Manuscript Details

Manuscript number	COMICR_2018_2
Title	A decade of genome sequencing has revolutionized studies of experimental evolution
Short title	Sequencing in experimental evolution
Article type	Review article

Abstract

Genome sequencing has revolutionized studies using experimental evolution of microbes because it readily provides comprehensive insight into the genetic bases of adaptation. In this perspective we discuss applications of sequencing-based technologies used to study evolution in microbes, including genomic sequencing of isolated evolved clones and mixed evolved populations, and also on the use of sequencing methods to follow the fate of introduced variations, whether neutral barcodes or variants introduced by genome editing. Collectively, these sequencing-based approaches have vastly advanced the examination of evolution in the lab, as well as begun to synthesize this work with examination of the genetic bases of adaptation and evolutionary dynamics within natural populations.

Keywords	genome sequencing; experimental evolution; adaptation; clonal interference; distribution of fitness effects

Corresponding Author	Christopher Marx
Corresponding Author's Institution	University of Idaho
Order of Authors	Eric Bruger, Christopher Marx
Suggested reviewers	Michael Desai, Dmitri Petrov, Jeffrey Barrick

Submission Files Included in this PDF

File Name [File Type]

Bruger and Marx review MS 180109.doc [Manuscript File]

Bruger and Marx review Figs 171220.ppt [Figure]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Abstract

Genome sequencing has revolutionized studies using experimental evolution of microbes because it readily provides comprehensive insight into the genetic bases of adaptation. In this perspective we discuss applications of sequencing-based technologies used to study evolution in microbes, including genomic sequencing of isolated evolved clones and mixed evolved populations, and also on the use of sequencing methods to follow the fate of introduced variations, whether neutral barcodes or variants introduced by genome editing. Collectively, these sequencing-based approaches have vastly advanced the examination of evolution in the lab, as well as begun to synthesize this work with examination of the genetic bases of adaptation and evolutionary dynamics within natural populations.

Highlights

- Genome sequencing of isolates remains the most straightforward way to determine evolved changes.
- Sequencing of mixed evolved population samples reveals existing variation and its dynamics.
- Active approaches using sequencing to track either neutral, lineage-tracking barcodes or methods to generate variation that can be selected upon have opened the door to novel questions at an unprecedented scale.
- Sequencing approaches have closed the gap between analyses of laboratory-evolved populations and evolution occurring in natural microbial populations.

Experimental evolution before the application of genome sequencing

Studies delving into microbial evolution date back to early experiments involving pond microbes conducted by the reverend William Henry Dallinger in the late 1800s [1-2]. In the second half of the 20th century, pioneered by researchers such as Bruce Levin, Dan Dykhuizen, and colleagues, the use of evolution experiments in the laboratory became increasingly popular [3-8]. The attraction to this approach was the ability to precisely control the selective environment, transfer regime, and initial genotype, thereby seeding replicate populations that could be cryopreserved as a living fossil record. Upon resuscitation, comparisons could then be made through time, between lineages, and across experiments. A tremendous amount was learned about changes in *phenotype* that occur during adaptation, best exemplified by incredibly fruitful series of discoveries from Rich Lenski's long-term evolution experiment (LTEE) with *Escherichia coli* [9-10]. Stepping back from particulars, some commonalities emerged from the LTEE and other similar experiments. Perhaps most

prominently, the rate of adaptation is almost always fastest early in the experiment, and slows as increasing generations accumulate [10,11]. Conversely, other phenomena were found to behave quite differently depending upon the organism and experiment in question, such as whether replicate populations would exhibit parallelism or divergence in phenotypic changes, or in the extent of tradeoffs between fitness in the selective environment versus alternative environments [12]. Unfortunately, there was generally an inability to link these changes in phenotype with mutations that occurred to the *genotype* [9,13].

Although these numerous experimental evolution studies constituted what was then called 'population genetics without the genetics' [14], in the slightly more than a decade since the first application of whole genome sequencing to experimental evolved populations [15] it is hard to imagine anything further from the truth. Genome sequencing and other related sequencing-based technologies have led to unprecedented progress in the study of microbial evolution in the lab [16], and increasingly have been extended to studying natural environments. Here we will first discuss the purely *passive*, observational role that sequencing has played in earlier investigations following changes in experimental populations (Figure 1A). We follow this with a discussion of how sequencing can provide the key output data for experimental designs where the researcher plays an *active* role in generating variation prior to the initiation of adaptation (Figure 1B).

Sequencing individual isolates reveal evolved genotypes

The most straightforward use of genome sequencing to understand evolution is to determine the complete genome sequence of individual evolved isolates. Researchers using viruses as model systems had been using standard Sanger sequencing for this purpose much earlier [17,18], but the use of 454 sequencing to determine the genetic basis of adaptation in an experiment with *Myxococcus xanthus* [15] was the first in a wave of papers using whole genome sequencing to uncover the genetic bases of adaptation in numerous bacterial systems. This approach provides the number, type, and targets of mutations, and it unambiguously reveals that these mutations are linked together as a genotype (Box 1). Assuming genetic manipulation is possible for the organism of interest, it is then possible to parse apart which of these mutations contribute to these phenotypes. These allelic exchange experiments reveal both specific answers about adaptation of a particular organism to a particular environment, and illuminate general trends about adaptation, such as that beneficial mutations are generally less and less beneficial when present upon backgrounds with higher fitness (i.e., diminishing returns epistasis [19,20]).

Whereas obtaining a single whole genome sequence for an evolved isolate was astonishing in 2006, this has

become absolutely trivial at this point, and the low hurdle for sequencing has tremendously altered the types of scientific questions that can be asked. One great advantage has been the ability to sequence isolates from a tremendous number of independent evolution experiments, thereby obtaining a reasonably-sized sample of what is possible for that strain placed in the selective conditions used. For example, by sequencing isolates from 120 separate populations of *Escherichia coli* evolved to grow at an elevated temperature, it became possible to use the occurrence (or nonoccurrence) of mutations together in the same genotype more (or less) frequently than random expectation to reveal positive (or negative) epistasis between them [21] (Figure 2A). This readily revealed multiple distinct evolutionary trajectories that were possible. If the power of sequencing many isolates is instead directed at multiple isolates from multiple timepoints in a single population, it becomes possible to loosely infer clonal dynamics of these populations [22]. Although it was once thought that beneficial alleles arise and escape drift rarely enough that they would rise in frequency and fix one at a time (i.e., periodic selection, [23]), genomic analyses of isolates (and populations, see below) have made it abundantly clear allele dynamics in populations are tremendously messy due to multiple lineages with beneficial mutations arising at the same time and competing with each other (i.e., clonal interference, [24]). Whereas many of these isolate sequencing studies represented the capstone analysis of already well-studied experimental systems, isolate sequencing has become sufficiently trivial that it can even be used as the first 'gateway' step to decide whether an evolution experiment is worth further experimental inquiry. Indeed, in our own group, multiple times we have only proceeded to pursue a story in greater depth after sequencing results revealed that there was an interesting, unexpected physiological basis to adaptation [25], or a surprising similarity between beneficial mutations arising in different environmental or genetic contexts [26]. Looking forward, assuming the relatively high error rates and large regions lacking coverage common to single-cell sequencing [27,28] can be improved, applying singe-cell approaches to evolved populations would constitute a more extreme version of isolate-based sequencing, in that it would still provide linkage information, but also simultaneously give information about the breadth of genetic diversity in the population.

Metagenomic sequencing of populations uncovers genetic diversity and its dynamics

Just as metagenomics has been applied to directly determine the genomic composition of mixed natural communities, it has become increasingly common to simply sequence the total genomic DNA of evolving populations to sample their diversity across time points and/or replicates. What began with analysis of a single population to determine what fraction of evolved diversity fixed or was lost [29] has matured greatly with increased sequencing depth and new analysis pipelines applied across an evolution experiment [e.g. 30]. The tremendous advantage of metagenomics is that it

samples all alleles present with sensitivity that depends upon depth of coverage and allele abundance, such that alleles that emerge to multiple percent of the population can be confidently identified as not simply being sequence errors. The great challenge, however, is that these data lack linkage information between the detected variants, and thus it is not directly clear which alleles are present on the same genetic background. Instead, additional indirect information – such as the correlation between time points in an evolution experiment – are required for the trajectories of alleles and linkage information to be inferred.

Metagenomic analyses of evolving populations have revealed many insights into the nature of evolutionary dynamics in microbes. For example, work by Lang *et al*. [31] demonstrated that selective sweeps often involved whole cohorts of mutations that had accumulated in a lineage, rather than a series of individual beneficial mutations that arose victorious from clonal interference. A recent paper on the Lenski LTEE populations highlighted lessons metagenomics can reveal over the tremendous timescale of 60,000 experimental generations [30] (Figure 2B). Using extremely fine-scaled temporal coverage of the dozen *E. coli* populations evolved in glucose medium allowed for the direct calculation of quantities such as total mutations along lineages, survival probabilities and transit times of fixed alleles, etc. Furthermore, in this experiment, as was examined over a shorter timescale in an earlier experiment with two growth substrates [32], there were abundant clues in the allele dynamics that it was possible to "sequence ecology" [33]. Despite the overlapping complexities of clonal interference there was evidence of adaptive diversification into two ecotypes [34] occupying separate niches due to the fact that selective sweeps were confined to separate subpopulations within the whole population, indicating non-transitivity and likely negative frequency-dependent fitness interactions.

Lineage tracking reveals fate of many subpopulations simultaneously

In order to uncover the distribution of fitness effects (DFE) possible for a given strain in an environment, it requires accurately quantifying a very large number of (initially) rare lineages across many time points, most of which never rise to more than tiny fractions of a percent of the population, even if it means sacrificing the ability to simultaneously identify the causal mutations that arose to generate those dynamics. In this case, rather than whole-genome sequencing, amplicon sequencing of neutral, barcoded loci focuses the available sequencing depth upon just those tagged sites. Sequencing only a 10^2-10^3 bp stretch containing a barcode signatures, implemented as shorthand for a 10^6-10^7 bp genome, increases the sensitivity of detection by 10^4-10^5 fold [35]. One recent influential paper by Levy *et al*. [36] utilized ~500,000 barcoded lineages to capture the DFE of 25,000 beneficial mutations that occurred during the initial adaptation of yeast to

rich media (Figure 2C). The standard expectation had been that the upper tail of the DFE of beneficial mutations would fall exponentially and monotonically, such that big benefit mutations are uniformly rarer than moderately beneficial ones [37]. Instead, their data suggested that the DFE for beneficial mutation was neither monotonic nor exponential, with the mutations rising to high frequency coming from a small number of discrete peaks in the fitness distribution that occur at substantially higher rates than the exponential expectation. The results of the study also suggested that early adaptive dynamics for the populations investigated were deterministic due to the huge crowd of modestly beneficial mutations that almost never gave rise to lineages that would ultimately be successful but regardless drive increases in mean population fitness, and only later on did stochastic effects become more important (drift, timing of large benefit mutations, occurrence of double mutant combinations). The barcode identifiers can also aid in pulling out the individual winning genotypes, from which standard genome sequencing can reveal the putative causative mutations [38]. Because of its utility for tracking many different variants in tandem, the use of lineage tracking with amplicon sequencing will certainly expand in the future, providing researchers unprecedented speed and depth for probing questions about population dynamics.

Sequencing the fate of variation introduced at sites under selection readily reveals genotype to phenotype mapping

Despite the many advantages to experimental evolution, there is generally no way to control several key features that may have inspired one to be interested in evolving the system that they study in the first place, such as which loci will contain the beneficial mutations that emerge, what types of variation will be exposed to selection, or the simultaneous ability to assess fitness consequences of beneficial, neutral, and deleterious alleles. To explore selection upon a target set of genetic variants, there now exist methods to introduce desired alleles and track their fate simultaneously via amplicon sequencing in a manner analogous to the neutral barcodes described above. At the level of individual genes, this combination of gene synthesis techniques and amplicon sequencing is known as “deep mutational scanning” [39]. This allows the fitness consequence of mutations or mutational combinations to be assessed in parallel via representation in sequencing reads before and after selection, and has been applied across all individual variants of entire proteins, or large subsets of possible mutational combinations [39,40]. Most such experiments run a limited number of generations to assay the fitness consequence of the initial variation that was introduced, but could be allowed to run longer to probe the differential ability to further adapt through additional mutations.

To expand analysis of selection to combinations of alleles at multiple genomic locations, *in vivo* gene editing techniques such as CRISPR/Cas, MAGE, MuGENT, and others can be used to change desired loci across bacterial or

archaeal chromosomes [41-46]. These techniques all allow the researcher to choose where in the genome changes are made, as well as what type of variation is introduced, which is a huge step forward from mutant generation techniques like error-prone PCR that produce random mutational changes. These editing technologies have allowed researchers to alter multiple genes in a single process and assess their effects (Figure 2D), or even to make large scale changes across the genome to unrelated genes and analyze their combined effects [47]. If an efficient screen is available for a phenotype other than fitness, such as the enhancing the production of an industrially valuable compound, such as lycopene [41] or PHB [44], then these techniques can have tremendous biotechnological potential. A current limitation in these techniques, however, is the lack of simple methods to obtain linkage information between the edited loci, although there are some promising approaches being developed [48]. If facile approaches to obtain linkage information from these studies emerge, there will be a tremendous potential to use this approach to map from genotypic to phenotypic landscapes [49], perhaps ultimately leading to greater predictability of evolutionary outcomes related to a given trait [50,51].

Sequencing has allowed evolution in natural populations to be tracked in the same way as laboratory experiments.

Historically, studying microbial evolution outside a laboratory setting has been much less tractable than within, and many of the experimental questions posed could not be effectively executed in natural environments. Sequencing technologies have helped to put studies of natural systems on essentially equal footing with those looking at experimental lab populations. Conversely, this also allows researchers in the lab to construct experiments that are more complex, thus beginning to resemble natural environments or communities [52,53]. One clear way in which this can be seen is in examples from the literature in which time series samples from infections have been taken from patients, where strikingly similar patterns to laboratory experiments have been found for phenomena such as parallelism, rates of molecular evolution, within population dynamics such as clonal interference, etc. [e.g 54-59]. This has frequently been done by sequencing multiple isolates in parallel, but can also be extended to sequencing whole populations obtained from patient samples. In terms of active approaches, barcoded and/or pooled variants can be generated and introduced into infection systems such as animal models to test which genotypes are favored in a host [60,61]. In this way, the inclusion of sequencing has the potential to improve our understanding of disease dynamics, as well as aid in diagnostic evaluations of infections in real time and possibly inform better therapeutic intervention strategies.

It was apparent to those of us who worked upon microbial evolution in the ‘pre-next-generation sequencing’ era that cheaper sequencing was coming, and it would be quite useful, but it would have been hard to envision just how

169 transformative it has been. Now the onus is upon us to design and interpret experiments that maximally utilize the wealth of
170 genomic data are available. Genotypes had been 'losing the battle' to phenotypes in terms of what could be learned, but
171 now genotype-based studies have a seemingly insurmountable lead. Let us hope for similarly revolutionary developments
172 in the ability to assay relevant phenotypes quantitatively and in a high-throughput manner so that it can at last become a fair
173 fight.

174

175 **Figure 1.** Different means for applying sequencing approaches to evolution experiments. A) Passive approaches include
176 isolate as well as metagenomic sequencing to capture information on the diversity of mutations that evolve in experimental
177 populations. Figure adapted from [22]. B) Active approaches arise from methods that allow the generation and/or
178 construction of large numbers of initial variants – neutral barcodes or at loci under selection – and tracking them over time.
179 A short experimental timeframe permits observation of the various rates at which deleterious mutations are lost and neutral
180 mutations will remain at steady frequencies, whereas a longer timeframe will see the neutral mutations begin to be squeezed
181 out by the rising mean fitness of the population, but the relative differences in the beneficial mutational effects become
182 more prominent.

183

184 **Figure 2**. Examples of how sequencing has advanced microbial evolution studies. Panel A: Sequencing of many isolates in
185 tandem allowed tracking of the co-occurrence of mutations, which suggested where genes interact epistatically on fitness
186 (reprinted from [21]). Panel B: Extensive metagenomic sequencing from the Lenski LTEE lineages through 60,000
187 generations of experimental evolution allowed the identification of different sublineages and quantification of selective
188 coefficients for evolved mutations (reprinted from [30]). The presence of co-occurring ecotypes (in bold) was evidenced by
189 a deep divergence between two lineages and selective sweeps only occurring within each of these lineages. Panel C: High-
190 throughput lineage tracking with barcode sequences allowed the quantification of fitness effects for many mutations in
191 parallel, giving a clearer picture of the general distribution of these effects (reprinted from [36]). Panel D: MAGE was
192 implemented to introduce multiplexed genotypic changes into *E. coli* promoters (bottom row) that demonstrated differential
193 effects upon both Indigo production ("phentoype" - middle row) and effects upon both growth ("fitness" - top row)
194 (reprinted from [42]).

195

196 **Box 1. What to expect when you sequence evolved isolates?**

Investigators new to using sequencing as part of their experimental studies are often (justifiably) curious about what they should expect to see from their experimental results. Years of isolate sequencing have provided ample information on a number of general trends that consistently crop up in evolution experiments (many of these were highlighted in [29]), including:

- Observed biases toward non-synonymous changes selected more commonly over synonymous changes within genes.

- More mutations in promoters than expected by chance.

- A high proportion of mutations caused by insertion sequence (IS) element transposition and/or homologous recombination between multiple copies of the same IS.

- Parallelism in the loci containing beneficial mutations between replicate lineages, but generally not to the same site/SNPs. This is especially true for loss-of-function alleles that are beneficial for fitness.

- Patterns of mutations and direct allelic exchange experiments indicate an overwhelming pattern of positive selection upon beneficial mutations, with the exception of strains that become mutators. Mutators display a much wider spectrum of mutational effects observed.

Acknowledgements

ELB and CJM recognize support from a NSF (MCB-1714949).

References

[1] Hass JW: **The Reverend Dr William Henry Dallinger, FRS (1839-1909).** *Notes and Records of the Royal Society* 2000, **54:**53-65.

[2] Haas JW: **The Rev. Dr. William H. Dallinger FRS: early advocate of theistic evolution and foe of spontaneous generation'.** *Perspectives on Science and Christian Faith* 2000, **52:**107-117.

[3] J. Monod: *Recherches sur la croissance des cultures bactériennes.* Hermann & Cie; 1942.

[4] Monod J: **La technique de culture continue: Théorie et applications.** *Ann. Inst. Pasteur* 1950, **79:**390-410.

[5] Levin BR, Stewart FM, Chao L: **Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage.** *The American Naturalist* 1977, **111:**3-24.

[6] Dykhuizen D, Hartl D: **Evolution of competitive ability in *Escherichia coli*.** *Evolution* 1981, **35:**581-594.

[7] Hall BG, Zuzel T: **Evolution of a new enzymatic function by recombination within a gene.** *Proceedings of the National Academy of Sciences* 1980, **77:**3529-3533.

[8] Horne MT: **Coevolution of *Escherichia coli* and bacteriophage in chemostat culture.** *Science* 1970, **168:**992-993.

*[9] Elena SF, Lenski RE: **Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation.** *Nature Reviews Genetics* 2003, **4:**457-469.

*Review encapsulating major ideas about the dynamics and genetics of adaptive changes gleaned from experimental evolution studies predating major recent advances in sequencing technologies.

*[10] Lenski RE, Rose MR, Simpson SC, Tadler SC: **Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations.** *The American Naturalist* 1991, **138:**1315-1341.

*A fundamental study examining the adaptive changes of multiple bacterial populations in parallel over a non-trivially long time period.

[11] Lenski RE, Travisano M: **Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations.** *Proceedings of the National Academy of Sciences* 1994, **91:**6808-6814.

[12] Kassen R: **The experimental evolution of specialists, generalists, and the maintenance of diversity.** *Journal of Evolutionary Biology* 2002, **15:**173-190.

[13] Pardo-Diaz C, Salazar C, Jiggins CD: **Towards the identification of the loci of adaptive evolution.** *Methods in*

Ecology and Evolution 2015, **6:**445-464.

[14] Marx CJ: **Evolution as an experimental tool in microbiology:'Bacterium, improve thyself!'.** *Environ Microbiol Rep* 2011, **3:**12-14.

*[15] 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.** *Proceedings of the National Academy of Sciences* 2006, **103:**8107-8112.

*First study to utilize whole-genome sequencing to investigate adaptive changes in clonal isolates from evolved *Myxococcus xanthus* popuations.

*[16] Barrick JE, Lenski RE: **Genome dynamics during experimental evolution.** *Nature Reviews Genetics* 2013, **14:**827-839.

*Important review that covers many concepts related to evolutionary and ecological forces driving genome adaptation that have been illuminated by genome sequencing approaches.

[17] Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ: **Exceptional convergent evolution in a virus.** *Genetics* 1997, **147:**1497-1507.

[18] Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ: **Different trajectories of parallel evolution during viral adaptation.** *Science* 1999, **285:**422-424.

[19] Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ: **Diminishing returns epistasis among beneficial mutations decelerates adaptation.** *Science* 2011, **332:**1190-1192.

[20] Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF: **Negative epistasis between beneficial mutations in an evolving bacterial population.** *Science* 2011, **332:**1193-1196.

[21] Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS: **The molecular diversity of adaptive convergence. *Science* 2012, **335:**457-461.

**A prominent example of the power of applying genome sequencing of isolates at a deep level to investigate fundamental biological questions, using statistical patterns from many independent populations to uncover epistasis and its effect on constraining adaptive trajectories, in this case in *Escherichia coli* populations in response to temperature stress.

*[22] Maddamsetti R, Lenski RE, Barrick JE: **Adaptation, clonal interference, and frequency-dependent interactions in**

a long-term evolution experiment with *Escherichia coli*. *Genetics* 2015, **200:**619-631.

*The application of isolate sequencing over the time course in a particular evolved population of *E. coli* reveals critical information on phenomena affecting the adaptive dynamics of this population.

[23] Atwood KC, Schneider LK, Ryan FJ: **Periodic selection in *Escherichia coli*.** *Proceedings of the National Academy of Sciences* 1951, **37:**146-155.

[24] Gerrish PJ, Lenski RE: **The fate of competing beneficial mutations in an asexual population.** *Genetica* 1998, **102:**127-144.

[25] Michener JK, Neves AA, Vuilleumier S, Bringel F, Marx CJ: **Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post–transfer refinement.** *elife* 2014, **3:**e04279.

[26] Nayak DD, Agashe D, Lee MC, Marx CJ: **Selection maintains apparently degenerate metabolic pathways due to tradeoffs in using methylamine for carbon versus nitrogen.** *Current Biology* 2016, **26:**1416-1426.

[27] Ning L, Liu G, Li G, Hou Y, Tong Y, He J: **Current challenges in the bioinformatics of single cell genomics.** *Frontiers in Oncology* 2014, **4**.

[28] Gawad C, Koh W, Quake SR: **Single-cell genome sequencing: current state of the science.** *Nature Reviews Genetics* 2016, **17:**175-188.

[29] Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF: **Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 2009, **461:**1243-1247.

**First detailed application of genomic sequencing of isolates to a long-term experimentally evolved population. The study also provides a detailed accounting of the adaptive changes encountered by the examined experimental population over the course of 40,000 generations.

[30] Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM: **The dynamics of molecular evolution over 60,000 generations. *Nature* 2017, **551:**45–50.

**Exceptional temporal sequencing depth is applied to LTEE *E. coli* populations to produce detailed trajectories of genetic changes occurring within these populations.

[31] Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM: **Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 2013, **500:**571-574.

**A principal example of the potential behind using of metagenomic sequencing for detecting within population changes, this paper revealed that lineages that selectively sweep populations are often composed of cohorts of beneficial mutations rather than alternating sweeps by single mutational changes.

[32] Herron MD, Doebeli M: **Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*.** *PLoS Biology* 2013, **11:**e1001490.

[33] Marx CJ: **Can you sequence ecology? Metagenomics of adaptive diversification.** *PLoS Biology* 2013, **11:**e1001487.

[34] Cohan FM: **Bacterial species and speciation.** *Systematic Biology* 2001, **50:**513-524.

[35] Chubiz LM, Lee MC, Delaney NF, Marx CJ: **FREQ-Seq: a rapid, cost-effective, sequencing-based method to determine allele frequencies directly from mixed populations.** *PLoS One* 2012, **7:**e47959

[36] Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G: **Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 2015, **519:**181-186.

**Demonstration of the utility of applying massively parallel barcoding to track early evolutionary dynamics, thereby uncovering the initial distribution of fitness effects of beneficial mutations in that environment.

[37] Orr HA: **The distribution of fitness effects among beneficial mutations.** *Genetics* 2003, **163:**1519-1526.

[38] Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Hérissant L, Blundell JR, Levy SF, Fisher DS: **Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast.** *Cell* 2016, **166:**1585-1596.

[39] Araya CL, Fowler DM: **Deep mutational scanning: assessing protein function on a massive scale.** *Trends in Biotechnology* 2011, **29:**435-442.

[40] Wrenbeck EE, Klesmith JR, Stapleton JA, Adeniran A, Tyo KE, Whitehead TA: **Plasmid-based one-pot saturation mutagenesis.** *Nature Methods* 2016, **13:**928-930.

[41] Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM: **Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 2009, **460:**894-898.

**Herein, multiplexed gene editing using MAGE approach in *E. coli* is used to rapidly introduce variation pertinent to the 1-deoxy-D-xylulose-5-phosphate (DXP) biosynthesis pathway.

*[42] Wang HH, Kim H, Cong L, Jeong J, Bang D, Church GM: **Genome-scale promoter engineering by coselection**

MAGE. *Nature Methods* 2012, **9:**591-593.

*Further demonstration of the use of MAGE to quickly introduce variation on the genome scale to introduce promoter changes at multiple operons.

*[43] Dalia AB, McDonough E, Camilli A: **Multiplex genome editing by natural transformation.** *Proceedings of the National Academy of Sciences* 2014, **111:**8937-8942.

*Another multiplexed genome editing technique that depends on natural competence, MuGENT, is introduced to rapidly introduce multiple mutational changes into the genome of *Vibrio cholerae*.

[44] Dalia TN, Hayes CA, Stolyar S, Marx CJ, McKinlay JB, Dalia AB: **Multiplex genome editing by natural transformation (MuGENT) for synthetic biology in *Vibrio natriegens*.** *ACS Synthetic Biology* 2017, **6:**1650–1655.

[45] Nayak DD, Metcalf WW: **Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*.** *Proceedings of the National Academy of Sciences* 2017, **114:**2976-2981.

[46] Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA: **RNA-guided editing of bacterial genomes using CRISPR-Cas systems.** *Nature Biotechnology* 2013, **31:**233-239.

[47] Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ: **Precise manipulation of chromosomes in vivo enables genome-wide codon replacement.** *Science* 2011, **333:**348-353.

[48] Zeitoun RI, Pines G, Grau WC, Gill RT: **Quantitative tracking of combinatorially engineered populations with multiplexed binary assemblies.** *ACS Synthetic Biology* 2017, **6:**619-627.

[49] Orgogozo V, Morizot B, Martin A: **The differential view of genotype–phenotype relationships.** *Frontiers in Genetics* 2015, **6**.

[50] De Visser JA, Krug J: **Empirical fitness landscapes and the predictability of evolution.** *Nature Reviews Genetics* 2014, **15:**480-490.

[51] Lässig M, Mustonen V, Walczak AM: **Predicting evolution.** *Nature Ecology & Evolution* 2017, **1:**0077.

[52] Datta MS, Sliwerska E, Gore J, Polz MF, Cordero OX: **Microbial interactions lead to rapid micro-scale successions on model marine particles.** *Nature Communications* 2016, **7:**11965.

[53] Cordero OX, Ventouras LA, DeLong EF, Polz MF: **Public good dynamics drive evolution of iron acquisition**

strategies in natural bacterioplankton populations. *Proceedings of the National Academy of Sciences* 2012, **109:**20059-20064.

[54] Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis Jr MR, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, McAdam AJ: **Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genetics* 2011, **43:**1275-1280.

**A premier example of how sequencing approaches can be leveraged to yield information about the evolution of bacterial infections within patients.

[55] Xue KS, Stevens-Ayers T, Campbell AP, Englund JA, Pergam SA, Boeckh M, Bloom JD: **Parallel evolution of influenza across multiple spatiotemporal scales.** *eLife* 2017, **6**.

[56] Silva IN, Santos PM, Santos MR, Zlosnik JE, Speert DP, Buskirk SW, Bruger EL, Waters CM, Cooper VS, Moreira LM: **Long-term evolution of *Burkholderia multivorans* during a chronic cystic fibrosis infection reveals shifting forces of selection.** *mSystems* 2016, **1:**e00029-16.

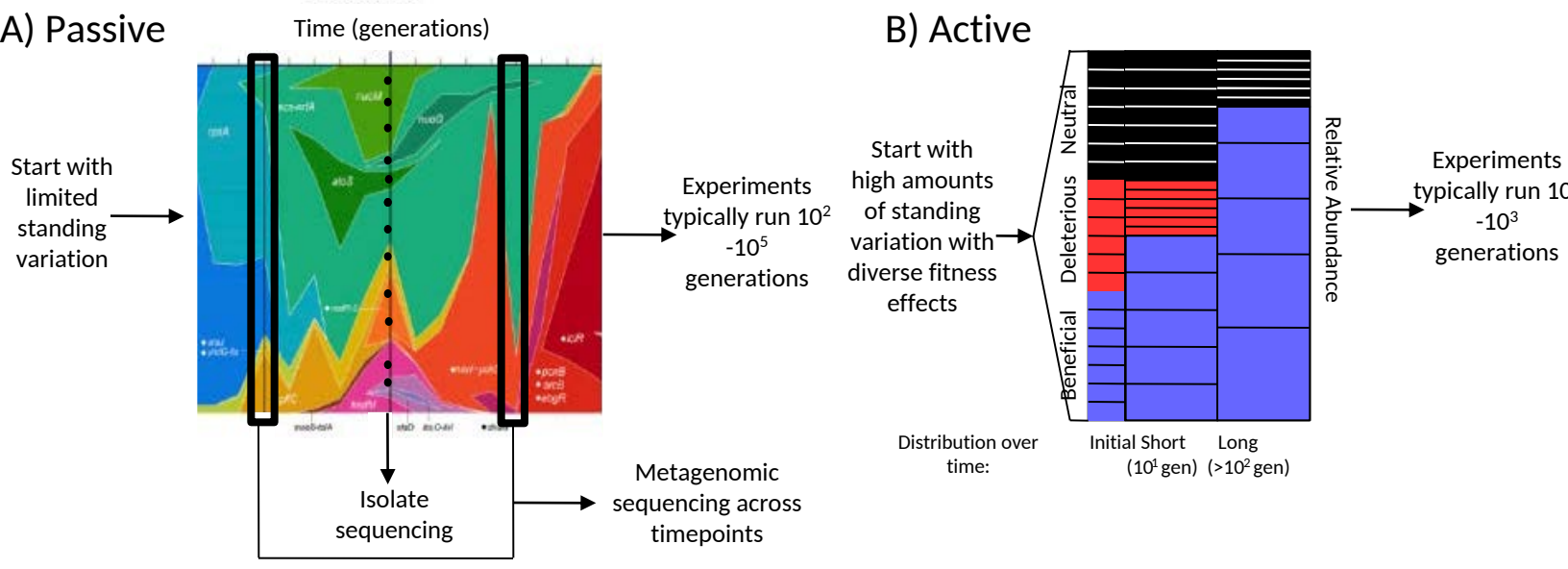
[57] Zhao S, Lieberman TD, Poyet M, Groussin M, Gibbons SM, Xavier RJ, Alm EJ: **Adaptive evolution within the gut microbiome of individual people.** *bioRxiv* 2017, Jan 1:208009.

[58] Marvig RL, Sommer LM, Molin S, Johansen HK: **Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis.** *Nature Genetics* 2015, **47:**57-64.

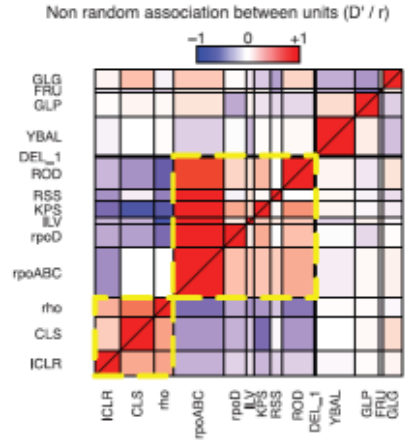
[59] Levade I, Terrat Y, Leducq JB, Weil AA, Mayo-Smith LM, Chowdhury F, Khan AI, Boncy J, Buteau J, Ivers LC, Ryan ET, Charles RC, Calderwood SB, Qadri F, Harris JB, LaRocque RC, Shapiro BJ: ***Vibrio cholerae* genomic diversity within and between patients.** *Microbial Genomics* 2017, **3**.

[60] Blundell JR, Levy SF: **Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer.** *Genomics* 2014, **104:**417-430.

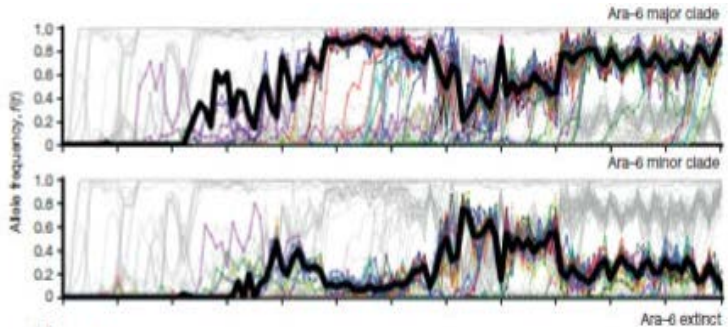
[61] Van Opijnen T, Camilli A: **Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms.** *Nature Reviews Microbiology* 2013, **11:**435-442.


A) Passive
Time (generations)
Start with limited standing variation
Experiments typically run 10^2 -10^5 generations
Isolate sequencing
Metagenomic sequencing across timepoints
B) Active
Start with high amounts of standing variation with diverse fitness effects
Neutral
Deleterious
Beneficial
Relative Abundance
Experiments typically run 10 -10^3 generations
Distribution over time:
Initial Short (10^1 gen)
Long (>10^2 gen)

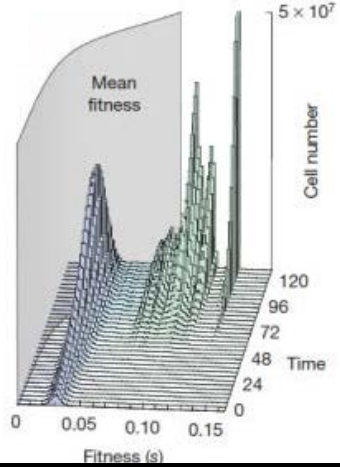
A) Isolate Sequencing



B) Metagenomic Sequencing



C) Lineage Tracking



D) Generation of Variation