

Primer

# Can You Sequence Ecology? Metagenomics of Adaptive Diversification

Christopher J. Marx<sup>1,2\*</sup>

**1** Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America, **2** Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge, Massachusetts, United States of America

**Abstract:** Few areas of science have benefited more from the expansion in sequencing capability than the study of microbial communities. Can sequence data, besides providing hypotheses of the functions the members possess, detect the evolutionary and ecological processes that are occurring? For example, can we determine if a species is adapting to one niche, or if it is diversifying into multiple specialists that inhabit distinct niches? Fortunately, adaptation of populations in the laboratory can serve as a model to test our ability to make such inferences about evolution and ecology from sequencing. Even adaptation to a single niche can give rise to complex temporal dynamics due to the transient presence of multiple competing lineages. If there are multiple niches, this complexity is augmented by segmentation of the population into multiple specialists that can each continue to evolve within their own niche. For a known example of parallel diversification that occurred in the laboratory, sequencing data gave surprisingly few obvious, unambiguous signs of the ecological complexity present. Whereas experimental systems are open to direct experimentation to test hypotheses of selection or ecological interaction, the difficulty in “seeing ecology” from sequencing for even such a simple system suggests translation to communities like the human microbiome will be quite challenging. This will require both improved empirical methods to enhance the depth and time resolution for the relevant polymorphisms and novel statistical approaches to rigorously examine time-series data for signs of various evolutionary and ecological phenomena within and between species.

## Introduction

The capacity of current sequencing technologies has revolutionized fields such as microbial ecology and evolution. Research projects and entire careers have been invented. For example, it has now become respectable, indeed fashionable, to sequence poop. Mouse poop, human poop: it is officially a cottage industry. Why? The microbial flora that outnumber our cells 10-fold and have a total gene content 100-fold greater than our own genome are finally getting the credit (or blame) they deserve for the diverse ways in which they affect our health.

But how much can be gleaned from sequencing alone? The direct sequencing of mixed communities (i.e., metagenomics) and subsequent annotation generates fantastic hypotheses of the functions various members are engaged in. From the perspective of population biology, it is thrilling to know that somewhere in the

petabytes of data are the mutations that underlie processes such as evolutionary adaptation or ecological interactions. But which ones? For example, which signals are present in time-course data that could distinguish typical adaptation of a microbe to a single niche from whether it had also diversified into multiple specialists occupying distinct niches? Given the tremendous layers of complexity in our gut community, the challenge is formidable.

## Experimental Evolution as a Model Approach to Understand Natural Communities

Analogous to how classical model systems like *Escherichia coli* and its phage helped unlock the basics of molecular biology, the same sorts of systems have been used to understand fundamental evolutionary processes during adaptation in the laboratory [1]. Most work has been necessarily phenomenological; the genetic basis of adaptation was nearly impossible to uncover prior to genome resequencing. A senior colleague of mine once quipped (and I have previously relayed [2]) that experimental evolution was “population genetics without the genetics.” Times have changed. As with the poop-omics described above, researchers can now sequence isolates [3,4] or mixed samples [5] of evolving populations, thereby uncovering the mutations that occur, as well as changes in their frequencies over time.

Which patterns should be expected from population sequencing in the simplest imaginable scenario: one (asexual) genotype of one species grown on one nutrient in a closed system (without migration)? If I had taken population genetics, I would have been told the gospel that past selection has already rendered most organisms to near perfection, thus almost all new mutations are neutral or deleterious. Beneficial ones are so incredibly rare (and mainly of small effect) that populations would have to wait a substantial time for something good enough to come along and escape random loss. Once established, however, that new rock-star genotype could rise to fixation (perhaps with other, more-or-less neutral mutations that could hitchhike with it), unchallenged as it outcompetes the homogenous sea of unimproved genotypes around it. The mutated genotype would become the new normal, destined to linger until the process repeats itself. This idealized model of steplike improvements is termed “periodic selection”

**Citation:** Marx CJ (2013) Can You Sequence Ecology? Metagenomics of Adaptive Diversification. *PLoS Biol* 11(2): e1001487. doi:10.1371/journal.pbio.1001487

**Published:** February 19, 2013

**Copyright:** © 2013 Christopher J. Marx. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The author received no specific funding for this work.

**Competing Interests:** The author has declared that no competing interests exist.

\* E-mail: cmarx@oeb.harvard.edu

Primers provide a concise introduction into an important aspect of biology highlighted by a current *PLOS Biology* research article.

and, until recently, formed the basis of much of evolutionary theory regarding adaptation [6]. Furthermore, depending upon how many ways a given genotype might improve, replicate populations may fix mutations in parallel functions, genes, or even nucleotides. Indeed, parallelism has been quite commonly observed in evolution experiments [7–12]. Periodic selection would give an extremely clear metagenomic signal: rarely a single new allele would rise in frequency exponentially through time, and after a while, a second one (on the background of the first, Figure 1A). It is a shame that reality does not live up to this ideal.

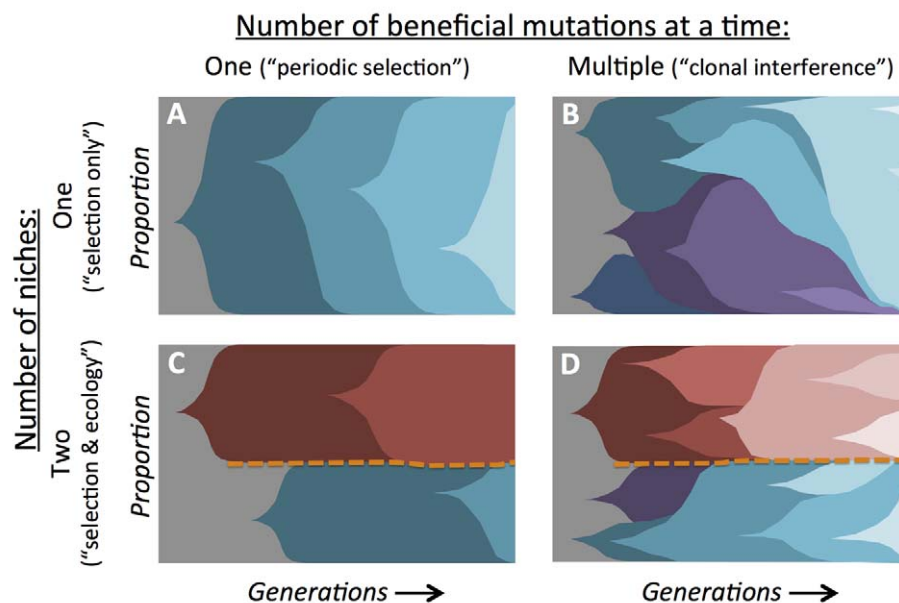
A first complication to periodic selection arises because typical experimental populations have been sufficiently large to have multiple beneficial mutations arise and vie for fixation simultaneously (Figure 1B). Just like when several new companies dive into a market at the same time, your business model has to be both viable and better than those of all of your competitors. Amongst asexual organisms this is known as “clonal interference” [13], and it biases winning mutations toward those with the largest selective effects likely to occur at that population size. Clonal interference also drags out fixation events, providing time for further beneficial variants to arise from competitors before any of them have fixed [14]. This will wreak havoc on metagenomic data. Although there will still be rapid changes in allele frequencies as expected for periodic selection, now there will be many lineages transiently rising and falling as they continue to mutate and compete. There is growing evidence from multiple approaches for exactly these sorts of dynamics [5,15–18].

The second major complication, even in the simple regime of well-mixed environments seeded with a single genotype, is that the ancestral strain can diversify into multiple coexisting ecological specialists. First, this will mean that although some mutations may be generally beneficial, others will only be useful in certain niches.

These may, however, occur repeatedly across replicate populations that diversify. Second, selection can drive a lineage to split into multiple specialists in what is called “adaptive diversification” [19]. This can occur when selection becomes “disruptive”, rewarding divergent phenotypes whose fitness is not absolute, but depends upon the frequency of both types. If this “frequency-dependent” selection is both negative (more fit when rare) and has regimes where either type is the best due to trade-offs, this generates a stable equilibrium between the genotypes. Over the long term, this may result in maintenance of multiple lineages that can each continue to adapt to their niche without eliminating lineages in the other niche(s) [20]. Diversification generates rather complicated metagenomic signatures (Figure 1C), all the more so given clonal interference would also be occurring (Figure 1D). The defining difference is whether alleles sweep through the whole species or only appear to affect some of the species’ lineages. If sequence data cannot distinguish “simple” competition in one niche from adaptive diversification for *E. coli* in a flask, what are our chances of understanding evolution and ecology in a gut from sequencing what comes out of it?

### Looking in Sequence Data for Signs of Ecological Diversification when You Know It Happened

In this issue of *PLOS Biology*, Herron and Doebeli [21] report metagenomic sequencing from 1,200 generations of adaptation and ecological specialization of *E. coli* in the laboratory. One of the key advantages of this study is the backdrop of a rich history of earlier papers that characterized parallel diversification across replicate populations that evolved in a mixture of glucose and acetate [22–25]. Their ancestral strain grows quickly on glucose,



**Figure 1. Dynamics of allele frequencies under different evolutionary and ecological scenarios.** These diagrams indicate the proportion of alleles through time, with each color series representing those that arose from a common first mutation upon the ancestral (gray) genotype. A) The canonical model for adaptation in a single niche has been one of periodic selection, whereby beneficial mutations occur rarely enough that only one ever rises through the population at a time. B) Experimental evolution has repeatedly shown that many beneficial mutations can occur simultaneously and compete with each other before any one of them fixes, a scenario known as clonal interference. C) If multiple ecological niches exist, selection can drive a lineage to split into multiple, coexisting phenotypes (i.e., adaptive diversification). Lineages in each niche are indicated by either warm or cool colors and are separated by an orange dashed line representing the apparent equilibrium. Fixation events occur within each niche without eliminating diversity in the other niche. D) Both clonal interference and ecological diversification can operate simultaneously, giving rise to multiple lineages competing within each niche.  
doi:10.1371/journal.pbio.1001487.g001

and then slowly switches to eating the much less desirable acetate. In each of the ten populations evolved on glucose and acetate, two distinct evolved phenotypes emerged: one that grows even more rapidly on glucose but takes longer to transition to acetate (slow-switchers, SS), and another that is not as fast on glucose as SS but can immediately adjust to grow on acetate (fast-switchers, FS) [24]. Either phenotype can invade the other when rare, coming to a stable equilibrium [22]. Furthermore, both the likelihood of FS emerging [24] and the benefit of particular mutations within this lineage [25] have been shown to depend upon whether the SS phenotype had already evolved. Some of the genetic basis of these phenotypes had been worked out previously [23,25], and this paper extends these analyses by first sequencing a dozen isolates representing known SS or FS phenotypes from three populations. The major data, however, was metagenomic sequencing of time series to test whether raw sequence data could capture that adaptive diversification took place.

Parallel beneficial mutations already gave some hint of adaptation to multiple ecological strategies. While a few genes were targets for beneficial mutations across populations and strategies (distinct deletions of the ribose operon, *Δrbs*, in all but one lineage), others seemed to be specific to each niche. In all three populations, the SS phenotype started with *Δrbs* and mutations in *spoT*, a global regulator of the transition from growth and starvation. The next mutation in the SS lineages was nearly always in *nadR*, which encodes a multifunctional enzyme/regulator of NAD biosynthesis. On the other hand, the FS phenotype always started with changes in acetate metabolism (mutations in one or more of *ackA*, *pta*, or *ptsG*). The repeated observation of the same pair of mutational patterns is consistent with the presence of two ways to improve in all replicate populations.

The temporal dynamics of allele frequencies showed many complicated rises and falls, a few of which clearly indicated ecological interactions. There were multiple lineages, reversals in the direction of allele-frequency changes, and no fixations over 1,200 generations; all of these are qualitatively indistinguishable from previous observations of clonal interference in single-resource environments [5,15–18]. The major signal of ecological diversification, however, came when genotypes rose in frequency to exclude some lineages, but then stabilized with respect to others that appeared to be “immune” to their advantage (like Figure 1C–D). This is a clear violation of transitivity for fitness expected in a single-niche environment, and thus indicates some sort of diversification into multiple niches.

One utility of sequencing is to unveil the evolved alleles that likely caused specialization and the resulting coexistence. A great advantage of laboratory experiments is the ease of directly testing these hypotheses by reconstructing communities with different genotypic (or species) composition. For example, the authors of the present study suggest *nadR* alleles in the SS lineages were beneficial only after the FS lineage arose. Alternatively, since the *nadR* alleles consistently rose after the *Δrbs* and *spoT* mutations occurred in

their own lineage, perhaps their benefit was modified by earlier mutations in their lineage, as has been found in other studies [17,26,27] including one of the authors’ own [25]. So did *nadR* alleles arise because of between-organism coevolution, within-genome epistasis, both of these effects, or neither of them? Thankfully, these sorts of questions can be answered definitively in resynthesized communities.

## Implications for Natural Communities and Future Challenges

For communities that can be observed but not easily manipulated—such as the human gut—can sequencing alone identify adaptation of its members or ecological interactions between them? Despite known adaptive diversification, it should be noted that surprisingly little of the temporal dynamics of the two-niche *E. coli* population unambiguously defy what is possible from simple selection. But are there further, more nuanced aspects of time-series data such as these that would not jibe with simple selection in a single niche? On the empirical side, such quantitative analyses would benefit tremendously from more precise data (more reads per timepoint for the polymorphisms in question) and greater temporal resolution of populations. For example, my laboratory recently developed FREQ-Seq, which barcodes samples and eliminates library preparation in a manner that can generate  $\sim 10^5$  reads per allele per timepoint for thousands of timepoints in a single Illumina lane [28]. On the computational front, there is a clear need for statistical models that can rigorously interpret the temporal dynamics for signs of selection and/or niche differentiation between genotypes of individual species within sequenced communities. These within-species analyses can then be integrated with methods that infer ecological dynamics between species from their correlated abundances [29].

A final fascinating, and somewhat sobering, lesson from Herron and Doebeli is that one species can rapidly evolve to behave like two due to just one or two mutations. Consider the converse situation: that multispecies communities sometimes have been characterized as a much smaller number of “guilds,” comprised of species with relatively similar niches [30]. Collectively, these two concepts would generate a quite fluid scenario whereby one species can quickly act like several; and many already present may act like one. This potential blurring of ecology and evolution implies that beneficial mutations in one species could drive an unrelated species (with a similar niche) to extinction, while sparing extremely closely-related, recently diverged genotypes of its own species. And if this was not headache enough, throw in horizontal gene transfer, which has been inferred to be particularly common in environments such as the gut [31]. It is clear that studies of microbial evolution and ecology in natural communities will remain challenging and interesting for a long time. It is equally clear that systems as simple as “just *E. coli* in a flask” have many lessons left to teach us.

## References

- Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 4: 457–469.
- Marx CJ (2011) Evolution as an experimental tool in microbiology: “Bacterium, improve thyself!” *Env Microbiol Rep* 3: 12–14.
- Velicer GJ, Raddatz G, Keller H, Deiss S, Lanz C, et al. (2006) Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. *Proc Natl Acad Sci USA* 103: 8107–8112.
- Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, et al. (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 38: 1406–1412.
- Barrick JE, Lenski RE (2009) Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harb Symp Quant Biol* 74: 119–129.
- Gillespie JH (2004) Population genetics: a concise guide. 2<sup>nd</sup> ed. Baltimore: Johns Hopkins Press.
- Notley-McRobb L, Ferenci T (1999) Adaptive *mgl*-regulatory mutations and genetic diversity evolving in glucose-limited *Escherichia coli* populations. *Environ Microbiol* 1: 33–43.
- Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ (1999) Different trajectories of parallel evolution during viral adaptation. *Science* 285: 422–424.
- Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, et al. (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99: 16144–16149.
- Zhong S, Khodursky A, Dykhuizen DE, Dean AM (2004) Evolutionary genomics of ecological specialization. *Proc Natl Acad Sci USA* 101: 11719–11724.
- Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE (2006) Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc Natl Acad Sci USA* 103: 9107–9112.

12. Chou H-H, Berthet J, Marx CJ (2009) Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. *PLoS Genet* 5: e1000652. doi:10.1371/journal.pgen.1000652
13. Gerrish PJ, Lenski RE (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* 102–103: 127–144.
14. Desai MM, Fisher DS, Murray AW (2007) The speed of evolution and maintenance of variation in asexual populations. *Curr Biol* 17: 385–394.
15. Rozen DE, de Visser JA, Gerrish PJ (2002) Fitness effects of fixed beneficial mutations in microbial populations. *Curr Biol* 12: 1040–1045.
16. Lang GI, Botstein D, Desai MM (2011) Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* 188: 647–661.
17. Kvitek DJ, Sherlock G (2011) Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet* 7: e1002056. doi:10.1371/journal.pgen.1002056
18. Lee M-C, Marx CJ (2013) Unprecedented clonal interference at a single locus: waves of parallel integrations of an introduced plasmid into the host genome during adaptation. *Genetics*. In press.
19. Doebeli M (2011) *Adaptive diversification*. Princeton, NJ: Princeton University Press.
20. Rozen DE, Schneider D, Lenski RE (2005) Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *J Mol Evol* 61: 171–180.
21. Herron MD, Doebeli M (2013) Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biol* 11: e1001490. doi:10.1371/journal.pbio.1001490
22. Friesen ML, Saxer G, Travisano M, Doebeli M (2004) Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution* 58: 245–260.
23. Spencer CC, Bertrand M, Travisano M, Doebeli M (2007) Adaptive diversification in genes that regulate resource use in *Escherichia coli*. *PLoS Genet* 3: e15. doi:10.1371/journal.pgen.0030015
24. Spencer CC, Tyerman J, Bertrand M, Doebeli M (2008) Adaptation increases the likelihood of diversification in an experimental bacterial lineage. *Proc Natl Acad Sci USA* 105: 1585–1589.
25. Le Gac M, Doebeli M (2010) Epistasis and frequency dependence influence the fitness of an adaptive mutation in a diversifying lineage. *Mol Ecol* 19: 2430–2438.
26. Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332: 1190–1192.
27. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332: 1193–1196.
28. Chubiz LM, Lee M-C, Delaney NF, Marx CJ (2012) *FREQ-Seq*: A rapid, cost-effective, sequencing-based method to determine allele frequencies directly from mixed populations. *PLoS ONE* 7: e47959. doi:10.1371/journal.pone.0047959
29. Friedman J, Alm EJ (2012) Inferring correlation networks from genomic survey data. *PLoS Comput Biol* 8: e1002687. doi:10.1371/journal.pcbi.1002687
30. Terborgh J, Robinson S (1986) Guilds and their utility in ecology. In: Kikkawa J, Anderson DJA, editors. *Community ecology: pattern and process*. Oxford: Blackwell Scientific Publications. pp 65–90.
31. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, et al. (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480: 241–244.