional selective benefit for three of the four loci consistently decreased when they were introduced onto more fit backgrounds. These three alleles all reduced morphological defects caused by expression of the foreign pathway. A simple theoretical model segregating the apparent contribution of individual alleles to benefits and costs effectively predicted the interactions between them. These results provide the first evidence that patterns of epistasis may differ for within- and between-gene interactions during adaptation and that diminishing returns epistasis contributes to the consistent observation of decelerating fitness gains during adaptation.

Epistasis describes genetic interactions in terms of how phenotypic effects of a mutation depend on other mutations in the genome. If two mutations act on a given phenotype independently, each would be expected to exert the same proportional effect regardless of whether the other allele was present, although other models can be applied (1–3). Deviations from this null expectation have been used to uncover interacting genes via genetic screens for

second mutations that suppress the effect of the first, to identify the order of enzymes in biochemical pathways, and to unravel systems-level interaction patterns characterized with genome-wide double-knockout libraries. One general trend has been that the detrimental effect of a lesion in a pathway (or module) (4) is greater alone than when there is already another deleterious mutation in that process (i.e., antagonistic epistasis). In contrast, lesions in parallel pathways

producing the same product tend to cause stronger phenotypes (synergistic epistasis) than expected; the extreme case of the latter, termed synthetic lethality, results in a nonviable genotype.

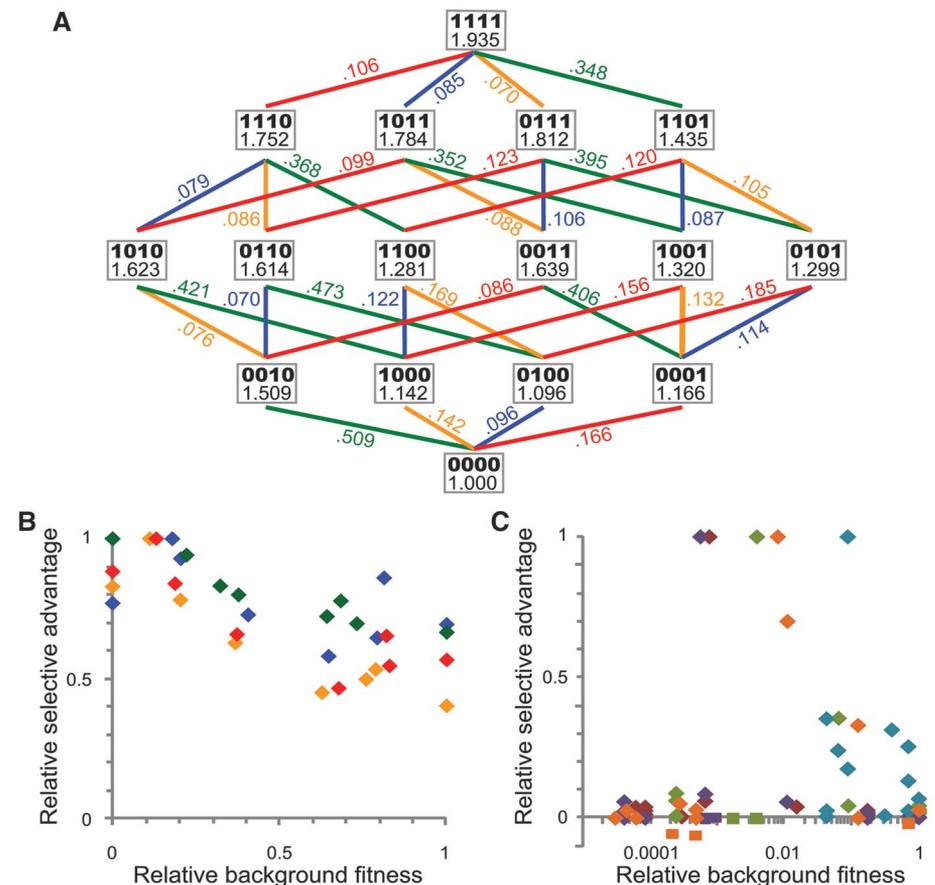
Epistasis between beneficial mutations remains largely unexplored. Previous studies examined epistasis between five amino acid (or promoter) substitutions within an allele of  $\beta$ -lactamase selected for cefotaxime resistance in *Escherichia coli* (5). By constructing all possible mutation combinations within the  $\beta$ -lactamase locus, a single-peaked fitness landscape was revealed with numerous cases where the identical mutation increased resistance on some backgrounds but decreased it on others (i.e., sign epistasis). Similar results have been found for cofactor use by isopropylmalate dehydrogenase (6) and for hormone receptors (7). In contrast, few studies have addressed interactions between beneficial mutations in different genes (8, 9).

<sup>1</sup>Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. <sup>2</sup>Graduate Program in Bioinformatics, Boston University, Boston, MA 02215, USA. <sup>3</sup>Department of Biology and Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA. <sup>4</sup>Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA.

\*Present address: Institute of Molecular Systems Biology, ETH Zürich, CH-8093 Zürich, Switzerland.

†To whom correspondence should be addressed. E-mail: cmarx@oeb.harvard.edu

**Fig. 1.** Mutational network and distinct patterns of epistasis for mutations between and within genes. **(A)** Each node displays the allelic composition (*fghA*, *pntAB*, *gshA*, *GB*) of a given genotype (bold) and its fitness. Ancestral and evolved alleles are indicated by 0 and 1, respectively, leading from the ancestral EM strain (0000) to the evolved EVO isolate (1111). Each edge indicates an allelic replacement (*fghA*, orange; *pntAB*, blue; *gshA*, green; *GB*, red) and the corresponding selective coefficient. Variation in relative selective effect (normalized to maximum  $s_i$ ) of each allele as a function of the fitness of the background it was introduced into is shown for: **(B)** between-gene epistasis in *Methylobacterium* [background fitness normalized from EM = 0 to maximum = 1; colors as in (A)] or **(C)** within-gene epistasis for *E. coli*  $\beta$ -lactamase [background fitness normalized to minimum inhibitory concentration maximum of 1; log scale for visualization; squares indicate deleterious effects]. [Data from (5)]



The distribution of epistatic interactions between mutations may greatly influence evolutionary outcomes—from the maintenance of sexual reproduction to the fixation rate of beneficial alleles—and hence the speed of adaptation itself. The most consistent finding across studies of laboratory-evolved populations has been a rapid deceleration of the rate of fitness increase (10). Theoretical analysis suggests that the observed dynamics of fitness increase and accumulation of substitutions (11) are best described by a class of fitness landscapes with antagonistic interactions between beneficial mutations (12).

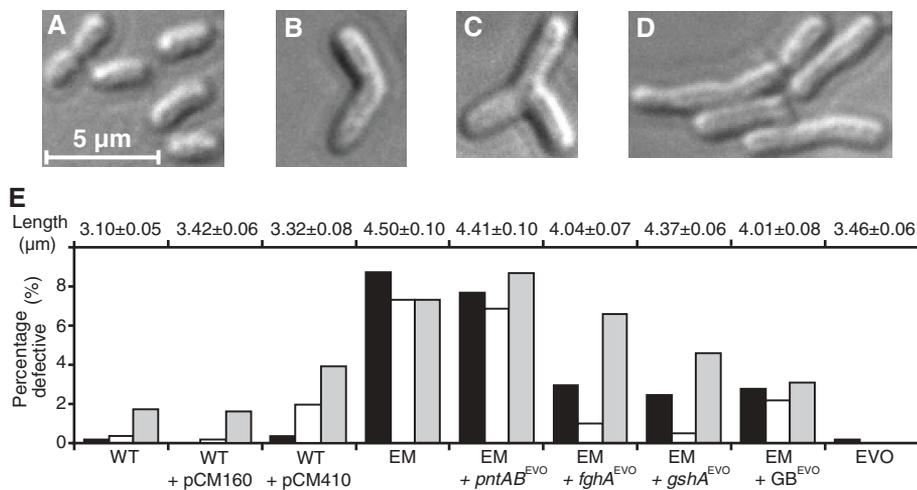
We took both experimental and theoretical approaches to investigate potential epistasis in populations that were initiated with an engineered strain of *Methylobacterium extorquens* AM1 (hereafter, EM) (table S1) and evolved in batch culture with methanol as the sole carbon source (3). In order to grow on methanol, *Methylobacterium* must oxidize formaldehyde into formate. Wild-type *Methylobacterium* (WT) performs this oxidation with a tetrahydromethanopterin-dependent pathway (13). In EM, this native

pathway was eliminated and replaced by a non-orthologous, glutathione (GSH)-dependent pathway from *Paracoccus denitrificans* (fig. S1) (14). As a result, the EM strain could grow on methanol, but at a rate one-third that of WT (fig. S2). Adaptation in eight replicate populations dependent on this engineered metabolic function (analogous to natural horizontal gene transfer) resulted in an average fitness increase after 600 generations of 66.8% (fig. S3), as determined by competition assays (3), and was largely carbon substrate-specific (fig. S4).

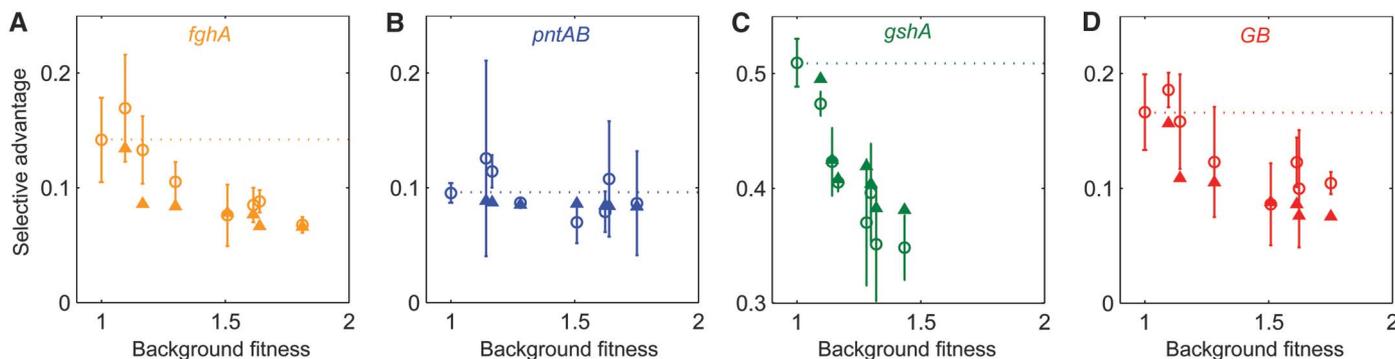
The genome of an evolved isolate from generation 600 (“EVO”) (9) with the highest fitness ( $W_{EVO} = 1.94$ ) (table S2) was sequenced to identify the genetic basis of adaptation in that lineage (3). In total, nine mutations were identified (fig. S5) (3). We found an 11-base pair (bp) deletion between the two genes that encode the GSH-dependent pathway, *flhA* and *fghA* (i.e., *fghA*<sup>EVO</sup>), in a plasmid specifically introduced into EM (pCM410) (fig. S6). This deletion removed the apparent ribosome-binding site for *fghA* and decreased expression of these enzymes by 55%

and 73% (3), respectively. This change, however, increased fitness by 14.2% (Fig. 1A), which suggested that production of these enzymes in the EM ancestor was higher than the optimum. In WT, where the GSH pathway is extraneous, a strain with an empty vector had a 14.1% fitness advantage relative to fitness when both genes were expressed. It therefore appeared that the primary advantage of the *fghA*<sup>EVO</sup> allele was to reduce the costs of protein overexpression (e.g., energy consumption, ribosome sequestering, and protein misfolding). We also identified a single-nucleotide polymorphism (SNP) in the promoter region of pyridine nucleotide transhydrogenase (*pntAB*<sup>EVO</sup>), and a 2-bp deletion in the promoter of the most rate-limiting enzyme of GSH biosynthesis,  $\gamma$ -glutamylcysteine synthetase (*gshA*<sup>EVO</sup>). These gene products have clear linkages to methanol utilization in EM (3). The remaining six genetic changes included a large deletion (fig. S7), a synonymous SNP, the loss of a plasmid, two transposon insertions, and a 6-bp insertion (3). The last-named six are difficult to reconstruct genetically, were individually neutral under our experimental conditions (15), or were deemed unlikely to greatly contribute to fitness. We thus treated them as a single collective locus, the “genetic background” (*GB*<sup>EVO</sup>), for the purpose of examining epistasis between beneficial mutations. All identified alleles, when present individually in the ancestral background conferred fitness benefits ranging from 10 to 51% (Fig. 1A).

In order to investigate epistasis between these beneficial mutations, strains with each allelic combination ( $2^4 = 16$ ) were constructed (3), and their fitness values were measured (Fig. 1A). The adaptive landscape of this genotypic space contained a single peak; each allele was universally beneficial across genetic backgrounds (i.e., showed no sign epistasis, but the degree of benefit conferred varied) (Fig. 1B). Except for *pntAB*<sup>EVO</sup>, the remaining three alleles exhibited a significant trend of diminishing returns: Their selective benefits declined in genetic backgrounds with higher fitness. In contrast, the resistance to cefotaxime conferred by each mutation within the *E. coli*  $\beta$ -lactamase gene (5) was idiosyncratic with regard to the



**Fig. 2.** Morphological aberrations caused by expression of the foreign pathway. Distinct cellular morphologies of (A) WT or an EM ancestor showing (B) curved, (C) branched, or (D) elongated cells. (E) Mean cell length and proportion of elongated (black), branched (white), and curved (gray) cells for various strains. Plasmid pCM410 expresses the foreign pathway; pCM160 is an empty control plasmid.



**Fig. 3.** Antagonistic trend of epistasis detected from the data and captured by the benefit-cost model. (A to D) Plots of measured (open circles) and predicted (solid triangles) selective coefficients  $s$  for each of the four evolved alleles,

respectively, versus the fitness of the background onto which the allele was introduced. Dashed lines indicate selective advantages for each allele on the ancestral background (i.e., expectation for no epistasis).

resistance of the background onto which it was introduced (Fig. 1C).

We found a connection between antagonistic epistasis and a physiological problem caused by protein overexpression in EM. Cells of the EM ancestor showed an increased length and aberrant morphologies relative to WT (Fig. 2 and fig. S8), similar to those commonly observed for protein overexpression (16). Reducing expression of the foreign pathway in EM via *fghA<sup>EVO</sup>* suppressed cellular abnormalities, whereas expressing it in WT (where it is redundant) induced similar defects. This confirmed that the morphological defects were caused by overexpression of the foreign pathway. In addition, the *gshA<sup>EVO</sup>* and *GB<sup>EVO</sup>* alleles (but not *pntAB<sup>EVO</sup>*) also individually reduced morphological defects (to about one-third), and when all the evolved alleles are present together (e.g., the EVO strain), abnormal cells were nearly absent (a finding recapitulated across all eight populations) (fig. S2D). These data suggest that part of the benefit conferred by the three alleles whose selective benefit wanes on fitter backgrounds resulted from directly or indirectly decreasing protein overexpression costs.

Epistasis has been often represented as the deviation from a null model in which individual mutations affect the ancestor's fitness ( $W_0 = 1.0$ ) with independent multiplicative factors  $\lambda_i$  (double mutant's fitness,  $W_{ij} = \lambda_i \lambda_j W_0$ ). However, in our system, rather than being captured by a single indivisible phenotype, cell growth seems to depend on at least one separately measurable component, i.e., the growth burden imposed by expressing the foreign pathway. As stated above, three of the four alleles identified in the EVO strain appear to increase fitness at least partly by reducing this cost. Therefore, in analogy to the contributions to fitness by a single enzyme (17), we developed a mathematical model that partitions fitness into two phenotypes: a "benefit" component  $b_0$ , analogous to a single conglomerate "enzyme activity" that sets the rate of energy extracted from the substrate to generate biomass; and a cost  $c_0$ , encompassing a fixed amount of energy diverted to deal with overexpression of the foreign pathway (3). Thus, the fitness of the ancestral strain can be written, as  $W_0 = b_0 - c_0 = 1$ . We hypothesize that a new allele  $i$  could modify the benefit and the cost of the ancestral background by certain multiplicative factors ( $\lambda_i$  and  $\theta_i$ , respectively), giving rise to a fitness  $W_i = \lambda_i b_0 - \theta_i c_0$ . A successive allele  $j$ , on top of the background of mutant  $i$ , is similarly assumed to act multiplicatively on the benefit and cost components, yielding a fitness  $W_{ij} = \lambda_i \lambda_j b_0 - \theta_i \theta_j c_0$ .

If we could determine experimentally the values of  $b_0$ ,  $c_0$ , and the  $\lambda_i$  and  $\theta_i$  for each allele, then the above model should provide predictions for the fitness of any multiallele strain, computable as

$$W_{\text{mutant}} = \prod_{i \in \text{Alleles}} \lambda_i b_0 - \prod_{i \in \text{Alleles}} \theta_i c_0$$

(3). We estimated these parameters: Cost was determined by expressing the foreign pathway in

WT, where its metabolic function was fully redundant ( $c_0 = 0.141$ ) (17). Setting  $W_0 = 1$ , results in  $b_0 = 1.141$ . The lowered cost of expression,  $\theta_i$ , for each allele was approximated as the decreased relative proportion of morphological defects (table S3) (3). Factors  $\lambda_i$  could then be estimated using the single-allele benefit-cost model.

Without specifying further information, this simple model partitioning fitness into benefit and cost outperformed the standard null model in predicting fitness values of multiallele combinations [coefficient of determination ( $R^2$ ) = 0.97 versus  $R^2 = 0.64$ ] (figs. S9 and S10) (3). It also recapitulated the antagonistic trend of epistasis among the three alleles affecting cell morphology and correctly predicted the consistent magnitude of benefit from *pntAB<sup>EVO</sup>* (Fig. 3). The agreement between our experimental data and model predictions supports our model assumptions and thus our hypothesis as to why diminishing returns epistasis was observed: Proportional reductions of a cost became successively less beneficial as the cost itself was alleviated.

Diminishing returns has been predicted (18) but because of different assumptions. The non-linearity of fitness increase in these models arises because it is assumed that a given trait is under stabilizing selection for an intermediate optimum, which explicitly considers fitness as being displaced from a fixed adaptive peak. Although the assumption of intermediate optimality holds well for many traits like body weight and length, fitness rises monotonically with increasing growth rate or decreasing protein expression burden. In this study, we considered a higher-level phenotype (growth rate) as the sum of two constituent phenotypes (metabolic rate and protein expression burden), which allowed us to generate a precise expectation for the fitness of multiallele strains without explicitly assuming stabilizing selection. The success of this approach suggests that it may be possible to generalize the idea of expressing higher-level phenotypes (such as fitness) as combinations of multiple underlying traits to provide quantitative predictions of epistasis.

An analogous study (19) of the interactions between beneficial mutations in *E. coli* evolved in minimal glucose medium found similar epistatic trends: four of five new alleles exhibit significant diminishing returns. The fifth such mutation, and a mutation present as a component of our *GB<sup>EVO</sup>* allele that is beneficial only in metal-poor media (15), showed the opposite trend: an increase in selective advantage with higher background fitness. Thus, across these two distinct model systems 7 of 10 alleles consistently showed antagonism, whereas only 2 exhibited synergy. This tendency toward diminishing returns between beneficial mutations was predicted from trajectories of fitness increase and substitution rate (12) but had never been tested directly. Furthermore, these results are in stark contrast to the epistatic effects seen among mutations within single proteins, which are varying and unpredictable

in their effect with regard to background activity (5, 7). This distinction between results from within- and between-gene epistasis suggests that the underlying causes of epistasis at different physiological scales (i.e., within-gene protein biophysics versus between-gene physiological networks) lead to categorically distinct, but reproducible, trends in genetic interactions that affect both the speed of adaptation and the degree to which possible trajectories are limited.

*Note added in proof:* Kvitck and Sherlock (20) recently reported an additional experimental system, evolution of *Saccharomyces cerevisiae* in a glucose-limited chemostat, that exhibited a general trend of diminishing returns for epistasis between pairs of beneficial mutations in different loci.

## References and Notes

1. P. C. Phillips, *Nat. Rev. Genet.* **9**, 855 (2008).
2. H. J. Cordell, *Hum. Mol. Genet.* **11**, 2463 (2002).
3. Materials and methods are available as supporting material on Science Online.
4. D. Segrè, A. Deluna, G. M. Church, R. Kishony, *Nat. Genet.* **37**, 77 (2005).
5. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science* **312**, 111 (2006).
6. M. Lunzer, S. P. Miller, R. Felsheim, A. M. Dean, *Science* **310**, 499 (2005).
7. J. T. Bridgham, S. M. Carroll, J. W. Thornton, *Science* **312**, 97 (2006).
8. R. Sanjuán, A. Moya, S. F. Elena, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15376 (2004).
9. M. K. Applebee, M. J. Herrgård, B. O. Palsson, *J. Bacteriol.* **190**, 5087 (2008).
10. S. F. Elena, R. E. Lenski, *Nat. Rev. Genet.* **4**, 457 (2003).
11. J. E. Barrick et al., *Nature* **461**, 1243 (2009).
12. S. Kryazhinskiy, G. Tkacik, J. B. Plotkin, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18638 (2009).
13. L. Chistoserdova, J. A. Vorholt, R. K. Thauer, M. E. Lidstrom, *Science* **281**, 99 (1998).
14. C. J. Marx, L. Chistoserdova, M. E. Lidstrom, *J. Bacteriol.* **185**, 7160 (2003).
15. H. H. Chou, J. Berthet, C. J. Marx, *PLoS Genet.* **5**, e1000652 (2009).
16. C. G. Kurland, H. Dong, *Mol. Microbiol.* **21**, 1 (1996).
17. E. Dekel, U. Alon, *Nature* **436**, 588 (2005).
18. G. Martin, S. F. Elena, T. Lenormand, *Nat. Genet.* **39**, 555 (2007).
19. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, *Science* **332**, 1193 (2011).
20. D. J. Kvitck, G. Sherlock, *PLoS Genet.* **7**, e1002056 (2011).

**Acknowledgments:** We thank S. Kryazhinskiy, D. Weinreich, J. Weitz, C. Wilke, R. Whitaker, and members of the Marx and Segrè labs for suggestions and discussions. Supported by NIH grant R01 GM078209, and NSF grant DEB-0845893. The resulting single-end 36-bp read data have been deposited in the NCBI Short Read Archive with accession no. SRA030695.1.

## Supporting Online Material

www.sciencemag.org/cgi/content/full/332/6034/1190/DC1  
Materials and Methods  
SOM Text  
Figs. S1 to S10  
Tables S1 to S3  
References

3 February 2011; accepted 25 April 2011  
10.1126/science.1203799