28. Z. Gedalof, D. L. Peterson, N. J. Mantua, *J. Am. Water Resour. Assoc.* **40**, 1579–1592 (2004).
29. E. R. Lutz, A. F. Hamlet, J. S. Littell, *Water Resour. Res.* **48**, W01525 (2012).
30. F. M. Ralph *et al.*, *Geophys. Res. Lett.* **33**, L13801 (2006).
31. P. A. O'Gorman, C. J. Muller, *Environ. Res. Lett.* **5**, 025207 (2010).
32. P. C. D. Milly, K. A. Dunne, A. V. Vecchia, *Nature* **438**, 347–350 (2005).
33. E. P. Salathé Jr., L. R. Leung, Y. Qian, Y. Zhang, *Clim. Change* **102**, 51–75 (2010).
34. D. D. Breshears *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15144–15148 (2005).
35. S. J. Wenger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14175–14180 (2011).
36. G. T. Pederson, J. L. Betancourt, G. J. McCabe, *Geophys. Res. Lett.* **40**, 1811–1816 (2013).
37. J. R. Minder, *J. Clim.* **23**, 2634–2650 (2010).

Acknowledgments: We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in table S1 of this paper) for producing and making available their model output. The U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support for CMIP and led the development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We thank the reviewers for their comments and insights, which substantially improved the paper. J.T.A. was supported by NSF's Experimental Program to Stimulate Competitive Research (EPSCoR) (EPS-0814387). This work was partially supported by NASA through NNH11ZDA001N-FIRES.

Supplementary Materials

www.sciencemag.org/content/342/6164/1360/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 and S2
References (*38*–*60*)

24 June 2013; accepted 8 November 2013
10.1126/science.1242335

Long-Term Dynamics of Adaptation in Asexual Populations

Michael J. Wiser,[1,2] Noah Ribeck,[1,3] Richard E. Lenski[1,2,3]*

Experimental studies of evolution have increased greatly in number in recent years, stimulated by the growing power of genomic tools. However, organismal fitness remains the ultimate metric for interpreting these experiments, and the dynamics of fitness remain poorly understood over long time scales. Here, we examine fitness trajectories for 12 *Escherichia coli* populations during 50,000 generations. Mean fitness appears to increase without bound, consistent with a power law. We also derive this power-law relation theoretically by incorporating clonal interference and diminishing-returns epistasis into a dynamical model of changes in mean fitness over time.

The dynamics of evolving populations are often discussed in terms of movement on an adaptive landscape, where peaks and valleys are states of high and low fitness, respectively. There is considerable interest in the structure of these landscapes (*1*–*7*). Recent decades have seen tremendous growth in experiments using microbes to address fundamental questions about evolution (*8*), but most have been short in duration. The Long-Term Evolution Experiment (LTEE) with *Escherichia coli* provides the opportunity to characterize the dynamics of adaptive evolution over long periods under constant conditions (*1*, *9*, *10*). Twelve populations were founded from a common ancestor in 1988 and have been evolving for >50,000 generations, with samples frozen every 500 generations. The frozen bacteria remain viable, and we use this "fossil record" to assess whether fitness continues to increase and to characterize mean fitness trajectories (*11*).

We first performed 108 competitions, in the same conditions as the LTEE, between samples from nine populations at 40,000 and 50,000 generations against marked 40,000-generation clones (*11*). Three populations were excluded for technical reasons (*11*). Fitness was quantified as the dimensionless ratio of the competitors' realized growth rates. Most populations experienced significant improvement (Fig. 1A), and the grand mean fitness increased by 3.0% (Fig. 1B).

To examine the shape of the fitness trajectory, we competed samples from all 12 populations and up to 41 time points against the ancestor (*11*). We compared the fit of two alternative models with the fitness trajectories. The hyperbolic model describes a decelerating trajectory with an asymptote. The power law also decelerates (provided the exponent is <1), but fitness has no upper limit.

Hyperbolic model

$$\overline{w} = 1 + at/(t + b)$$

Power law

$$\overline{w} = (bt + 1)^a$$

Mean fitness is $\overline{w}$, time in generations is t, and each model has two parameters, a and b. Both models are constrained such that the ancestral fitness is 1, hence the offset of +1 in the power law. The hyperbolic model was fit to the first 10,000 generations of the LTEE (*9*), but others suggested an alternative nonasymptotic trajectory (*12*). The grand mean fitness values and the trajectory for each model are shown in Fig. 2A and the individual populations in fig. S1. Both models fit the data very well; the correlation coefficients for the grand means and model trajectories are 0.969 and 0.986 for the hyperbolic and power-law models, respectively. When Bayesian information criterion scores (*11*) are used, the power law outperforms the hyperbolic model with a posterior odds ratio of ~30 million (table S1). The superior performance of the power law also holds when populations are excluded because of incomplete time series or evolved hypermutability (table S1). The power law provides a better fit to the grand-mean fitness than the hyperbolic model in early, middle, and late generations (fig. S2). The power law is supported (odds ratios >10) in six individual populations, whereas none supports the hyperbolic model to that degree (table S2). The power law also predicts fitness gains more accurately than the hyperbolic model. When fit to data for the first 20,000 generations

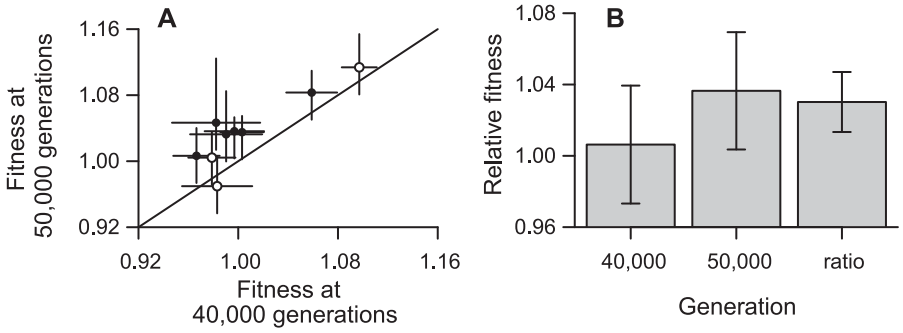

Fig. 1. Fitness changes in nine *E. coli* populations between 40,000 and 50,000 generations. (A) Filled symbols: six populations whose improvement was significant ($P < 0.05$); open symbols: three populations without significant improvement. **(B)** Grand-mean fitness at 40,000 and 50,000 generations relative to 40,000-generation competitor and the ratio of means showing overall gain. Error bars are 95% confidence limits based on replicate assays (A) or populations (B).

[1]BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824, USA. [2]Department of Zoology, Michigan State University, East Lansing, MI 48824, USA. [3]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA.

*Corresponding author. E-mail: lenski@msu.edu

only, the hyperbolic model badly underestimates later measurements, whereas the power-law trajectory predicts them accurately (Fig. 2B and fig. S3).

The power law describes the fitness trajectories well, but it is not explanatory. We have derived a dynamical model of asexual populations with clonal interference and diminishing-returns epistasis, which generates mean-fitness trajectories that agree well with the experimental data. Clonal interference refers to competition among organisms with different beneficial mutations, which impedes their spread in asexual populations (*13*–*16*). Diminishing-returns epistasis occurs when the marginal improvement from a beneficial mutation declines with increasing fitness (*5*, *6*). We outline key points of the model below (*11*).

We used a coarse-grained approach that describes the magnitudes and time scales of fixation events (*13*). Beneficial mutations of advantage s are exponentially distributed with probability density $\alpha e^{-\alpha s}$, where $1/\alpha$ is the mean advantage. This distribution is for mathematical convenience; the theory of clonal interference is robust to the form of the distribution (*13*). We assume that deleterious mutations do not appreciably affect the dynamics; deleterious mutations occur at a higher rate than beneficial mutations, but the resulting load is very small relative to the fitness increase measured over the course of the LTEE (*17*).

We assume the distribution of available benefits declines after a mutation with advantage $\langle s \rangle$ fixes, such that α increases by a factor linearly related to $\langle s \rangle$:

$$\alpha_{n+1} = \alpha_n(1 + g\langle s_{n+1} \rangle)$$

where $g > 0$ is the diminishing-returns parameter, $\langle s_n \rangle$ is beneficial effect of the *nth* fixed mutation, and α_n is α after n fixations. Then, the mean fitness of an asexual population adapting to a constant environment is approximated by (*11*):

$$\overline{w} \approx \left(2g\langle s_1 \rangle e^{g\langle s_1 \rangle}\frac{t}{\langle t_1 \rangle} + 1\right)^{1/2g}$$

where $\langle s_1 \rangle$ and $\langle t_1 \rangle$ are the beneficial effect and fixation time, respectively, for the first fixed mutation.

Comparing this formula with the power law, $g = 1/2a$. The value of g estimated for the six populations that retained the low ancestral mutation rate throughout 50,000 generations is 6.0 (95% confidence interval 5.3 to 6.9). In the LTEE, the beneficial effect of the first fixation, $\langle s_1 \rangle$, is typically ~0.1 (*1*, *9*, *10*). It follows that the distribution of beneficial effects immediately after the first fixation is shifted such that the mean advantage is $1/(1 + g\langle s_1 \rangle) \approx 63\%$ of its initial value (*11*). This estimate of g also accords well with epistasis observed for early mutations in one of the populations (fig. S4). In principle, g might vary among populations if some fixed mutations lead to regions of the fitness landscape with different epistatic tendencies (*18*). However, an analysis of variance shows no significant heterogeneity in g among the six populations that maintained the ancestral mutation rate ($P = 0.3478$) (table S3). The g values tend to be lower for several populations that evolved hypermutability (table S4). However, these fits are confounded by the change in mutation rate; we show below that it is not necessary to invoke a difference in diminishing-returns epistasis between the hypermutable populations and those that retained the low ancestral mutation rate.

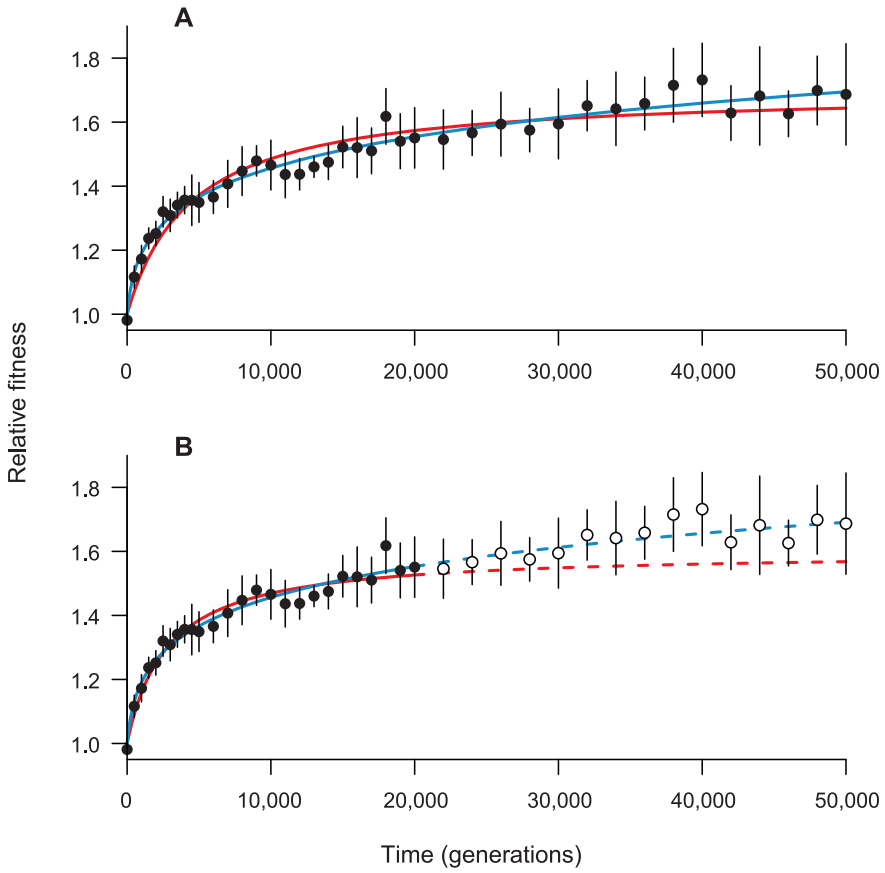

Fig. 2. Comparison of hyperbolic and power-law models. (**A**) Hyperbolic (red) and power-law (blue) models fit to the set of mean fitness values (black symbols) from all 12 populations. (**B**) Fit of hyperbolic (solid red) and power-law (solid blue) models to data from first 20,000 generations only (filled symbols), with model predictions (dashed lines) and later data (open symbols). Error bars are 95% confidence limits based on the replicate populations.

Diminishing-returns epistasis generates the power-law dynamics through the relation between a and g. Clonal interference affects the dynamics through the parameter b, which depends on $\langle s_1 \rangle$ and $\langle t_1 \rangle$, which in turn are functions of the population size N, beneficial mutation rate μ, and initial mean beneficial effect $1/\alpha_0$ (*11*). For the LTEE, $N = 3.3 \times 10^7$, which takes into account the daily dilutions and regrowth (*1*). However, μ and α_0 are unknown. Pairs of values that all match the best fit to the populations that retained the low mutation rate are shown in Fig. 3A. The expected values for beneficial effects and fixation times across a range of pairs are shown in Fig. 3B. The dynamics are similar among pairs with high beneficial mutation rates ($\mu > 10^{-8}$), giving $\langle s_1 \rangle \approx 0.1$ and $\langle t_1 \rangle \approx 300$ generations for the first fixation, which agree well with observations from the LTEE (*1*, *9*, *10*). At lower values of μ, adaptation becomes limited by the supply of beneficial mutations, and fixation times are inconsistent with the LTEE. This model also predicts that the rate of adaptation decelerates more sharply than the rate of genomic evolution (fig. S5), which is qualitatively consistent with observations (*10*, *11*). The model assumes that individual beneficial mutations sweep sequentially, although "cohorts" of beneficial mutations may co-occur, especially at high μ (*11*, *15*, *16*, *19*). However, the inferred role of diminishing returns in generating population mean-fitness dynamics is unaffected by this complication, because the power-law exponent is independent of μ. Moreover, we have verified by numerical simulations that co-occurring beneficial mutations have no appreciable affect on long-term fitness trajectories over the range of parameters considered here (fig. S6).

Six populations evolved hypermutator phenotypes that increased their point-mutation rates by ~100-fold (*11*). Three of them became hypermutable early in the LTEE (between ~2500 and ~8500 generations) and had measurable fitness trajectories through at least 30,000 generations (table S2). Our model predicts these populations should adapt faster than those that retained the

ancestral mutation rate. We pooled the data from these early hypermutators and confirmed that their composite fitness trajectory was substantially higher than that of the populations with the low mutation rate (Fig. 4). If the hypermutators' beneficial mutation rate also increased by ~100-fold, the difference in trajectories is best fit by an ancestral rate $\mu = 1.7 \times 10^{-6}$ (95% confidence interval 2.5×10^{-7} to 6.1×10^{-5}), although higher values cannot be ruled out (*11*). Note that this fit was obtained by using the same initial distribution of fitness effects, α_0, and epistasis parameter, g, for the hypermutators and the populations that retained the ancestral mutation rate.

Both our empirical and theoretical analyses imply that adaptation can continue for a long time for asexual organisms, even in a constant environment. The 50,000 generations studied here occurred in one scientist's laboratory in ~21 years. Now imagine that the experiment continues for 50,000 generations of scientists, each overseeing 50,000 bacterial generations, for 2.5 billion generations total. At that time, the predicted fitness relative to the ancestor is ~4.7 based on the power-law parameters estimated from all 12 populations (table S4). The ancestor's doubling time in the glucose-limited minimal medium of the LTEE was ~55 min, and its growth commenced after a lag phase of ~90 min (*20*). If the bacteria eliminate the lag, a fitness of 4.7 implies a doubling time of ~23 min (fig. S7). Although that is fast for a minimal medium where cells must synthesize most constituents, it is slower than the 10 min that some species can achieve in nutrient-rich media (*21*). At some distant time, biophysical constraints may come into play, but the power-law fit to the LTEE does not predict implausible growth rates even far into the future. Also, some equilibrium might eventually be reached between the fitness-increasing effects of beneficial mutations and fitness-reducing effects of deleterious mutations (*22*), although it is impossible to predict when for realistic scenarios with heterogeneous selection coefficients, compensatory mutations, reversions, and changing mutation rates.

Fitness may continue to increase because even very small advantages become important over very long time scales in large populations. Consider a mutation with an advantage $s = 10^{-6}$. The probability that this mutation escapes drift loss is ~$4s$ for asexual binary fission (*13*), so it would typically have to occur 2.5×10^5 times before finally taking hold. Given a mutation rate of 10^{-10} per base pair per generation (*23*) and effective population size of $\sim 3.3 \times 10^{-7}$, it would require $\sim 10^8$ generations for that mutation to escape drift and millions more to fix. Also, pleiotropy and epistasis might allow a sustained supply of advantageous mutations, because many net-beneficial mutations have maladaptive side effects that create opportunities for compensatory mutations to ameliorate those effects.

The LTEE uses a simple, constant environment to minimize complications and thus illuminates the fundamental dynamics of adaptation by natural selection in asexual populations. The medium has one limiting resource and supports low population densities (for bacteria) to minimize the potential for cross-feeding on, or inhibition by, secreted by-products. Frequency-dependent interactions are weak in most populations, although stronger in some others (*24*). Also, such interactions should favor organisms that are more fit than their immediate predecessor, but they are not expected to amplify gains relative to a distant ancestor, as fitness was measured here. In fact, such interactions may cause fitness to fall relative to a distant ancestor (*25*). In any case, small-effect beneficial mutations should allow fitness to increase far into the future.

At present, the evidence that fitness can increase for tens of thousands of generations in a constant environment is limited to the LTEE, but these findings have broader implications for understanding evolutionary dynamics and the structure of fitness landscapes. It might be worthwhile to examine fitness trajectories from other evolution experiments in light of our results, although data from short-term experiments may not suffice to discriminate between asymptotic and non-asymptotic trajectories. We hope other teams will perform long experiments similar to the LTEE and that theoreticians will refine our models as appropriate.

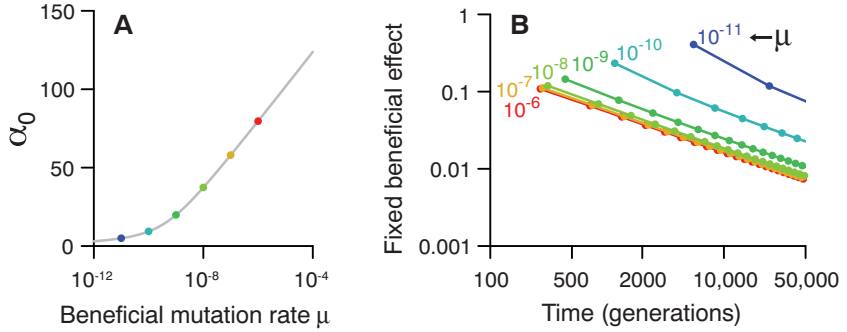

Fig. 3. Theoretical model generating power-law dynamics. (**A**) Parameter pairs for μ and α_0 that match best fit of power law to fitness trajectories for populations that retained ancestral mutation rate for 50,000 generations. (**B**) Expected times and beneficial effects of successive fixations for different pairs that match the best fit. The α_0 values corresponding to each μ are shown in (A). In both panels $g = 6.0$, and $N = 3.3 \times 10^7$.

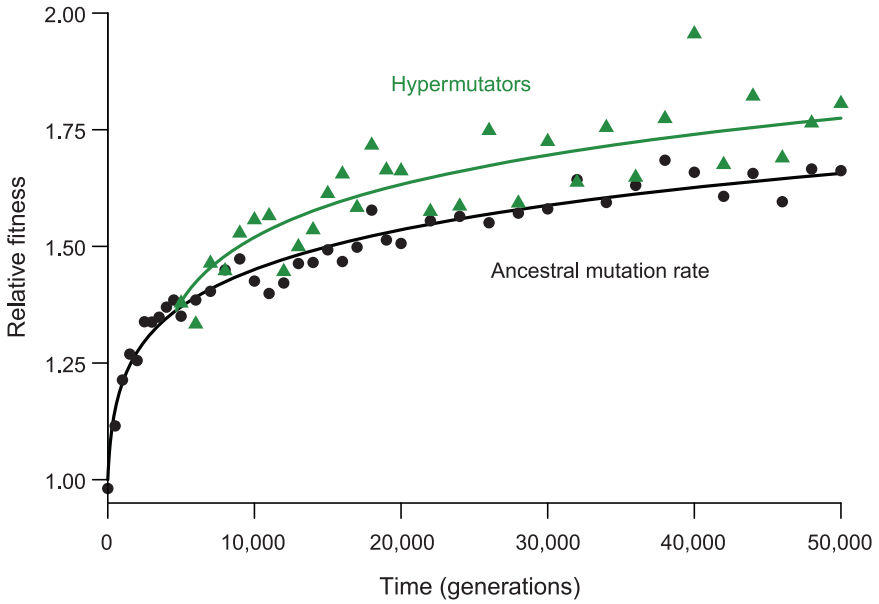

Fig. 4. Effect of hypermutability on observed and predicted fitness trajectories. Black circles: mean fitness of six populations that retained low ancestral mutation rate. Green triangles: mean fitness of three populations that evolved hypermutability early in the LTEE, including one with measurable values through 30,000 generations only. The hypermutators have higher mean fitness at 28 of 31 time points from 5000 to 50,000 generations. Black curve: Predicted trajectory of dynamic model with $\mu = 1.7 \times 10^{-6}$, $\alpha_0 = 85$, $g = 6.0$, and $N = 3.3 \times 10^7$. Green curve: Predicted trajectory with μ increased 100-fold starting at 4667 generations and all other parameters unchanged.

References and Notes

1. R. E. Lenski, M. R. Rose, S. C. Simpson, S. C. Tadler, *Am. Nat.* **138**, 1315–1341 (1991).
2. C. L. Burch, L. Chao, *Genetics* **151**, 921–927 (1999).
3. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science* **312**, 111–114 (2006).
4. S. Kryazhimskiy, G. Tkačik, J. B. Plotkin, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18638–18643 (2009).
5. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, *Science* **332**, 1190–1192 (2011).
6. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, *Science* **332**, 1193–1196 (2011).
7. I. G. Szendro, M. F. Schenk, J. Franke, J. Krug, J. A. G. M. de Visser, *J. Stat. Mech.* **2013**, P01005 (2013).
8. T. J. Kawecki *et al*., *Trends Ecol. Evol.* **27**, 547–560 (2012).
9. R. E. Lenski, M. Travisano, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6808–6814 (1994).
10. J. E. Barrick *et al*., *Nature* **461**, 1243–1247 (2009).
11. Materials and methods and supplementary text are available as supporting material on *Science* Online.
12. P. Sibani, M. Brandt, P. Alstrøm, *Intl. J. Mod. Phys.* **12**, 361–391 (1998).
13. P. J. Gerrish, R. E. Lenski, *Genetica* **102-103**, 127–144 (1998).
14. M. Hegreness, N. Shoresh, D. Hartl, R. Kishony, *Science* **311**, 1615–1617 (2006).
15. S.-C. Park, J. Krug, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18135–18140 (2007).
16. G. I. Lang *et al*., *Nature* **500**, 571–574 (2013).
17. S. Wielgoss *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 222–227 (2013).
18. R. J. Woods *et al*., *Science* **331**, 1433–1436 (2011).
19. M. M. Desai, D. S. Fisher, A. W. Murray, *Curr. Biol.* **17**, 385–394 (2007).
20. F. Vasi, M. Travisano, R. E. Lenski, *Am. Nat.* **144**, 432–456 (1994).
21. R. G. Eagon, *J. Bacteriol.* **83**, 736–737 (1962).
22. S. Goyal *et al*., *Genetics* **191**, 1309–1319 (2012).
23. S. Wielgoss *et al*., *G3 (Bethesda)* **1**, 183–186 (2011).
24. S. F. Elena, R. E. Lenski, *Evolution* **51**, 1058–1067 (1997).
25. C. E. Paquin, J. Adams, *Nature* **306**, 368–370 (1983).

Acknowledgments: This work was supported by grants from the National Science Foundation (DEB-1019989) including the BEACON Center for the Study of Evolution in Action (DBI-0939454), and by funds from the Hannah Chair Endowment at Michigan State University. We thank three reviewers for comments; I. Dworkin, J. Krug, A. McAdam, C. Wilke, and L. Zaman for discussions; and N. Hajela for technical assistance. R.E.L. will make strains available to qualified recipients, subject to completion of a material transfer agreement that can be found at www.technologies.msu.edu/inventors/mta-cda/mta/mta-forms. Datasets and analysis scripts are available at the Dryad Digital Repository (doi:10.5061/dryad.0hc2m).

Supplementary Materials

www.sciencemag.org/content/342/6164/1364/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S4
References (*26*–*40*)

17 July 2013; accepted 4 November 2013
Published online 14 November 2013;
10.1126/science.1243357

Exonic Transcription Factor Binding Directs Codon Choice and Affects Protein Evolution

Andrew B. Stergachis,[1] Eric Haugen,[1] Anthony Shafer,[1] Wenqing Fu,[1] Benjamin Vernot,[1] Alex Reynolds,[1] Anthony Raubitschek,[2,3] Steven Ziegler,[3] Emily M. LeProust,[4]* Joshua M. Akey,[1] John A. Stamatoyannopoulos[1,5]†

Genomes contain both a genetic code specifying amino acids and a regulatory code specifying transcription factor (TF) recognition sequences. We used genomic deoxyribonuclease I footprinting to map nucleotide resolution TF occupancy across the human exome in 81 diverse cell types. We found that ~15% of human codons are dual-use codons ("duons") that simultaneously specify both amino acids and TF recognition sites. Duons are highly conserved and have shaped protein evolution, and TF-imposed constraint appears to be a major driver of codon usage bias. Conversely, the regulatory code has been selectively depleted of TFs that recognize stop codons. More than 17% of single-nucleotide variants within duons directly alter TF binding. Pervasive dual encoding of amino acid and regulatory information appears to be a fundamental feature of genome evolution.

The genetic code, common to all organisms, contains extensive redundancy, in which most amino acids can be specified by two to six synonymous codons. The observed ratios of synonymous codons are highly nonrandom, and codon usage biases are fixtures of both prokaryotic and eukaryotic genomes (*1*). In organisms with short life spans and large effective population sizes, codon biases have been linked to translation efficiency and mRNA stability (*2*–*7*). However, these mechanisms explain only a small fraction of observed codon preferences in mammalian genomes (*7*–*11*), which appear to be under selection (*12*).

Genomes also contain a parallel regulatory code specifying recognition sequences for transcription factors (TFs) (*13*), and the genetic and regulatory codes have been assumed to operate independently of one another and to be segregated physically into the coding and noncoding genomic compartments. However, the potential for some coding exons to accommodate transcriptional enhancers or splicing signals has long been recognized (*14*–*18*).

To define intersections between the regulatory and genetic codes, we generated nucleotide-resolution maps of TF occupancy in 81 diverse human cell types using genomic deoxyribonuclease I (DNaseI) footprinting (*19*). Collectively, we defined 11,598,043 distinct 6– to 40–base pair (bp) footprints genome-wide (~1,018,514 per cell type), 216,304 of which localized completely within protein-coding exons (~24,842 per cell type) (Fig. 1, A and B; fig. S1A; and table S1). Approximately 14% of all human coding bases contact a TF in at least one cell type (average 1.1% per cell type) (Fig. 1C and fig. S1B), and 86.9% of genes contained coding TF footprints (average 33% per cell type) (fig. S1, C and D).

The exonic TF footprints we observed likely underestimate the true fraction of protein-coding bases that contact TFs because (i) TF footprint detection increases substantially with sequencing depth (*13*), and (ii) the 81 cell types sampled, although extensive, is far from complete; we saw little evidence of saturation of coding TF footprint discovery (fig. S2).

To ascertain coding footprints more completely, we developed an approach for targeted exonic footprinting via solution-phase capture of DNaseI-seq libraries using RNA probes complementary to human exons (*19*). Targeted capture footprinting of exons from abdominal skin and mammary stromal fibroblasts yielded ~10-fold increases in DNaseI cleavage—equivalent to sequencing >4 billion reads per sample by using conventional genomic footprinting (fig. S3A)—quantitatively exposing many additional TF footprints (fig. S3, B to D). Overall, we identified an average of ~175,000 coding footprints per cell type (fig. S1E), which is 7- to 12-fold more than with conventional footprinting.

Although coding sequences are densely occupied by TFs in vivo, the density of TF footprints at different genic positions varied widely, with many genes exhibiting sharply increased density in the translated portion of their first coding exon (Fig. 1D and fig. S4A). In contrast, internal coding exons were as likely as flanking intronic sequences to harbor TF footprints (Fig. 1D). The total number of coding DNaseI footprints within a gene was related both to the length of the gene and to its expression level (fig. S4, B to D).

Given their abundance, we sought to determine whether exonic TF binding elements were under evolutionary selection. Fourfold degenerate coding bases are frequently used as a model of neutral (or nearly neutral) evolution (*20*) but

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [2]Department of Immunology, University of Washington, Seattle, WA 98109, USA. [3]Benaroya Research Institute, Seattle, WA 98101, USA. [4]Agilent Technologies, Santa Clara, CA 95051, USA. [5]Department of Medicine, University of Washington, Seattle, WA 98195, USA.

*Present address: Twist Bioscience, San Francisco, CA 94158, USA.

†Corresponding author. E-mail: jstam@uw.edu