Minireview

# Analysis of genetic systems using experimental evolution and whole-genome sequencing
## Matthew Hegreness*† and Roy Kishony*‡

Addresses: *Department of Systems Biology, Harvard Medical School, Longwood Avenue, Boston, MA 02115, USA, †Department of Organismic and Evolutionary Biology and ‡School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.

Correspondence: Roy Kishony. Email: roy_kishony@hms.harvard.edu

## Abstract

The application of whole-genome sequencing to the study of microbial evolution promises to reveal the complex functional networks of mutations that underlie adaptation. A recent study of parallel evolution in populations of *Escherichia coli* shows how adaptation involves both functional changes to specific proteins as well as global changes in regulation.

The comparative study of extant genomes has revolutionized biology, shedding light not only on evolution but also on physiology, genetics and medicine. But the utility of comparisons among naturally evolved isolates is lessened by incomplete knowledge of the environment to which the organisms adapted. Precise knowledge of conditions is attainable only in comparative genomic studies of organisms that have diverged under the controlled conditions of the laboratory, where it is possible to run replicate experiments that distinguish which outcomes are inevitable and which the result of mere chance.

Advanced sequencing and mutation-detection technologies now make it possible to reveal the complete genetic basis for an adaptive trait that separates an evolved clone from a reference strain [1-4]. The first whole-genome sequencing of cellular organisms adapted to controlled laboratory conditions has already revealed mutations that contribute to symbiosis [1] and cooperative behavior [5-7]. A new study by Herring *et al.* [8] takes whole-genome sequencing a significant step further by exploring parallel evolution and its dynamics in replicate populations of *Escherichia coli*. They also provide direct characterizations of the effects of the detected mutations using site-directed mutagenesis. Their results offer clues to how complex biological systems function and evolve, suggesting that adaptive regulation can occur not only at the loci of genes that are directly involved in the adaptive trait but also in distant areas of the network. Whole-genome sequencing of parallel evolved strains promises to reveal novel functional links among genes and genetic modules. Future studies may be able to use genome-sequencing technologies to answer a range of pressing questions in biology and evolution: how biological networks are constructed, constrained, and modified; how clonal interference shapes the outcomes of evolution; and what is the complete spectrum of genetic mutations available to selection.

## The advantages of bacteria for experimental evolution
In 1893, HL Russell, a bacteriologist at the University of Wisconsin, enumerated some of the "evident advantages that bacteria possess for experimental research in evolutionary biology" [9]. These included how the "physical and chemical environment [in which bacteria grow] can be so rigidly controlled that the variability of conditions ...is practically excluded", as well as how, by virtue of short generation times, a "rapid successive transference of cultures to fresh media can secure the effect of an experiment covering an immense number of generations within a limited space of time" [9]. Russell's ideas appear to have remained unrealized for nearly a century, but the field of experimental evolution finally emerged as a vibrant and independent discipline towards the end of the twentieth

century [10]. With advances in the culture and genetic manipulation of microbes, investigators in the 1980s began to let microbes compete and evolve in the laboratory. Early studies used the ability to obtain precise fitness measurements in chemostats to reveal subtle fitness differences associated with natural, induced, and engineered mutations [11], demonstrating the direct link between metabolic flux and fitness [12]. Later experiments, using long-term laboratory evolution of parallel lines, were aimed at the key evolutionary question of how much variability we would expect were we to replay the 'tape of life' [13]; that is, how reproducible are evolutionary outcomes. The most celebrated long-term parallel experiment was begun by RE Lenski in 1988 with 12 replicate populations of *Escherichia coli* and has been running continuously for almost 20 years and more than 40,000 generations in glucose-limited medium [14]. These long-term lines have shed much light on the inherent variability of the evolutionary process at a range of phenotypic levels [14,15]. Now, with recent advances in genomic technologies, these questions have begun to be addressed at the genotypic level [14,16-18]. With whole-genome sequencing, all genetic changes underlying an adaptive trait can be revealed and their dynamics tracked over time. The new study by Herring *et al.* [8] suggests some of the ways in which whole-genome sequencing will provide deeper insight into the connections between parallel evolution at the genotypic and phenotypic levels.

## Parallel adaptation in functional modules

One salient finding that has emerged from laboratory studies of evolving microbes is that parallel evolutionary changes are often seen in replicate populations adapting to a novel environment. Parallel evolution is a hallmark of natural selection: identical or very similar changes reach high frequency or fixation in independent lineages evolving under identical conditions. The use of parallel evolution to infer that adaptation had occurred was first applied to morphological traits [19], but it has been even more convincing in the world of molecules [4,14,20-25]. With their spartan genomes, RNA and DNA viruses were the first organisms for which individual genomes from replicate laboratory populations were fully sequenced. Although whole-genome sequencing reveals all the mutations between an evolved strain and its ancestor, further experimentation is needed to show whether any of these mutations are neutral and how the various mutations combine to form an adaptive trait. In addition to the revelation that the vast majority of mutations that reach appreciable frequency in viral populations are beneficial, such sequencing studies produced striking examples of parallel evolution - often exactly the same change in the same amino acid [26].

It is perhaps not so surprising that we should find a limited set of changes and pervasive parallel evolution in viruses, whose genomes are very small and which lack the complicated regulatory networks of the higher forms of life. Evolution acts on biological function, and in viruses functions are often mapped to single genes. In complex cellular life forms such as bacteria and yeast, however, complex functions are typically attributed to modules of genes [27]. We might expect, therefore, that parallel evolution for cellular life does not necessarily mean similar changes in the same genes but rather similar changes in related modules. For example, recent studies have found that a phenotype under significant positive selection in Lenski's long-term lines is the degree of DNA supercoiling [21]. Although a candidate-gene approach revealed the genes responsible for the changes in supercoiling in some of the evolving strains, the genetic causes underlying the same phenotype in many of his other strains remain obscure [21]. Whole-genome sequencing of these lines could undoubtedly reveal the many different genetic changes that can produce the same parallel phenotypic change in DNA topology, and it could thus unmask the supercoiling gene-module under selection. Through the revelation of parallel cellular phenotypes produced by seemingly dissimilar genetic changes, functional connections within and between genetic modules [28-29] can now be revealed by experimental evolution coupled with whole-genome sequencing.

The current study by Herring *et al.* [8] focuses on metabolism and its regulation. Metabolism provides perhaps the best example of a large cellular network (comprising hundreds of genes) that is relatively amenable to quantitative phenotypic predictions at the whole-cell level [30-33]. Although the overall optimal fitness of adapting populations limited by a given single metabolic resource can be predicted [34,35], only some of the mutations underlying the actual fitness changes appear in the list of candidate genes [34]. Presumably, a regulatory change in a remote location of the network can have far-reaching and unexpected effects.

Using a new microarray-based technology of whole-genome resequencing for identifying the changes between a known and reference strain, Herring *et al.* [8] explore the parallel changes in metabolic and regulatory networks that appeared in five replicate *E. coli* populations that evolved separately in glycerol minimal media for 44 days. This study provides new examples of parallel evolution in candidate genes, but also, as a consequence of the comprehensive sequence information obtained, begins to provide examples of how remote changes might propagate through complex networks and how seemingly disparate changes can have similar phenotypic effects.

Herring *et al.* [8] observed parallel changes in both global regulation patterns and local protein sequences. Resequencing five clones - one clone from each of the replicate populations - the authors identified 13 mutations. A single gene, *glpK* (encoding glycerol kinase), was mutated in all five lineages. The protein synthesized by this gene catalyzes the first step in glycerol catabolism, and the mutations in this gene led to more than 50% increases in the reaction rate of glycerol kinase.

This is an exceptional example of parallel evolution that resonates with the results from experimental viral evolution.

Apart from the glycerol kinase mutations, the most significant mutations identified affected global transcription patterns. The largest fitness changes (representing roughly half of the total increase in growth rate) in any of the five populations resulted from mutations in genes encoding the two major subunits of RNA polymerase (*rpoB* and *rpoC*). In three of the five populations, natural selection fixed a mutant variant of *rpoB* or *rpoC* within 25 days from the start of the experiment. The reason that these changes were beneficial is unknown. Two of these populations subsequently experienced a sweep of secondary mutations that were only conditionally beneficial; these may represent compensatory changes that might have been needed to alleviate the presumed deleterious effects of global changes in gene expression. One of the populations that did not have mutations in RNA polymerase had an 82 base-pair deletion adjacent to *crr*, which encodes critical component in catabolite repression. Herring *et al.* [8] suggest that attenuation of *crr* expression, as well as the mutations in the genes encoding RNA polymerase, may reduce the expression of genes that lead to catabolite repression, which inhibits growth on glycerol. The basis of these effects is still to be identified.

Whole-genome sequencing coupled with the careful control of conditions that is possible in laboratory evolution thus allowed Herring *et al.* [8] to demonstrate how molecular evolution proceeds both *in cis* and *in trans*: that is, adaptation involves local changes to specific proteins (for example, *glpK*) as well as remote regulatory changes.

## Studying the basis of clonal interference by whole-genome sequencing

Herring *et al.* [8] sequenced clonal samples from their populations after 44 days. Sequencing of many clones from each population is still technologically unfeasible. How different would the results have been if it was possible to sequence many different individuals from each evolving population? Bacterial populations invariably show some degree of genetic variability as a result of spontaneous mutation rates and genetic drift of neutral and deleterious alleles. But beneficial mutations are particularly important to population heterogeneity, especially on laboratory timescales. Microbial evolution invariably includes clonal interference among competing lineages, that is, multiple distinct beneficial mutations spread through a population at a given time [25,36-45]. On laboratory timescales during which horizontal transfer of mutations is negligible, beneficial mutations remain linked to the genome in which they appeared, and so the spread of one beneficial mutation can impede the spread of others.

Herring *et al.* [8] recognized that clonal interference shaped the evolution of their populations, and they attempted to discover competing clonal lineages by sequencing the handful of candidate genes suggested by their whole-genome sequencing in search of alternative alleles. They found four alternative *glpK* alleles in two populations. Furthermore, their time course of allele frequency measurements shows several telltale signs of clonal interference, such as transient or permanent decreases in frequency of particular beneficial alleles after an initial rise, indicating competition with a fitter lineage. The independent appearances of mutations in *glpK* and *rpoC* in replicate populations is a less obvious consequence of clonal interference - many beneficial mutations of small effect are probably spreading through the population but do not reach high frequency by the time the strong mutations in *glpK* and *rpoC* spread through the population.

As whole-genome sequencing becomes cheaper and more reliable, it will be easier to study clonal interference as a mechanism affecting the overall rate of adaptation. One question is whether clonal interference happens most frequently between clones with roughly the same phenotype - that is, competition between different genotypic changes in the same specific genes, pathways, or regulatory networks - or whether different phenotypic changes are competing instead.

## The raw material of evolution

In the relatively brief evolutionary timescales and moderate population sizes of studies such as that of Herring *et al.* [8], neutral mutations would not have had time to spread appreciably in the population by genetic drift. Furthermore, although new neutral and deleterious mutations would arise every generation, deleterious and neutral alleles are pushed toward extinction as lineages carrying strongly beneficial alleles spread to fixation. Thus, it is not surprising that the few mutations discovered by Herring *et al.* [8] were all beneficial, and typically of large effect. In addition to studying adaptive mutations, whole-genome sequencing could be used to explore better the actual underlying genotypic spectrum of mutations before selection's winnowing; that is, it could be used to look at what the raw material presented to natural selection is and how it varies across organisms and environments.

Whole-genome sequencing can elucidate the nature of spontaneous mutations when coupled with experimental evolution in mutation accumulation (MA) lines. MA lines are evolved for many generations with as little selection as experimentally feasible [46-49]. This is typically achieved by putting a population through consecutive one-organism bottlenecks every few generations. This allows one to observe how deleterious, neutral and beneficial mutations accumulate without selection. Whole-genome sequencing of MA lines offers considerable promise for seeing the types of mutations that arise in such selection-less experiments. This

will enable geneticists to go from the most basal level, the mutations that compose the raw material for evolution, all the way through gene function and regulation to the ultimate evolutionary phenotype - fitness.

When we see how much whole-genome sequencing has revealed about evolution in nature [50,51], we can imagine how much more can be learned about evolution on a laboratory timescale. By sequencing clones from populations that have evolved in identical laboratory conditions, Herring *et al.* [8] provide further evidence for the ubiquity of parallel evolution on the genotypic level, and their study suggests that remote changes are propagated through genetic systems. Experimental evolution, coupled with genomic technologies, is poised to answer many important questions at the interface between cellular processes and observed evolutionary consequences. Evolution is becoming a powerful tool for studying biological processes, principles and systems.

## Acknowledgements

## References
1. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM: **Accurate multiplex polony sequencing of an evolved bacterial genome.** *Science* 2005, **309:**1728-1732.
2. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen ZT, *et al.*: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437:**376-380.
3. Braslavsky I, Hebert B, Kartalov E, Quake SR: **Sequence information can be obtained from single DNA molecules.** *Proc Natl Acad Sci USA* 2003, **100:**3960-3964.
4. Segre AV, Murray AW, Leu JY: **High-resolution mutation mapping reveals parallel experimental evolution in yeast.** *PLoS Biol* 2006, **4:**1372-1385.
5. Velicer GJ, Raddatz G, Keller H, Deiss S, Lanz C, Dinkelacker I, Schuster SC: **Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor.** *Proc Natl Acad Sci USA* 2006, **103:**8107-8112.
6. Fiegna F, Yu YTN, Kadam SV, Velicer GJ: **Evolution of an obligate social cheater to a superior cooperator.** *Nature* 2006, **441:**310-314.
7. Foster KR: **Sociobiology: the phoenix effect.** *Nature* 2006, **441:** 291-292.
8. Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, *et al.*: **Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale.** *Nat Genet* 2006, **38:**1406-1412.
9. Russell HL: **Bacteriology in its general relation (continued).** *Am Nat* 1893, **27:**1050-1065.
10. Elena SF, Lenski RE: **Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation.** *Nat Rev Genet* 2003, **4:**457-469.
11. Dykhuizen DE, Hartl DL: **Selection in chemostats**. *Microbiol Rev* 1983, **47:**150-168.
12. Dykhuizen DE, Dean AM, Hartl DL: **Metabolic flux and fitness.** *Genetics* 1987, **115:**25-31.
13. Gould SJ: *Wonderful Life: The Burgess Shale and the Nature of History.* New York: WW Norton; 1990.
14. Lenski RE: **Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli.*** *Plant Breeding Rev* 2004, **24:**225-265.
15. Pelosi L, Kuhn L, Guetta D, Garin J, Geiselmann J, Lenski RE, Schneider D: **Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli.*** *Genetics* 2006, **173:**1851-1869.
16. Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M: **Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements.** *Genetics* 2000, **156:**477-488.
17. Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M: **Genomic evolution during a 10,000-generation experiment with bacteria.** *Proc Natl Acad Sci USA* 1999, **96:**3807-3812.
18. Lenski RE, Winkworth CL, Riley MA: **Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations.** *J Mol Evol* 2003, **56:**498-508.
19. Simpson GG: *The Major Features of Evolution.* New York: Columbia University Press; 1953.
20. Cooper TF, Rozen DE, Lenski RE: **Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli.*** *Proc Natl Acad Sci USA* 2003, **100:**1072-1077.
21. Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D: **Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection.** *Genetics* 2005, **169:**523-532.
22. Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE: **Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli.*** *Proc Natl Acad Sci USA* 2006, **103:**9107-9112.
23. Honisch C, Raghunathan A, Cantor CR, Palsson BO, van den Boom D: **High-throughput mutation detection underlying adaptive evolution of *Escherichia coli*-K12.** *Genome Res* 2004, **14:**2495-2502.
24. Ferea TL, Botstein D, Brown PO, Rosenzweig RF: **Systematic changes in gene expression patterns following adaptive evolution in yeast.** *Proc Natl Acad Sci USA* 1999, **96:**9721-9726.
25. Notley-McRobb L, Ferenci T: **Adaptive mgl-regulatory mutations and genetic diversity evolving in glucose-limited *Escherichia coli* populations.** *Environ Microbiol* 1999, **1:**33-43.
26. Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ: **Different trajectories of parallel evolution during viral adaptation.** *Science* 1999, **285:**422-424.
27. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology.** *Nature* 1999, **402:**C47-C52.
28. Segré D, DeLuna A, Church GM, Kishony R: **Modular epistasis in yeast metabolism** *Nat Genet* 2005, **37:**77-83.
29. Yeh P, Tschumi AI, Kishony R: **Functional classification of drugs by properties of their pairwise interactions.** *Nat Genet* 2006, **38:**489-494.
30. Kacser H, Burns JA: **The control of flux.** *Symp Soc Exp Biol* 1973, **27:**65-104.
31. Heinrich R, Schuster S: *The Regulation of Cellular Systems.* Dordrecht: Chapman & Hall; 1996.
32. Kauffman KJ, Prakash P, Edwards JS: **Advances in flux balance analysis.** *Curr Opin Biotechnol* 2003, **14:**491-496.
33. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110.** *Appl Environ Microbiol* 1994, **60:**3724-3731.
34. Dekel E, Alon U: **Optimality and evolutionary tuning of the expression level of a protein.** *Nature* 2005, **436:**588-592.
35. Fong SS, Palsson BO: **Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes.** *Nat Genet* 2004, **36:**1056-1058.
36. Gerrish PJ, Lenski RE: **The fate of competing beneficial mutations in an asexual population.** *Genetica* 1998, 102-**103:**127-144.
37. Muller HJ: **Some genetic aspects of sex.** *Am Nat* 1932, **66:**118-138.
38. Crow JF, Kimura M: **Evolution in sexual and asexual populations.** *Am Nat* 1965, **99:**439-450.
39. Lenski RE, Rose MR, Simpson SC, Tadler SC: **Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2,000 generations.** *Am Nat* 1991, **138:**1315-1341.
40. Miralles R, Gerrish PJ, Moya A, Elena SF: **Clonal interference and the evolution of RNA viruses.** *Science* 1999, **285:**1745-1747.
41. Arjan JA, Visser M, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE: **Diminishing returns from mutation supply rate in asexual populations.** *Science* 1999, **283:**404-406.
42. Hegreness M, Shoresh N, Hartl D, Kishony R: **An equivalence principle for the incorporation of favorable mutations in asexual populations.** *Science* 2006, **311:**1615-1617.
43. Colegrave N: **Sex releases the speed limit on evolution.** *Nature* 2002, **420:**664-666.

44. Imhof M, Schlotterer C: **Fitness effects of advantageous mutations in evolving *Escherichia coli* populations.** *Proc Natl Acad Sci USA* 2001, **98:**1113-1117.

45. de Visser JA, Rozen DE: **Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*.** *Genetics* 2006, **172:**2093-2100.

46. Mukai T, Chigusa SI, Crow JF, Mettler LE: **Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*.** *Genetics* 1972, **72:**335-355.

47. Denver DR, Morris K, Lynch M, Thomas WK: **High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome.** *Nature* 2004, **430:**679-682.

48. Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK: **High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*.** *Science* 2000, **289:**2342-2344.

49. de la Pena M, Elena SF, Moya A: **Effect of deleterious mutation-accumulation on the fitness of RNA bacteriophage MS2.** *Evolution Int J Org Evolution* 2000, **54:**686-691.

50. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, *et al.*: **Analysis of one million base pairs of Neanderthal DNA.** *Nature* 2006, **444:**330-336.

51. Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*.** *Nature* 2004, **428:**617-624.