Experimental evolution of *Methylobacterium*: fifteen years of planned experiments and surprise findings

Christopher J. Marx

Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, Center for Modeling Complex Interactions, University of Idaho, Moscow, ID, 83844. USA

cmarx@uidaho.edu

Abstract

Experimental evolution has become an increasingly common approach for studying evolutionary phenomena, as well as uncovering physiological connections in a manner complementary to traditional genetics. Here I describe the development of *Methylobacterium* as a model system for using experimental evolution to study questions at the intersection of metabolism and evolution. Each experiment was initiated to address a particular question inspired by patterns in natural methylotrophs, such as tradeoffs between single-carbon and multi-carbon growth, or the challenges involved in incorporating novel metabolic pathways or genes with poor codon usage that are acquired via horizontal gene transfer. What I could not have appreciated initially, however, was just how many fortuitous surprise findings would emerge. These have ranged from the repeatability of evolution, complex dynamics within populations, epistasis between beneficial mutations, and even the ability to use simple mathematical models to generate testable, quantitative hypotheses about the fitness landscape.

Running title: Experimental evolution of *Methylobacterium*

Experimental evolution of populations in the laboratory allows a researcher to simultaneously address evolutionary and physiological questions. The great advantage from an evolutionary perspective is that – in a typical experimental design – replicate populations initially have no within- or between-population genetic variation, and the selective conditions are under the control of the experimenter. This allows for a "reductionist" approach to evolutionary questions, whereby the influence of one or a few individual factors upon the outcome can be ascertained (reviewed in Lenski, 2017). From a physiological perspective, experimental evolution is simply a patient version of a genetic selection experiment (reviewed in Marx, 2011). Rather than requiring a discrete change in phenotype to be immediately apparent upon plating, the continued transfers of the experiment permit mutations of "modest" effect – such as a 10% increase in growth rate – to occur, escape drift, and rise toward fixation. Furthermore, the advent of high-throughput sequencing has revolutionized the ability to address both the evolutionary and physiological questions (reviewed in Bruger and Marx, 2018).

The kind invitation I received to write this chapter was a request to specifically describe the work in my laboratory where we have repeatedly used experimental evolution with *Methylobacterium extorquens* to address both evolutionary and physiological questions. Given that charge, I will shamelessly describe themes arising from our own work, but my primary goal is to highlight two broader messages. The first message is that experimental evolution can be utilized as a complementary approach to address some of the same questions that emerge from studying natural populations. Second, the unexpected, surprise results from experimental evolution can be as impactful as answering the questions that were part of the intended experimental design. As such, I have divided the topics along these lines – the planned questions and the answers we uncovered (section ***1***), and then the several categories of surprise findings (section ***2***) – but have tried to illuminate the connections between these as best as I can.

1. The planned: experimental evolution to address questions inspired by natural methylotrophs

Thus far, my laboratory has published experimental evolution stories using *M. extorquens* that were designed to address three major topics: evolution of metabolic tradeoffs, adaptation to utilize a foreign metabolic pathway, and the use of experimental evolution to illuminate how codon usage can evolve.

1.1 Tradeoffs between C_1 and multi-C metabolism.

My initial foray into experimental evolution was to explore metabolic tradeoffs between C_1 and multi-C metabolism. Many methylotrophs have a fairly limited breadth of multi-C use. In particular, although *Methylobacterium* are a member of Bradyrhizobiales, which generally have a wide substrate breadth, the major model systems of *M. extorquens*, strains AM1 and PA1 (Peel and Quayle, 1961; Knief et al., 2010; Nayak and Marx, 2014a), have not been reported to use many common multi-C substrates, such as sugars. Having studied the function of C_1 pathways as a grad student in Mary Lidstrom's lab, when I applied for an NSF postdoctoral fellowship to work with Rich Lenski, I focused half of my proposal upon exploring whether C_1 vs. multi-C tradeoffs would emerge over the short term in the laboratory, if they would be consistent and symmetric between substrates, and whether they would be driven by selection or arise neutrally due to a lack of selection.

Before moving to our results, how was this experimental evolution – and the other ones mentioned below – actually carried out? For this experiment (Lee et al., 2009), I set up eight populations to evolve on methanol (15 mM) in minimal medium, and eight to evolve on succinate (3.5 mM). These somewhat lower concentrations than usual were chosen for two reasons: to stay within the regime where final cell density scales linearly with substrate concentration, and to prevent large changes in medium pH (which would exert a second selective pressure). Critically, each population was initiated from a single colony of the given ancestral strain, thus ensuring that if any identical mutations emerged between replicates, these could be interpreted as having risen independently, rather than from shared ancestry, as in a Luria-Delbrück experiment (Luria and Delbrück, 1943). Culturing took place in 9.6 ml of liquid medium in 50 ml flasks, shaken, at 30 °C. A volume of 150 µl was transferred to fresh medium

every two days, which permitted a 2^6 = 64-fold net increase in cells, and thus six generations per transfer. These particular populations were carried forward for 1500 generations in this manner. There are a couple differences in experimental design I will note later, but otherwise, this style of batch transfer experiments were the norm throughout what I present here. Unsurprisingly, given where I did my postdoc, this style of experimental evolution is analogous to that used during the very well-studied Lenski Long-term experimental evolution populations of *Escherichia coli* (Lenski et al., 1991; Lenski, 2017).

The second generic aspect to convey is how fitness is assayed, as natural selection depends upon the fitness (i.e., relative net reproductive success) of genotypes. The classical method for assaying fitness is a pairwise competition between the strain of interest and a common competitor, generally the ancestor of the populations. These strains must be distinguishable in order to enumerate each of them in a mixture, and we did possess the ability to use pigmentation and count colonies (Van Dien et al., 2003). Given the time intensity of plating and counting, however, as well as the large measurement noise inherent in counting relatively small numbers of colonies, for our first paper on experimental evolution we invested time in developing a flow cytometry-based method (Lee et al., 2009). This allowed us to use a fluorescently-labeled ancestor as a common competitor against a non-fluorescent evolved isolates (or a genetically-constructed strain). If possible, even better resolution can be obtained by using two different fluorescent proteins, one in each competitor, but this was not yet established in our early work. We found that either Venus or mCherry expressed at a modest level from a chromosomal locus was effectively neutral (fitnesses between 0.995-1.005) and could be measured accurately (95% confidence intervals of less than 0.3% of the mean).

Finally, it is critical to realize that fitness is not synonymous with maximal growth rate, but in this sort of evolution regime the two are often very highly correlated. Under batch culture conditions in well-mixed medium, exponential growth rate is the primary fitness component, with lag or survival in stationary phase generally contributing to a lesser extent (Vasi et al., 1994). Note that final yield (in terms of biomass converted or cfu/ml) is not a fitness component during batch growth (Vasi et al., 1994). The Malthusian fitness parameter, *W*, is calculated from the change in the ratio of competitors from the initial to final timepoint, as well as the number of net generations. As such, *W* is a relative, "per generation" fitness value averaged over the entire growth cycle. For example, a fitness of 1.2 would indicate the strain in question is 20% more fit than the one it is being compared to, such as its ancestor. If the two strains only differ in growth rate, this would also equate to a 20% difference in growth rates. Indeed, in many experiments we have observed a tremendously tight correlation between fitness measured in a competition and growth rate measured individually (e.g., R^2 of 0.98; Agashe et al., 2013). If, however, there are interactions between strains beyond competition, such as toxins or cross-fed metabolites, this makes fitness both non-transitive across multiple strains, and frequency-dependent between strains (i.e., different values depending upon the ratios of competitors). Furthermore, the fitness can even depend upon the absolute densities of competitors rather than their ratio, as discussed below (***1.2.1***). There was very little of these sort of complex non-transitive fitness effects for most of the work I describe here on *M. extorquens* alone, but it can

be quite common in other scenarios. For example, non-transitive fitness interactions between genotypes were the norm in as a series of experiments we performed with two- and three-species consortia that exchanged metabolites as public goods in spatially-structured environments (Harcombe et al., 2014; Harcombe et al., 2016; Douglas et al., 2016; Douglas et al., 2017).

Having covered how the experimental evolution was performed, what did we uncover regarding tradeoffs in performance on C_1 vs. multi-C substrates? This work, led by a former graduate student, Ming-Chun "Miki" Lee, generated two primary findings (Lee et al., 2009). First, the tradeoffs were asymmetric: succinate-evolved became worse in methanol, but the methanol-evolved became better in succinate. Correlated improvements in substrates other than the one experienced in the selective environment are in fact quite common in such experiments, even over quite long timescales (e.g., Leiby and Marx, 2014). There are many dimensions to the selective environment (temperature, media, gas exchange, timescale of transferring) that are identical even when testing different substrates. In this regard, the methanol-evolved populations simply improved to the overall lab conditions. The second primary finding was that, although the succinate-evolved populations on *average* decreased in performance on methanol, this obscured a bimodal evolutionary response. Approximately half of the populations actually improved in growth on methanol, whereas the other half entirely lost the ability to grow on methanol at all!

Was the loss of ability to grow on methanol a side-product of a mutation that was beneficial during succinate growth (and was thus driven by selection, a phenomenon known as antagonistic pleiotropy), or did it occur due to lack of purifying selection to maintain methylotrophy in a succinate-only world (and thus occurred neutrally, a phenomenon known as mutation accumulation). Subsequent genome sequencing of one isolate revealed a loss-of-function mutation in formate-tetrahydrofolate ligase (encoded by *ftfL*; Marx et al., 2003a; Marx et al., 2003b). Miki Lee discovered that all of the methanol-minus isolates – across independent populations – had non-identical mutations that caused loss-of-function of *ftfL*. So were these *ftfL* null alleles beneficial during succinate growth? A former postdoc, Sean Carroll, picked up this work and found that there was not a simple answer. A loss-of-function *ftfL* allele was beneficial in the strains in which it emerged during succinate growth, but the same allele was neutral, or even deleterious, in strains that had retained the ability to grow on methanol. This dependence of a mutation's phenotypic effect upon other alleles – i.e., epistasis – indicated that one (or more) of the other beneficial alleles that had arisen during succinate adaptation either opened the door to selection-driven loss of methylotrophy, or would prevent it, at least from this mechanism. We thus had a story that began due to a desire to explore tradeoffs, and ended up uncovering a key role for epistasis in the modest level of evolutionary repeatability, themes covered in more detail in ***2.3*** and ***2.1***, respectively.

1.2 Adaptation requiring use of an introduced enzyme or metabolic pathway.

The second half of my postdoctoral fellowship proposal focused upon the concept of using an "engineered horizontal gene transfer" to study how bacteria adapt to use foreign metabolic

pathways. This was written in 2002, when the appreciation of horizontal gene transfer from phylogenetic and genomic analyses was sky-rocketing, but we had very little understanding of the evolutionary steps that proceed immediately after gene acquisition in the cases that turn out to be successful (which I co-authored a review about; Michener and Marx, 2015). Beyond my initial project involving formaldehyde oxidation, three other systems involving primary oxidation enzymes for C_1 compounds were later established by members of the lab, and I will discuss each of these in turn.

1.2.1 Swapping formaldehyde oxidation pathways.

The first example of adaptation with a foreign pathway involved swapping unrelated formaldehyde oxidation pathways. During my graduate work, I introduced a pathway with known function to try to decipher the roles of two apparently redundant pathways with unclear functions (Marx et al., 2003c). The glutathione- (GSH)-dependent formaldehyde oxidation pathway from a distantly-related methylotroph, *Paracoccus denitrificans*, is known to function irreversibly in the dissimilatory direction (Ras et al., 1995; Harms et al., 1996). *M. extorquens* , on the other hand, has two pathways that were thought to function in formaldehyde oxidation: the tetrahydromethanopterin- (H_4MPT)-dependent C_1 transfer pathway and the tetrahydrofolate- (H_4F)-dependent pathway. Mutants with lesions in either pathway failed to grow on methanol, despite that fact that either pathway was thought at the time to be sufficient for formaldehyde oxidation (Chistoserdova and Lidstrom, 1994; Chistoserdova et al., 1998; Marx et al., 2003b; Marx et al., 2003c; Marx and Lidstrom, 2004). When I introduced the foreign GSH-dependent pathway on an expression plasmid into mutants defective for one of these two pathways, I found that it could complement the H_4MPT-dependent pathway mutants, but not the H_4F-dependent pathway ones (Marx et al., 2003c). This, combined with analytical chemistry experiments to follow radio- or stable-isotopes through the cell led to the realization that only the H4MPT-dependent pathway functions oxidatively in vivo; the H_4F-dependent pathway functions in the reverse direction to feed the serine cycle, and thus the flow of carbon splits to assimilation versus dissimilation at the level of formate, not formaldehyde (Marx et al., 2003b; Marx et al., 2005; Crowther et al., 2008). The replacement of the H_4MPT-dependent pathway with the GSH-dependent one meant that an unrelated set of enzymes could handle the central flow of carbon in the cell, yet it did so substantially slower (3-fold) than wild-type, suggesting there may be clear room for evolutionary improvement.

What I proposed in my fellowship application was that the slowed growth of the GSH-dependent *M. extorquens* strains presented a remarkable opportunity to study adaptation subsequent to acquiring a new pathway – as often occurs via horizontal gene transfer – because it focuses selection upon a particular module of the cell. Due to this, I hypothesized that the biggest benefit mutations – which would be the most likely to escape drift and rise in frequency in the populations – would be constrained to occur within the introduced pathway, or functions that were directly connected to it.

As in the above experiment, I set up eight populations that were passaged through 64-fold dilutions and evolved these for a period of time. Owing to their very slow initial growth (t_D = 11

h), they were passaged every four days for the first 300 generations, before bumping all populations up to the two day schedule I was using with my tradeoff work. The majority of our work was on isolates taken from 600 generations, although some of the dynamics were followed for 900 generations (see ***2.2***).

Could *M. extorquens* substantially improve over this timescale in terms of its ability to utilize such a different pathway for formaldehyde oxidation? I had not really even considered this as a "hypothesis" for the project when I started it, but looking back I had implicitly assumed they would improve, and likely due to multiple mutations, rather than just a single quick-fix. My graduate student, Hsin-Hung "David" Chou spearheaded the majority of this project. He found tremendous improvements in fitness: over 600 generations the average improvement was to nearly double fitness (Chou et al., 2011). And unlike wild-type having evolved on methanol, improvement of this strain was specific to methanol, and not generic adaptation to lab conditions. Similar gains were noted for methylamine growth, which also requires formaldehyde oxidation, but not for formate growth, which is downstream of that step (Carroll and Marx, 2013). Adaptation was also quite rapid, with a full 50% of the observed improvement occurring in just the first 150 generations. Some of the populations improved much more than others, however, and the shape of fitness increases suggested that some may have gotten stuck on a lower trajectory of improvement than others (see more about epistasis below; Lee and Marx, 2013).

Was adaptation largely due to mutations in either the GSH-dependent pathway or one stage removed from it? As it turned out, yes! David collaborated with another former graduate student, Nigel Delaney, to use Illumina to generate the first resequenced genome from an evolved isolate from our lab. This strain, CM1145, will come up several times below. We identified a total of nine mutations that had occurred over 600 generations in CM1145 (Chou et al., 2011). Three of these fit our expectations, and using allelic exchange (Marx, 2008) to generate various mutational combinations, these collectively accounted for ~85% of the total improvement. These three loci were: *fghA*, which encodes the second of the two GSH-dependent pathway enzymes; *gshA*, which encodes the first step of GSH biosynthesis; and *pntAB*, which encodes transhydrogenase (the GSH-dependent pathway cannot make NADPH directly). This makes quite clear the point that use of a novel pathway can require coordinated mutational changes in both the new genes and in the rest of the recipient genome (Michener and Marx, 2015). Much more emerged from this experiment, but these surprise findings about repeatability (***2.1.2***), evolutionary dynamics (***2.2***), epistasis (***2.3***), and modeling the fitness landscape (***2.4***) are discussed below.

As a side note, the eight populations mentioned above were not the first ones initiated from the GSH-dependent *M. extorquens* strain, but rather were a re-start to correct an initial "failure" that turned out to have its own interesting story (Marx, 2012). During my postdoc, the first 8 populations I started behaved quite strangely. Despite having tested what dilution scheme would work with this slow strain, when I started these populations, 6 of the 8 crashed to final cell densities 60- to 400-fold lower than their cohorts. Five of these recovered soon after, but

one population persisted with final densities 10-fold lower than expected for over 100 generations. It turns out that specifically for this strain growing on methanol, there is an absolute density threshold for growth. The one population that remained at low density rapidly adapted to this scenario: by generation 84 it had a fitness more than 3-fold greater than the ancestor if competed at a starting density 16-fold lower than normal, but had a fitness less than half the ancestor if started at the typical density of ~1×10^6 ml^{-1}. By the time this population recovered to a typical final density by generation 180, its advantage at low density had waned. These data suggest that adaptation to low densities may come with tradeoffs to growth at high densities. This whole mistake occurred because I went straight from single colonies – to eliminate the possibility of shared mutations – to growth in methanol media. Due to the very slow growth on methanol, four days was not sufficient for the tiny inoculum of a single colony to reach full density, and thus the cultures remained too dilute after the next 1/64 dilution. I avoided this for the second set of populations with a simple change in protocol: for the first transfer cycle, each colony was inoculated into a mixture of succinate and methanol in the first flask, and then grew on methanol only thereafter. This allowed the populations to rapidly use succinate to pop up from a single colony to half the typical final density in the flask (for this strain, ~1×10^6 ml^{-1}), and then the remaining time before the first transfer they could switch metabolism to utilization of methanol.

1.2.2 Use of the novel, toxic C_1 compound dichloromethane.

The second experimental example of this sort involved introduction of a metabolic capacity not present in *M. extorquens* AM1. A couple *M. extorquens* strains have the ability to utilize dichloromethane (DCM), an industrial solvent, as a growth substrate. In strains such as *M. extorquens* DM4, DCM is converted – via a genotoxic intermediate – to one formaldehyde and two hydrochloric acids (Muller et al., 2011). What a way to live! A former postdoc, Josh Michener, developed a collaboration with two experts on DCM growth, Stéphane Vuilleumier and Françoise Bringel from Université de Strasbourg, to explore how easily this novel trait could be moved into other *Methylobacterium* strains. In principle, DCM use only needs a single enzyme, DCM dehalogenase (encoded by *dcmA*), to generate formaldehyde, and thus feed into metabolism in the same manner as methanol would be utilized. Expressing *dcmA* in various *Methylobacterium* strains within and beyond the *M. extorquens* clade, he found that strains were quite variable in their ability to use DCM, and it did not correlate with phylogenetic relatedness (Michener et al., 2014a). Inspired by this finding, Josh initiated replicate populations for each of five initial strains bearing the same *dcmA*-expressing plasmid (Michener et al., 2014b). Populations from all but one of these initial ancestors showed substantial improvement in just 150 generations. Genome sequencing revealed no mutations on the *dcmA*-expressing plasmid, but rather, each genome had one mutation in a known or putative transporter. As it turns out, there were four separate genomic targets revealed that could lead to increased Cl^- export, which appears to be the primary physiological challenge for these cells (more about repeatability in ***2.1.2***). By making a construct that expressed *dcmA* and one of these identified exporters, *clcA*, he demonstrated across a wide swath of natural isolates that providing both the enzyme and a "solution" to the primary physiological challenge – Cl^- export – strains could utilize

DCM much better than with *dcmA* alone. Furthermore, when later testing chloromethane growth after having introduced the necessary "*cmu*" gene cluster (Vannelli et al., 1999; Vuilleumier et al., 2011), to our surprise, there was no correlation between the ability of these strains to use DCM and chloromethane, despite the fact that both lead to HCl production (Michener et al., 2016).

1.2.3 Utilization of methylamine as a growth substrate by novel pathways.

The third example of novel metabolic pathways involved two very different pathways of methylamine utilization. Another former graduate student, Dipti Nayak, initiated these projects due to curiosity as to why some strains of *M. extorquens* grow phenomenally faster on methylamine than others ($t_D \approx 4$ h; versus $t_D \geq 24$ h). It turns out that the major difference came down to whether the cells were using the well-studied methylamine dehydrogenase system (the fast ones), or were using the *N*-methylglutamate (NMG) pathway (the slow ones; Nayak et al., 2015). Besides having uncovered that the NMG pathway directs C_1 units through metabolism in a unique manner (Nayak and Marx, 2014b), and that all *Methylobacterium* genomes encode the NMG pathway, whereas only select few appear to have acquired methylamine dehydrogenase (Nayak et al., 2015), she decided to test the ability to improve the methylamine use of strain in two ways. First, she tried introducing methylamine dehydrogenase (from *M. extorquens* AM1) into a strain without it (*M. extorquens* PA1). Remarkably, unlike all examples above, this immediately resulted in growth equivalent to that in *M. extorquens* AM1 without any further mutations (Nayak et al., 2015)! Second, she tried to evolve strains of *M. extorquens* AM1 or PA1 that were left to utilize their NMG pathway (Nayak et al., 2016). In this case, she observed only modest improvement. Why was adaptation so stunted? The beneficial mutations that Dipti identified were quite illuminating. Besides changes in the NMG pathway itself to increase its expression, she identified mutations that appeared to have relieved the challenges of cytoplasmic production of ammonium. These occurred in either a K^+/H^+ antiporter (KefB), or in a urea transporter (and urea was detected in the supernatant of this strain). Use of methylamine dehydrogenase obviates this challenge because it generates ammonium in the periplasm. This finding also hinted at a possible *advantage* to cytoplasmic production: to prevent ammonium loss when using methylamine as a nitrogen source. Indeed, she demonstrated that strains without the NMG pathway perform worse when using low concentrations of methylamine as a nitrogen source during growth on succinate. Thus, the outcome of experimental evolution provided the key hint as to selective pressures in nature to use different pathways for optimal use of methylamine as a carbon versus a nitrogen source could maintain selection upon what would otherwise appear to be degenerate pathways.

1.3 Adaptation to ameliorate the negative effects of synonymous mutations.

The final example in this section involves a study by a former postdoc, Deepa Agashe, to examine what selective pressures act upon synonymous codons, and how adaptation may proceed to overcome these challenges. Many C_1 pathways appear to have been acquired via horizontal gene transfer, and thus may initially be encoded with poor codons (e.g., Kalyuzhnaya et al., 2005). As a target protein for this work we chose the small, highly expressed (1-2% total

protein) enzyme FAE (formaldehyde-activating enzyme), which catalyzes the condensation of formaldehyde with H_4MPT (Vorholt et al., 2000). We designed seven versions of *fae* to be synthesized that only differed in their codon usage (Agashe et al., 2013). These ranged from using the most frequent codon at all sites, to using the rarest codon at all sites. Remarkably, there were huge fitness differences between these versions, and they did not correlate with the proportion of frequent codons. The only commonality was that all versions led to insufficient FAE protein levels. Thus, although frequent codons are generally selected for across the genome, they may not always be advantageous in a particular gene.

Strains with these versions swapped into the chromosomal locus were used as ancestors for experimental evolution (Agashe et al., 2016). Despite the fact that these variants contained up to 150 synonymous mutations, Deepa observed rapid adaptation in all lineages. Some of these beneficial mutations were in the promoter region of *fae*, but most were single SNPs in its coding sequence, including some synonymous mutations (more about repeatability in this system in ***2.1.3***). This indicates that even single mutations can rapidly lead to improved usage of genes that have poor codon usage in their new host.

2. Surprises: unexpected lessons emerging from evolution experiments

Although all of the above experiments were initiated with clear goals to examine a particular question inspired by natural methylotrophs, one of the great benefits of experimental evolution is the rich set of unintended outcomes. Indeed, I would have to say that these collateral benefits have, in many cases, had a greater level of impact than the answers to the questions we intended to ask. To try to help maintain connection to the previous sections, within each theme below I have covered the model systems in the same order as they were described above.

2.1 Repeatability in adaptation.

One of the most common outcomes to emerge from experimental evolution is the ability to ascertain whether – for a particular strain in the chosen selective environment – adaptation tends to proceed similarly across independent replicates, or rather differently. This question can be addressed at the level of the genetic basis of adaptation, or at the level of repeated, parallel changes in phenotype. Repeatability will be affected by the relative rates of different beneficial mutations, as well as by their selective coefficients. The latter is particular important due to the fact that beneficial mutations tend not to be alone in a population: these populations are plenty big to have many beneficial mutations occurring and escaping drift simultaneously. Those with a greater selective benefit not only escape drift and "establish" more easily, but also have an increased ability to outcompete the other beneficial mutations that have established. This process is known as "clonal interference" (Gerrish and Lenski, 1998), and is in contrast to the traditional image of one beneficial mutation establishing and sweeping through a population is known as "periodic selection" (more below in ***2.2***).

2.1.1 Repeatability during evolution of tradeoffs between C_1 and multi-C metabolism.

Although we never broadly applied whole genome resequencing to uncover the genetic basis of adaptation from our tradeoff experiment. The repeated occurrence of loss-of-function

mutations in *ftfL* for succinate-evolved lineages that lost the ability to grow on methanol was already briefly described above in ***1.1***, and two other stories also emerged.

The first route to uncovering repeatability in adaptation originated in an odd way: from trying to uncover what was causing the inability to produce reliable media when I first started by own laboratory. I had no troubles growing *M. extorquens* as a postdoc at Michigan State, but when I moved to Harvard in 2005, we were besieged with inconsistent growth. Our first major hint came from one of the first beneficial mutations David Chou uncovered (Chou et al., 2009). An insertion sequence (IS) element had transposed right upstream of a putative metal transporter, and kicked up expression of this locus. Upon examination of other populations, 30 of 32 that had evolved with methanol as a carbon source in their media had mutations at the same locus, but none of the 8 populations that grew on succinate alone. Most remarkably, all of these mutations were caused by the same ISMex4 element, and it inserted into one of only two locations upstream of this gene. After a series of elegant experiments, David was able to demonstrate that this gene encoded a cobalt transporter, which we named *icuAB* (for increased cobalt uptake). Why was cobalt a problem, and only for the methanol-evolved populations? As it turns out, the recipe for the Vishniac trace metal mix that was used in our medium had changed compared to its original formulation in terms of strength and ratio of metals (Vishniac and Santer, 1957; Delaney et al., 2013). Notably, the total metal concentration was now similar in strength as the EDTA chelator present. Due to the light sensitivity of EDTA, this led to inconsistencies: David noted good growth on sunny days when the metal mix was a paler purple, and worse growth on cloudy days when the metal mix was more intensely colored. Why did $icuAB^{EVO}$ alleles only arise as beneficial mutations in medium containing methanol? Cobalt is required for B_{12}, and there are a couple of enzymatic steps unique to the ethylmalonyl-CoA pathway for glyoxylate regeneration in this organism that use B_{12} (Peyraud et al., 2009). Mutants in this pathway have a defect on C_1 or C_2 compounds, which mirrored the scenarios that the medium was problematic, and were the media in which $icuAB^{EVO}$ alleles were beneficial. Accordingly, just like for glyoxylate regeneration mutants, the addition of glyoxylate – the end product of this pathway – alleviated the problem. The laboratory of Julia Vorholt at ETH Zürich independently found the same cobalt problem in the medium, but their clue came from metabolomics. They noted accumulation of intermediates immediately upstream of the vitamin B_{12}-requiring reactions of the ethylmalonyl-CoA pathway (Kiefer et al., 2009). Thankfully, the new minimal medium developed by Nigel Delaney and others avoids these issues and has become increasingly utilized in the field (Delaney et al., 2013).

The second story of repeatability from these populations emerged in a different manner: use of comparative genomic hybridization to microarrays (remember those?) to identify potential gene amplification or deletions. Many other selection experiments have identified key selective targets due to amplifications that increased gene dosage, and thus expression (Dunham et al., 2002; Maisnier-Patin and Roth, 2015). Thus, prior to today's ability to inexpensively sequence genomes, we decided to try it out. Miki Lee looked at a total of 44 isolates across 32 different populations that had evolved for 1500 generations, including the 16 populations described above in the tradeoffs story (***1.1***), and found much more than we could have guessed (Lee and

Marx, 2012). About 80% of these populations experienced a series of similar deletions that removed 5-10% of the genome. This region was present on a 1.3 Mb megaplasmid unique to *M. extorquens* AM1 versus the other *M. extorquens* isolates (Vuilleumier et al., 2009; Marx et al., 2012). All of the deletions involved homologous recombination between pairs of matching IS elements, but the precise end points of these events were different between populations. Most importantly, Miki was able to generate clean deletions of these regions and demonstrated that they were selectively beneficial, but not in a manner that scaled with the length of the deletion. As such, it is the loss of particular gene products that was beneficial, rather than directly due to having a smaller genome to replicate.

2.1.2 Repeatability during adaptation to use of an introduced enzyme or metabolic pathway.

Several stories of repeatability emerged from the multiple examples of evolving engineered ancestral strains that depended upon a novel, introduced metabolic capacity (***1.2***). To see parallelism when selection acts upon a particular physiological challenge is not surprising, but the types and targets of adaptation were beyond what we could have expected.

Our richest information comes from the populations that evolved on methanol with the GSH-dependent pathway in place of the native H_4MPT-dependent one. Perhaps unsurprisingly, looking across populations, every winning lineage possessed mutations on the plasmid expressing the GSH-dependent pathway (Chou and Marx, 2012). David found that these were of three mutational classes: mutations that affected expression of the two introduced genes that encode the GSH-dependent pathway (*flhA* and *fghA*), mutations that affected the copy number of the plasmid, or integrations of the introduced plasmid into the host genome via incomplete transposition events. It was not a shock that all of these mutations, despite the molecular details being so different, all affected levels of FlhA and FghA in the same direction, but we were rather surprised that they all led to *decreases* in expression! As it turns out, the high level of FlhA and FghA arising from using the strong P_{mxaF} promoter on a multi-copy plasmid (Marx and Lidstrom, 2001) led to expression costs (Chou et al., 2011). Although the expression phenotype was similar across lineages, the mutational paths to get there were quite distinct. Strangely, the one "obvious" class of mutations – mutations to decrease the strength of the promoter – was never observed. More on that below in ***2.4***.

Unlike the universal presence of beneficial mutations in the introduced plasmid expressing the GSH-dependent pathway in these strains, chromosomal mutational targets was much more variable between lineages (Carroll et al., 2015). Sean Carroll examined the genome sequences for one strain from generation 600 for each of the eight populations. There was great variety in the total number of mutations present, from 4 to 18. The chromosomal locus most commonly mutated was *icuAB* described above in ***2.1.1*** (6 of 8), followed by *gshA* mentioned above in ***1.2*** (5 of 8), and then seven other loci with mutations in 2 or 3 of the 8 populations. This modest overlap of mutational targets is perhaps consistent with the relatively large spread in fitness values achieved by these lineages, ranging from 70% to 160%, and the corresponding maintenance of among-population variation in fitness through time (Lee and Marx, 2013).

Although the beneficial mutations may have been only modestly repeatable, analysis of gene expression and some currency metabolites were much more consistent across lineages. Sean took these same eight strains and analyzed mRNA levels during exponential phase (Carroll and Marx, 2013). There were huge changes in gene expression from the ancestor to each of the eight evolved isolates examined, and these were remarkably parallel. How to rationalize this compared to the fairly different set of underlying mutations? Comparison back to wild-type, which grows quickly using the H_4MPT-dependent pathway was tremendously useful. Nearly all the genes that went up (such as stress responses) or down (metabolism and ribosomes) due to swapping the pathways in the original strain were those with a reversed pattern during adaptation. It is clear that most of these represent the indirect consequences of poor growth versus rapid growth, and this aspect changed in all lineages. Steady-state ratios of the concentrations of the pyridine nucleotides – NAD(P)(H) – also changed in this manner: increases in the ratio of NADPH/$NADP^+$ and NADH/NAD^+ occurred in the engineered ancestor compared to wild-type, and these were reversed during adaptation. Indeed, the exact ratios for each pair of nucleotides correlated quite well with the fitness of the strains.

This pattern of modest repeatability was also seen for the DCM adaptation by Josh Michener (Michener et al., 2014b). Although Cl^- export appeared to be the central challenge in all cases, he found four distinct loci could ameliorate this challenge. Interestingly, the same locus was often observed as a beneficial mutation across different strains or species used as ancestors. Beneficial alleles in *secY* (subunit of protein secretion system) were found in *M. extorquens* AM1, *M. extorquens* BJ001 (formerly *M. populi*), and *M. nodulans*; *clcA* (Cl^-/H^+ antiporter) in *M. extorquens* PA1 and BJ001; and *edgA* (new locus named for evolved DCM growth) in *M. extorquens* PA1 and *M. nodulans*. This pattern seemed to imply that the same loci could solve this common problem regardless of the genomic background of the strain. Indeed, beneficial mutations from one strain could be introduced into another strain where none of the beneficial mutations were at that locus, and that allele was still beneficial in the new genomic context.

Perhaps the most exciting part of this study was from Josh's examination of how *M. extorquens* DM4 first may have evolved originally to use *dcmA* in nature (Michener et al., 2014b). The amino acid sequences for all four loci identified above to solve the Cl^- export problem were identical between the natural DCM utilizer (DM4), and the other *M. extorquens* strains (AM1, BJ001, CM4, PA1). There were, however, mutations unique to *M. extorquens* DM4 in the *clcA* promoter. Swapping just P_{clcA} between DM4 and the DCM-naïve PA1 strain was sufficient to completely swap the phenotype! Furthermore, as there happens to be only one other strain of *M. extorquens* known to utilize DCM, Josh obtained this strain and sequenced P_{clcA}. A beneficial mutation (another IS insertion) was also found in this strain that increased *clcA* expression. Thus, not only was adaptation repeatable in the lab across different strains and species within a genus, we were able to identify the chromosomal locus that permitted effective DCM utilization in two independent natural isolates after they had acquired *dcmA* in the past.

2.1.3 Repeatability during adaptation to ameliorate the negative effects of synonymous mutations.

Deepa Agashe's work on rapid adaptation to various synonymous recodings of *fae* produced another remarkable pattern of repeatability (Agashe et al., 2016). Although all of the strains initially grew poorly on methanol due to insufficient levels of FAE, they evolved in a manner that was highly consistent per strain, but very different between them. Replicate populations for three of the six variants always had beneficial mutations in the promoter of *fae* that increased gene expression, but never had coding mutations. On the other hand, for the other three synonymous variants the pattern was reversed: always coding mutations and never promoter mutations. For this latter set, there were commonly repeated mutations to the exact same residue of FAE, such as three distinct mutations to Ile12, two different synonymous mutations and one nonsynonymous one. Furthermore, to briefly mention epistasis in this system (the theme below in ***3***), when 12 different beneficial coding mutations were moved to alternative synonymous ancestral versions, only two of these conferred benefit in the new context. This version-specific pattern of adaptation corroborated her earlier work (Agashe et al., 2013) that suggested that each recoded version of *fae* was problematic to express in a unique way.

2.2 Complex allele dynamics during adaptation.

Besides examination of the collection of alleles or phenotypes that have emerged in single evolved isolates, a complementary aspect of the evolutionary process is to uncover the trajectories of alleles in the populations. As I have commented upon before (Marx, 2013), these dynamics alone can give clues as to aspects such as multiple ecological niches. Recent work has greatly expanded the quantitative inferences that can be made in this regard (e.g., Lang et al., 2013; Good et al., 2017), but two simple stories have emerged from our work on adaptation of *M. extorquens* dependent upon the GSH-dependent pathway for formaldehyde oxidation.

The first story of allele dynamics built upon our desire to develop a simple, inexpensive method to follow allele frequencies in populations via amplicon sequencing. Before then, the standard method to follow allele dynamics was to isolate many (10-100) isolates per timepoint to interrogate, and then amplify each allele from each isolate at each timepoint, and submit these for standard Sanger sequencing. Although simple in design, this is rather laborious and expensive. A former postdoc, Lon Chubiz, worked with Miki Lee to develop a cheap method we named "FREQ-Seq" (Chubiz et al., 2012). The primary innovation was to recognize that we wanted to barcode individual samples, yet did not want to order tens or hundreds of barcoded primers for every locus in question. FREQ-Seq uses two rounds of PCR, beginning with a single set of allele-specific primers to amplify a locus in question from a mixed population timepoint (using primers with generic 5' extensions that match the 3' extensions of primers in step two). This product is then re-amplified in a mixture containing three primers: two universal primers (that are also the Illumina A and B adapters, eliminating the need for further sample prep) and a small amount of a long bridging primer that carries the sample-specific barcode. This worked amazingly well. The very first time we tested this in full, Lon sacrificed an opportunity to join the lab for a well-earned beer, and instead spent ~3 h performing two rounds of PCR for each of three loci (*fghA*, *gshA*, *pntAB*; see ***1.2***) at 27 time-points of the population that CM1145 was isolated from (Chou et al., 2011). These samples – as well as controls for each locus – were then

put in to 3% of an Illumina control lane. Even at the modest sequencing depth available at the time, this resulted in an average coverage of 150,000-fold per timepoint per locus, with almost no PCR bias. What did we learn? First, it became clear that, of these three mutations, the *gshA*^{EVO} mutation occurred first, and then *fghA*^{EVO}, then *pntAB*^{EVO}. This matched the order of their selective benefits from largest to smallest (Chou et al., 2011), which was on its own unsurprising. What was fascinating, however, was that the *gshA*^{EVO} had rapidly risen to ~40% of the population by generation 108, and then fell to <10% at generation 180, before rapidly rising again to near fixation by generation 210. This pattern is indicative of clonal interference: the *gshA*^{EVO} allele did not become less beneficial, but was temporarily passed by one (or more) lineages that were even more fit, before being able to rise to eventual fixation once coupled with the *fghA*^{EVO} allele. In this case we were lucky, and in the next story we figured out just what mutation(s) temporarily pushed *gshA*^{EVO} down in frequency.

The second window into allele dynamics arose due to a particular type of allele mentioned above in ***1.2***: the beneficial integration of our introduced plasmid encoding the GSH-dependent pathway into the host genome (Chou and Marx, 2012). As bizarre of a genetic event as this would seem to be, it occurred multiple times in the small number of isolates David examined. These integrations were not due to traditional homologous recombination, but, rather, were caused by incomplete transposition of ISMex25 into our plasmid, pCM410, resulting in a cointegrate. Every one of these events that we obtained was an insertion into *trfA*, which encodes the essential replication initiation gene of this small IncP plasmid backbone (Marx and Lidstrom, 2001). Miki decided to follow-up on this, and devised a simple semi-quantitative PCR approach to identify such insertions. Her data were striking: not only could she follow the dynamics of these bizarre alleles in the populations where they had been previously identified, it turns out every population contained multiple (up to 17!) such insertions. This was possible to discern because distinct insertion sites or orientations of ISMex25 in *trfA* generated PCR products of different lengths. If these were beneficial (~20% when tested alone) and so easy to produce, how could they have failed to become the winning lineage in all populations? In the five populations where these alleles did not fix, the whole set of alleles became detectable at the same time, rose in frequency together, and then fell in synchrony with each other. They were thus victims of clonal interference. In the population studied with FREQ-Seq and described above, we could clearly see that it was the 17 genotypes bearing one these insertion alleles (along with other mutations) that rose in frequency and peaked precisely when the *gshA*^{EVO}-containing genotype was at its lowest point, and then the insertion alleles fell in frequency together when the *gshA*^{EVO} rapidly rose towards eventual fixation. It should be noted that the complex dynamics of clonal interference are not just seen in laboratory-based experimental evolution, but also occur in natural situations such as persistent infections (Lieberman et al., 2011; Yang et al., 2011; Silva et al., 2016; Levade et al., 2017; Xue et al., 2017).

2.3 Epistatic interactions between beneficial mutations.

Once one has found the beneficial mutations that were targets of adaptation, and perhaps having uncovered their complex temporal dynamics, this opens the door to considering how the

mutations affected each other along an adaptive trajectory. If the selective effect of an allele were independent of the genotype it arose upon, the order of beneficial mutations observed would solely be due to the balance between the likelihood of a mutation happening, and the probability a mutation of that (fixed) selective coefficient would escape drift and beat out other mutations it was clonally interfering with. It would also imply a perfectly smooth, single-peaked adaptive landscape. But if the fitness effect of a mutation is not fixed and depends upon the genetic background, then all bets are off. For a pair of mutations *i* and *j*, we typically consider epistasis (ε_{ij}) as the difference between the phenotype (here fitness, *W*) actually observed (W_{ij}) and the null expectation if there were no "extra" effect of the two alleles having been combined. This null hypothesis is generally the product of the two individual effects ($W_i \times W_j$), consistent with the idea of independently acting allele having a constant proportional affect upon fitness in the absence of epistasis. Prior to taking on these studies, this type of effect had been examined for beneficial mutations within a single gene (e.g., Weinreich et al., 2006), but had been little explored for mutations that had occurred across the genome. Having "snuck in" brief mentions of epistasis emerging from studies of metabolic tradeoffs and selection upon codons, here I concentrate on three examples from work by David Chou on adaptation with the foreign, GSH-dependent pathway.

The first example of epistasis we uncovered involved the *icuAB* cobalt transporter we discovered (Chou et al., 2009; see ***2.1.1***). As mentioned above, beneficial mutations in *icuAB* occurred in 30 of 32 populations examined that evolved using methanol as a substrate, and this included 6 of the 8 populations with the GSH-dependent pathway. In the wild-type context, $icuAB^{EVO}$ conferred an 18% advantage during growth on methanol in the original (unintentionally) cobalt-limiting media. To David's great surprise, when he introduced $icuAB^{EVO}$ into the GSH-dependent ancestor, there was only a very slight increase in fitness (<3%). If that was so, how could it have emerged repeatedly? He recognized that $icuAB^{EVO}$ likely was not the first allele to arise, however, and thus could have arisen on a fitter version of that strain. Having begun to re-construct strains bearing $fghA^{EVO}$, $gshA^{EVO}$, or $pntAB^{EVO}$, he introduced $icuAB^{EVO}$ into each of these somewhat faster strains. Plotting the selective coefficient of $icuAB^{EVO}$ against the baseline growth rate of the genotypes it was introduced into, there was a straight line with positive slope. This indicates that the beneficial effect of $icuAB^{EVO}$ was synergistic with the other mutations, and thus conferred a higher proportional benefit the fitter the strain background was. David then had the clever idea to test whether changes in the environment that would affect growth rate would generate a similar effect as epistasis. By simply growing wild-type at 16, 20, and 25 °C (rather than the standard 30 °C), he could generate a similar range of growth rates as the GSH-dependent strains grown at 30 °C. Indeed, the fitness effect of these manipulations plotted right on top of the epistasis data. This indicates that it really does not seem to matter what makes the cells grow more slowly, but simply doing so results in a consistent decrease in the fitness effect of $icuAB^{EVO}$. What underlies this commonality between epistasis (also known as genotype × genotype, or G × G interactions) and genotype × environment (i.e., G × E) interactions? This requires thinking about where cobalt "goes" during cell growth. Unlike carbon, it is not breathed off, but simply becomes a part of the B_{12} pool. The

only thing that makes the B_{12} concentration go down is dilution by cell growth. As such, the need for increased cobalt uptake only becomes increasingly severe as the cell grows faster and faster.

The second example of epistasis in this system came from analyzing a network of allele combinations from the best-studied evolved isolate, CM1145, whose genome was first sequenced (Chou et al., 2011), and has been mentioned in multiple sections above. Although this strain contained nine mutations, many of them were either hard to reproduce (like a huge deletion of 10% of the genome, like what Miki had uncovered in other populations; Lee and Marx, 2012), of questionable effect (synonymous mutation or in a gene of unknown function), or clearly had to do with the cobalt-limitation in the medium (*icuAB*; Chou et al., 2009). Thus, rather than labor to make all $2^9 = 512$ strains, David wisely chose to concentrate upon the loci that were directly related to the introduced GSH-dependent formaldehyde oxidation pathway (*fghA*, *gshA*, *pntAB*), and lump the rest of the six mutations as a single allele, the "genetic background" (*GB*). The fitness value for all 16 combinations of these alleles was determined, and a clear pattern emerged. For three of the four alleles, the selective benefit of that allele steadily decreased with the fitness of the background it was introduced into. For example, the $gshA^{EVO}$ allele had a 51% benefit when present alone, but only conferred a 34% benefit if introduced into a strain already containing the three other evolved alleles. A similar trend was shown by my colleague Tim Cooper (now at Massey University, New Zealand) for the first mutational steps of one of the Lenski lineages of *E. coli* evolved to utilize glucose (Khan et al., 2011). Tim and I coordinated our submissions, and these ended up being the first papers to suggest a generic trend for "diminishing returns" epistasis between beneficial mutations. Mutations becoming less and less valuable when stacked on top of each other plays a critical role in the well-established tendency of experimental evolution experiments to decelerate their rate of adapation (Lenski et al., 1991; Wiser et al., 2013). This trend has also been observed in many other biological systems, suggesting one or more general principles at cause (Kvitek and Sherlock, 2011; Rokyta et al., 2011; Kryazhimskiy et al., 2014).

Our final example of epistasis came from an orthogonal approach that David took: combine beneficial mutations from separate evolved lineages that affected the same pathway (Chou et al., 2014). He made a series of allele combinations where one mutation decreased expression of the GSH-dependent pathway due to lowering the expression per plasmid, and the second mutation was one that decreased plasmid copy number. His rationale was that these mutations should affect expression in independent ways, and thus be free from epistasis at the level of enzyme expression. This turned out to be the case; the expression level of the double mutant was well-predicted by the product of the individual mutation effects. When it came to fitness, however, $W_{ij} = W_i \times W_j$ failed miserably, with combinations of beneficial alleles consistently providing less benefit than expected. In some cases, two highly beneficial alleles were tremendously deleterious when combined for reasons described below.

2.4 Simple mathematical models to understand epistasis between beneficial mutations.

Perhaps the aspect of all this work that I am the most excited about, and guides several of the current efforts in the lab, are efforts to propose mathematical models of biochemistry as

hypotheses for how mutations should affect fitness, alone or together. My interest in this area was shaped tremendously by the pioneering work of Dan Dykhuizen, Tony Dean, and Dan Hartl throughout the 1980s (e.g., Hartl et al., 1985; Dean et al., 1986; Dykhuizen et al., 1987). They used theory from Metabolic Control Analysis (Kascer and Burns, 1973; Heinrich and Rapoport, 1974) to understand the fitness effects of mutations in *lacZ* or *lacY* upon growth of *E. coli* in lactose-limited chemostats. This was evolutionary systems biology long before the term was ever used (or was popular). Chemostats are great for the ability to focus selection largely upon a single limiting resource, but this is also their limitation; it is unclear how well biochemistry could predict growth rate in the absence of a single external limitation. On the other hand, my GSH-dependent *M. extorquens* has a single internal limitation, so I hoped there was a chance for this approach to work.

The first form of epistasis we modeled was the diminishing returns observed when combining alleles that arose in a single evolved isolate (Chou et al., 2011). For this, we teamed up with my good friend and collaborator, Daniel Segrè, and his former graduate student, Hsuan-Chao Chiu (Boston University). David Chou had noted that the expression costs for synthesizing the two enzymes of the GSH-dependent pathway visually manifested as larger, morphologically abnormal cells. Mutations that directly reduced expression of the pathway (e.g., $fghA^{EVO}$) led to fewer such cells, however, even the $gshA^{EVO}$ and GB^{EVO} alleles led to fewer abnormal cells (but not $pntAB^{EVO}$). This suggested that the benefit of $gshA^{EVO}$ and GB^{EVO} might at least partially be due to reducing the protein expression costs of the two enzymes of the introduced GSH pathway (whereas the benefit of $pntAB^{EVO}$ is unrelated to expression costs). These three alleles that appeared to reduce expression costs – $fghA^{EVO}$, $gshA^{EVO}$ and GB^{EVO} – were also the three of the four that exhibited diminishing returns ($pntAB^{EVO}$ imparted a ~10% benefit regardless of background). This led us to speculate a connection between expression costs and diminishing returns epistasis. We formulated a simple model in which we treated protein expression costs as a separate phenotype from everything else that affects growth rate. Then we imagined each mutation could affect one or both of these phenotypes and interact independently (i.e., no epistasis) upon that phenotype, yet fitness is the intrinsic growth rate minus expression costs. By using the proportion of morphological abnormalities as a proxy to partition the degree to which each single mutation affected benefits or costs, we were able to predict the fitness values of double, triple, and quadruple mutation combinations from the phenotypes of the single mutants with a surprising degree of success ($R^2 = 0.97$). The intuition for diminishing returns in this example of protein expression costs is easy to illustrate. As a thought experiment, imagine three mutations that each cut protein expression costs in half in distinct ways, and that these expression costs reduce fitness by 20%. The first of these beneficial mutations to occur will cut the cost from 20% to 10%, and will thus produce a fitness benefit of (1.2-0.1)/(1.2-0.2) = 10%. The next mutation, however, would only reduce the expression cost a further 5%, and would have a smaller benefit of (1.2-0.05)/(1.2-0.1) = 4.5%. The third mutation would have even less remaining effect, cutting the costs by 2.5%, thus a (1.2-0.025)/(1.2-0.05) = 2.2% fitness benefit. As developed in a follow-up theory paper led by Hsuan-Chao Chiu, continued proportional reductions to a single phenotype inherently leads to diminishing returns (Chiu et al., 2012).

The second type of epistasis modeled was that of multiple mutations that reduced expression of the GSH-dependent pathway via independent molecular mechanisms (Chou et al., 2014). As mentioned above in ***2.3***, enzyme expression was well-predicted via a simple multiplicative model, but fitness was not. Rather than build a full metabolic control analysis model with many parameters that were never manipulated in this experiment, we generated a simplified algebraic expression for the "control curve" that describes the consequences of changing activity of a single enzyme at a time upon the steady-state flux through a pathway (which in a case like this is proportional to fitness). As these control curves are inherently hyperbolic, they can be approximated in a manner analogous to a Michaelis-Menten relationship: the "v_{max}" parameter is the maximal steady-state flux of the pathway given infinite activity of the enzyme in question, and the "K_M"-like saturation parameter is the level of that enzyme needed to provide half-maximal steady-state flux through the pathway. Two more modifications to this logic were suggested by the data: linear costs of the enzymes, and an offset from the origin due to the apparent need for a threshold value of enzyme activity to avoid formaldehyde toxicity (and thus no growth at all). All of this led to single equation with five parameters. We fit these five parameters using 27 data points from strains with either inducible promoters driving known changes in enzyme levels, or single mutants. We then used this model to extrapolate to 17 mutational combinations not used in the fitting. Remarkably, despite the many simplifications involved, this model provided a fairly precise prediction of these fitness values (R^2 = 0.98). This demonstrates that the fitness landscape of enzyme expression is smooth and can be easily modeled if the data are available.

Most importantly, several aspects of the adaptation that had occurred were illuminated by the shape of the resulting fitness landscape. First, it became clear that the winning beneficial mutations affecting the levels of the two enzymes of the GSH-dependent pathway made it to the fitness peak in a single mutational step, rather than a long series of smaller mutations stacked on top of each other. Second, the lack of promoter mutations was now explained: although turning down transcription via the inducible promoter could be beneficial, the highest fitness benefit obtainable by doing so is ~20%. The particular mutations that did occur generated benefit up to ~40%, and thus would have generally outcompeted the lesser benefit solutions to this particular physiological challenge. Third, it was now clear why some mutations interacted epistatically so poorly with other mutations; these were the ones that sat closest to the "cliff edge" of the fitness landscape, such that any further decrease in expression would lead to a massive decrease in growth.

Conclusions.

Combining experimental evolution, microbial physiology, and mathematical biology has been an enjoyable and reasonably fruitful approach for my lab to address questions about *M. extorquens* that lie at the intersection of these fields. Although the original questions being asked with each experiment outlined here turned up answers, I strongly feel the "collateral" findings that were not predicted from the start have been the most rewarding and broadly applicable to other systems. This balance between planned versus surprising results perhaps mirrors the tension of

evolution itself: selection upon traits in a given environment is deterministic and (in principle) predictable, whereas the arrival and escape from loss due to drift of the very mutations that affect these traits is inherently stochastic.

Acknowledgements.

CJM recognizes an NSF award (MCB-1714949) that supports work on the experimental evolution of *Methylobacterium*.

References

Agashe, D., Martinez-Gomez, N.C., Drummond, D.A., and Marx, C.J. (2013) Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous variants in a key enzyme. Mol. Biol. Evol. 30, 549-560. doi.org/10.1093/molbev/mss273

Agashe, D., Sane, M., Phalnikar, K., Diwan, G.D., Habibullah, A., Martinez-Gomez, N.C., Sahasrabuddhe, V., Polachek, W., Wang, J., Chubiz, L.M., and Marx, C.J. (2016) Large-effect beneficial synonymous mutations mediate rapid and parallel adaptation in a bacterium. Mol. Biol. Evol. 33, 1542-1553. doi.org/10.1093/molbev/msw035

Bruger, E.L., and Marx, C.J. (2018) A decade of genome sequencing has revolutionized studies of experimental evolution. Curr. Opin. Microbiol. 45, 149-155. doi.org/10.1016/j.mib.2018.03.002

Carroll, S.M., and Marx, C.J. (2013) Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. PLoS Genet. 9, e1003427. doi.org/10.1371/journal.pgen.1003427

Carroll, S.M., Lee, M.-C., and Marx, C.J. (2014) Sign epistasis limits evolutionary trade-offs at the confluence of single- and multi-carbon metabolism in *Methylobacterium extorquens* AM1. Evolution. 68, 760-771. doi.org/10.1111/evo.12301

Carroll, S.M., Chubiz, L.M., Agashe, D., and Marx, C.J. (2015) Parallel and divergent evolutionary solutions for the optimization of an engineered central metabolism in *Methylobacterium extorquens* AM1. Microorganisms. 3, 152-174. doi.org/10.3390/microorganisms3020152

Chistoserdova, L.V., and Lidstrom, M.E. (1994) Genetics of the serine cycle in *Methylobacterium extorquens* AM1: identification of *sgaA* and *mtdA* and sequences of *sgaA*, *hprA*, and *mtdA*. J. Bacteriol. 176, 1957–1968.

Chistoserdova, L., Vorholt, J.A., Thauer, R.K., and Lidstrom, M.E. (1998) C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea. Science. 281, 99–102.

Chiu, H.-C., Marx, C.J., and Segrè, D. (2012) Epistasis from functional dependence of fitness on underlying traits. Proc. Royal Soc. B. 279, 4156-4164. doi.org/10.1098/rspb.2012.1449

Chou, H.-H., and Marx, C.J. (2012) Optimization of gene expression through divergent mutational paths. Cell Rep. 1, 133-140. doi.org/10.1016/j.celrep.2011.12.003

Chou, H.-H., Berthet, J., and Marx, C.J. (2009) Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. PLoS Genet. 5, e1000652. doi.org/10.1371/journal.pgen.1000652

Chou, H.-H., Chiu, H.-C., Delaney, N.F., Segrè, D, and Marx, C.J. (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science. 332, 1190-1192. doi.org/10.1126/science.1203799

Chou, H.-H., Delaney, N.F., Draghi, J.A., and Marx, C.J. (2014) Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. PLoS Genet. 10, e1004149. doi.org/10.1371/journal.pgen.1004149

Chubiz, L.M., Lee, M.-C., Delaney, N.F., and Marx, C.J. (2012) FREQ-Seq: A rapid, cost-effective, sequencing-based method to determine allele frequencies directly from mixed populations. PLoS One. 7, e47959. doi.org/10.1371/journal.pone.0047959

Crowther, G.J., Kosály, G., and Lidstrom, M.E. (2008) Formate as the main branch point for methylotrophic metabolism in *Methylobacterium extorquens* AM1. J. Bacteriol. 190, 5057-5062. doi.org/10.1128/JB.00228-08

Dean, A.M., Dykhuizen, D.E., and Hartl, D.E. (1986) Fitness as a function of beta-galactosidase activity in *Escherichia coli*. Genet. Res. 48, 1-8.

Delaney, N.F., Kaczmarek, M.E., Ward, L.M., Swanson, P.K., Lee, M.-C., and Marx, C.J. (2013) Development of an optimized medium, strain and high-throughput culturing methods for *Methylobacterium extorquens.* PLoS One. 8, e62957. doi.org/10.1371/journal.pone.0062957

Douglas, S.M., Chubiz, L.M., Harcombe, W.R., Ytreberg, F.M., and Marx, C.J. (2016) Parallel mutations result in a wide range of cooperation and community consequences in a two-species bacterial consortium. PLoS One. 11, e0161837. doi.org/10.1371/journal.pone.0161837

Douglas, S.M., Chubiz, L.M., Harcombe, W.R., and Marx, C.J. (2017) Identification of the potentiating mutations and synergistic epistasis that enabled the evolution of inter-species cooperation. PLoS One. 12, e0174345. doi.org/10.1371/journal.pone.0174345

Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., and Botstein, D. (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 99, 16144-16149. doi.org/10.1073/pnas.242624799

Dykhuizen, D.E., Dean, A.M., and Hartl, D.E. (1987) Metabolic flux and fitness. Genetics. 115, 25-31.

Gerrish, P.J., and Lenski, R.E. (1998) The fate of competing beneficial mutations in an asexual population. Genetica. 102-103, 127-144.

Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., and Desai, M.M. (2017) The dynamics of molecular evolution over 60,000 generations. Nature. 551, 45-50. doi.org/10.1038/nature24287

Harcombe, W.R., Riehl, W.J., Dukovski, I., Granger, B.R., Betts, A., Lang, A.H., Bonilla, G., Kar, A., Leiby, N., Mehta, P., Marx, C.J., and Segrè, D. (2014) Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. Cell Rep. 7, 1-12. doi.org/10.1016/j.cellrep.2014.03.070

Harcombe, W.R., Betts, A., Shapiro, J.W., and Marx, C.J. (2016) Adding biotic complexity alters the metabolic benefits of mutualism. Evolution. 70, 1871-1881. doi.org/10.1111/evo.12973

Harms, N., Ras, J., Reijnders, W.N., van Spanning, R.J., and Stouthamer, A.H. (1996) *S*-Formylglutathione hydrolase of *Paracoccus denitrificans* is homologous to human esterase D: a universal pathway for formaldehyde detoxification? J. Bacteriol. 178, 6296-6299.

Hartl, D.L., Dykhuizen, D.E., and Dean, A.M. (1985) Limits of adaptation: the evolution of selective neutrality. Genetics. 111, 655-674.

Heinrich, R., and Rapoport, T.A. (1974) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur. J. Biochem. 42, 89-95.

Kalyuzhnaya, M.G., Korotkova, N., Crowther, G., Marx, C.J., Lidstrom, M.E., and Chistoserdova, L. (2005) Analysis of gene islands involved in methanopterin-linked C_1 transfer reactions reveals new functions and provides evolutionary insights. J. Bacteriol. 187, 4607-4614. doi.org/10.1128/JB.187.13.4607-4614.2005

Kascer, H., and Burns, J.A. (1973) The control of flux. Symp. Soc. Exp. Biol. 27, 65-104.

Kiefer, P., Buchhaupt, M., Christen, P., Kaup, B., Schrader, J., and Vorholt, J.A. (2009) Metabolite profiling uncovers plasmid-induced cobalt limitation under methylotrophic growth conditions. PLoS One. 4, e7831. doi.org/10.1371/journal.pone.0007831

Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E., and Cooper, T.F. (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. Science. 332, 1193-1196. doi.org/10.1126/science.1203801

Knief, C., Frances, L., and Vorholt, J.A. (2010) Competitiveness of diverse *Methylobacterium* strains in the phyllosphere of *Arabidopsis thaliana* and identification of representative models, including *M. extorquens* PA1. Microb. Ecol. 60, 440-452. doi.org/10.1007/s00248-010-9725-3

Kryazhimskiy, S., Rice, D.P., Jerison, E.R., and Desai, M.M. (2014) Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science. 344, 1519-1522. doi.org/10.1126/science.1250939

Kvitek, D.J., and Sherlock, G. (2011) Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet. 7, e1002056. doi.org/10.1371/journal.pgen.1002056

Lang, G.I., Rice, D.P., Hickman, M.J., Sodergren, E., Weinstock, G.M., Botstein, D., and Desai, M.M. (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 500, 571-574. doi.org/10.1038/nature12344

Lee, M.-C., and Marx, C.J. (2012) Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS Genet. 8, e1002651. doi.org/10.1371/journal.pgen.1002651

Lee, M.-C., Chou, H.-H., and Marx, C.J. (2009) Asymmetric, bi-modal tradeoffs during adaptation of *Methylobacterium* to distinct substrates. Evolution. 63, 2816-2830. doi.org/10.1111/j.1558-5646.2009.00757.x

Lee, M.-C., and Marx, C.J. (2013) Synchronous waves of failed soft sweeps in the laboratory: remarkably rampant clonal interference of alleles at a single locus. Genetics. 193, 943-952. doi.org/10.1534/genetics.112.148502

Leiby, N., and Marx, C.J. (2014) Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. PLoS Biol. 12, e1001789. doi.org/10.1371/journal.pbio.1001789

Lenski, R.E. (2017) Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J. 11, 2181-2194. doi.org/10.1038/ismej.2017.69

Lenski, R.E., Rose, M.R., Simpson, S.C., and Tadler, S.C. (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138, 1315-1341.

Levade, I., Terrat, Y., Leducq, J.B., Weil, A.A., Mayo-Smith, L.M., Chowdhury, F., Khan, A.I., Boncy, J., Buteau, J., Ivers, L.C., Ryan, E.T., Charles, R.C., Calderwood, S.B., Qadri, F., Harris, J.B., LaRocque, R.C., and Shapiro, B.J. (2017) *Vibrio cholerae* genomic diversity within and between patients. Microb. Genom. 3, 000142. doi.org/10.1099/mgen.0.000142

Lieberman, T.D., Michel, J.B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis Jr., M.R., Skurnik, D., Leiby, N., LiPuma, J.J., Goldberg, J.B., McAdam, A.J., Priebe, G.P., and Kishony, R. (2011) Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat. Genet. 43, 1275-1280. doi.org/10.1038/ng.997

Luria, S.E., and Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics. 28, 491-511.

Maisnier-Patin, S., and Roth, J.R. (2015) The origin of mutants under selection: how natural selection mimics mutagenesis (adaptive mutation). Cold Spring Harb Perspect Biol. 7, a018176. doi.org/10.1101/cshperspect.a018176

Marx, C.J. (2008) Development of a broad-host-range *sacB*-based vector for unmarked allelic exchange. BMC Res. Notes. 1, 1. doi.org/10.1186/1756-0500-1-1

Marx, C.J. (2011) Evolution as an experimental tool in microbiology: “Bacterium, improve thyself!” Environ. Microbiol. Rep. 3, 12-14.

Marx, C.J. (2012) Recovering from a bad start: rapid adaptation and tradeoffs to growth below a threshold density. BMC Evol. Biol. 12, 109. doi.org/10.1186/1471-2148-12-109

Marx, C.J. (2013) Can you sequence ecology? Metagenomics of adaptive diversification. PLoS Biol. 11, e1001487. doi.org/10.1371/journal.pbio.1001487

Marx, C.J., and Lidstrom, M.E. (2001) Development of improved versatile broad-host-range vectors for use in methylotrophs and other Gram-negative bacteria. Microbiology. 147, 2065-2075.

Marx, C.J., and Lidstrom, M.E. (2004) Development of an insertional vector system for *Methylobacterium extorquens* AM1 and generation of null mutants lacking *mtdA* and/or *fch*. Microbiology. 150, 9-19. doi.org/10.1099/mic.0.26587-0

Marx, C.J., O’Brien, B.N., Breezee, J., and Lidstrom, M.E. (2003a) Novel methylotrophy genes of *Methylobacterium extorquens* AM1 identified by using transposon mutagenesis including a putative dihydromethanopterin reductase. J. Bacteriol. 185, 669-673. doi.org/10.1128/JB.185.2.669-673.2003

Marx, C.J., Chistoserdova, L., and Lidstrom, M.E. (2003b) The formaldehyde-detoxifying role of the tetrahydromethanopterin-linked pathway in *Methylobacterium extorquens* AM1. J. Bacteriol. 185, 7160-7168. doi.org/10.1128/JB.185.24.7160-7168.2003

Marx, C.J., Laukel, M, Vorholt, J.A., and Lidstrom, M.E. (2003c) Purification of the formate-tetrahydrofolate ligase from *Methylobacterium extorquens* AM1 and demonstration of its requirement for methylotrophic growth. J. Bacteriol. 185, 7169-7175. doi.org/10.1128/JB.185.24.7169-7175.2003

Marx, C.J., Van Dien, S.J., and Lidstrom, M.E. (2005) Flux analysis uncovers key role of functional redundancy in formaldehyde pathways. PLoS Biol. 3, e16. doi.org/10.1371/journal.pbio.0030016

Marx, C.J., Bringel, F., Chistoserdova, L., Moulin, L., Ul Haque, M.F., Fleischman, D.E., Gruffaz, C., Jourand, P., Kneif, C., Lee, M.-C., Muller, E.E.L., Nadalig, T., Peyraud, R., Roselli, S., Russ, L., Goodwin, L.A., Ivanova, N., Kyrpides, N., Lajus, A., Land, M.L., Médigue, C., Mikhailova, N., Nolan, M., Woyke, T., Stolyar, S., Vorholt, J.A., and Vuilleumier, S. (2012) Complete genome sequences of six strains of the genus *Methylobacterium*. J. Bacteriol. 194, 4746-4748. doi.org/10.1128/JB.01009-12

Michener, J.K., and Marx, C.J. (2015) After horizontal gene transfer, metabolic pathways may need further optimization. Microbe. 10, 61-67.

Michener, J.K., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014a) Phylogeny poorly predicts the utility of a challenging horizontally-transferred gene in *Methylobacterium* strains. J. Bacteriol. 196, 2101-2107. doi.org/10.1128/JB.00034-14

Michener, J.K., Camargo Neves, A.A., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014b) Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post-transfer refinement of the host. eLife. 3, e04279. doi.org/10.7554/eLife.04279

Michener, J.K., Vuilleumier, S., Bringel, F., and Marx, C.J. (2016) Transfer of a catabolic pathway for chloromethane in *Methylobacterium* strains highlights different limitations for growth with chloromethane or with dichloromethane. Front. Microbiol. 7, 1116. doi.org/10.3389/fmicb.2016.01116

Muller, E.E.L., Bringel, F., and Vuilleumier, S. (2011) Dichloromethane-degrading bacteria in the genomic age. Res. Microbiol. 162, 869-876. doi.org/10.1016/j.remic.2011.01.008

Nayak, D.D., and Marx, C.J. (2014a) Genetic and phenotypic comparison of facultative methylotrophy between *Methylobacterium extorquens* strains PA1 and AM1. PLoS One. 9, e107887. doi.org/10.1371/journal.pone.0107887

Nayak, D.D., and Marx, C.J. (2014b) Methylamine utilization via the *N*-methylglutamate pathway in *Methylobacterium extorquens* PA1 involves a novel flow of carbon through C_1 assimilation and dissimilation pathways. J. Bacteriol. 196, 4130-4139. doi.org/10.1128/JB.02026-14

Nayak, D.D., and Marx, C.J. (2015) Experimental horizontal gene transfer of methylamine dehydrogenase mimics prevalent exchange in nature and overcomes the methylamine growth constraints posed by the sub-optimal *N*-methylglutamate pathway. Microorganisms. 3, 60-79. doi.org/10.3390/microorganisms3010060

Nayak, D.D., Agashe, D., Lee, M.-C., and Marx, C.J. (2016) Selection maintains apparently degenerate metabolic pathways due to tradeoffs in using methylamine for carbon versus nitrogen. Curr. Biol. 26, 1416-1426. doi.org/10.1016/j.cub.2016.04.029

Peel, D., and Quayle, J.R. (1961) Microbial growth on C_1 compounds. 1. Isolation and characterization of *Pseudomonas* AM1. Biochem J. 81, 465-469.

Peyraud, R., Kiefer, P., Christen, P., Massou, S., Portais, J.C., and Vorholt, J.A. (2009) Demonstration of the ethylmalonyl-CoA pathway by using ^{13}C metabolomics. Proc Natl Acad Sci USA. 106, 4846-4851. doi.org/10.1073/pnas.0810932106

Ras, J., Van Ophem, P.W., Reijnders, W.N., Van Spanning, R.J., Duine, J.A., Stouthamer, A.H., and Harms, N. (1995) Isolation, sequencing, and mutagenesis of the gene encoding NAD- and glutathione-dependent formaldehyde dehydrogenase (GD-FALDH) from *Paracoccus denitrificans*, in which GD-FALDH is essential for methylotrophic growth. J. Bacteriol. 177, 247–251.

Rokyta, D.R., Joyce, P., Caudle, S.B., Miller, C., Beisel, C.J., and Wichman, H.A. (2011) Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. PLoS Genet. 7, e1002075. doi.org/10.1371/journal.pgen.1002075

Silva, I.N., Santos, P.M., Santos, M.R., Zlosnik, J.E., Speert, D.P., Buskirk, S.W., Bruger, E.L., Waters, C.M., Cooper, V.S., and Moreira, L.M. (2016) Long-term evolution of *Burkholderia multivorans* during a chronic cystic fibrosis infection reveals shifting forces of selection. mSystems. 1, e00029-16. doi.org/10.1128/mSystems.00029-16

Van Dien, S.J., Marx, C.J., O'Brien, B.N., and Lidstrom, M.E. (2003) Genetic characterization of the carotenoid biosynthetic pathway in *Methylobacterium extorquens* AM1 and isolation of a colorless mutant. Appl. Environ. Microbiol. 69, 7563-7566. doi.org/10.1128/AEM.69.12.7563-7566.2003

Vannelli, T., Messmer, M., Studer, A., Vuilleumier, S., and Leisinger, T. (1999) A corrinoid-dependent catabolic pathway for growth of a *Methylobacterium* strain with chloromethane. Proc Natl Acad Sci USA. 96, 4615-4620.

Vasi, F., Travisano, M., and Lenski, R.E. (1994) Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. Am. Nat. 144, 432-456.

Vishniac W., and Santer, M. (1957) The thiobacilli. Bacteriol. Rev. 21, 195-213.

Vorholt, J.A., Marx, C.J., Lidstrom, M.E., and Thauer, R.K. (2000) Novel formaldehyde-activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol. J. Bacteriol. 182, 6645-6650.

Vuilleumier, S., Chistoserdova, L., Lee, M.-C., Bringel, F., Lajus, A., Zhou, Y., Gourion, B., Barbe, V., Chang, J., Cruveiller, S., Dossat, C., Gillett, W., Gruffaz, C., Haugen, E., Hourcade, E., Levy, R., Mangenot, S., Muller, E., Nadalig, T., Pangi, M., Penny, C., Peyraud, R., Robinson, D., Roche, D., Rouy, Z., Saenampechek, C., Salvignol, G., Vallenet, D., Wu, Z., Marx, C.J., Vorholt, J.A., Olson, M.V., Kaul, R., Weissenbach, J., Medigue, C., and Lidstrom, M.E. (2009) *Methylobacterium* genome sequences: a reference blueprint to investigate microbial metabolism of C1 compounds from natural and industrial sources. PLoS One. 4, e5584. doi.org/10.1371/journal.pone.0005584

Vuilleumier, S., Nadalig, T., Ul Haque, M.F., Magdelenat, G., Lajus, A., Roselli, S., Muller, E.E., Gruffaz, C., Barbe, V., Médigue, C., and Bringel, F. (2011) Complete genome sequence of the chloromethane-degrading *Hyphomicrobium* sp. Strain MC1. J. Bacteriol. 193, 5035-5036. doi.org/10.1128/JB.05627-11

Weinreich, D.M., Delaney, N.F., DePristo, M.A., and Hartl, D.L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 312, 111-114. doi.org/10.1126/science.1123539

Wiser, M.J., Ribeck, N., and Lenski, R.E. (2013) Long-term dynamics of adaptation in asexual populations. Science. 342, 1364-1367. doi.org/10.1126/science.1243357

Xue, K.S., Stevens-Ayers, T., Campbell, A.P., Englund, J.A., Pergam, S.A., Boeckh, M., and Bloom, J.D. (2017) Parallel evolution of influenza across multiple spatiotemporal scales. eLife. 6, e26875. doi.org/10.7554/eLife.26875

Yang, L., Jelsbak, L., Marvig, R.L., Damkiær, S., Workman, C.T., Rau, M.H., Hansen, S.K., Folkesson, A., Johansen, H.K., Ciofu, O., Høiby, N., Sommer, M.O., and Molin, S. (2011)

Evolutionary dynamics of bacteria in a human host environment. Proc Natl Acad Sci USA. 108, 7481-7486. doi.org/10.1073.pnas.1018249108